CMRJack

!Mikro

Table of contents

Introduction

CMRJack
Software for estimation of
child mortality measures with standard errors
Version 3.5 Saturday, January 7, 2023

Welcome

CMRJack produces child mortality estimates and estimates of their standard errors from micro data of full and summary birth histories as well as pregnancy histories. Full birth histories are birth histories were women have been asked about the time of birth, survival status and in the case of death the time or age of death of all their children.
Summary birth histories are birth histories that include only the total number of children ever born, and their survival status classified according to age of women or time since first birth.
Pregnacy histories are full birth histories that also reord all pregnacies women have had.

Full birth histories

In the case of full birth histories the program has been designed for accepting data of the type produced by the Demographic and Health Surveys (DHS) and the World Fertility Surveys (WFS) , while in the case of summary birth histories it has been designed to read data as produced by the Multiple Indicator Cluster Surveys (MICS).

When using full birth histories the program estimates neonatal mortality, post neonatal mortality, early neonatal mortality, infant mortality, ${}_2q$, ${}_3q$, ${}_4q$, ${}_5q$ (Under five mortality) and ${}_4q_1$ as well as gender ratios (for neonatal, infant mortality, ${}_5q_0$ and ${}_4q_1$) and their associated standard errors. It also produces the ratios (and their standard errors) of the neonatal mortality rate to the infant mortality rate and under five mortality. Early neonatal mortality is only estimated if the data file contains number of days lived for the first month of life.

The program can produce estimates based in calendar years, or based on years before the survey took place (i.e. when the individual woman was interviewed).

The program can estimate standard errors as done in the DHS reports (i.e. ignoring stratification) or by taking stratification into account.

The estimation period for the child mortality may be one, two, three, four, five or ten years, and it may also produce an optimal time series, which is a series where all the estimates have a coefficient of variation less than 10 (if that is possible, given the data).

Summary birth histories

Estimating child mortality from summary birth historiy is also called indirect estimation.
When using summary birth histories either estimates based on classification by age of women or time since last birth can be used(Hill, 2013). When data are classified by age of women the program will produce estimates of $_5q$ employing indirect estimation techniques of either the Trussel version of the Brass indirect method using Coale-Demeny model life tables, or the Palloni-Heligman version based on the United Nations model life tables. Several versions of the life tables are available.

Pregnancy histories

CMRJack can estimate stillbirth rates from pregnancy histories, and also stillbirth rates by gestational age if data are available.

What's new

Version 3.5.0 - 7.1.2023

Major changes

- Completely redesigned the setting of options. Options are now found under the menu heading "settings" in the main menu line. The settings are now divided into three pages: Calculation, Data, Output, and Quality.
- Increased ability to use memory for 32-bit CMRJack running on 64-bit windows.
- Added the ability to produce data quality related graphs for full birth histories
- Added graphs in presentation output
- Added the ability to output R source files that can be used to run the age pattern model developed by Guillot, Priteto, Verhulst and Garland (2022) on output from CMRJack.
- Introduced choice between consistency with DHS and consistency with definitions of neonatal rates when day-level data on age of death are available.

Minor changes

- Fixes to display behavior when switching between monitors or resizing window.
- Allowed for choice between <32 days (full month) and <28 days (correct definition) for neonatal rates when day-level age of death data are available.
- Fixed possible inconsistency between Excel version and the file type (extension) of Excel output files.
- Added a note field on the main screen. It can be used to enter information that one wants to include in the documentation output sheet or in the note to presentation tables.
- Changed output format file name conventions slightly to avoid too long names.
- Added type of time period to the documentation sheet output.
- Added time period and variance estimation options to SPSS, R, python and command line calls. Note that this will change defaults to what is explicitly or implicitly specified by the call rather than the setting current for CMRJack.
- Added Python 3 support for CALLJACK (for SPSS 25+)
- Fixed that CMRjack did not output gender ratios for summary age group 25-34 in raw output.
- Added command in file menu to save and open Excel file with results in one click.
- Minor changes in core computation of birth histories to increase speed.
- Added date of the creation of output to the Excel documentation sheet.
- Added elapsed time and memory use in log window
- Fixed that the setting for inclusion of the first year of surveys that span two or more years did not get stored.
- Removed option that CMRJack would divide a data set into random groups by itself. (CMRJack can analyze such data but it is better that a user constructs the groups - it is difficult for CMRJack to guess at the proper construction)

- Minor changes in some messages
- Changed hyphen to en-dash in periods and removed spaces (i.e. 2005 - 2009 is now written 2005–2009).
- Changed naming of worksheets in presentation output so that they do not contain spaces (some programs don't like spaces in worksheet names)
- Removed the "Tools" menu item (nobody was using it to merge files which was the only function).

Version 3.2.1 - 2.5.2020

- Added type of period used (calendar vs years before survey) to output file name.
- Removed requirement to provide current age of the child as input data for the full birth history.
- Fixed stupid error message about having reached maximum number of error messages when there are no errors and the maximum number of errors is set to 0.
- Fixed crash when when trying to save output before results are created.
- Fixed note giving number of births and deaths as 0 for indirect estimates presentation output.The note now reports the number of women.

Version 3.2.0 - 26.2.2020

- Added code to protect against crash when trying to run after having canceled file opening.
- Fixed the error that was fixed in the previous update some more (crash with births that have taken place a long time ago).
- Cleaned up some code and made memory management more efficient for the core estimation procedure.
- Made more space for each label on the output tabs.
- Fixed incorrect midpoint reference time and period labels for time before survey that were shown for some surveys that span a new year. As a consequence the calculation of the midpoint time reference is slightly different and may be up to one month different from the previous one in some cases.

Version 3.1.9 - 5.1.2020

- Fixed an error that led to crash when data allow for estimation far back in time and short time periods were used.

Version 3.1.8 - 26.11.2019

- Added link to CMRJack.org on HelpAbout form and in help menu.
- Added example files to setup package.
- When reading an SPSS file with stillbirths CMRJack will not suggest an initial factor variable since the default factor variable is gender that is usually not observed for stillbirths. Factor variables can be added if they are meaningful.
- Fixed error in output specification when CMRJack was called from SPSS. Keywords in SPSS syntax did not match actual output.
- Fixed wrong error message when CMRJack encountered decimal wealth index scores.
- Fixed error in display of presentation results for early mortality broken down by factors. Header line was displayed one cell displaced.
- Fixed crash that occurred when trying to compute optimal gender ratios when no gender variable was defined.
- Fixed error in reading some csv-files in some locales.

Version 3.1.7 - 19-10-2019

- Fixed error in displayed count of stillbirths (it accumulated each time a file was opened)
- Fixed error in time reference and display of periods for full birth history estimates calculated as time before the survey.
- Fixed problem that option "print documentation tables" was not saved.
- Fixed problem that some impossible combinations of output tables were accepted as valid and resulted in generation of invalid Excel files when the user tried to save.
- Fixed: Duplicate presentation tables were generated when there was no factor variable.

- Fixed: After installation of CMRJack the full birth history periods were not defined. This resulted in the program not doing anything when one tried to run an analysis.
- Fixed: Crash when trying to run a data file with stillbirths but using a factor variable with only missing for stillbirths (typically sex of child).
- Added warning for data files with too many missing values for the factor variable for still births.
- Added option to control how many data error messages that are displayed in the log window/file.
- Changed handling of screen resizing, high resolution screens and multiple monitors (This remains to some extent work in progress).
- Some improvements in documentation

Version 3.1.6 - 12-07-2019

- Fixed error introduced in 3.1.5 that sometimes gave wrong error message about too many missing.
- Fixed exception occurring when time span between first and last birth in file were more than 30 years (and 1 year periods were asked for).
- Fixed unreasonable default for number of maximum missing values.
- Added application version to date/time indication in log memo.
- Added stillbirth option in the CALLJACK SPSS syntax to call CMRJack from SPSS. (Also works when calling CMRJack from batch files, R and Python)
- Added DOCUMENTATION output to CALLJACK SPSS syntax. (Also works when calling CMRJack from batch files, R and Python)
- Changed behavior of the OPTIMAL output when called from SPSS. Before OPTIMAL had to be specified in the print subcommand. Now it will be implied whenever OPTIMAL = YES is given as an option in the full birth history specification.

Version 3.1.5 - 29-05-2019

- Added run option to run with more errors and missings than the set limit. The option is located in the file sub-menu. This is mainly useful for stillbirth estimates where there are no gestational age recorded.
- Added error message when a factor variable have more than 90% missing (especially relevant for stillbirth data).

Version 3.1.4 - 3-4-2019

- Fixed error in saving of raw files from indirect estimates

Version 3.1.3 - 2-4-2019

- Fixed more irritating errors in file opening introduced in 3.1. Added warning if one asks for output without any output specified.

Version 3.1.2 - 2-4-2019

- Fixed file opening error introduced in 3.1. (CMRJack would not open a second file)

Version 3.1. 31-3-2019

- Fixed memory issue introduced in 3.1

Version 3.1 27-3-2019

- Added the possibility of calculating stillbirth rates from full pregnancy histories.
- Changed reporting on errors and missing values in data to be clearer.
- Changed the variable name "Period" in raw output to "interval". The general idea is that "Period" should mean a range of years, e.g. 2015-2019, while interval should mean the number of years in the period (i.e. 5 in the example). Also fixed similar inconsistency in the output grids on the screen. However, more must be done here to achieve total consistency.
- Added a sheet with documentation of the processing to the output to the Excel output.
- The installation program has been rewritten.
- Fixed error where data file name was forgotten in some output.

Version 3.0.8 - 23-8-2018

- Added "Presentation tables" in output. Presentation tables are intended to be read, rather than further processed with some other program.
- Added option to open the most recently used data files from the files menu.
- Documented way to call CMRJack from command line, batch files, R and Python. Added support for

reading CSV and Excel files from command line.
- Added: Shortcut CTRL-E for exit
- Fixed: CMRJack kept old file name for output of results when the factor variable was changed, leading to overwriting of output and rather confusing naming.
- Fixed: Error in output in CMEfiles (estimate rather than SE output in some cases)

Version 3.0.7
- Fixed: Crash for some surveys when one year estimation periods were used.
- Fixed: Exception message that occurred when on the file definition screen the factor variable was already entered and the user double clicked on a variable in the variable list. A side effect of the change is that one can double click on a variable to get it entered as a factor variable even if the factor variable is already defined. For other variables the variable field must first be cleared.
- Added: Output files also receive the factor variable name and label as part of their name.

Version 3.0.6 - 26.2.2018
- Added support for summary birth histories without separate questions for boys and girls.
- Fixed: Crash when too high factor variable value was not recognized as missing.
- Fixed: Misleading file name for output files.
- Fixed: Option for limiting error messages when lots of errors in files did not work.
- Added: Warning message when more than maximum errors to be printed have been reached.

Version 3.0.5
Fixed: Calculation of average time of survey went wrong because of change introduced in 3.0.3. Also fixed display formatting error in 3.04.

Version 3.0.4
Added web update. CMRJack will now check for available updates when it starts and offer the option to update. Updates can be deferred, and later be started from the help menu.

Version 3.0.3
- Changed: Error reporting for summary birth histories when the data file contains both Age of Woman classification and Time since first birth classification.
- Fixed: Now detects weights with a value of 0 (cases are omitted).
- Fixed: Wrong dates for direct estimates that use time since survey based periods.Changed output to years for years before survey to be years before survey rather than calendar years.
- Fixed: Setting for type of time period (calendar years vs time before survey) was not stored.

Version 3.0.2
- Fixed: Missing output of gender ratios for indirect measures
- Fixed: Some categories of missing data were retained, leading to strange estimates.
- Fixed: Optimal estimates got default maximum period value of 0 leading to strange estimates (if not manually changed).
- Fixed: Output of optimal estimates was scaled wrong, logits were mixed with real and other embarrassing stuff .

Version 3.0.1
Fixed: Output of 1q0 for indirect estimates was missing

Version 3.0

This version of CMRjack is a major rewrite of the program.

The most important change is that the program now reads SPSS sav files in addition to Excel and comma

separated files.
It also accepts factor variables with any number of values (within reason and your computer's memory). Ratios and bivariate inequity measures can also be produced for any factor variable.

Another important change is that CMRJack can be run from SPSS using a SPSS command (CALLJACK).

Other changes include:

- Choice of how to match indirect estimates to life tables.
- Printout of results
- Children ever born and surviving table can now be saved and printed
- Changes to formats of output
- The presentation of options has been changed
- A help file has been added
- Different versions of life tables and life table matching can be used

If you find errors

CMRjack has become a rather complex piece of software, and there is bound to be errors.

If the program crashes or the output looks strange, please report the error to the mail address supplied in the help-about form, or use the forum on CMRJack.org.
If possible attach the data file that created the problem.

If the program does not crash, but things are still strange, please consult the troubleshooting guide in this document. If that does not help, please consult the website or send a mail.

Getting Started

To install CMRJack use the setup program or it can be done manually if you have not got the setup program. (Not all versions have the setup program).

The install program will assign reasonable folders to the installation. It will check for the presence of SPSS and install the SPSS extension command into the proper directory. If you have more than one version of SPSS the install program will use the folder belonging to the highest version.

Manual installation
You can copy all the program files into a folder of your choice. CMRJack is a 32 bit program so it should be installed in the Program Files(X86) folder if you want to put it in a standard location.

To be able to run CMRJack from SPSS you have to run it once as an ordinary program. This is because it needs to set the registry entry that tells SPSS where it is located.

Two files must be copied to the SPSS extension folder for CMRJack to be called from SPSS. That is the CALLJACK.xml and the CALLJACK.py.

The SPSS extension folder may be located several places, depending on the version of SPSS.

In SPSS versions 23 to 25 (and probably higher) the folder will be
C:\ProgramData\IBM\SPSS\Statistics\25\extensions
or
C:\Program Files\IBM\SPSS\Statistics\25\extensions
'although it seems that the first is the preferred one. The number (25) should correspond to the SPSS version.

In earlier versions of SPSS the ProgramData folder is not used, only the Program files one.

If you have a 32 bit version of SPSS the SPSS program (and the extension folder under the program itself) will be located in Program Files (X86).

System requirements

CMRJack runs under Windows 7 and above. More memory is better than little, especially if you have very large surveys (with many clusters, the actual number of cases does not matter that much).
A speedy processor is much better than a slow one, but the program will run with a slow processor as well.

CMRJAck is designed for a minimum display size of 1024x768 pixels. However, displays that are scaled to more than 100 percent (for example the common 125 or 150 percent) may still face the problem that the form is larger than the screen even if the number of pixels on the screen is larger than the CMRJack design limit.
See CMRJack forms takes up too much space on the display

The 32-bit version of CMRJack may run out of memory for very large data files. This may happen even if your computer has a lot of memory. The reason is twofold: fFrst a program may not use more than 2 - 4 GB in total. Second, CMRJack needs to allocate fairly large chunks of memory. Since memory on a windows machine is typically fragmented, it may not be able to do so, even if the total memory requirement is not exceeded. Adding more memory will not help. Restarting the machine, and then running CMRJack immediately after restart may help somewhat.

Getting help

When running CMRJack as a standalone program you can get help by hitting F1 or using the help menu.

When running from SPSS you can use the command CALLJACK /HELP.

A very quick how to

The following steps are needed to obtain mortality estimates:

Prepare a data file
1. Prepare a data file - it can be either an Excel, comma separated or SPSS file, but SPSS may be best. Data files can be prepared in nearly any statistics package.
2. Ensure that the data file has the correct variables. Which variables you need depends on the analysis: See either full birth histories, summary birth histories or pregnancy histories for details
3. Sort the data file by stratum and cluster

Run the program
1. Start CMRJack and open the data file (file, open). CMRJakc will figure out the file type (Full, summary or pregnancy history)
2. Tell CMRJakc what variables to use (it will make educated guesses).
3. Select time periods (1,2,3,4,5 or 10 year)
4. Run (Run in the menu or CTRL-R)

Save output
1. Ensure that the options for saving output to Excel has been ticked off (see options in the menu)
2. Choose File, Save from the menu (or hit CTRL-S)
3. Open the Excel file by clicking on the button "Show output"

Running CMRJack

CMRJack can be run in several different ways:

1. As a standalone program - see Running CMRJack as a standard program
2. As a command in SPSS - see Running CMRJack from SPSS
3. As a command in R, Python or from the command line - see Running CMRJack from the DOS command line, batch file, script, R or Python

The differentways of running the program will give the same results. However, running CMRJack as a standalone gives more control over the options. Note, however, that CMRJack options are sticky. Thus, if you set options interactively with CMRJack they will be retained also when the program is run from another program or from the command line provided the option is not defined as command line option.

Data requirements

CMRJack can, at present, read SPSS binary files ('*.sav'), Excel files (*.xls and *xlsx) as well as comma separated files (*.csv). The comma separated files may be separated by comma, tab or ; but the file extension must be csv.

In contrast to previous versions of CMRJack there is no requirement on any particular order of the variables in the file, but, given the nature of the analysis certain variables must be present.

The variable requirements vary between summary and full birth histories. For details see:

- Full birth histories
- Summary birth histories
- Pregnancy histories

Labels for Factor variables will be taken from the SPSS file variable and value labels, but in the case of the Excel and csv files the labels must be supplied in a separate file. See The label file

In SPSS files any SPSS missing value is accepted as a missing value.That is, defined user missing values as well as the system missing is considered missing. In addition -99 is treated as missing.

In CSV files and Excel files only -99 is accepted as missing. Blank is not accepted as a value and will produce an error.

See Treatment of missing values

Note: CMRJack is fairly lenient with the coding of CSV-files. The program reads both UTF-8 and ASCII-files. If you can bring the file up in Notepad then CMRJack should be able to read it. The delimeter between values may be , (comma), ; (semicolon) or the tab character. Decimal point and decimal comma are both accepted.

Sort the data file

The data file must be sorted according to stratum and cluster identification. The program does not care if the stratum identifiers are contiguous and also does not care if cluster numbers are. (Thus, a sequence of stratum identifiers like 1001, 1101, 1201 is quite OK).

At least two clusters in each stratum

Every stratum must contain at least two clusters. If a stratum has only one cluster the program will stop, and one should collapse strata, i.e. merge two strata into one. If DHS variance estimation is selected you will still get a warning about single cluster strata, but you will be allowed to make the estimates. (This is because the estimation method ignores the stratification).

Note that many DHS and MICS data files do not have a proper stratum identifier. Although the standard recode file stratum identifier is V022 in DHS, it is sometimes a constant. In MICS the stratum identifier is often missing altogether. The stratification may often be reproduced by consulting the report for the survey. DHS surveys are often implicitly stratified.

If your data do not use CMC coding, then dates must be converted to that format. The DHS CMC coding is simply the number of months since 1900. Thus if year and month are both known, the formula for converting is:

$$CMC = (Year - 1900)12 + Month$$

Making subsets of the data in order to report for particular sub domains should be done with great care, and only if you are very sure about what you are doing. The only safe way to make subsets is to make subsets that span entire strata. Subsets that cut across strata may invalidate the variance estimates.

Variable coding

Note that CMRJack is somewhat more restrictive on the coding of variables than some statistics packages. Thus, for coding of categories on a variable, one must use consecutive codes from 1 to n. The value 0 should not be used as a code. Thus, 1=Male, 2= Female is ok, while 0=Male, 1=Female is not. Similarly, 1=Male, 3=Female will make a mess of things. It is possible that the restriction will be lifted eventually, but that does not have high priority on the list of possible changes and improvements to CMRJack. Note also that there is one exception. Because DHS use 0,1 coding for whether or not a child is still alive, this is accepted by CMRJack (and is in fact mandatory).

In general, the restriction should not pose a problem, as most variables are coded this way anyway. Nevertheless, note that statistical offices in many countries code regions in groups, with for example the first digit representing the group, while the second and third represents the region. Thus 101,102 are the first and second regions in the first group, 201,202,203 are the three regions in the second group and so on. Such regional coding must be changed to consecutive numbering, so that 1=Original 101, 2= original 102, 3=Original 201, 4=Original 202 and so on.

Coding for ratios

Also note that when CMRJack calculates ratios it assumes that the numerator is identified by code 1 for a variable, and the denominator by code 2. Thus, as gender is usually (in DHS and MICS) coded as 1=Male, 2=Female, the gender ratio is calculated as Male/Female. Errors here will be computationally correct but easily misleading when it comes to interpretation.
Missing values

Wealth variables

In the case of wealth type variables (or more formally variables used for bivariate inequity measures) it is assumed that 1 is the most deprived category. Ratios are calculated as most deprived/least deprived, and differences as most deprived – least deprived. Thus, larger ratios and differences indicate large inequities.

Very large numbers

Some statistical offices use very large integers for stratum and cluster coding. This is usually because the coding reflects the administrative geographical hierarchy employed for identifying strata or enumeration areas. If the number is very large (more than 2,147,483,647, and yes, that happens) CMRJack cannot handle it. The solution is then to recode the variable in question, typically by assigning sequential numbering to the strata or clusters.

Below is one way to do the recoding in SPSS, assuming the variable V021 contains the cluster with very large numbers. V021B is the new variable.

```
do if missing(lag(V021)).
   compute V021B=1.
else.
  compute V021B=lag(V021B)+(V021 <>lag(V021)).
end if.
EXECUTE.
Formats V021B (F5.0).
```

The result will be:

```
V021        V021B
10150001001 1
10150001001 1
10150001001 1
10150001001 1
10150001001 1
10150001001 1
10150001001 1
10150001001 1
10150001001 1
10150001001 1
10150001001 1
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150003005 2
10150006002 3
10150006002 3
10150006002 3
10150006002 3
10150006002 3
10150006004 4
10150006004 4
10150006004 4
```

It is also possible to use the `autorecode` command (which will have the same result but also add a large number of value labels that you don't really need).

In Stata the following code will do the same trick:

```
gen V021B=1 if V021[_n-1]==.
replace V021B=cond(V021[_n-1]==V021, V021B[_n-1],V021B[_n-1]+1) if V021[_n-1]
<.
```

In R you could use the following, assuming that tm is the data frame with the data

```
for (i in 1:length(tm$V021))
{
```

```
  if (i==1) tm$V021B[i]<-1 else {
   if (tm$V021[i-1]==tm$V021[i]) tm$V021B[i]<-tm$V021B[i-1] else
     tm$V021B[i]<-tm$V021B[i-1]+1
   }
}
```

Date coding

Note that dates are in CMC - Century Month Coding - as used in the DHS.

```
CMC(month,year)=12(year-1900)+month
```

and to obtain the year and month from a CMC:

```
Year = 1900+ trunc((cmc-1)/12)
```

(where `trunc` means rounding down to nearest whole number, that is the integer portion of the number. In SPSS, Stata and R this function is called `trunc` and some other languages `int`. Note that ordinary rounding cannot be used.)

```
Month = cmc-12(Year-1900)
```

or

```
Month = cmc-12*trunc((cmc-1)/12)
```

See: IUSSP Tools for demographic estimation (with some examples and a reference to the DHS manual)

See also the discussion about calendars in this help file.

Treatment of missing values

Missing values are treated in a listwise fashion. That is, if a variable has a missing value, then the whole case will be missing.

Note that this is particularly important when the program needs to consider both exposure and events. Thus, if the event is missing (e.g. when age of death is not recorded, then the child will also not contribute to the denominator of the mortality rate. Similarly, if the number of children dead is missing for indirect measure, then the woman and her total number of children will not be considered.

SPSS files

In SPSS files any SPSS missing value is accepted as a missing value.That is, defined user missing values as well as the system missing is considered missing. In addition -99 is treated as missing.

Note that you should delete value labels of the missing values for factor variables. If you for example have the variable Gender which is coded 1=Male, 2=Female, 9=Missing gender, then delete the missing gender value. (A simple way is by redefining the value. Otherwise CMRJack will produce tables for the missing value (which most likely will be meaningless).

CSV/Excel files

In CSV files and Excel files only -99 is accepted as missing. Blank is not accepted as a value and will produce an error.

There is an option that defines how many missing CMRJack will accept. See Options

Sample description variables

Four variables describe the sampling. They must be present in the data file. The variables, with their standard DHS names, are the following:

Variable	Content
V005	Estimation weight
V008	Time of survey. This should be in standard DHS CMC (Century Month Coding)
V021	Cluster identifier
V022	Stratum identifier

The estimation weight may be either the expansion weight (that adds to the population size) or the relative weight, which adds to the sample size.
The V005 variable in DHS files is the relative weight multiplied by 1,000,000. Thus, when preparing the file it is conventional to divide V005 by 1,000,000 to obtain the actual weight.

The time of the survey should be coded in DHS Century Month Coding (CMC) . It is significant in summary birth history analysis for computing the time reference, and in full birth history analysis for computing number of years before the survey. If the analysis is in terms of calendar years the variable is only used in order to fix the calendar years of analysis. Some surveys use national calendars. These should be transformed to the standard Gregorian (unless you only plan to carry out analysis in terms of time before the survey, in which case it is irrelevant unless a calendar with a non-solar year is chosen). See [Calendars]

The cluster identifier must uniquely define clusters within strata. They do not need to be single series of adjacent numbers. Thus, cluster numbers within strata of e.f. 1,2,3,4,5... ; 10, 20, 30, 50...; or 23, 34, 87, 98.... are all acceptable. However, they must be sorted within each stratum.

The strata must be uniquely identified, and the file sorted according to stratum. As in the case of the clusters, the actual coding does not matter.

There must be at least two clusters within each stratum. Some surveys are designed with only one. If that is the case, the strata must be collapsed, i.e. two strata merged into one. see Kish 1995:232; Heeringa, West & Berglund 2010:101). One should note that in most cases one does not need the full fledged collapsed stratum designs as discussed by Wolter (1985:47-54 and Kish 1995:283-286) as these are designs with only one cluster in all strata. Demographic surveys typically do not use such designs, and single cluster strata usually only arise from problems of allocation of clusters to the strata. However, the general result that collapsing strata tends to overestimate variance is valid, but is no cause for concern because the effect is very small.

In surveys that use more than two stages of selection the cluster should be the first random selection, not the area unit at the penultimate stage of selection. The first random selection is called the "ultimate cluster" and is the aggregate of all subsequent sampling units (see Kish 1995:155, Heeringa, West & Berglund 2010:67). Some DHS files have a variable named "ultimate cluster" that should be used.

Data files without clusters

Some data files do not have cluster and stratification information. This may be because it has been deleted, or because the information never existed,
A typical example of data with no cluster and strata information is census data, since census data are complete and do not have clustering and stratification.
The same would be true of subsets of census data from for example IMPUMS (see https://international.ipums.org/international/).

In the case of census data a possible approach to the variance estimation is to use the Delete-a-Group-Jackknife (Kott 2001). This involves creating clusters. The procedure is simply assigning clusters to cases randomly, so that each clusters can be seen as a sample of the population. A simple way to do so is to assign a uniformly distributed random number to each case, sort the data file by the number and then assign clusters sequentially. Cluster sizes of 100 -300 seem to work fine. The stratum variable should be set to a constant, e.g. 1.

Below is some SPSS code that achieves this (for the IPUMS Kenya 2009 sample, but that is irrelevant except for the line of code that avoid a small cluster at the end).

```
compute stratum=1.
compute  clustpos=RV.UNIFORM(0,5000000).
sort cases by clustpos.
compute  clust100=trunc($casenum/100)+1.
compute  clust200=trunc($casenum/200)+1.
compute  clust300=trunc($casenum/300)+1.
* The following line is for avoiding a small cluster at the end of the file.
Must be checked manually in each case.
if clust100=9350 clust100=9349.
freq clust100 clust200 clust300.
```

Full birth histories

Full birth histories must have one child on each data record (line). Thus, a woman that has five children will be represented with five records, one for each child.

For full birth histories the following variables are needed:

Variable	Content
V005	Estimation weight (The program has only been tested with normalized weights. In principle expansion weights should work just as well)
V008	Time of survey. (Date of interview) This should be in standard DHS CMC (Century Month Coding)
V021	Cluster identifier (see note on maximum size under variable coding)
V022	Stratum identifier (see note on maximum size under variable coding)
B3	Date of birth CMC
B5	Alive or dead at time of interview (0=dead, 1 alive). Note that this is the only place where the use of 0 for coding of a discrete variable is allowed. The only reason it is allowed is that all DHS files are coded this way.
B7	Age at death. Three formats are possible 1. The age at death simply coded as the number of completed months lived. 2. As the first but with with a decimal part indicating number of days for deaths within the first month of life. Thus 0.17 would be for 17 days, and 0.05 would be five days. For completed months from 1 and above, only the whole number part should be coded. 3. As 1 or 2 for live births, but for full pregnancy histories stillbirths coded as number of weeks of gestational age.
Factor	Optional factor variable. Common ones are B4 (Gender, which must be coded 1=Male, 2=Female (to get gender ratios in the conventional form of male/female) V190 is the Wealth index in DHS files expressed as quintiles. V191 is the raw index. CMRJack can split the raw index into groups.

If the file is supposed to be analyzed using the form recommended in the Tools For Demographic Estimation, the Date of Birth (B3)and Age at Death (B7) must be coded in decimal form. The decimal form is similar to the CMC, but the days within the month is added as a decimal. Thus, the second day may be coded as 2/30.44 (30.44 being the average length of the month). The Tools for Demographic estimation has a suggestion for converting standard DHS files to the file decimal format.

Note that the data file must be sorted by V022 (stratum) and V021 (Cluster). See Data requirements.

The coding of age at death in days for the first month of life may cause a problem. DHS, MICS, and WFS estimate neonatal mortality with the number of deaths during the first month as numerator and number of births and denominator. The reason is that these surveys do not use days in the estimates.
Strictly speaking, neonatal mortality is defined with deaths during the first 28 days as the numerator, i.e., with age at death less than 28 days (since day 0 is possible). See, for example, WHO or Measure Evaluation. When there is a coding of age at death in days in a data file, then CMRJack by default uses the strict definition of neonatal mortality.
See the section on neonatal mortality estimation for more details and options.

Note: There is an example file "ExampleFBH 2015.sav" in the example data folder. While somewhat based on real data, the data is not representative of any country and is not intended as anything but an example.

Note: The variable names are not significant in themselves, but if you use the standard names commonly used in DHS and MICS then CMRJack will probably guess right when it opens the file.

Note: It may be useful to include identification variables for household, woman and birth. Then, if you want to add more covariates later you can just match them in.

Note: Versions prior to 3.3.0 also used B8, i.e. Current age of child. That was a remnant of a previous estimation syntax that just got left there when the code was changed. Therefore it was removed.

Example syntax for SPSS.

```
if missing(b7) b7=-99.
* Create a variable with day level data for age at death.
compute B7A=B7.
if Range(B6,100,130) B7A=(B6-100)/100.
Variable labels B7A "Age at death with day level data for 0-30 days".
* Create a normalized weight variable.
COMPUTE v005=V005/1000000.
formats v005 (F10.7).
sort cases by v022, v021.  /* sort by stratum and cluster.
* Change name of file to correct one.
SAVE OUTFILE="MyCountry.sav"
/KEEP  v005 v008 v021 v022 b3 b4 b5 b7 B7A V101.
```

Pregnancy histories

Pregnancy histories are handled in the same way as full birth histories, i.e. with the same variables. There are two coding differences:

The status variable for the child being alive or dead (typically B5) is coded differently. The coding has been extended so that the possible values are the following:

Code	Meaning
0	The child was born alive, but is now dead
1	The child was born alive, and is still alive
2	Stillbirth
3	Miscarriage
4	Provoked termination of pregnancy (abortions)

The age at death variable (typically B7) is interpreted as gestational age *in weeks* for children with code 2,3 and 4 (stillbirths, miscarriages and abortions). Some surveys use a more elaborate coding. For example, some code the reason for miscarriage or stillbirths. Also some distinguish between abortions carried out by health personnel and self induced abortions. CMRJack cannot use more than the codes listed above.

One should note that the definition between stillbirths and miscarriages may differ between countries and between surveys. Some use gestational age as criterion for distinguishing stillbirths and miscarriages, with various cutoffs. Classifying stillbirths as fetal deaths in or after the 22 week of pregnancy or in or after the 28 week is more common, and CMRJack will produce statistics for both if data are available.

Some surveys do not collect weeks of gestation. The program will accept missing weeks for all cases (as of version 3.1.5), but note that one then either has to set acceptable missing cases to a very large number or use the "Run with errors" option in the file menu.

Surveys that collect pregnancy histories rarely collect the gender of stillbirths. Since CMRJack automatically may suggest gender as a factor variable one will probably need to remove that variable from the variable definition form. When reading SPSS files with stillbirth data CMRJack will try not to suggest a factor variable.

Note that the data file must be sorted by V022 (stratum) and V021 (Cluster). See Data requirements.

Note: There is an example file "Example FBH SB 2016.sav" in the example data folder. While somewhat based on real data, the data is not represenative of any country and is not intended as anything but an example.

Note: The variable names are not significant in themselves, but if you use the standard names commonly used in DHS and MICS then CMRJack will probably guess right when it opens the file.

Summary birth histories

For summary birth histories the following variables are needed:

Variable	Content

Ivdate	Time of survey (interview). This should be in standard DHS CMC (Century Month Coding).
Weight	Estimation weight (The program has only been tested with normalized weights. In principle expansion weights should work just as well)
Stratum	Stratum identifier (see note on maximum size under variable coding)
Cluster	Cluster identifier (see note on maximum size under variable coding)
Womage	Classification by age: Current age of women, coded from 1 to 7
Wombiry	Classification by time since last birth: Time since last birth, 5 year groups coded from 1 to 5
deadB	Dead boys
deadG	Dead girls
birthB	Boys ever born (i.e. including dead ones)
birthG	Girls ever born (i.e. including dead ones)

See below for what to do if you have data that do not distinguish between boys and girls.

As is the case for Full birth histories, one or more factor variables can be included. Either the Age of Woman classification or the Birth year classification will be used in any given run of the program, but it does not matter if both variables are present.

The classification of women must be coded as follows

Code	Age group (years)
1	15-19
2	20-24
3	25-29
4	30-34
5	35-39
6	40-44
7	45-49

The classification of time since first birth must be coded as follows:

Code	Time since first birth (years)
1	0-4
2	5-9
3	10-14
4	15-19
5	20-24

Data without gender distinction

Some surveys only collect data for children regardless of gender. If you have such a survey use either the dead boys, boys ever born without the girls, or dead girls, girls ever born without the boys. The single gender will then be interpreted as both genders.

Note that the data file must be sorted by V022 (stratum) and V021 (Cluster). See Data requirements.

Note: There is an example file "ExampleSBHAOW 2005.sav" that illustrates analysis by age of woman in the example data folder. There is also another, "ExampleSBHTSFB 2009.sav" that does the same for Time Since First Birth analysis. While somewhat based on real data, the data files are not representative of any country and are not intended as anything but examples.

i Note: The variable names are not significant in themselves, but if you use the standard names commonly used in DHS and MICS then CMRJack will probably guess right when it opens the file.

i Note: It may be useful to include identification variables for household, woman and birth. Then, if you want to add more covariates later you can just match them in.

Factor variables

Factor variables are variables that are used to provide separate estimates of mortality by sub groups. Examples include residence (Urban/rural), Region or Education.

There are some restrictions on factor variables:

- Only one variable can be used at a time (but this can be circumvented by creating merged variables in SPSS)
- They must be numeric
- They must be coded from 1 and use adjacent codes, i.e. 1,2,3,4...n. Do not use 0 as a code.

There are no particular restriction on how many categories there may be in a factor variable, - the program allocates space as needed. However, memory and sample size may pose problems.

When reading SPSS files CMRJack will use the value labels in the SPSS input file. If the file has no value labels, the values themselves will be for labeling the output. That is likely to be confusing. It is, however, possible to supply a label file for comma separated and Excel files. See The label file

Factor variables may be

Simple factor variables

Ratio factor variables

Inequity factor variables

All produce results by the value of the factor variable, but ratio and inequity in addition produce other output.

Simple factor variables

Simple factor variables are simply used to define sub groups of variables. CMRJack will make simple tabulations for such variables, without any derived measures

Ratio factor variables

Ratio factor variables are used to produce various types of ratios, such as gender ratios, neonatal mortality rate to infant mortality rate.

Ratio factor variables are used to produce various types of ratios. The underlying assumption is that the variable has two values (coded 1 and 2) and that the lower value serves as the numerator in the ratio, and the higher value provides the denominator. Thus, the standard sex ratio is (male rate)/(female rate) since male is usually coded 1 and female 2.
The following ratios are produced

Ratio	Calculation
General ratios	

Neonatal	$r_{{}_nq_x} = \frac{{}_nq_{x\|y-1}}{{}_nq_{x\|y-2}}$
1q0	as above
5q0	as above
4q1	as above
Early Neonatal Mortality	as above
Late Neonatal Mortality	as above
Other ratios	
Neonatal to 1q0	Neonatal/1q0
Neonatal to 5q0	Neonatal/1q0
Early Neonatal to Neonatal	ENMR/NMR
Infant to 5q0	IMR/U5MR

See Transformations for the logit and log transformations of the ratios.

Inequity factor variables

Defining a factor variable as an inequity variable leads CMRjack to Calculate bivariate inequity measures. Such measures assume that the factor variable has ranked values such as, for example quintiles of a wealth index. CMRJack will then tabulate the estimates for each value, and then calculate the absolute range; the relative range; the ratio between the lowest and the highest value; the relative risk; and the relative risk percentage.

CMRJack is also able to use a continuous variable and rank it. See below (Weights and inequity measures)

Weights and inequity measures

The DHS and MICS surveys calculate wealth indices based on material belongings and infrastructure available to the household. The index is constructed with household level data and weights for the index are calculated using Principal Components Analysis on the household level(Rutstein & Johnson, 2004). Wealth quintiles are then constructed on the individual level. Thus, the quintiles are constructed so that each contains 20 percent of the household population of individuals. The procedure consequently makes the upper (rich) quintiles likely to contain fewer births than the lower (poor) quintiles, or put in another way: the population of births are not divided into five equal sized groups.

There are two main consequences of the way the quintiles have been constructed. The first is that the variance of the measures based on the quintiles is likely to be higher than it would have been if the quintiles had been ranked according to births rather than household members.

The second consequence is that one must be careful about the interpretation of the measures. Thus, the child mortality estimate for, say, the lowest quintile should be interpreted as “the mortality rate of the births occurring to women living in households that make up the 20 poorest of the population of individuals”. That interpretation is perhaps not exceedingly easy to communicate.
CMRJack allows an alternative way of computing bivariate inequity indices, in that it can compute ntiles from a continuous ranking variable (containing, for example, the object score output from a principal component analysis). The ranking can use from two to ten groups, but defaults to five, thus producing quintiles. The ranking is by weighted by the number of births in the relevant group. Thus each wealth group will consist of the same weighted number of births.

In the case of summary birth histories the above means that a separate ranking is calculate for every age group of women, while for complete birth histories it means that a separate ranking is computed for every period of estimation (i.e. the 10 year group closest to the time of the survey has another ranking than the 10 year group that ends 10 years before the survey).
In the case of summary birth histories there are two options that control how the weighting is carried out.

One option uses weighting based in women rather than births. Thus, the interpretation is in relation to equal sized groups of women.

The other option is a bit subtle. When the wealth groups are constructed based in births in the summary

birth history, women without children (0-parity women) poses a problem. By default this is dealt with by ranking them along with women with children, but assigning them 0 weight in the cumulative distribution used for the ranking. Thus, the women are given the rank of the group they belong to in terms of their household wealth, but they are not counted in relation to the size of the groups.
This can be changed by an option giving the women their original estimation weight. This is like assuming they each have one birth. The interpretation of this option is not clear-cut, - the ranking is basically in relation to a mix of births and women.   Unless you have some good reason, this option should probably not be chosen. (We mainly included it as part of the process of experimenting with such measures.)

Also see Options

The label file

If CMRJack is reading an SPSS file, then the variable and value labels in that file will be used.

If, on the other hand, a comma separated (*.csv) or Excel file is being read, then that file will not have any labeling. Value labels will not be displayed, and in their place there will be simply the value of the variable. Thus, CMRjack will label Male with 1, and Female with 2 (if that is the coding). This is fine if one remembers the coding, and there are not too many codes.

A way to solve the problem is to construct a label file. The label file should have the extension lbl and be placed int the same folder as the datafile itself. The name of the file must be the exactly same as the name of the datafile.

Thus if the datafile is called "Cambodia2005.csv", then the label file must be called "Cambodia2005.lbl".

The file itself should be constructed as a typical windows inifile, as shown in the example below:

[B4]
Label="Gender"
1=Male
2="Female"

[V190]
Label="Standard DHS Wealth quintiles"
1="Poorest"
2="Second poorest"
3="Middle"
4="Second richest"
5="Richest"

Thus, the variable name should be in brackets on its own line, The label should be indicated with the text "Label" followed by an equal sign and text.
Each value should be indicated by its value, and equal sign and its value label.

If the text is more than one word it should be enclosed in quotes. If it is only one word it does not matter if there are quotes. (see B4 above - both of the value labels are valid).

There can be an unlimited (within reason) variables in the label file. If they are not used in a specific analysis they will be ignored.

Note: There is an example file "ExampleFBH 2015.lbl" in the example data folder. The file contains the same information as given for B4 and V190 above.

Example data files

There are five different data sets in the example data folder. Each data set has three different formats: comma separated (csv), SPSS (sav) and Excel (xlsx). The data sets are constructed, and does not represent any particular country or time. The year identifcation of the files are simply because it is convenient and because both full and summary birth histories require information about time of the survey.

The files are as follows:

File	Content
Example FBH 2015.csv	Full birth history file in csv format, with coding for early and late neonatal mortality
Example FBH 2015.lbl	Label file to support Example FBH 2015.csv/xlsx
Example FBH 2015.sav	Full birth history file in SPSS (sav) format, with coding for early and late neonatal mortality
Example FBH 2015.xlsx	Full birth history file in Excel (xlsx) format, with coding for early and late neonatal mortality
Example FBH NEM 2015.csv	Full birth history file in csv format, but WITHOUT coding for early and late neonatal mortality
Example FBH NEM 2015.sav	Full birth history file in SPSS (sav) format, but WITHOUT coding for early and late neonatal mortality
Example FBH NEM 2015.xlsx	Full birth history file in Excel (xlsx) format, but WITHOUT coding for early and late neonatal mortality
Example SBH AOW 2005.csv	Summary birth history classified by age of woman, csv format
Example SBH AOW 2005.sav	Summary birth history classified by age of woman, SPSS (sav) format
Example SBH AOW 2005.xlsx	Summary birth history classified by age of woman, Excel (xlsx) format
Example SBH TSFB 2009.csv	Summary birth history classified by time since first birth, csv format
Example SBH TSFB 2009.sav	Summary birth history classified by time since first birth, SPSS (sav) format
Example SBH TSFB 2009.xlsx	Summary birth history classified by time since first birth, Excel (xlsx) format
Example FPH SB 2016.csv	Full pregnancy history file in csv format with stillbirth coding
Example FPH SB 2016.sav	Full pregnancy history file in SPSS (sav) format with stillbirth coding
Example FPH SB 2016.xlsx	Full pregnancy history file in Excel (xlsx) format with stillbirth coding

The SPSS or Excel (if you don't have SPSS or PSPP) are probably the most easy to look at in order to understand the file format.

Running CMRJack as a standard program

CMRJack can be run as a standalone program by simply starting the program in the usual fashion in Windows. One should then set the required options, open a file, and then choose run from the menu. You can also choose "Run with errors" from the file menu. This option overrides the limits set in the options menu for maximum number of missings and errors allowed. (If that is sensible or not is entirely up to your judgment)

Running CMRjack from SPSS

CMRJack can be executed from SPSS through an SPSS extension command.

If CMRJack is properly installed, the extension command behaves like an ordinary SPSS command, except that it saves the output in an Excel file rather than to the output window of SPSS.

The syntax for the command is as follows:

```
CALLJACK
  [/OUTPUT FILE  = {file}]
  /SAMPLE WEIGHT = {Var} STRATUM = {Var} CLUSTER = {Var} DATEINT = {Var}
  [/FBH BIRTHDATE = {Var} ALIVE = {Var} AGEDEATH = {Var} CURRAGE = {Var}]
    [PERIODS={list of 1,2,3,4,5**,10}]
    [OPTIMAL = {YES|NO**}]
    [TIME = {CALENDAR**|BEFORESURVEY}]
    [STILLBIRTHS = {NONE**|SBONLY|ALL}]
  [/SBHAGE AGEWOMAN = {Var} DEADBOYS = {Var}DEADGIRLS = {Var} LIVEBOYS = {Var} LIVEGIRLS
= {Var}]
  [/SBHTSFB TSFB = {Var} DEADBOYS = {Var} DEADGIRLS = {Var} LIVEBOYS = {Var} LIVEGIRLS =
{Var}]
  [/LIFETABLE MODEL=[UNLATINAMERICAN | UNCHILEAN | UNSOUTHASIAN | UNFAREAST |UNGENERAl |
NORTH | EAST |SOUTH | WEST **]
          [PRINT = {SELECTEDMODEL** | ALLMODELS}]]
  [/FACTOR VARIABLE= {Var} TYPE={SIMPLE**|RATIO|INEQUITY}]
  [/TRANSFORM SCALE = {REAL** | LOG |LOGIT}]
  [/VARIANCE ESTIMATOR = {STRATIFIED**| NOSTRATA}]
  [/SURVEY [NAME={text}] [YEAR={text}] [TYPE={text}] [COUNTRY={text}]]
  [/PRINT FORMAT=[PRESENTATION] [CMEINFO] [RAW**] [OPTIMAL] [CEBCSTABLE][DOCUMENTATION]]
  [/HELP].
```

** Default if the subcommand is omitted.

OUTPUT controls the filename of the Excel file with results

SAMPLE describes the sample structure with three variables WEIGHT STRATUM and CLUSTER and also the Date of interview (DATEINT)

Use only one of the commands FBH, SBHAGE or TBHTSFB.

FBH Full birth history.
Requires a list of variables that provide birth date of the child, gender, whether the child is currently living, the age of death, and the current age.
PERIODS allows you to supply the periods required for estimation
OPTIMAL turns calculation of optimal periods on (OPTIMAL=YES) or off (OPTIMAL=NO).
STILLBIRTHS controls generation of fetal loss estimates NONE means no estimates (even if data are present)
SBONLY means estimation of stillbirth rates only, while all estimates stillbirths, miscarriages and other terminations.
SBHAGE Summary birth history with classification after age of mother.
The subcommand requires a list of variables:
AGEWOMAN, DEADBOYS, DEADGIRLS, LIVEBOYS, LIVEGIRLS

SBTSFB Summary birth history with classification after time since first birth
Summary birth history with classification after age of mother.
The subcommand requires a list of variables: ,
TSFB, DEADBOYS, DEADGIRLS, LIVEBOYS, LIVEGIRLS

LIFTEABLE specifies the life table model.
LIFETABLE are only used in the context of summary birth histories.
Summary birth histories using the Time Since First Birth approach (SBTSFB) can only use the Coale-Demeny
North, East, South and West.

FACTOR allow you to supply a variable that is used to provide subgroups in the tabulation.
The keyword SIMPLE to the FACTOR command means that simple estimates should be produce; RATIO means that ratios based
on the variable should be produced and INEQUITY produces bivariate inequity measures.

TRANSFORM and its subcommand SCALE determines if the estimates and their standard errors should be on the REAL scale, LOG or LOGIT scale. See Transformations

SURVEY allows you to give a title to the output.
example SURVEY NAME="DR Congo" YEAR="2013-14" TYPE="DHS" COUNTRY="Democratic Republic of Congo"

PRINT controls the output. Choices of FORMAT are (see Save results)
PRESENTATION (easily readable tables)
CMEINFO (Output used to input into the CME Info database – little use for other users)
RAW (File with the estimates, each as a column)
CEBCSTABLE (Table with estimates of children ever born, children surviving)
OPTIMAL Output from optimal estimates
DOCUMENTATION Documentation sheet of settings, file etc.

Example syntax

```
file handle dataloc /name='......'.    /* Add datalocation here
get file 'dataloc\senegal2015.sav'.
dataset name senegal.

calljack
  /output file='c:\users\USER\documents\cmrtest\nytest.xlsx'
/SAMPLE WEIGHT= V005 STRATUM= V022 CLUSTER=V021 DATEINT= V008
/FBH BIRTHDATE= B3 ALIVE=B5 AGEDEATH=B7 CURRAGE=B8 PERIODS=5 OPTIMAL=YES
/FACTOR VARIABLE=B4 TYPE=RATIO
/SURVEY NAME= "Senegal DHS 2015" COUNTRY="Senegal"
YEAR="2015" TYPE="Demographic and Health Survey"
/PRINT FORMAT=CMEINFO RAW OPTIMAL.



get file 'dataloc\car 2006 Sum.sav'.
dataset name car.
CALLJACK /output file='c:\users\USER\documents\cmrtest\testind.xlsx'
/SAMPLE WEIGHT=WEIGHT STRATUM=STRATUM CLUSTER=CLUSTER DATEINT=IVDATE
/SBHAGE AGEWOMAN= womage LIVEBOYS= birthB LIVEGIRLS=birthG DEADBOYS= deadB
DEADGIRLS= DEADG
/FACTOR VARIABLE=WEALIDX TYPE=INEQUITY
/SURVEY COUNTRY="CAR" YEAR="2006" TYPE="MICS" NAME="CAR 2006"
/PRINT FORMAT=CMEINFO RAW.
```

Running CMRJack from the DOS command line, batch file, script, R or Python

CMRJack can be run from the DOS command line, or by extension from a batch file, or a script file. It can also be called with a command from several programming languages (for example R or Python).

Basically, what is needed is the command line required to invoke CMRJack, and a ton of required and optional options.
Note that some options that are possible from the windows interface have not been implemented. However, the options will be set as they were when you last changed them with the windows interface.

A BAT-file consisting of:

```
C:\CMRJack\CMRJack
 /o:"C:\Users\YOU\Documents\Wakanda 2018.xlsx"
/a:FBH /1:V005 /2:V022 /3:V022 /4:V001 /5:B3 /6:B5 /7:B7 /8:B8 /FA:B4 /FT:RATI
O /P:5 /OPT:YES /SB:NONE /SV:2018 /SVC:"Wakanda" /PT:21 /s:"C:\\Users\\YOU\
\Documents\\cmrtest\\Wakanda.csv\"
```

will call CMRJack provided CMRJack is located in the c:`CMRJack directory. (The command should go on a single line).

To call CMRJack from R the following code will do:

```
# Run CMRJack from R
fp <- file.path("SOFTWARE", "CMRJack", "CMRJAck", fsep="\\")
registryentries<-readRegistry(fp, "HCU")
system2(command=registryentries$ExeDir,args="/o:\"c:\\users\\YOU\\Documents\
\CMRTest\\Wakanda
2018.xlsx\" /a:FBH /1:V005 /2:V022 /3:V021 /4:V008 /5:B3 /6:B5 /7:B7 /8:B8 /FA
:B4 /FT:RATIO /P:5 /OPT:NO /SB:NONE /SV:2018 /SVC:\"Wakanda\" /PT:21 /s:\"C:\
\Users\\YOU\\Documents\\cmrtest\\Wakanda.csv\""  ,invisible=FALSE)
```

Note that the the quotes must be escaped with \ and \\ is needed for separators in paths.

To call CMRJack from Python (the example is using Python 3.4.4):

```
import subprocess,os, winreg
hkey=winreg.OpenKey(winreg.HKEY_CURRENT_USER,"SOFTWARE\\CMRJack\\CMRJack\\")
ProgramDir=winreg.QueryValueEx(hkey,"ExeDir")
inp=os.path.normpath("C:\\Users\\YOU\\Documents\\cmrtest\\Cambodia2005.csv")
outp=os.path.normpath("c:\\users\\YOU\\Documents\\CMRTest\\Cambodia 2005A.xlsx")
subprocess.call([ProgramDir[0],'/o:"'+outp+'"', r"/a:FBH", r"/1:V005",
r"/2:V022",r"/3:V021", r"/4:V008", r"/5:B3", r"/6:B5", r"/7:B7", r"/8:B8", r"/FA:B4",
r"/FT:RATIO", r"/P:5", r"/OPT:NO", r"/SB:NONE", r"/SV:2018", r'/SVC:"Cambodia"',
r"/PT:21", '/s:"'+inp+'"'])
```

Note that the arguments are split into a list in the Python code. Also note the use of raw strings. Actually Python insists on adding slashes before and after the file names, but CMRjack takes care of that.

Both the R and Python code depends on using the windows registry to find the location of CMRJack.

Note that there should be no space between an option specifier (e.g. /1:) and the specification. /1:V005 is correct, while /1: V005 will provoke an error.

You can use input files that are either SPSS, CSV or Excel. CSV files will only be recognized if they have the extension CSV or TXT.

Output is stored in the Excel output file.

The options are as follows:

/o:"outputfile"   optional, default is cmrjackout.xlsx

/s:"Savfilename"   the input file, required.

/a:ANALYSESTYPE Required. should be either FBH or AOW or TSFB

The following options are always required:

/1:Weight variable
/2:Stratum variable

/3:Cluster variable
/4:Date of interview variable

The following options are all required for a full birth history.

/5:Date of birth variable.
/6:Alive/Dead status variable
/7:Age at death variable

The following options are optional for full birth histories:

/OPT:Carry out optimal period estimation (YES or NO) (Only relevant for full birth histories)
/SB: Use full pregnancy history if present and estimate stillbirths (NONE: do not do it, SBONLY: only stillbirths, ALL: Stillbirths, miscarriages and other terminations)
/P:Periods to estimate Either 1,2,3,4,5,10 or any combination
/TI: Type of time used: CAL - calendar years, BEF - years before survey

The following are relevant for summary birth histories

/9:Age of woman classifier
/10:Time since first birth classifier
/11:Dead Boys
/12:Dead Girls
/13:Live Boys
/14:Live Girls

The following options are optional:

/FA:Factor variable (for breakdowns)
/FT:Factor variable type (SIMPLE, RATIO or INEQUITY)
/L:Scale of output (REAL,LOG, LOGIT, default is real)
/SV:Survey year
/SVC:Survey country
/SVT:Survey Type (descriptive text for survey type)
/VA: Type of variance estimator: STR - Stratified, NOS - Not stratified

/PT:Print 1 Presentation 2: CMEINFO 4 Raw 8 CEBSTABLE 16 OPTIMAL 32 DOCUMENTATION
The options for printing are additive. That is 17 (1+16) will produce a presentation table and optimal tabulation, while 21 in addition will produce a raw table (1+4+16).
Note that OPTIMAL output will be generated if OPTIMAL =YES is specified under the full birth history option regardless of how the print command is given.

Controlling CMRjack

How CMRJAck runs and what output it produces are controlled by several sets of options.

The settings of the various options are stored in the system registry of the computer and will be kept between sessions.

Some of the options can be set when the program is started from SPSS. Those that cannot be controlled from SPSS will use the value that is stored in the registry.

The options are divided into three groups that are found in different locations

Survey information group

The survey information group is located in the main window, and is usually filled in by the program, usually from the file name.
They comprise the country, the country code (which follows after country) the survey label and the survey

type. All of these can be edited.

Full Birth History group

The Full Birth History group is located on the full birth history panel and comprises the key choices of:

1. Period of estimation
2. Whether or not to make optimal summary
3. The variance estimation method
4. The type of time period (i.e. DHS type years before the survey, or calendar years)

For details see Method used for direct estimates and Optimal estimates

General options group

The general options group are located in the options menu and control a variety of things.
For details, see Options

The menu

The menu controls the main aspects of CMRJack. Those include opening data files; setting options; running the analysis, and saving results.

File

Open

This menu item is used to open data files. One may also use the keyboard shortcut CTRL-O. The menu item shows an open file dialog, and when a suitable file is chosen a dialog for describing the data will appear. When a data file has been defined, the program will display the main screen again. It will then either wait for more commands, or run automatically if that option has been selected.

The dialog that is used to determine the variables used in the analysis will look more or less like the one below.

On the left hand side, there is the variable list for the file. If you right click on a variable, value labels, if any, will be shown. The program will attempt to fill in the column ion the right hand side, which defines which variable contain what information for the analysis. You can drag the variable from the left to the right, and use the red crosses to delete a variable.
It is also possible to double-click on a variable in the file, and it will be moved to the first empty slot on the right side that makes sense.

CMRJack will attempt to fill in the analysis variables from the file, based on the names commonly used in DHS and MICS files, but the procedure is far from perfect.

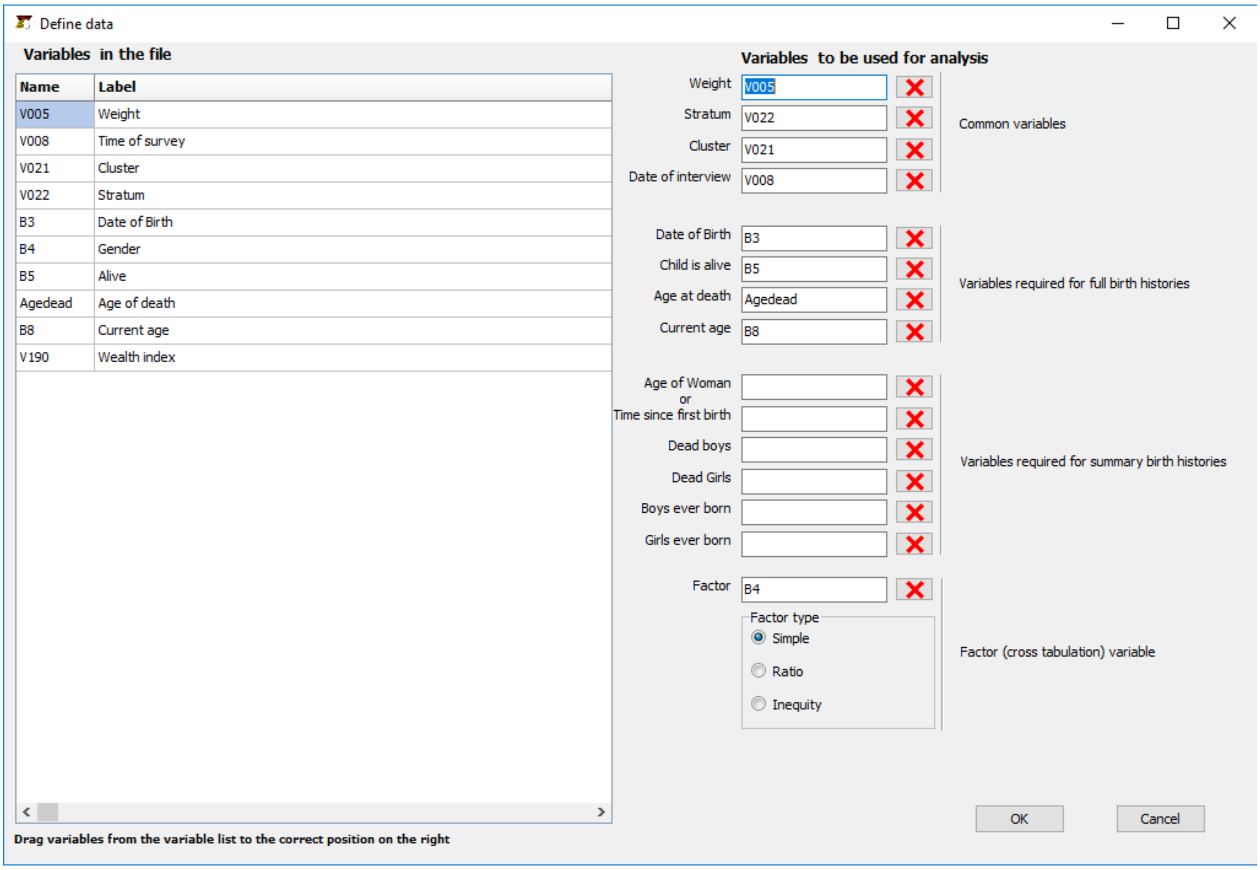
The group called "Common variables" consists of variables needed to define the survey. Those are the sampling weight, the stratum, the cluster and the date of interview.

The content of the group labeled "variables required for full birth histories" should be self evident.

The "Variables required for summary birth histories" is similar. Only one of the variables "Age of woman" or "Time since first birth" is needed.

The factor (or cross tabulation) variable is the variable used for defining sub groups for analysis. The factor variable may be of one of three kinds:

- Simple - using this options will result in only sub group estimates (and the total) being produced
- Ratio - using this option will lead to computation of ratios based on the sub groups, such as gender ratios. The program assumes that the variable is dichotomous. (If it is not, then the highest and lowest value will be used for gender type ratios)
- Inequity - using this options leads to the computation of bivariate inequity measures.



When one has pressed OK the program will read the data file into memory and checks the contents. Errors and missing values will be reported.

Errors are data values that means that the case cannot be used for estimation. This includes all of the "common variables (see above). and the variables used to define the birth history.
Missings are missing values in factor variables. They do not necessarily mean that the value will be totally deleted from the analysis.

Run with errors

The quaintly named "Run with errors" menu item runs the analysis regardless of how many errors or missings that were found when the file was opened. This is mainly used for stillbirth analysis when gestational age is missing, and may also be used when one wants an initial impression from a data file that needs cleaning.

Save results

The effect of choosing save results depends on the options. There are seven types of output that can be selected:

- CMEInfo output are used to input data into the CMEInfo system, and are also a fairly readable format. It can also be useful for plotting data in Tableau or R.
- Raw output provides a compact list of results with one line per period, and all estimates and standard errors in columns. It can be used for plotting data in Excel or a similar program.

- Optimal output is the result of the optimization procedure. The format is somewhat similar to the CMEInfo one.
- CEBCS table output provides the raw table of children ever born and surviving/dead.
- Presentation output provides relatively readable tables (at least that is the intention) as well as graphs.
- Output to mortality age pattern analysis using the R Logquad5q0 package.

If output to Excel is selected then the output can either take the form of a single workbook with several sheets, or several files, each with one sheet. (If the program is run from SPSS the output will always be in one workbook). The Excel output can be shown by clicking the "Show output" button under the logging window after having saved.

One may also use the keyboard shortcut CTRL-S

Save and show output

The command is a combination of the "Show Results" command and "show output" button.

It can be activated with the keyboard combination CTRL-Alt+S

Print children ever born table

Save log file

Saves the log file, i.e. the report of what CMRJack has done that is displayed in the window to the right on the screen. Note that it is possible to add text to the log manually.

Printer settings

Calls up the printer settings dialog. This works like any other Windows printer settings dialog.

Exit

Exits the program

One may also use the keyboard shortcut CTRL-E

Recent files

Below the exit menu CMRJack maintains a list of data files that have been opened recently.
Selecting one of the files causes CMRJack to reread that file.
One can use keyboard shortcuts Alt-1, Alt-2, Alt-3 etc, where the number indicates the position in the list.
Most likely Alt-1 is the most useful, since it reopens the last file used

Run

This menu item simply starts the calculations. If the option "Run automatically after open" is ticked off (see Options), then the program will start calculations immediately after the opening of a file. Then it will also save the results, provided the save options have been set.

One may also use the keyboard shortcut CTRL-R

Details about running CMRJack can be found in:

Running CMRJack

Running CMRJack as a standard program

Running CMRjack from SPSS

Running CMRJack from the DOS command line, batch file, script, R or Python

Quality

The quality menu item is for generating data quality plots. It is active for full birth histories. The menu contains a submenu with the entries:

- Standard Lifetable IMR CMR match
- Distribution of age at death, Days
- Distribution of age at death, Months
- Sex ratio at birth
- Neonatal mortality vs infant mortality
- Early neonatal mortality vs neonatal mortality
- Infant mortality compared to other surveys
- Under-five mortality compared to other surveys

To see the plots you must have processed the data file.

When appropriate, comparative estimates from the IGMEs database, the human mortality database, the MICS surveys, and the DHS surveys can be added to the plots.
The characteristics of the plots can be controlled from the settings page. The prepared plots may be copied to the clipboard, saved in various formats, or printed.

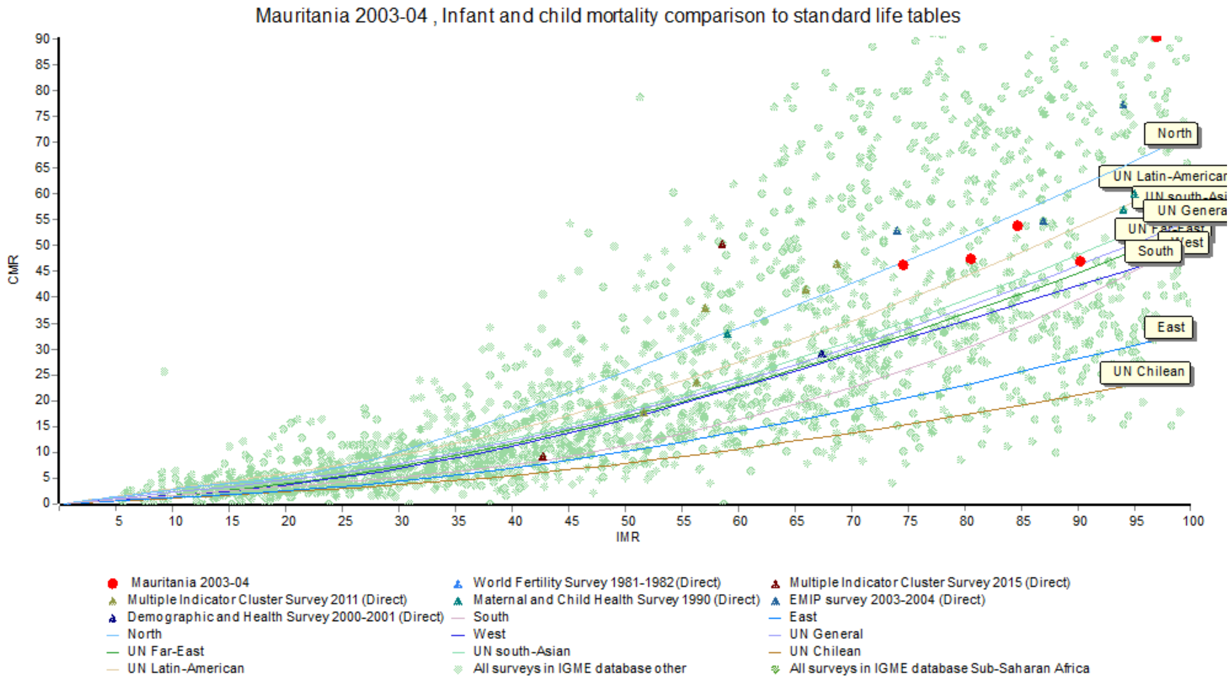
The window with the generated graph allows one to generate all the graph types.

Standard Lifetable IMR CMR match

The Standard Lifetable IMR CMR match produces a plot of the relationship between infant (1q0) on the x-axis and child mortality (4q1) on the y-axis. In addition to the observed values for the current survey, the lines for each of the UN model life tables and each of the Coale-Demeny (Princeton) life tables are shown. Ideally, one would expect the observed data points from the survey to follow one of the model life tables closely. However, in practice, this will often not be the case. The lack of correspondence with the model life tables can be observed if one plots the data available in Human Mortality Database (HMD) on the same plot (there is an option for this in Settings and on the plot page). The HMD data has a tendency to show higher child mortality than the model life tables. Similar observations can be made if one plots recent Demographic and Health survey data or the data available in the UN IGME database of sources.

In general, it is better to look for consistency, i.e. that the survey of interest has a smooth declining ratio of child to infant mortality as infant mortality declines. Erratic jumps across the model life table lines may be a sign of problems.

One should be aware that the statistical uncertainty on, especially child mortality, can be significant.

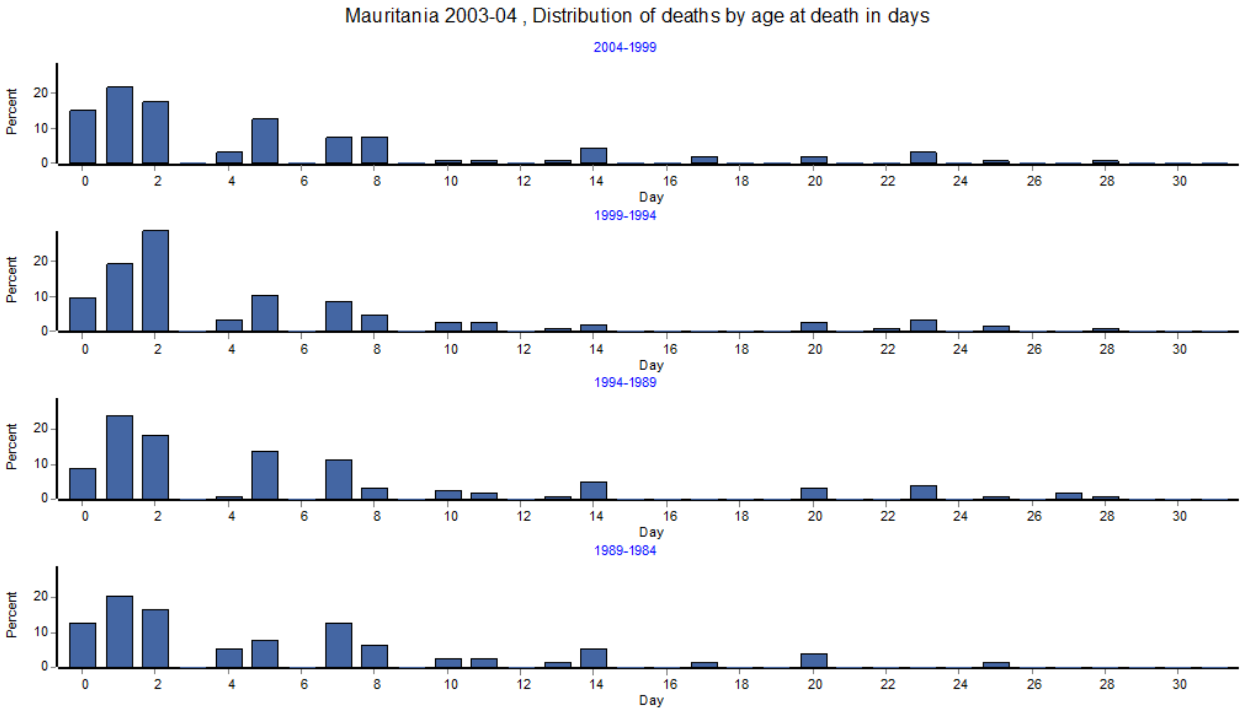


Distribution of age at death in days

The distribution of age at death in days shows this distribution by single days for day 0 to day 31 after birth for four periods before the survey. For most surveys, it is recommended to use five-year periods since, otherwise, the sampling uncertainty likely will be too large.

One may choose between absolute numbers or percentages. If one uses percentages, the percentage may be based on all deaths under five or only those occurring within the first 31 days.

One may use weighted or unweighted estimates. Weighted estimates would refer to the sampled finite population of child deaths, while unweighted estimates refer to the sample obtained.

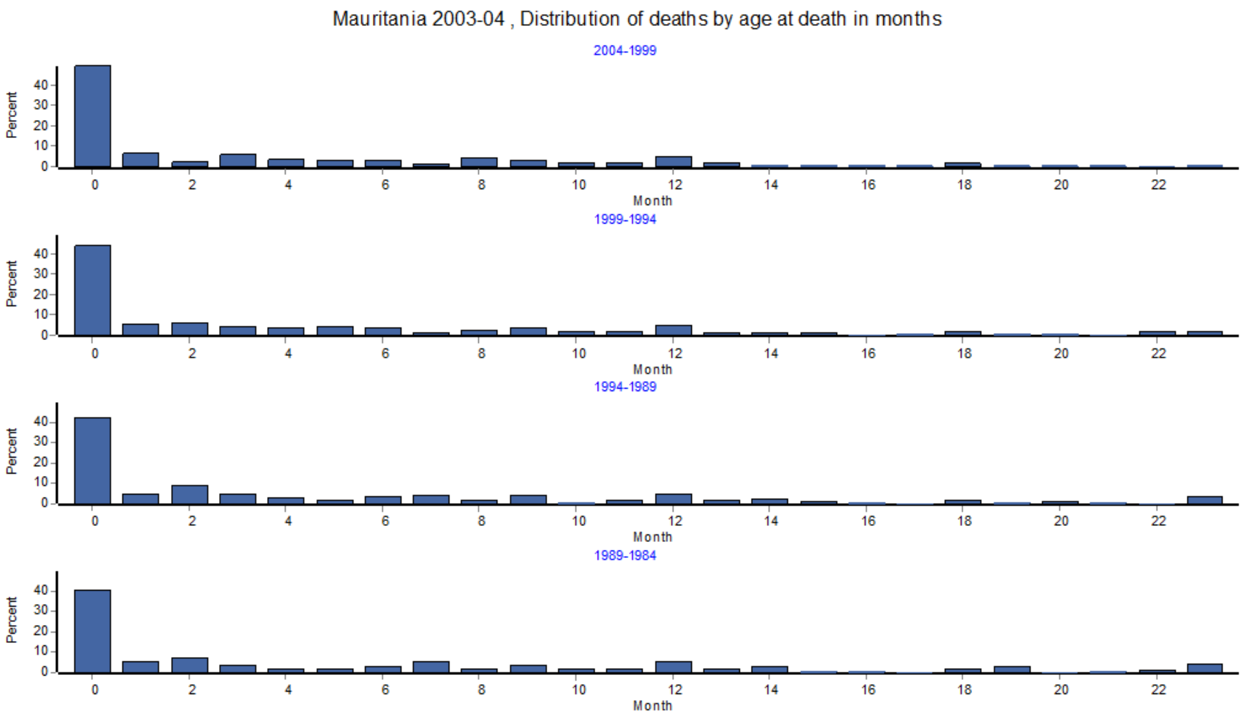


Distribution of age at death in months

The distribution of age at death in months shows this distribution by single months for month 0 to month 23 after birth for four periods before the survey. For most surveys, it is recommended to use five-year periods since, otherwise, the sampling uncertainty likely will be too large.

One may choose between absolute numbers or percentages. If one uses percentages, the percentage may be based on all deaths under five or only those occurring within the first two years.

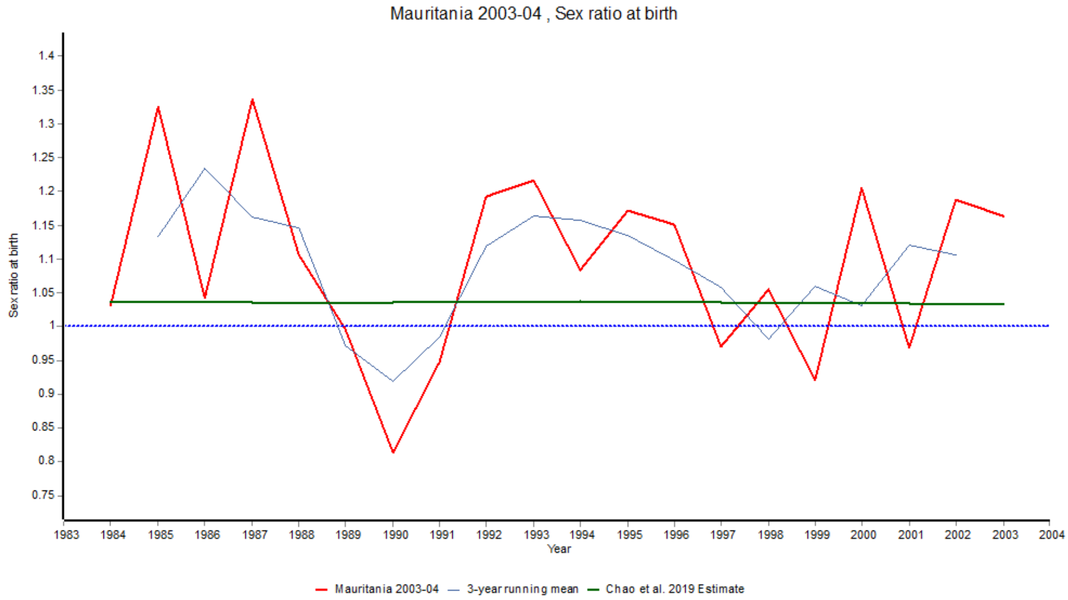
One may use weighted or unweighted estimates. Weighted estimates would refer to the sampled finite population of child deaths, while unweighted estimates refer to the sample obtained.



Sex ratio at birth

The sex ratio at birth graph plots the sex ratio at birth for each of 20 years before the survey. The sex ratio at birth (male births/female births) typically varies between 1.03 to 1.07 in the absence of sex selective abortion. (see Chao et al. 2019a,b)

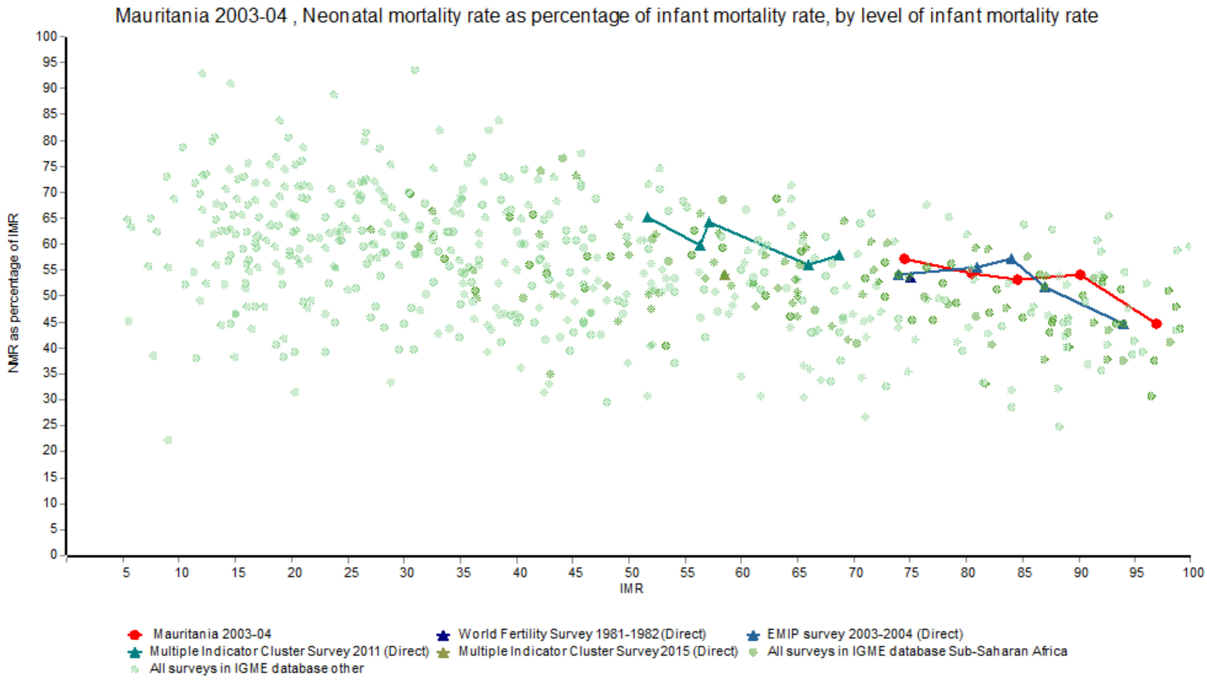
The sex-ratio plot may include other surveys from the same country and the estimated sex ratio from Chao et al 2019.



The sex ratio at birth is single-year estimates. Therefore, the sex ratios may be relatively variable. A running mean with adjustable window and a loess smooth may also be included in the plot.

Neonatal mortality rate as percenatge of infant mortality rate

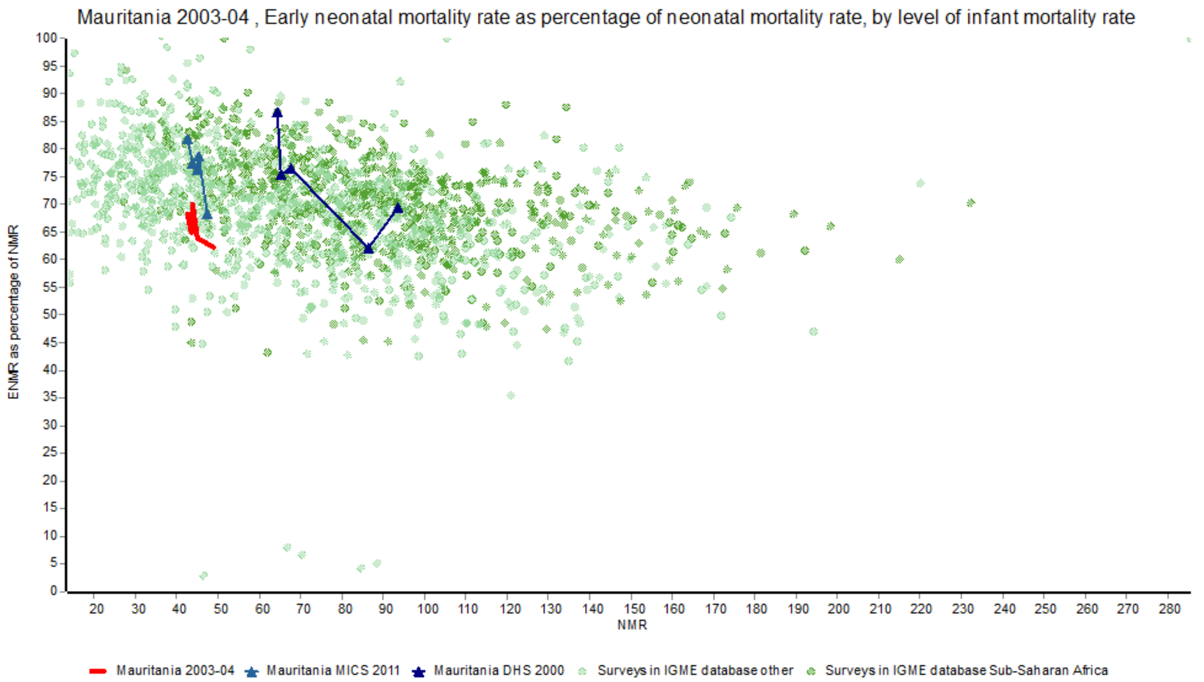
There is a tendency that the percentage of deaths occuring early increases as mortality decreases. Also, there is a tendency in surveys that early deaths are omitted. The graph provides the opportunity to gauge the realism of collected data on early mortality.



Early neonatal mortality rate as percentage of neonatal mortrality rate

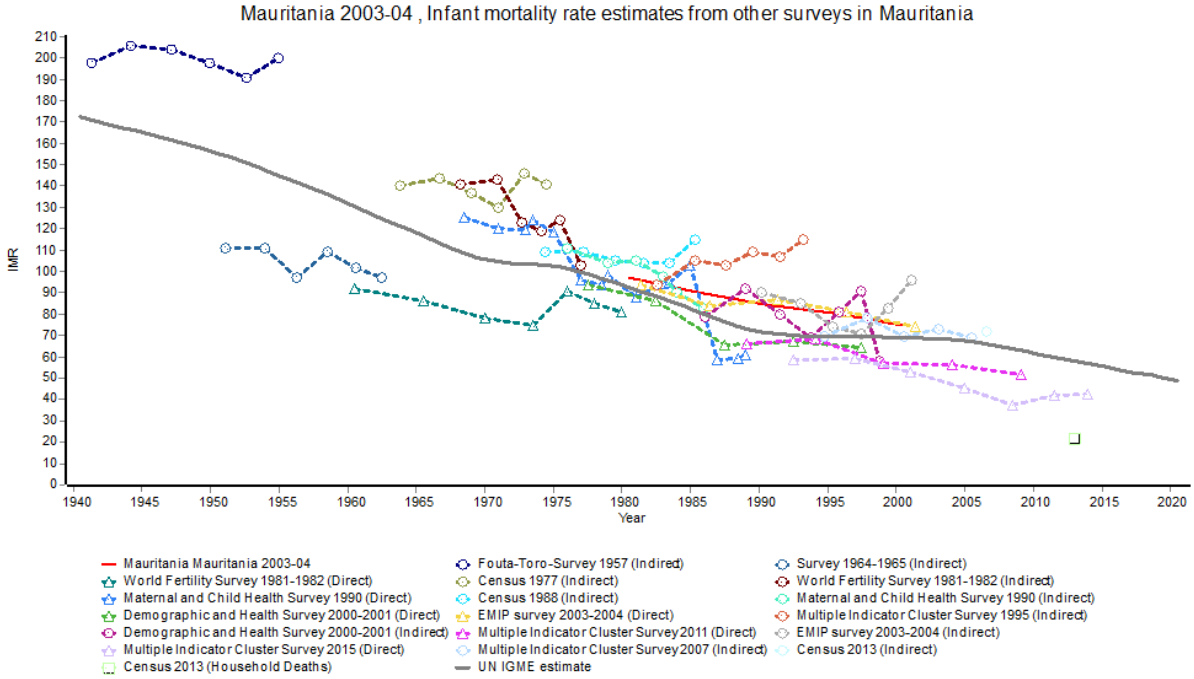
SImilar to neonatal mortality as a percentage of infant mortality one would expect the early neonatal mortality to increase relative to neonatal mortality as infant mortality decreases. The graph allows one to asesses if collecetd data are realistic.

Note that the comparative data may have some very high early neonatal mortality rates (som are in fact higher than the neonatal mortality rate). That is due to inconsistent definitions between the early and full neontal mortality rate. See the discussion in the section on neonatal mortality rate estimation



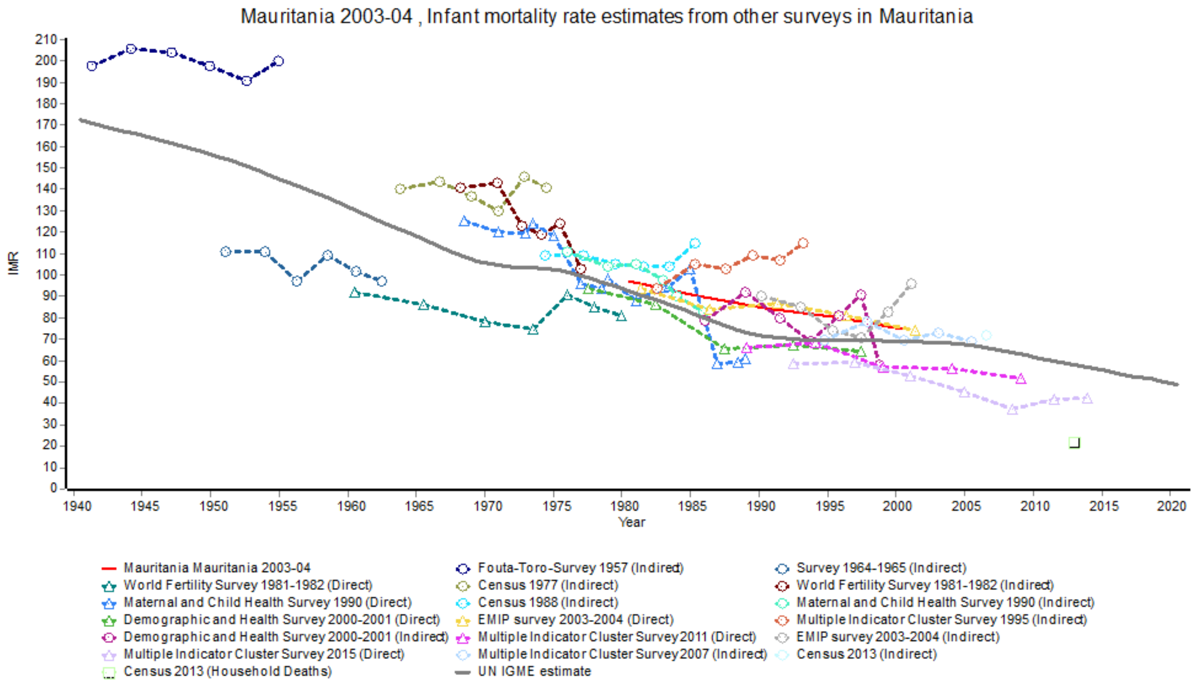
Infant mortality compared to other surveys

The graph is simply a graph of all available sources for the country in question. The graph also shows the most recent IGME estimate of infant mortality.



Under-five mortality compared to other surveys

The graph is simply a graph of all available sources for the country in question. The graph also shows the most recent IGME estimate of under-five mortality.



Settings

Use multiplesheets in one file rather than separate files for output

If the output is to Excel then all the output (i.e.e CME, raw etc) may be put in one Excel file. Alternative, one file may be created for each kind of output.

Calculation

Calculation methods

Method of calculation of direct estimates.

This can be either DHS or the IUSSP Tools for Demographic Estimation Method. See Method used for direct estimates

Include first year of survey period for surveys that span two years (calendar year estimation only).

This option allow you to use data from the first year of a survey that has field work that spans two years even when estimating for calendar years. (Normally the complete field work period is omitted from estimation). This option is meaningless for estimation based in years before the survey. Note that at present the reference date for the estimate is slightly wrong for calendar year estimates that include an incomplete year. It should be decreased by ((13-Month when survey started)/12)/2.

Note that the option may introduce some estimation bias because the last clusters completed in a survey are typically different from the other clusters. Standard errors may also increase.

Maximum period for optimal period.

This option can be used to set the longest period that the program will try to use. This is also the period that will be used if there is no period that can be found with sufficiently low value coefficient of variation. For example, if the value is set to 5, then the optimization routine will return the estimates for five year periods if no shorter can be found. The value must be between 2 and 10.

Type of scaling of estimates.

The estimates can either be on the real scale (i.e. not transformed in any way), log transformed or logit transformed. See Transformations

Stillbirth estimation: Produce Stillbirth estimates when data are available

CMRJack will detect whether or not stillbirth data are available.Using this control determines if stillbirth data will be produced. one can chose between:

- None
- Complete
- Summary only (no weekly estimates, but (i.e. only 22-27, 22+, 28+ and missing weeks of gestation, as well as stillbirth to neonatal ratios)
- Complete omitting termination and miscarriage
- Summary omitting termination and miscarriage

Child mortality estimates will be produced in any case.

Neonatal and early neonatal estimates

There are two choices for how neonatal estimates can be produced

- Cohort type day resolution
- DHS type estimates

See Neonatal mortality estimates for a discussion of the details of this

Use month as limit for neonatal estimates based in days

Ticking off this option will lead to calculation of the neonatal mortality rate with a numeratar derived from all deaths during the first month, i.e. more compatible with DHS. See Neonatal mortality estimates for a discussion of the details of this

Life tables

Choice of life tables for indirect measures.

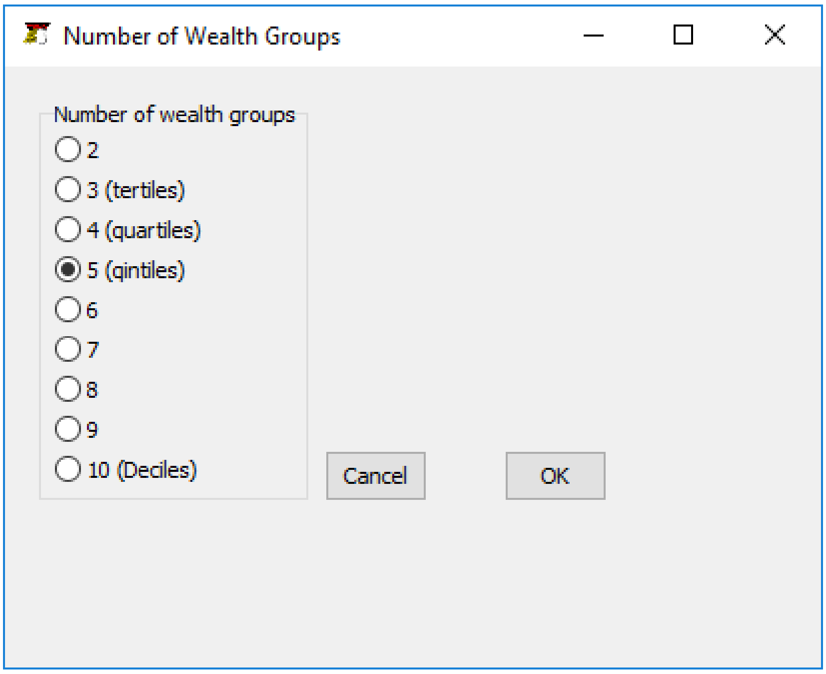
See Model life tables.
You can choose between the UN recalculation of the UN and Coale Demeny (Princeton) life tables, the old versions of the same and thye logit transformation.

Inequity

Number of constructed wealth groups.

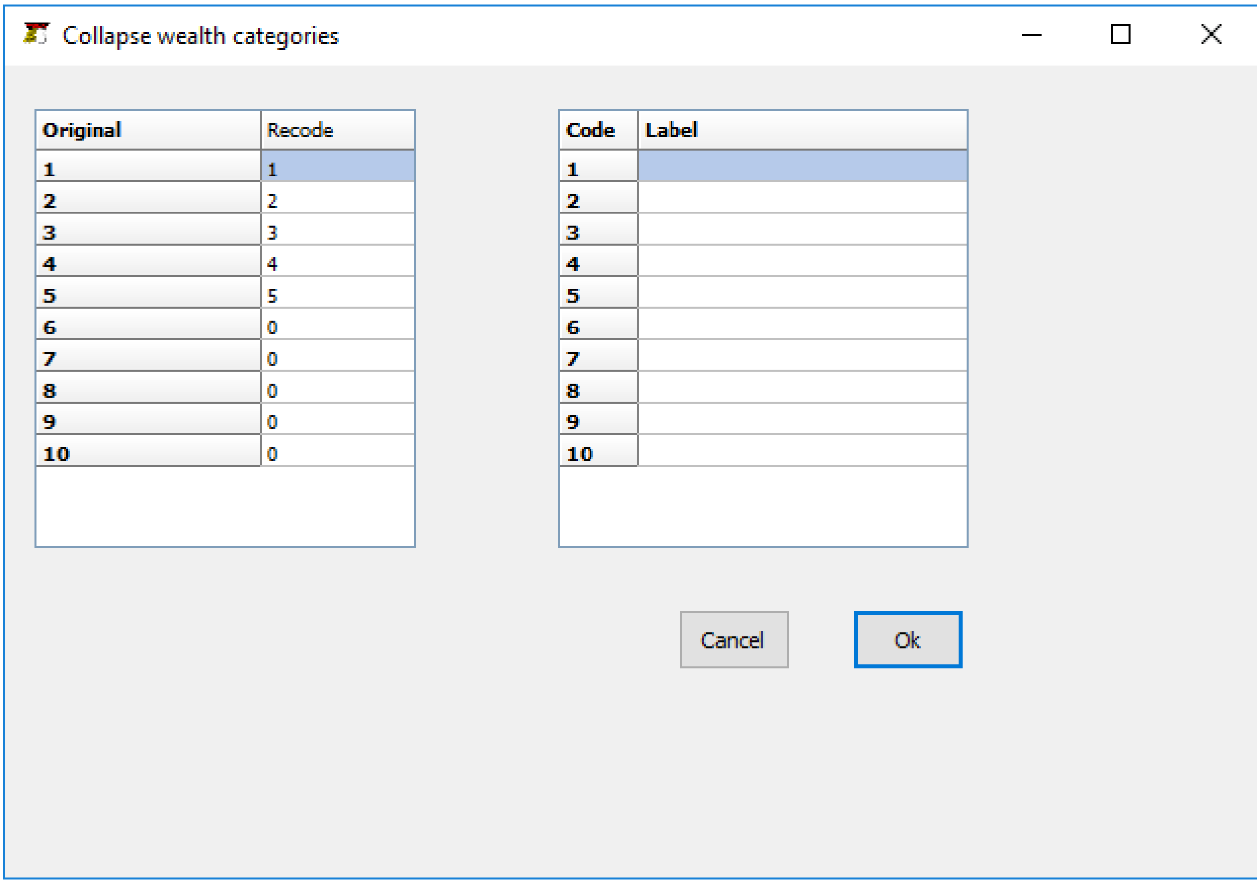
If the wealth index data are supplied as object scores (or other non-grouped form) the option controls how many groups the wealth ranking will have. Each group will have an equal number of weighted cases. See Weights and inequity measures

The choice will lead to the following window popping up.



Collapse wealth categories.

This options allows one to recode wealth categories on the fly. It works both for object score data and data that are already grouped.



As the picture illustrates, one fills out how the recoding should be done. Codes that are not involved in the recoding should be set to 0. An explanatory text can be given. (Fill in the label section)

Use 0-weight for 0-parity women in wealth ranking in summary birth histories

This option controls how 0-parity women are handled in the wealth ranking. See Weights and inequity measures

Use women as basis for weights in wealth ranking in summary birth histories

This option controls how 0-parity women are handled in the wealth ranking. See Weights and inequity measures

Run control

Run and automatically save after open.

This option makes it unnecessary to give an explicit run order to the software. This is mostly useful if one plans to do work with many files, and when one has decided which options to use.

Data

Location and format

Data folder.

This option is used to select a default location for data.

Default data format
The combo-box offers the options SPSS, Excel and CSV

Error handling

Delete cases with missing/bad data.

This option allows the user to do listwise deletion (i.e. the whole case is deleted if there is one missing or bad data point) of data records with missing data or data outside of accepted ranges. Note that the program will still stop at some types of illegal values, for example strata with only one cluster.

Maximum number of data errors accepted

This can be set to any number between 0 and 500, but will only be operative if the Delete cases with bad or missing data option is set.

Maximum number of data warnings accepted

This can be set to any number between 0 and 500, but will only be operative if the Delete cases with bad or missing data option is set.

Maximum number of data warnings/errors in logfile

This is simply the number of errors and warnings that will be reported in the logfile window. It is smart to keep the number low. most files do not have very many errors and large numbers of errors typically mean that you have chosen a vrong variable or a variable is improperly coded.

.

Output

ISO code options

ISO codes for country. The combo box can be used to choose what ISO code should be used in the output. In all cases, the codes are that of the ISO3661 standard. The possibilities are C2 (two characters), C3 (three characters) or M49 (three digits). The UN system typically uses the M49, while the character codes are perhaps more common for other applications.

The ISO code will be shown in the edit box to the left of the country identification on the main form. The ISO can be changed, but the change will only be reflected in files written after the change and before a new data file is read. For example, if the program does not find the code for Democratic Republic of the Congo (a likely occurrence because of the many possible ways the name of the country can be abbreviated in a file name), then one can edit the code field and enter the right code (i.e. CD, COD or 180). A change of the ISO code for a country will not be stored in the internal ISO-code database of the program.

Output file format

Choices are between XLSX, XLS, CSV and PDF (PDF is not currently working)

For Excel: Use multiple sheets in one file rather than several files

If this checkbox is ticked, then

Output produced

Presentation tables - tables intended for using more or less directly in reports etc

New style CME - format used for input to the CME-info system. The output is somewhat specialized, but contains some metadata. It also may be useful for reading data into R or Tableau for plotting.

Raw estimates - the estimates in a compact format. Mainly intended for direct input into another program. The data can be easily read by R, Stata, SPSS or nearly any other program that is supposed to process data.

Optimal estimates - the optimal estimates

CEBCS tables - children ever born / children surviving tables.

Documentation

If this option is checked, then a sheet with documentation about the data and processing options will be produced.

R-code for Lagrange 5q0 match

If this option is ticked, CMRJack will produce a text file with the needed R-code to run Lagrange 5q0 match. See Lagrange5q0 output to R

Graphs in presentation tables

Adds simple graphs to presentation tables in Excel.

Details

Only export optimized results to CME.

If this option is checked, only the optimized results will be saved in the new style CME output file, and not the estimates with fixed periods.

Include five years sex ratios in optimized output

If this option is checked sex ratios for five-year periods are output even if the “Only export optimized results to CME”-option is checked.

Output inequity measures for U5MR only

The program calculates bivariate inequity measures for most measures. However, their standard errors might be quite large, so the results may be useless.

Export infant mortality rates from indirect measures

The option allows the inclusion of infant mortality rates from the indirect measures. They are constructed by matching from the under five rate using the appropriate life table.

Generate mortality rates beyond age 5

Does not work at the moment

Export results from all life table families

One can either only output the default family or all the life table families.

Export 25-34 year average for summary birth histories

This option exports the average of the rates for women aged 25-25 years and women aged 30-34 years for indirect estimates based on classifcation by the age of women, or 10-19 years since first birth (i.e. average of 10-14 and 15-19) for classification by time since first birth. In both cases, the estimate is simply the average of the two rates. No weighting is performed.

Invert period order in graphs relative to table order

The periods in presentation output tables are sorted so that the most recent period is first. This may look strange in graphs because one is used to reading time scales from left to right.

Quality

General

The general options control the characteristics that are common to most of the plots produced.

Default interval to use for comparison.

CMRJack produces estimates for periods of 1,2,3,4,5 or 10 years. Any of these can be used in the plots, provided it has been generated. If the default period has not been generated CMRJack will set the period to the longest available.

Note that most of the comparative data uses five year periods. If several periods are available, then comparative data will match the default period.

Comparative data are available for the life-table match, sex-ratio at birth, Neonatal vs IMR, Early Neonatal/Neonatal vs IMR, and IMR, U5MR plots. The latter two shows only data from the same country.

Show other surveys from the same country

The option allows you to add comparative data from the country being analysed. Note that if you are analysing old data the current survey may be shown two times.

Show DHS data

These are the most recent five year periods before the survey. The data have been downloaded from the DHS Statcompiler.

Show HMD data

Shows the Human Mortality Database data available (five-year periods)

Show IGME-data

Shows the most recent UN-Interagency group for child mortality estimation data.

Mark common region for comparative data

When comparative data are used the plot may distinghuish between data points from a neighboring region to the country of convern and all other countries.

Region classification to use

CMRJack can use several regions for comparison:

Use titles for charts

Titles may be included on the quality charts. That may be usefulk for some purposes, but not for others. The checkbox gives the option of including them or not.

Life table match

Automatic setting of IMR for life table match

The checkbox controls the maximum infant mortality rate on the x-axis. One can choose in the Maximum IMR box if it is not ticked off. The reason for having the option is that including some of the comparative data may lead to very high IMR data points being included, thus making it difficult to view the data of interest.

Lagrange $_5q_0$

The options here pertain to how the export of R-code for the log-quad life table match occurs. See [Lagrange $_5q_0$ output to R].

Lagrange $_5q_0$ match value controls if the match is forced to be consistent with U5MR or IMR.

The Lagrange $_5q_0$ weight controls whether or not weights will be used and how. The options are:

No weights
Proportional to length of period
Proportional to inverse variance
Proportional to inverse standard error

It is, at the time of writing, not clear what the best option is. All of the options except for no weights will give more emphasis toward the end of the cumulative series of mortality estimates.

Age of death distribution and sex ratios

Use weights

If ticked, the age of death distribution graphs and the sex ratio graphs will used weighted data. Using weighted data can be generalized to the finite population from which the sample has been drawn, while the unweighted results will be descriptive of the sample.

Given the sampling methods commonly used in demographic surveys one would not expect large differences between the weighted and unweighted charts, except in countries with both
a) unequal geographic allocation of the sample, and b) large differences in mortality by geographic area.

Sex ratio at birth plot

Since the sex ratio is defined as male births divided by female births and thus is a relative measure, it is appropriate to use a log scale. (see Hosseinpoor and AbouZahr 2010) However, many people prefer real scale. Consider a situation with 200 male births for every 100 females. That results in a sex ratio of 2. If 200 females are born for every 100 males, that is a sex ratio of 0.5. On a real scale graph, the perceived difference from an even sex ratio (1.0) will be twice as long when the sex ratio is 2 as when it is 0.5. On a graph where the y-axis is on the log scale, the distance between equality and either sex ratio will be the same.

Smoothing of sex ratios

The sex ratios are calculated for single years. Thus, they may be variable and a better impression may be had by smoothing the series. CMRJack gives you the option of smoothing by simply drawing the average, using a running mean with a window of your choice, or a loess smooth with span and degree.
0 or 1 in the running mean box means that a running mean will not be produced. 2 or higher will produce running means. Odd numbers will centre the running mean on each value, while even numbers will fix the right edge of the window to each value. Thus, for the year 2015 and a window of 3, the smoothed value for 2015 will be calculated as $(y_{2014} + y_{2015} + y_{2016})/3$, while with a window of 4 and the smoothed value for 2015 will be calculated as $(y_{2014} + y_{2013} + y_{2014} + y_{2015})/4$
For loess smooth, a higher span smooths more, while a higher degree smooths less.

Distribution of age at death

Plots of age at death can be created based in months or days. In both cases, one may present the figures in percent or absolute numbers.
If percentages are chosen, one may base the percentages in all under-five deaths or limit the base to the total under one month or 0 to 23 months.

The chart type can be either a line chart or a bar-chart

Help

The help menu has four entries:

About - which provide copyright and version info
Contents - Which shows this help file with the table of contents active
Index - which shows this help file with the keyword search active
Search - which shows this help file with the general search active

Keyboard shortcuts

The following shortcuts can be used instead of the mouse in the menus:
CTRL-O: Open file

CTRL-R: Run
CRTL-S: Save results in Excel file
CTRL-E: Exit, leave the program

Alt-1..5: Open files in the recent files list.

Other buttons

Clear log

This button clears the CMRJack operation log. It does not delete any log that has been saved.

Show output

Pushing this button loads the output that just has been produced by CMRJack. It requires Excel to work.

In some cases Excel cannot open the file and responds with an error message. See "Excel cannot find the output file"

Choice of model family for visible output

Estimates for all available model families are generated when summary birth history analysis are run. However, only the default model family is shown on the screen. The default is defined in the MLT.CSV file for many countries. If another model family is wanted, then one can choose that with the drop-down selector.

Note that this only pertains to the model family within a set of life tables. If one wants to run the analysis with, say, logit life table matching rather than using the direct match to the UN 2011 set of life tables one has to rerun the analysis (see Model life tables.).

Quality graphs

CMRJack can produce several types of charts related to the quality of the data. The charges include the following;

1. Comparison between the Infant mortality rate and the child mortality rate in the data and model life tables.
2. The distribution of age of death In days and months.
3. The sex ratio at birth.
4. Neonatal mortality versus infant mortality.
5. Early in the unnatural mortality versus neonatal mortality.
6. Infant mortality compared to other surveys.
7. Under-five mortality compared to other surveys.

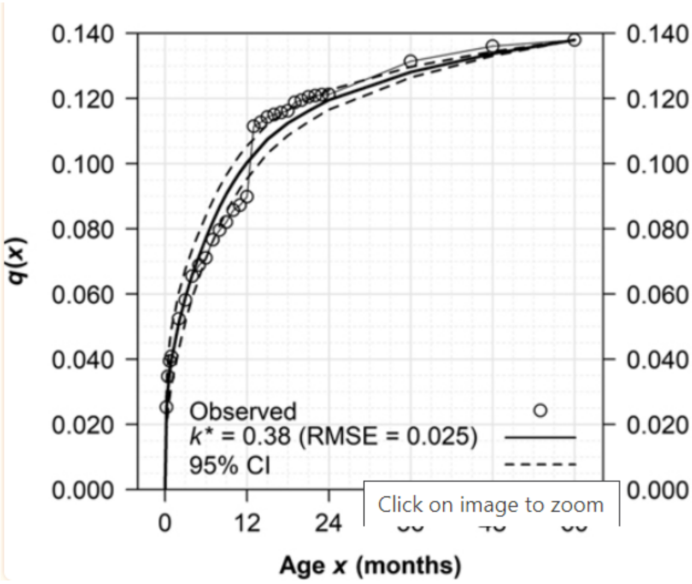
Lagrange5q0 output to R

CMRJack can produce R-code that can be used to test how well the observed mortality pattern generated by CMRJack from the micro-data fits with the log-quadratic model of under-5 mortality described by Guillot et al. 2022. The model is based on the log-quadratic relationship between the cumulative probabilities of dying at given ages and the under-five mortality rate.

$$ln({}_xq_0) = a_x + b_x + ln({}_5q_0) + ln({}_5q_0)^2 + v_x k$$

Here, the various coefficients have been estimated with training data from countries with good data (principally OECD countries) and the factor k is principally a shape parameter that controls how early or late the mortality pattern is in relation to the average pattern observed in the training data. The observed 5q0

determines the overall level.
The model can be used to indicate if a given observed age pattern is reasonable. The below graph is from the Bolivia 1989 DHS.



source: Guillot et al. 2022

It shows a smooth progression of xq0 from 01 to 11 months, a clear jump, and then a smooth progression. The pattern likely indicates age heaping at age 12 months, and it also indicates what a likely infant mortality rate is.
Since the current model is based on OECD countries, one cannot expect the model to work well in countries that do not fit their age patterns of mortality. Sub-Saharan Africa and south Asia are good examples (see Verhulst et al. 2022).

Use
When the output is asked for (see settings) CMRJack will generate R-code that can be used to create graphs of the match between the log-quadratic model and the observed data. It is able to use factor variables for the plots.
In order to use the code, it has to be loaded into R. In addition, one must have installed the modelling software into R. Instructions for doing that can be found in the code generated.

Transformations

CMRJack may transform the estimates into their logged form, logit form or may let them remain on the real scale. If they are transformed, then standard errors will be calculated on the transformed measure.

Both direct and indirect estimates may be scaled.

The following transformations are used:

Transformation	Formula
real (no transformation)	g=q
log	g=ln(q)
logit	g=ln(q/(1-q))

where q is the measure on the real scale measure, and g is the transformed measure

The above pertains to simple mortality rates. Ratios are scaled depending on the ratio in question. There are no difference between the logged and logit option for ratios. The log transformation is done using natural logarithms.

Ratio	Real Calculation	Logged/logit calculation
Sex ratios		
Neonatal	Male/Female	ln(Male/Female)
1q0	Male/Female	ln(Male/Female)
5q0	Male/Female	ln(Male/Female)
4q1	Male/Female	ln(Male/Female)
Early Neonatal Mortality	Male/Female	ln(Male/Female)
Late Neonatal Mortality	Male/Female	ln(Male/Female)
Other ratios		
Neonatal to 1q0	Neonatal/1q0	ln(Neonatal/(1q0-Neonatal))
Neonatal to 5q0	Neonatal/5q0	ln(Neonatal/(5q0-Neonatal))
Early Neonatal to Neonatal	ENMR/NMR	ln(ENMR/(NMR-ENMR))
Infant to 5q0	IMR/U5MR	Ln(IMR/(U5MR-IMR))

Method used for direct estimates

CMRJack estimates child mortality from full birth histories for time periods spanning from close to the survey to approximately 25 years before the survey.

The method used to produce the direct child mortality estimates may be either

- DHS
- Tools for Demographic Estimation

DHS method

The DHS method is a synthetic cohort method, using data with a month resolution (since DHS traditionally records only the month and year for birth dates and dates of death).
It calculates central death rates (M in life table terms) for the age groups 0-1 months, 1-3 months, 3-6 months, 6-12 months and yearly for 1 year to 5 years.
The central death rates are used as probabilities of dying, and the probabilities are linked to obtain the desired mortality rate. Generally, and somewhat simplified, children that live only partly through the period are counted by half of their weight.
See http://dhsprogram.com/publications/publication-DHSG1-DHS-Questionnaires-and-Manuals.cfm

Tools for Demographic Estimation method

The Tools for Demographic Estimation method uses a day resolution for dates and durations. This is the responsibility of the user to provide. In the case of the DHS, one must impute the values.*
It calculates central death rates (M) by month up to age five, and then by year. The death rates are then converted to q (probability of dying in the interval) assuming constant number of deaths over the interval. Then the child mortality rates for the ages of interest are constructed by linking the probabilities as in the DHS method. See http://demographicestimation.iussp.org/content/direct-estimation-child-mortality-birth-histories

Assumptions

There are relatively few assumptions associated with the direct estimation from complete birth histories. The most important one is that the mortality of children should be uncorrelated with that of their mothers. The most important situations where this may be an issue are areas of high HIV-infection rates, and areas of conflict.

It is also important that there usually is an age limit for interviews, i.e. women 50 years and above are usually not interviewed. Therefore, the farther back in time one tries to estimate child mortality, the population of women reporting on the deaths and survival of their children were progressively younger. For example, if the time period is 15-19 years before the surveys, the women reporting will have been 34 years or less at the time the deaths actually occurred. This may or may not lead to a bias, depending on the age pattern of fertility.

Apart from the assumptions inherent in the method, one should also be aware of data quality issues. Thus, reporting of deaths and births should have the same accuracy, and this should also be the case for sub groups, such as gender.

Dates of births and deaths should also have sufficient accuracy. Traditionally DHS and similar surveys have only collected data on births and deaths with month resolution (except for age at death less than one month) but increasing literacy and birth registration now allows for collecting dates directly in many countries.

Key choices for direct estimation

There are three basic choices that must be made for the estimation:

1. Variance estimation method;
2. Type of time reference (calendar years vs years before the survey),
3. Length of estimation period.

The variance estimation method: One can choose between the one employed in the DHS-reports that disregards stratification, and another method that takes stratification into account. The latter method is recommended (see below on variance estimation).

The type of time period used. One can again choose between DHS-style estimates and other estimates. The DHS-style estimates are based in years before the survey. Thus the time reference is the interview of each individual woman. For a survey where the field work period is 6 months, a “0-4 years before survey” is in reality a 0-5.5 years before the survey.
In contrast, the calendar year type of estimation, which is the other choice, uses the calendar years as basis. The UN Inter Agency Group for Child Mortality Estimation uses calendar years as standard.

The length of the estimation period. CMRJack can produce estimates for 1,2,3,4,5, and 10 years duration. It can also produce so called "Optimal periods", that is periods where the program itself chooses the time period based on the coefficient of variation. See Optimal estimates

Neonatal mortality estimates

Neonatal mortality is mortality that occurs during the first 28 days after birth. Thus, if the day of birth is day zero, any death occurring before day 28 counts. Early neonatal mortality measures mortality during the first seven days after Late neonatal mortality comprises deaths that occur between 7 and 28 days after birth. (https://www.who.int/data/gho/indicator-metadata-registry/imr-details/67) Finally, post-neonatal mortality is mortality between 28 days and one year.
The early neonatal, late neonatal, neonatal, and post-neonatal mortality rates use a denominator of births, rather than persons surviving to a given age. See https://reproductive-health-journal.biomedcentral.com/articles/10.1186/s12978-020-01028-0 . Thus, only the neonatal and early neonatal mortality rates are true probabilities of dying from one point in time to another, while the late neonatal mortality rate and the post-neonatal mortality rate are differences between probabilities.
While the definitions are clear, estimation practices do not follow the definitions strictly. DHS collects data on the age at death in days for deaths occurring during the first month after birth. However, the standard DHS tabulation discards the day-level information, and neonatal mortality is estimated based on deaths occurring during the first month following the birth, i.e., month zero. That leads to a slight overestimation of neonatal mortality since deaths on days 28,29,30 and 31 are also counted. Since the resolution of the age measurement is one month, DHS does not estimate early neonatal mortality.
When day-level information is available in the data file, one may estimate neonatal mortality based on the

formal definition and similarly estimate early neonatal mortality. How to do it raises some issues.

The lexis diagram below shows example life-lines of persons born in 2020, 2021 and 2022.

If the estimate of interest is neonatal mortality in 2021, one can see differences between the approaches and data used. For example, the person with the life-line marked D will be included in the DHS synthetic cohort and in a cohort approach with a one-month resolution of time, but not in a cohort approach with day resolution since one is then able to determine that the death occurred before a complete month but after 28 days. The life-line A would be included in the DHS approach, but not in the two other approaches. A summary of the differences is given in the table below.

Approach	Lifelines for dead (numerator)	Lifelines for births (denominator)
Cohort approach, day resolution (persons being born in 2021 who dies before reaching age 28 days:	C, E,F	Living: b,c, D Dead: C, E,F
Cohort approach, month resolution (persons being born in 2021 who dies before reaching age 1 month:	C,D,E,F	Living: b,c Dead: C,D,E,F
DHS synthethic cohort	½ A, ½ B, C, D, E, ½ F	½ a, b, c, ½ d ½ A, ½ B, C, D, E, ½ F

The DHS synthetic cohort method includes all births occurring in the shaded triangle indicated by births in the period month -1 to month 12, while the strict cohort approach only includes births between months 0 and 12. A consequence of the DHS approach is that deaths that occur before the period of interest are included in the estimate, although with a lower weight than deaths within the period of interest. This occurs because the resolution for age at
As one can gather from the table, there are some inconsistencies between formal definitions and typical DHS estimates.

For calculating the associated neonatal rates CMRJack has the option to use a cohort approach, i.e., calculate the rates in relation to the persons born during a time period. That is not entirely consistent with the overall synthetic cohort estimation strategy derived from DHS, but it is a) more accurate and b) easy to implement since the ages at death under consideration are so small.

However, the approach leads to some inconsistencies. The following table summarises the calculations.

	DHS		Cohort	
	Calculation	Example rates	Calculation	Example rates
Early neonatal	NA		(Deaths < day 7)/births	29.17
Late neonatal	NA		Neonatal-Early neonatal	13.52
Neonatal	Deaths under 1 month	43.39	(Deaths <day 28)/births	42.68
Post neonatal	IMR-Neonatal	31.5	IMR-Neonatal	31.87
IMR	1-((1-1q0)(1-2q1)(1-3q3)(1-6q6))	74.55	1-((1-1q0)(1-2q1)(1-3q3)(1-6q6))	74.55

(The example figures in the table is from Mauritania DHS 2000)

If one wants to have estimates of early and late neonatal mortality consistent with neonatal and post neonatal mortality rates, there are several options for more or less compatible with using the DHS estimation approach in general.

One is to estimate the three neonatal rates using day-level information and then calculate the post-neonatal rate. This will result in some minor differences in the neonatal rates from the DHS-type estimates, as shown in the table.

Another option is to keep the DHS neonatal rate as estimated, calculate the early neonatal rate as estimated in the cohort approach, and the late neonatal rate as the difference between the DHS-type neonatal rate and the cohort-type early neonatal rate. The neonatal and post-neonatal rate is then unaffected compared to DHS, but the neonatal rate uses a wrong age of death range.

CMRJack allows both estimation methods. The choice is made in the settings->calculation form. The first option is «Cohort type day resolution» while the second is «DHS type estimates». The way the estimates are arrived at in the two cases are shown below.

	DHS type estimates	Cohort-type day resolution
Early neonatal	Cohort	Cohort
Late neonatal	Neonatal-Early neonatal	Neonatal-Early neonatal
Neonatal	Neonatal (under 1 month, DHS)	Cohort
Post neonatal	IMR-Neonatal	IMR-Neonatal
IMR	Standard DHS	Standard DHS

It is also possible to include deaths between 28 days and the full month when using the cohort-type approach. That will bring the cohort type estimate closer to the DHS one, but otherwise, the utility of the option is not obvious.

Methods used for stillbirth estimates

Stillbirth rates are calculated as the ratio of stillbirths to the number of live births+stillbirths.

Stillbirth rates for given gestational ages are calculated as prospective stillbirth rates, i.e. as the ratio of the number stillbirths at a given gestational age to the number of live births plus stillbirths at that gestational age or greater.

A useful summary of definitions used in the literature to calculate stillbirth rates by gestational age can be found at https://evidencebasedbirth.com/studies-that-calculate-risk-of-stillbirth-by-gestational-age/

Methods used for indirect estimates

CMRJack can make use of summary birth histories, i.e. women's aggregate reports on children ever born, and children surviving
Estimates based on classification by age of women or time since last birth can be used (Hill, 2013b). When data are classified by age of women the program will produce estimates of ${}_5q_0$ employing indirect estimation techniques of either the Trussel version of the Brass indirect method using Coale-Demeny model life tables, or the Palloni-Heligman version based on the United Nations model life tables. The Time Since first birth method can only use the Coale-Demeny (Princeton) life tables.

The estimation of standard errors is the same as for the direct methods. See Variance estimation methods

Assumptions

The assumptions inherent in indirect estimation of child mortality from summary birth histories are much more stringent than for the direct estimates. This is due to the use of modeling as well as model life tables tables to obtain the estimate.

1. Mortality of children must be uncorrelated from that of their mothers
2. Reporting on children ever born and children surviving must be of similar accuracy
3. The actual age pattern of fertility and child mortality in the population must be sufficiently well represented by the models used.
4. Mortality of children should not vary by the classification of mothers by age or time since first birth for any time period.
5. Changes in child mortality in the past should have been gradual and unidirectional. Thus, indirect methods are not suited to estimates of child mortality due to sudden events.
6. The parities (average number of children ever born) observed in the survey by time since first birth or age of woman should correspond to the cohort parities.

Classification by age of women

The methods used for analysis of summary birth histories by age of women are described in Hill (2013a).

CMRJack allows for estimation based on the Coale-Demeny (Princeton) life table set and the UN life table set. In general the UN tables are rarely used. The default life table in CMRJack (which reflects the use of life tables in MICS and UN-IGME) is from the UN family in only two countries (Bolivia - UN General, Peru - UN-Latin American) out of 133 countries life tables are used for.

Classification by Time Since Last Birth

The methods used for analysis of summary birth histories by time since first birth are described in Hill (2013a). Note that the coefficients used in the description from 2013 is different from the original paper of Hill and Figueroa (2001).

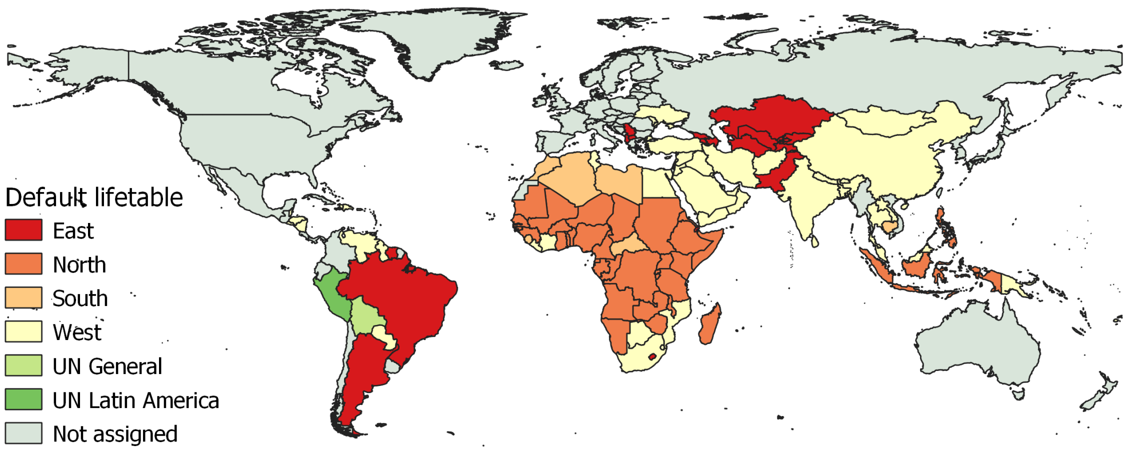
The time since first birth method has only been developed using the Coale-Demeny (Princeton) life tables.

Model life tables

Estimation from summary birth histories depend on the use of model life tables. The model life tables are used in two stages for estimation. First, the basic estimates are calculated using the appropriate coefficients specific to each model life table family. Second, the resulting estimates are matched to the model life table families in to obtain the final estimates

In order to use the indirect estimation method for summary birth histories a choice of life table family is necessary. The program includes a file “MLT.csv” that defines the default life table families for many countries of the world.

The default life table family has been chosen from the following sources in order of priority: a) The life table used for producing the MICS mortality estimates given in the relevant MICS report for a country, and used by the Inter Agency Group for Child Mortality Estimation (IGME). b) The life table used by UNPD c) The life table used by IGME for estimating ${}_1q_0$ from ${}_5q_0$. One should not assume that these sources are not necessarily consistent, although they are so for most cases. In some cases MICS mortality estimates have used different life tables for different surveys.



When there are large differences between CMRJack results and results from the same data for a country, one first should make sure that the differences are not due to the choice of model life table.

Logit life tables used for calculating $_1q_0$ and $_5q_0$ from indirect estimation are the the standard model life tables calculated by the UN population division (June 18, 2010 revision) downloaded from http://demographicestimation.iussp.org/content/relational-standards-used-project

Note that there are three versions of the standard life table families in relatively common use:

1. The classical sets of UN life tables and Coale Demeny. These life tables are the ones used in early versions of Mortpak, and published in Manual X and elsewhere.

2. The recalculated life tables of the UN population Devision, which includes both the UN Tables and the Coale Demeny tables. The one used by for example MICS (in the version 4 and 5 tabulations) is the 2011, version 1.3. These are used in Mortpak 4.3.

3. Logit approximation to the UN2011 recalculation is used in the [Tools for Demographic Estimation] recommendations, but the calculation is from a slightly earlier version than 1.3.

There are some differences in the results from the different versions of the tables, but seldom larger than about 1/1000. One should note that the classical sets (in particular Coale-Demeny) will not give estimates for low mortality levels.

CMRJack allows one to choose between the different tables sets. The different versions of life table sets only make a difference when matching desired mortality measures from estimated measures, i.e. the second stage of the indirect estimation. While they in principle would make a difference for the first part, the required coefficients have not been calculated, but remains the original ones.

Variance estimation methods

CMRJack uses a basic stratified JKn Jackknife variance estimation, as is also used in Westat's WesVar software(Westat, 2002), SUDAAN(RTI, 2001) and as described by Wolter(1985). A good description can also be found in Valliant and Dever (2017). Lu and Lohr (2022) have some discussion on computation.

Jackknife variance estimation is one of several sampling reuse methods. The assumption is that the observed distribution in a survey is a close replica of the sampling distribution. Then, since variance can be thought of as a function of the different values one will observe for an estimator if one makes many samples from a population, one can sample from the observed values instead. In the case of the Jackknife this is done by omitting one sampling unit from the observed distribution, calculating the estimate, replacing the sampling unit into the distribution and omitting the following one, calculating a new estimate, and so on until one has made a number of estimates corresponding to the number of sampling units. Then the variance is calculated as the variance between the different estimates. In the case of complex surveys the sampling units used are clusters, rather than ultimate sampling units (such as births).

The stratified JKn procedure implements the variance estimate very similarly to the unstratified version, i.e. by deleting and replacing each ultimate cluster in turn. However, when a cluster is omitted, the weights of the other clusters in the stratum are adjusted by a factor of $m_h/(m_h-1)$, where m_h is the number of sampled clusters in the stratum. Then, the actual variance estimate is calculated as given below, and r is the estimate from the full sample; and r_i is one estimate; θ is the estimate from the full sample, and θ_{hj} is the estimate from each replicate.

$$w_{i(hc)} = \begin{cases} w_i & \text{if observation } i \text{ is not in stratum } h \\ 0 & \text{if observation } i \text{ is in psu } c \text{ of stratum } h \\ \frac{m_h}{m_h-1} w_i & \text{if observation } i \text{ is in stratum } h \text{ but not in psu } c \end{cases}$$

The variance is then

$$\hat{V}_{JKn} = \sum_{h=1}^{H} \frac{m_h - 1}{m_h} \sum_{j=1}^{m_h} \left(\hat{\theta}_{hj} - \hat{\theta}\right)^2$$

The estimator used in DHS reports does not take stratification into account, and also does not make the weight corrections. The formula is the following:

$$\hat{V}_{JKdhs} = \frac{1}{m(m-1)} \sum_{j=1}^{k} (\hat{r}_j - \hat{r})^2$$

$$r_j = mr - (m-1)r_{(j)}$$

where m counts the number of clusters, r(j) is the estimate from the j-th replicate

Neither the DHS version of the variance estimate, nor the stratified version includes a finite population correction. In general, the first stage sampling rates in DHS and similar surveys are too small for this to matter.

Period of estimation

For direct estimates, CMRJack can produce estimates for time periods of different durations. One can either choose between 1,2,3,4,5 or 10 year for each period, or let CMRjack choose varying durations depending on the standard errors of the estimates (so called optimal estimates). The interval can be chosen for both types of time periods (i.e. years before the survey and calendar years).

See Pedersen and Liu (2012) for more on estimation periods for full birth histories.

In the case of summary birth histories the time location is determined by the data, and is estimate as a point, rather than as an interval. Therefore there is no choice of interval for summary birth histories.

Optimal estimates

CMRjack can produce so called optimal estimates.

Optimal estimates are estimates where the series of mortality estimates for a survey is for periods as small as possible while retaining a relative standard error (often called coefficient of variation) smaller than 0.1 (or 10 percent).

The estimates are constructed by starting at the most recent period for a given survey and trying first with a period of one year. If it turns out that the relative standard error is larger than 10%, then the estimation is tried again with the period of two years and so one until the duration with a sufficiently low relative standard error is found. Then the same procedure is repeated for the second most recent period, followed by the third most recent period, and so on.

By default CMRJack will not try longer durations than five years, but this can be changed. See Options

See Pedersen and Liu (2012) for more on estimation periods.

Calendars

CMRJack assumes that dates are in Century Month Coding (CMC) using a Gregorian calendar. The base of the CMC is year 1900.

Some surveys use calendars that are not Gregorian ("Western"). This does not pose a serious problem if estimation based in years before the survey is used, since the dates, at least in DHS, are coded as if they were Gregorian, i.e. with the standard Century Month Coding, and there is an underlying assumption that the year is solar. However, if actual calendar years are used as a basis for the calculation, then the dates will be wrong if there are no corrections made to all dates. Dates in lunar calendars, e.g. the Islamic one, must be transformed to Gregorian style.

Surveys carried out in Ethiopia (but not Eritrea) use the Ethiopian calendar, and surveys from Nepal use the "Bikram Samwat", the official Nepalese calendar. The Thai MICS uses the Thai calendar.
Completely accurate translation of the Ethiopian and Nepalese calendars is not possible for DHS and MICS since these surveys use month resolution for dates rather than days, but approximate results can be obtained. (CMCs are also not entirely comparable, because the length of the months differs between the Ethiopian, Nepalese and Gregorian calendars.)

Nepalese calendar

Translation of dates from the Nepalese calendar to the Gregorian can be simply done by subtracting 56.7 years from the CMC code, or 680 months.

Ethiopian calendar

An approximate conversion from the Ethiopian calendar to the Gregorian is the following SPSS macro :

```
set mprint off.
define  ETHGREG  (INCMC=!TOKENS(1)/OUTCMC=!TOKENS(1)).

compute  #cmc=!INCMC.

* get an Ethiopian year and date.
compute #EY=trunc((#CMC - 1)/12) + 1900.
compute  #EM=#cmc-(#EY-1900)*12.
compute #ED=1.

* convert to Julian.
compute #joffset = 1723856.
* Coptic : compute joffset = 1824665.
compute #n = 30 *(#EM - 1) + #ED - 1.   /* #ED-1 is actually 0.*/
compute #jd = #joffset+365 + 365*(#EY - 1) + trunc(#EY/4) + #n.

* convert to year and month.
compute #Z = #JD+0.5.
compute  #W = trunc((#Z - 1867216.25)/36524.25).
compute  #X = trunc(#W/4).
compute  #A = #Z+1+#W-#X.
compute  #B = #A+1524.
compute  #C = trunc((#B-122.1)/365.25).
compute  #D = trunc(365.25*#C).
compute  #E = trunc((#B-#D)/30.6001).
compute  #F = trunc(30.6001*#E).
compute #day = #B-#D-#F.
```

```
compute  #Month = #E-1 .
if #Month> 12 #Month=#E-13.
compute  #Year = #C-4715.
if #month ge 3 #Year=#C-4716.
compute   !OUTCMC=12*(#Year-1900)+#Month.
exe.

!ENDDEFINE.

* Example (a bit dangerous, reads and writes to same variable).
*.
ETHGREG incmc=b3 outcmc=b3.
```

The macro is based on the code found at http://ethiopic.org/calendars (the Ethiopian to Julian conversion) and the Julian to Gregorian conversion is based on http://www.hermetic.ch/cal_stud/jdn.htm.

Thai calendar

CMCs in the Thai calendar can be converted to Gregorian CMCs by simply subtracting 543*12 from the Thai CMC value (since the start of the Thai calendar is the Buddhist Era, as defined in Thailand).

Persian (Afghani) calendar

```
*    ===================================================================.
* File name:  Persian.sps.
* Purpose:    Convert from Persian/Afghan (Shamsi) calendar to Gregorian.
* Author:     Jon Pedersen.
* Date:       23.2.2017 (gregorian).
* Based on:    Algorithms at http://members.casema.nl/couprie/calmath/persian/index.html.
* Notes:      This is a macro.
* Input:      iYear, iMonth and iDay: Persian/Afghan year, month and day.
* Output:     GYEAR, GMONTH, GDAY: Corresponding Gregorian values.
*======================================================================.

set mprint off.
define DateFromPersian( iYear =!TOKENS(1)
   /iMonth =!TOKENS(1)
   /iDay =!TOKENS(1)
   /GYEAR = !TOKENS(1)
   /GMONTH = !TOKENS(1)
   /GDAY=  !TOKENS(1) )

compute #PERSIAN_EPOCH = 1948321.
* The JDN of 1 Farvardin 1.
do if !iYEar ge 0.
   compute #epbase=!iYear - 474.
ELSE.
   compute #epbase=!iYear - 473.
end if.
compute #epyear = 474 + Mod(#epbase ,2820).
DO IF !iMonth <= 7.
     compute #mdays = (!iMonth - 1) * 31.
    * Note original algorithm uses CLNg(Imonth) in Vbasic. SHould not matter.
 ELSE.
     compute  #mdays = (!iMonth - 1) * 30 + 6.
End If.
compute #jdn = !iDay + #mdays + Trunc(((#epyear * 682) - 110) / 2816) +
           (#epyear - 1) * 365 + Trunc(#epbase / 2820) * 1029983 + (#PERSIAN_EPOCH - 1).
* then convert julian day to date.
compute  #l = #jdn + 68569.
compute  #n = trunc((4 * #l) / 146097).
compute  #l = #l - trunc((146097 * #n + 3) /4).
compute  #i = trunc((4000 * (#l + 1)) / 1461001).
compute  #l =#l - trunc((1461 *#i) / 4) + 31.
compute  #j = trunc((80 * #l)/2447).
compute  !Gday = #l - trunc((2447 * #j) / 80).
compute   #l = trunc(#j / 11).
```

```
compute   !Gmonth = #j + 2 - 12 * #l.
compute  !Gyear = 100 * (#n - 49) + #i + #l.
execute.
!enddefine.

* example.
DateFromPersian  iYear=AYEAR  iMonth=AMONTH  iDay=ADAY  GYEAR=GregY  GMONTH=GregMonth
GDAY=GregDay.
```

A Stata version of the same macro is the following:

```
program define DateFromPersian
args iYear iMonth iDay gy gm gd
tempvar EPOCH epbase epyear mdays jdn n l i j
generate `EPOCH' = 1948321.
generate `epbase'=`iYear' - 474 if `iYear'>=0
replace `epbase'=`iYear' - 473 if `iYear'<0
generate `epyear' = 474 + mod(`epbase' ,2820)
generate `mdays' = (`iMonth' - 1) * 31 if `iMonth'<=7
replace  `mdays' = (`iMonth' - 1) * 30 + 6 if `iMonth'>7
generate `jdn' = `iDay' + `mdays' + trunc(((`epyear' * 682) - 110) / 2816) +
(`epyear' - 1) * 365 + trunc(`epbase' / 2820) * 1029983 + `EPOCH' - 1
generate  `l' = `jdn' + 68569
generate  `n' = trunc((4 * `l') / 146097)
replace  `l' = `l' - trunc((146097 * `n' + 3) /4)
generate  `i' = trunc((4000 * (`l' + 1)) / 1461001)
replace  `l' =`l' - trunc((1461 *`i') / 4) + 31
generate  `j' = trunc((80 * `l')/2447)
generate  `gd' = `l' - trunc((2447 * `j') / 80)
replace   `l' = trunc(`j' / 11)
generate  `gm' = `j' + 2 - 12 * `l'
generate  `gy' = 100 * (`n' - 49) + `i' + `l'
end

// example
DateFromPersian pyear pmonth pday gyear gmonth gday
```

Here `pyear`, `pmonth` and `pday` are Persian date input values, while `gyear`, `gmonth` and `gday` are variables that will be created by the do-file.

Differences in results from other programs

Full birth histories

As far as we know, CMRJack gives the same results as the code used to produce mortality estimates using full birth histories from DHS surveys (provided the options are set so that they use the same time references).

Other programs may use different methods to estimate direct estimates, and while they will give similar results they will rarely be the same.

The variance estimates produced by DHS and the variance estimates produced by CMRJack may differ substantially, because CMRJack by default takes account of stratification, while DHS does not. While stratification often reduces variance, in some cases it increases the variance. A common case of increase is when the sample is disproportionally allocated to strata.

Summary birth histories

Some differences may be seen between CMRJack and the results of the SPSS code that is used to produce results from summary birth history data for the MICS surveys as well as from Mortpak. The same reasoning pertains to direct calculation of the estimates with a spreadsheet.

The difference relates to five issues.

- Missing values. The MICS code, for example, aggregates values independently, while CMRJack drops a case completely if data are missing. For surveys with a lot of missing data or some missing data and low fertility and mortality) the two different ways of treating missing values may make a difference. In the case of Mortpak it is up to the analyst to tabulate the data first, and then the treatment of missing values depends on that tabulation. See also Treatment of missing values
- Numerical accuracy. In general, the numerical accuracy differences between CMRJack, SPSS and Mortpak are irrelevant. In some cases, again especially in cases with low fertility and mortality, numerical accuracy may matter.
  SPSS and CMRjack also use higher precision in the input to the estimating equations than do Mortpak because the programs themselves construct the input tables.
  When estimates are computed using Mortpak from finished tables, then the numerical precision used in the tables matter.
- Time location estimates for the mortality estimates. CMRJack, the SPSS MICS code and Mortpak all use the same estimation equations. However, the initial time of the survey may be arrived out differently. CMRJack estimates the time of the survey as the average interview time in the dataset. In the SPSS MICS code the date of the survey is explicitly entered into the code, and in Mortpak the user also have to enter the year and month of the survey. If the user entered dates differ from what CMRJack calculates the resulting time locations of the estimates will differ. In nearly all cases the difference will not be larger than one or two months.

- Differences between life table approximations and exactly what life table is used (there are different versions around). Such differences are generally small.

- Different model life table family. This comes more in the category of user error than actual difference.

In all cases, *except for that of selection of a different model life table*, the differences are substantially irrelevant, but may lead to considerable uncertainty about what is right. However, treatment of missing values may also be relevant.

Strange output

Sometimes the program will come up with very strange estimates. Most often this will be standard errors that are very high, estimates that are 0, or 9999.

Too little data
Such occurrences are not really errors, but rather indications that there for some measures are too few data points. For example, this may easily happen if one tries to make wealth estimates by one year groups.

Sex ratios
Note that sex ratios are relatively unstable in most surveys. The instability is particularly noticeable the ${}_4q_1$ sex ratios that may be based on very few observed deaths.

Indirect estimates and low fertility and mortality
The indirect estimates from summary birth histories tend to become numerically unstable when there are few children in an age group. In particular, the number of children (and women) with high education in the 15-19 year age group of women is likely to be very low. This may lead to underflow or overflow in the computations. In such cases the program will assign an arbitrary high value to the measure in question, resulting in strange estimates and strange standard errors. The problem should be readily detectable, but a future version of the program will perhaps handle the error reporting more gracefully than at present.

Negative mortality
Note that in some cases the model used to create indirect estimates grouped by age of mother will estimate negative mortality. This is particularly the case for the 15-19 year group of mothers, and happens when the parity ratio between 15-19 and 20-24 is large relative to the parity ratio 20-24 and 25-29. This happens when there has been a recent fertility increase. The situation is reflected by the value of 9999 as the estimate, and an error message.

While the estimates for the other age groups may seem plausible, one should keep in mind that the reason for the pattern of parity ratios is changing fertility. Therefore, one of the key assumptions of the method may be violated, and the results should be interpreted with considerable caution.

While theoretically a similar situation as that described above for indirect estimation based on classification by age of mother also can occur for models based on time since first birth, it is less likely. That does not mean, however, that one can not just as easily get into situations where the assumptions are violated.

Troubleshooting

Below are some common ways that things may go wrong

Nothing happens when I save results

If you save a file and nothing is saved, then the most likely reason is that you have not provided the options for what kind of file to save. The options are found towards the end of the options list and comprise CME info output, optimal output, raw output, presentation, and output of children ever born children surviving tables. If none of those options are marked, there will be no output.

In principle CMRJack warns you if no output is chosen.

Troubleshooting SPSS setup

CMRjack does not start when called from SPSS and the message:

Warnings
2,The system cannot find the file specified

is displayed in the SPSS output.
This is caused by SPSS not finding CMRJack where it is supposed to. The error can most likely be fixed by running CMRJack once outside of SPSS (This results in the information about the location of CMRJack being recorded into the system registry)

SPSS does not recognise the CALLJACK command

If you write CALLJACK in the syntax editor of SPSS everything should function as with a normal command. That is, SPSS should recognise the command, and suggest likely syntax.

If SPSS does not recognize CALLJACK then the most likely reason is that the files CALLJACK.py and CALLJACK.xml is not located in the extension folder of SPSS.

The error message that is likely to show is:

```
Error # 1.  Command name: CALLJACK
The first word in the line is not recognized as an SPSS Statistics command.
```

```
Execution of this command stops.
```

Find the two files and copy them into the appropriate directory, for example: C:
\ProgramData\IBM\SPSS\Statistics\25\extensions
(This depends on the version of SPSS.)

The two CALLJACK-files may be found in the zip file that can be downloaded from the CMRJack website. (

Missing ratios or inequity

If inequity measures or ratios are missing from the output, check if you selected ratio or inequity calculation when you opened the file.

The CMRJack forms takes up too much space on the display

On some computers CMRJack's forms will be larger than the display.
CMRJAck is designed for a minimum display size of 1024x768 pixels, so if the number of pixels on your display is larger than that, you should be OK.
(Check Settings, System, Display Resolution in Windows 10)

However, displays that are scaled to more than 100 percent (for example the common 125 or 150 percent) may still face the problem that the form is larger than the screen even if the number of pixels on the screen is larger than the CMRJack design limit.

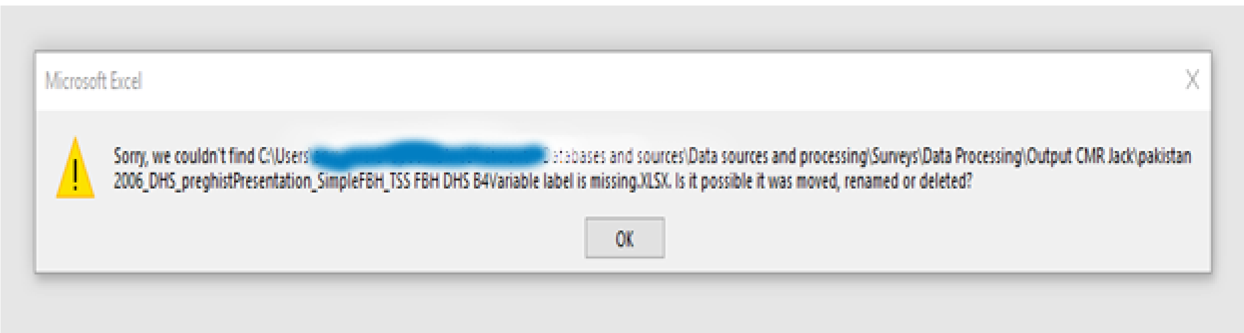
The way to resolve the issue is to set the display scaling to 100 percent (or some other value that is not too large).

On **Windows 10** this is set by choosing Settings, System, Display, Scale and Layout - Change the size of text, apps and other items.
The simplest solution is to set the scaling to 100%, but if you have a high resolution screen it is likely that you can use something larger.
(1024*scaling/100 must be less than your screen width, and 768*scaling/100 must be less than your screen height).

On **Windows 7** the way to change the scaling is to choose. Start Menu, Control Panel, Appearance and Personalization, Adjust Screen Resolution. Click on option saying "make text and other items larger and smaller" Set scaling to 100% (but see above)

Excel cannot find the output file

It may happen that when the Show output button is pressed Excel will start but indicate that it cannot find the file. The message will look something like this:



The error is caused by the path (folder description + file name) being longer than 260 characters. The limitation is unavoidable in Windows versions earlier than Windows 10, release 1067 (around May 2016) but can be changed with some difficulty in later versions.

The simple fix is to change the save folder to something that has a shorter path.

No administrative rights - cannot install CMRJack

The CMRJack install program requires that the Windows account you are installing from have administrative rights. This is in order to allow installation into the Program Files folder. If you do not have administrative rights either ask the person that has rights to install for you or cheat by downloading the zip file from the CMRJack website, and unzip it to a directory of your choice (for example Documents\CMRjack). Then run the program from there.

Cannot read Excel file exported from SPSS

In some cases CMRJack cannot read an Excel data file (xlsx). It will tell you with an obscure message like "Cannot find workbook in OPC".
The solution to the problem is to open the file in Excel, save it and then open it in CMRJack. This works because the reason for the problem is that some programs (such as SPSS, several versions) produce Excel files that are borderline compatible with Excel. Excel itself is quite good at reading files with problems, and will fix the problem when it saves.

Out of memory

If your data file is too large, CMRJack may run out of memory. It should tell you about it if that is the case. In general, the shorter period for estimation you use, the more memory will be used. It is difficult to run out of memory if you use five year periods for estimation. Optimal summaries also use a lot of memory, since they start with estimating for one year periods.
Many categories for factor variables also increase memory requirements.
The estimation of several periods at the same time does not increase the memory requirements much.
At the moment CMRJack can use in excess of 3GB of memory for computation and intermediate data storage. The absolute limit on a 64 bit windows system is 4 GB, and on a 32 bit windows system 2 GB. These limits include operating system needs.
If you have run many programs on your computer before starting a CMRJack session then you are more likely to run out of memory. This is because that even if the memory needed by CMRJack is available it may be fragmented, and CMRJack will then not be able to allocate the large chunks of memory it needs. If you have large data files it may be useful to restart your machine before running.
(The intricacies of memory management under windows is an arcane art that is unfortunately not taught at Galtworth or local wizard's colleges)

Note that out of memory errors may occur if the file is not sorted correctly according strata and clusters. If the incorrect sorting results in single cluster in a stratum then CMRJack will notice. However, if there still is at least two clusters in every stratum CMRJack will have no way of detecting the problem. In these cases CMRJack is also likely to run very slowly.

Only one cluster in stratum

The error message is something like "Only one cluster in stratum 151 (This may be because you have forgotten to sort the input file by stratum and cluster)"

The variance estimation method requires that there are at least two clusters in every stratum. Otherwise, the between cluster variance is not defined. If there is only one cluster you can merge the cluster into a similar stratum. See the description and references in the discussion of the sample description variables.

Sometimes the problem is simply that you have forgotten to sort the data file according to stratum and cluster.

If you run the analysis with the error you will like see standard errors that are just given as "INF".

In some data files there are more than one single cluster stratum. CMRJack typically warns only about the first.
One way to detect single cluster strata is the following SPSS code (stratum is typically V022 and cluster V021 in a DHS file):

```
dataset declare strata.
AGGREGATE outfile strata
```

```
  /break stratum cluster
  /households=N.
dataset activate strata.
compute single=0.
if lag(stratum) <> stratum single=1.
sort cases by stratum (D) cluster (D).
if lag(stratum)=stratum single=0.
sort cases by stratum cluster.
temporary.
select if single.
list stratum cluster.
```

Standard errors come out as "INF".

A likely reason for standard errors that show as "INF" is that you have forgotten to sort the data file by stratum and cluster, or have a stratum with only one cluster.
See the entry for "Only one cluster in stratum".

CMRJack runs very slowly

In general the running time of CMRJack depends on the number of strata and clusters present in the data file,
If CMRJack runs very slowly it may be because the file is sorted incorrectly. It should be sorted according to strata and cluster. If it is sorted according to something else, one of two things may happen:

1. CMRJack detects strata with only one cluster and stops.
2. CMRJack does not detect anything wrong. In that case the incorrect sorting may have constructed many more clusters than is actually the case. Then CMRJack will either crash, or finish very slowly with completely wrong standard errors.

Other file formats

As noted CMRJack only reads SPSS, comma separated and Excel files.

If you have data in Stata format, you can export them using the following code

```
* fix weight
replace  v005=v005/1000000
* Fix missing
replace b7=-99 if b7==.
replace b8=-99 if b8==.
* Sort according to strata and cluster
sort v022 v021
* Export data (v 12 and below)
outsheet v005 v008 v021 v022 b3 b4 b5 b7 b8 using
"myCountryMyYear.csv",delimiter(",") replace nolabel
* or use (from v 13)
export delimited v005 v008 v021 v022 b3 b4 b5 b7 b8 using
"myCountryMyYear.csv", delimiter(",") replace nolabel
```

Note that you will have to produce a label file manually. See The label file.

If you cannot afford SPSS, but want to draw upon variable transformations etc using SPSS and SPSS files a good option is to download the free PSPP. CMRjack reads SPSS-type files produced by PSPP.

The excellent program StatTransfer (which you have to pay for) can transfer nearly any data format used for statistics programs to nearly any other format.

Another possibility is to use the R-packages "haven" or "foreign" to read and write different file formats.

There is also the R-package "rio" that is able to read and write a large number of statistical file formats.

References

- Chao, Fengqing, Patrick Gerland, Alex R. Cook, and Leontine Alkema. 2019. 'Systematic Assessment of the Sex Ratio at Birth for All Countries and Estimation of National Imbalances and Regional Reference Levels'. *Proceedings of the National Academy of Sciences* 116 (19): 9303–11. https://doi.org/10.1073/pnas.1812593116.
- 'Correction for Chao et al., Systematic Assessment of the Sex Ratio at Birth for All Countries and Estimation of National Imbalances and Regional Reference Levels'. 2019. *Proceedings of the National Academy of Sciences* 116 (27): 13700–13700. https://doi.org/10.1073/pnas.1908359116.
- Guillot, Michel, Julio Romero Prieto, Andrea Verhulst, and Patrick Gerland. 2022. 'Modeling Age Patterns of Under-5 Mortality: Results From a Log-Quadratic Model Applied to High-Quality Vital Registration Data'. *Demography* 59 (1): 321–47. https://doi.org/10.1215/00703370-9709538.
- Heeringa, Steven, Brady T West, and Patricia A Berglund. 2010. *Applied Survey Data Analysis*. Boca Raton: Chapman and Hall/CRC. https://doi.org/10.1201/9781420080674.
- Hill, K, and M-E Figueroa. 2001. 'Child Mortality Estimation by Time since First Birth'. In *Brass Tacks: Essays in Medical Demography*, edited by B Zaba and J Blacker, 9–19. London: Athlone.
- Hill, Kenneth. 2013a. 'Direct Estimation of Child Mortality from Birth Histories'. In *Tools for Demographic Estimation*, edited by T.A. Moultrie, R Dorrington, A Hill, K Hill, I Timæus, and B Zaba, 2nd ed. Paris: IUSSP. http://demographicestimation.iussp.org/.
- ———. 2013b. 'Indirect Estimation of Child Mortality'. In *Tools for Demographic Estimation*, edited by T.A. Moultrie, R Dorrington, A Hill, K Hill, I Timæus, and B Zaba, 2nd ed. Paris: IUSSP. http://demographicestimation.iussp.org/.
- Hosseinpoor, Ahmad Reza, and Carla AbouZahr. 2010. 'Graphical Presentation of Relative Measures of Association'. *The Lancet* 375 (9722): 1254. https://doi.org/10.1016/S0140-6736(10)60541-7.
- Kish, Leslie. 1965. *Survey Sampling*. New York,: J. Wiley.
- Lu, Yan, and Sharon L. Lohr. 2021. *:: Design and Analysis, Third Edition*. New York: Chapman and Hall/CRC. https://doi.org/10.1201/9781003228196.
- Pedersen, J., and J. Liu. 2012. 'Child Mortality Estimation: Appropriate Time Periods for Child Mortality Estimates from Full Birth Histories'. *PLoS Med* 9 (8): e1001289. https://doi.org/10.1371/journal.pmed.1001289.
- RTI. 2001. *SUDAAN User's Manual, Release 8.0*. Research Triangle Park,NC: Reseaqrch Triangle Institute.
- Valliant, Richard, and Jill Dever. 2018. *Survey Weights: A Step-by-Step Guide to Calculation*. College Station: Stata Press.
- Verhulst, Andrea, Julio Romero Prieto, Nurul Alam, Hallie Eilerts-Spinelli, Daniel J. Erchick, Patrick Gerland, Joanne Katz, et al. 2022. 'Divergent Age Patterns of Under-5 Mortality in South Asia and Sub-Saharan Africa: A Modelling Study'. *The Lancet Global Health* 10 (11): e1566–74. https://doi.org/10.1016/S2214-109X(22)00337-0.
- Wagstaff, A. 2000. 'Socioeconomic Inequalities in Child Mortality: Comparisons across Nine Developing Countries.' *Bulletin of the World Health Organization* 78 (1): 19–29.
- Westat. 2002. *WesVar® 4.2 User's Guide*. Rockville: Westat.
- Wolter, Kirk M. 1985. *Introduction to Variance Estimatiom*. New York: Springer-Verlag.