

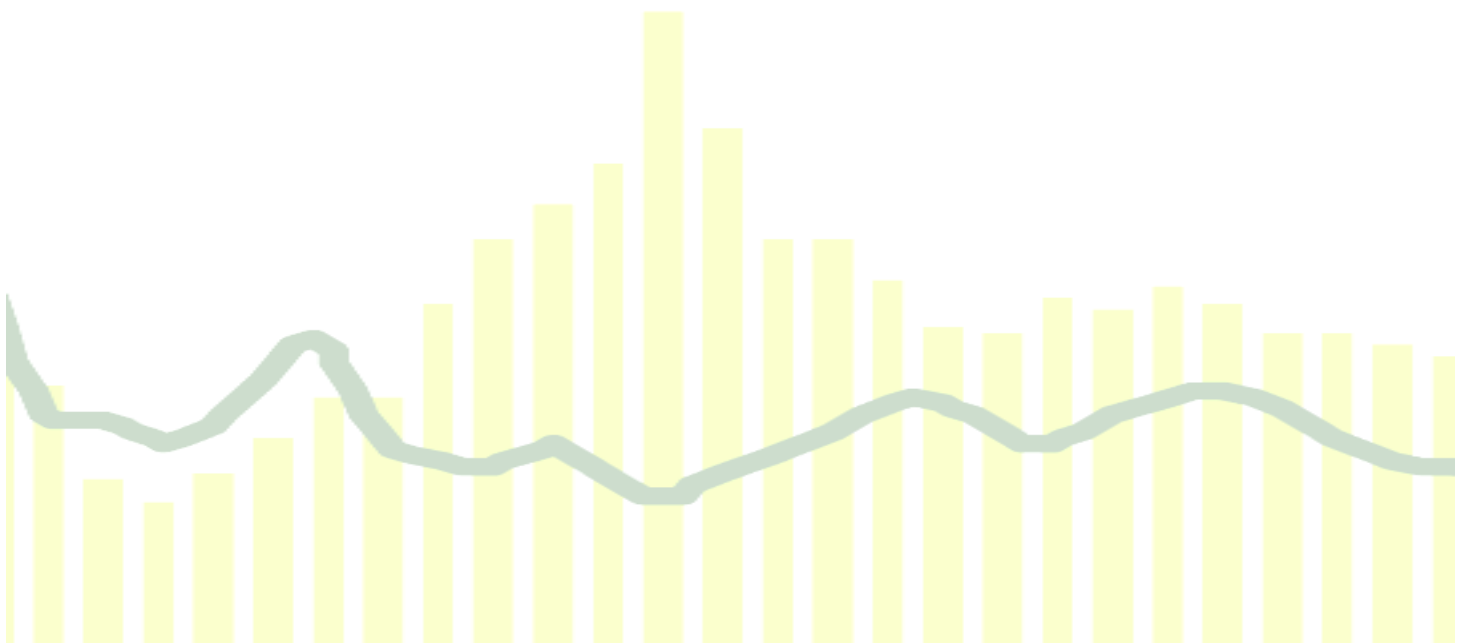


United Nations  
Department of Economic and Social Affairs

Population Division

Technical Paper  
No. 2019/2

## **R Scripts for computing adult and maternal mortality from DHS sibling survival histories**



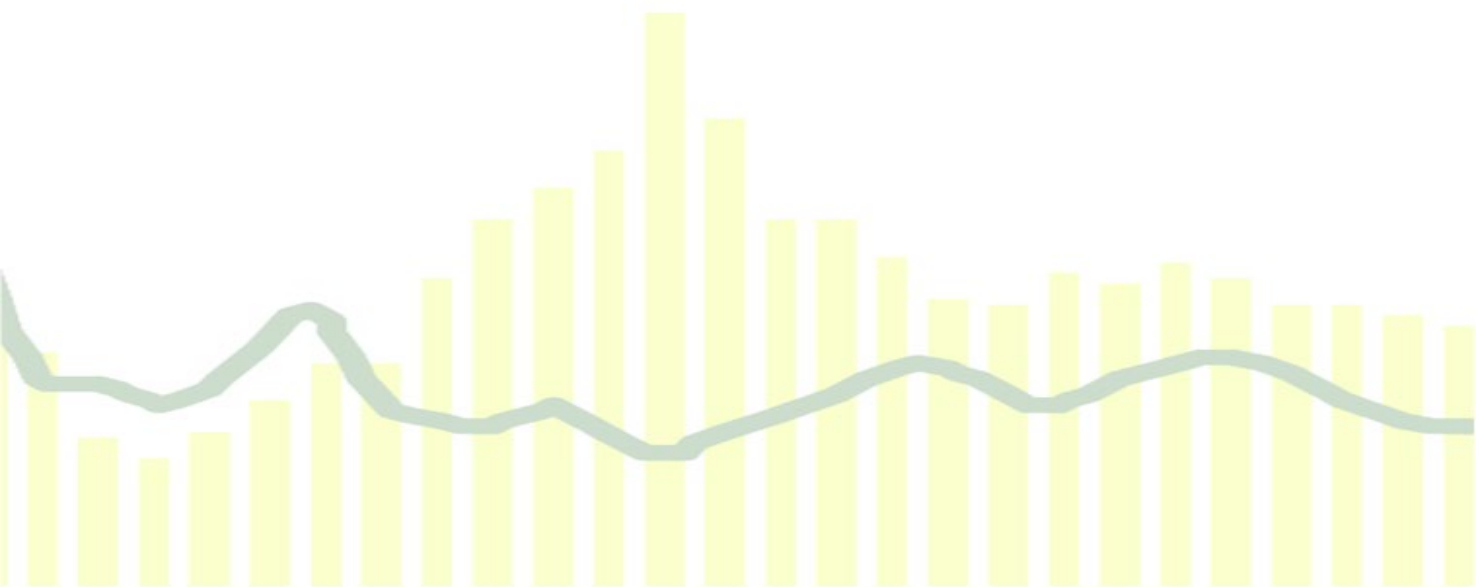
This page is intentionally left blank

Population Division

Technical Paper  
No. 2019/2

**R scripts for computing adult and maternal  
mortality from DHS sibling survival histories**

*Sara Hertog*



United Nations • New York, 2019

The Department of Economic and Social Affairs of the United Nations Secretariat is a vital interface between global policies in the economic, social and environmental spheres and national action. The Department works in three main interlinked areas: (i) it compiles, generates and analyses a wide range of economic, social and environmental data and information on which States Members of the United Nations draw to review common problems and take stock of policy options; (ii) it facilitates the negotiations of Member States in many intergovernmental bodies on joint courses of action to address ongoing or emerging global challenges; and (iii) it advises interested Governments on the ways and means of translating policy frameworks developed in United Nations conferences and summits into programmes at the country level and, through technical assistance, helps build national capacities.

The Population Division of the Department of Economic and Social Affairs provides the international community with timely and accessible population data and analysis of population trends and development outcomes for all countries and areas of the world. To this end, the Division undertakes regular studies of population size and characteristics and of all three components of population change (fertility, mortality and migration). Founded in 1946, the Population Division provides substantive support on population and development issues to the United Nations General Assembly, the Economic and Social Council and the Commission on Population and Development. It also leads or participates in various interagency coordination mechanisms of the United Nations system. The work of the Division also contributes to strengthening the capacity of Member States to monitor population trends and to address current and emerging population issues.

## Notes

The designations employed in this report and the material presented in it do not imply the expression of any opinions whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The term “country” as used in this paper also refers, as appropriate, to territories or areas.

This publication has been issued without formal editing. The views expressed in the paper do not imply the expression of any opinion on the part of the United Nations Secretariat.

This technical paper is available in electronic format on the Division’s website at [www.unpopulation.org](http://www.unpopulation.org). For further information about this technical paper, please contact the Population Division, Department of Economic and Social Affairs, Two United Nations Plaza, DC2-1950, New York, 10017, USA; phone: +1 212-963-3209; e-mail: [population@un.org](mailto:population@un.org).

### Suggested citation:

Hertog, Sara. (2019). R Scripts for computing adult and maternal mortality from DHS sibling survival histories. United Nations, Department of Economics and Social Affairs, Population Division, Technical Paper No. 2019/2.

Copyright © United Nations, 2019, made available under a Creative Commons license (CC BY 3.0 IGO) <http://creativecommons.org/licenses/by/3.0/igo/>

## PREFACE

The Population Division of the Department of Economic and Social Affairs provides the international community with timely and accessible population data and analysis of population trends and development outcomes for all countries and areas of the world. To this end, the Division undertakes regular studies of population size and characteristics and of all three components of population change (fertility, mortality and migration). Founded in 1946, the Population Division provides substantive support on population and development issues to the United Nations General Assembly, the Economic and Social Council and the Commission on Population and Development. It also leads or participates in various interagency coordination mechanisms of the United Nations system. The work of the Division also contributes to strengthening the capacity of Member States to monitor population trends and to address current and emerging population issues.

The purpose of the *Technical Paper* series is to publish substantive and methodological research on population issues carried out by experts within and outside the United Nations system. The series promotes scientific understanding of population issues among Governments, national and international organizations, research institutions and individuals engaged in social and economic planning, research and training.

This paper provides and describes a set of R scripts that compute adult and maternal mortality using the sibling survival histories collected in in Demographic and Health Surveys. This paper and R scripts were prepared by Sara Hertog and benefited from comments received from Patrick Gerland, Victor Gaigbe-Togbe and Thomas Spoorenberg.

The *Technical Paper* series as well as other population information can be accessed on the Population Division's website at [www.unpopulation.org](http://www.unpopulation.org). For further information concerning this publication, please contact the office of the Director, Population Division, Department of Economic and Social Affairs, United Nations, New York, 10017, USA, telephone +1 (212) 963-3209, fax +1 (212) 963-2147, email: [population@un.org](mailto:population@un.org).

This page is intentionally left blank

## CONTENTS

PREFACE.....	V
1. INTRODUCTION.....	1
2. THE SIBLING SURVIVAL METHOD IN BRIEF.....	2
3. SOFTWARE REQUIREMENT .....	2
4. REQUIRED DATA FROM DHS SIBLING SURVIVAL HISTORIES .....	2
5. CREATING A PERSON-PERIOD DATA FILE FROM DHS SIBLING SURVIVAL HISTORIES .....	4
6. ESTIMATING ADULT AND MATERNAL MORTALITY FROM THE PERSON-PERIOD DATA FILE .....	6
7. RECOMMENDATIONS FOR USE.....	9
APPENDIX: R SCRIPTS.....	10
REFERENCES .....	11

# R SCRIPTS FOR COMPUTING ADULT AND MATERNAL MORTALITY FROM DHS SIBLING SURVIVAL HISTORIES

*Sara Hertog*

## 1. INTRODUCTION

Estimates of the risks of mortality by age and sex are among the critical inputs that inform the demographic estimates and projections published in the *World Population Prospects* (WPP), which is widely considered to be the most authoritative resource describing trends in population around the world. Death rates computed from complete registration of vital events together with reliable population estimates are the preferred source of information about mortality risks by age and sex. However, for more than one third of countries included in the WPP, mostly in the less developed regions, the death registration systems are not yet able to capture all deaths or are deficient in accurately capturing the age of the deceased, thus alternative sources of information on mortality risks are sought.

Sample demographic surveys constitute one important alternative source of information on mortality risks. The Demographic and Health Surveys (DHS) is the world's largest programme of household demographic surveys, having begun in the 1980's and since collected data for more than 90 countries. The collection of data needed to estimate child mortality (the probability of dying between birth and age 5 years) is standard to the DHS and, over time, a growing number of these surveys have included modules that facilitate the estimation of mortality among adults as well.

A 2011 technical paper prepared by Trevor Croft, an expert consultant, for the United Nations Population Division (Croft, 2011), described the various approaches typically used to estimate mortality from major sample surveys and offered recommendations for specific computations and storage of mortality indicators. Regarding adult mortality, Croft identified the sibling history approach as the main direct estimation method from household surveys. In brief, this method uses information provided by adult respondents on the ages and survival status of all of the children born to their natural mother, as well as the years that have passed since death for deceased siblings. Croft described the data needs, assumptions, the calculation algorithm and offered suggestions for the selection of time periods, the use of survey weights, and strategies for data storage.

This paper details an implementation in R of Croft's recommendations for the direct estimation of adult and maternal mortality using the microdata of sibling survival histories collected in the DHS, through a series of R scripts developed explicitly for this purpose for the Population Division. In addition to estimating the mortality rates and probabilities of dying among men and women age 15 to 49 years as described by Croft, the scripts produce estimates of the maternal mortality rates and proportion maternal among deaths to females that are used to describe the mortality risks to women associated with pregnancy and childbirth. Moreover, the scripts support the computation of standard errors associated with each of the output adult and maternal mortality indicators.



## 2. THE SIBLING SURVIVAL METHOD IN BRIEF

The method explained by Croft is an elaboration of the recommendations provided in the DHS Guide to Statistics (Rutstein and Rojas, 2006, p. 130). In the full sibling survival history required by this approach, each adult respondent is questioned about the name, sex, age, survival status and, if dead, age at and year of death of each of their siblings born to the same mother. Because sibling histories in the DHS are designed to collect information for the estimation of maternal mortality, they additionally include information for female siblings as to whether the sibling was pregnant at the time of death, death was during childbirth, or within two months or 42 days of delivery. Presuming that the siblings of respondents currently living are representative of the adult population alive during the recent past, the information collected on the sibling histories can be used to estimate the deaths that occurred relative to the person-years of exposure to mortality in each age group, and thus to estimate the age-specific mortality risks faced by the reference population.

## 3. SOFTWARE REQUIREMENT

The scripts presented here are developed in version 3.6.1 of R, which is a United Nations-supported standard software application, used by the Population Division for various demographic computational tasks involving demographic estimates and projections, and data visualization. It has become a de-facto standard among statisticians for the development of statistical software, as well as for data analysis. It is a free, open source, software that can be obtained via the Comprehensive R Archive Network (CRAN) at <https://cran.r-project.org/>.

The scripts require the following R packages, which, upon first execution, will download from CRAN (requires internet connectivity) and install to a user's computer if they are not installed already:

“foreign” for reading Stata files into the R environment

“dplyr” for the tidyverse pipe operator and multiple functions called to manipulate the data

“tidyr” for the spread() and gather() functions used to transform the data from wide to long

## 4. REQUIRED DATA FROM DHS SIBLING SURVIVAL HISTORIES

The DHS collects reports of the survival of a respondent's siblings through the “maternal mortality module”. Female respondents' answers to the questions on this module are stored in the individual recode file (identified with “IR” in the file name) associated with each survey. Fewer than a dozen DHS to date administered the maternal mortality module to male respondents as well and these data are stored on the male recode file associated with each survey. The R scripts described in this paper rely only upon responses collected from females and ignore those collected from males.

For most DHS, microdata are publicly available through <https://dhsprogram.com/Data/> (registration required).

R scripts described in this paper utilize the microdata stored in Stata format and require the following standard variables that describe the respondent and her interview from the individual recode file:

```
v001: the cluster number associated with the survey sampling design.  
This field is necessary for the computation of standard errors  
associated with the output estimates  
v005: sample weight
```

v008: date respondent was interviewed, recorded by DHS in century-month code  
v011: respondent's date of birth, recorded by DHS in century-month code  
v015: indicator of whether interview was completed (completed=1). Only records associated with completed interviews are retained for estimation.  
v016: date of month of interview (integer from 1 to 31)

In addition, the scripts require the following variables that describe the respondent's siblings, also from the individual recode file:

mm1: sex of sibling (male=1; female=2)  
mm2: survival status of sibling (dead=0; alive=1; don't know=8; missing=NA)  
mm3: current age of sibling in integer years for siblings still alive  
mm4: sibling's date of birth imputed by DHS in century-month code  
mm6: number of integer years that have passed since death for deceased siblings  
mm7: sibling's age at death in integer years for deceased siblings  
mm8: sibling's date of death for deceased siblings imputed by DHS in century-month code  
mm9: pregnancy status at death for deceased sisters (various coding schema across surveys)  
mmidx: sibling number in sibship

Sibling histories are stored in the individual recode file in wide format, such that there is one record per DHS respondent and the sibling descriptor variables prefaced with "mm" described above are repeated for each sibling with an extension "\_xx" where xx uniquely identifies each of the respondent's siblings. To illustrate, the variable mm1\_01 describes the sex of each respondent's first reported sibling, mm2\_01 describes that sibling's survival status, and so on. The variable mm1\_02 then describes the sex of the respondent's second reported sibling and mm2\_02 that sibling's survival status. Most DHS capture information for up to 20 siblings of each respondent, thus the extensions on the names of the "mm" prefaced variables range from "\_01" to "\_20".

The R scripts developed here facilitate age standardization of output mortality rates if so desired by the user. This may be preferred if the age distribution of siblings reported by respondents differs substantially from the age distribution of the survey reference population. The procedure for age standardization requires that the microdata be stored in the DHS household member recode file (identified with "PR" in the file name) in Stata format. The following standard variables about each household member are required from the household member recode file:

hv005: sample weight  
hv104: sex of household member (male=1; female=2)  
hv105: current age of household member in integer years

If age standardization is not desired, then it is not necessary to download the household member recode file in order to execute the scripts.

The DHS program makes model datasets available without registration. These datasets are provided for practice and do not represent any country's actual data. The description of the R scripts and example implementation that follows uses the model individual recode file in Stata format, ZZIR62FL.DTA, as well as the model household member recode file in Stata format, ZZPR62FL.DTA. These models are based on the DHS6 round questionnaire. These files are available at the DHS website: [www.dhsprogram.com/data/Download-Model-Datasets.cfm](http://www.dhsprogram.com/data/Download-Model-Datasets.cfm) (accessed 2 October 2019).

## 5. CREATING A PERSON-PERIOD DATA FILE FROM DHS SIBLING SURVIVAL HISTORIES

The approach to direct estimation of adult and maternal mortality from DHS sibling survival histories presented here is broken down into two separate processes: 1) wrangling the data into the form required for analysis; and 2) performing the mortality estimation according to a set of options available to the user. This section describes the first step, which entails reading the individual recode file into R and then transforming it from the wide format with one record per respondent to a long format with one record per person (respondent and siblings) per five-year age group of exposure. The process is carried out through a function called `PrepForSiblingsDHS`. The function can be loaded into the R environment by executing the following line of script:

```
source("FUNCTION.PrepForSiblingsDHS.r")
```

The above line sources from the working directory the R file named `FUNCTION.PrepForSiblingsDHS.r`. Users can identify their working directory with the command `getwd()` and can change the working directory with `setwd()`. Alternatively, users can specify the full file path of the R script that contains the function to be sourced. For example:

```
source("C:/My folder/FUNCTION.PrepForSiblingsDHS.r")
```

To execute the `PrepForSiblingsDHS` function for a given survey, users must provide two pieces of information: the location of the individual recode file and the survey reference date in dd/mm/yyyy international format (i.e., dd=day, mm=month, yyyy=year).

```
df <- PrepForSiblingsDHS(ind.recode="ZZIR62FL.DTA",  
                        SurveyRefDate="01/07/2015")
```

The above example presumes that the individual recode file named `ZZIR62FL.DTA` is stored in the working directory. As with the source call above, users may choose to instead specify a file path. The survey reference date is the mid-point of the fieldwork period, which is described in the DHS report associated with each survey and available on the DHS program website.

The `PrepForSiblingsDHS` function carries out the following steps:

1. Reads the individual recode microdata into the R environment and retains only those variables required for the estimation of adult and maternal mortality described in section D above. If any of the required variables is missing from the input file, the function will return an error message listing the name(s) of the missing fields and it will fail to return a person-period dataset.
2. Checks if interview date is missing for any respondents and, if yes, then replaces missing values with the survey reference date.
3. Assigns random days of the month to the dates of birth and dates of death that DHS has imputed as century-month codes. This is done with a random uniform distribution of dates across 31-day months.
4. Checks for inconsistencies in date fields and adjusts where necessary:
  - a. If a respondent's date of birth is more than 50 years prior to the interview date, then that date of birth is adjusted to be exactly 50 years prior to the interview date (because only women 15-49 years are selected for the interview).
  - b. If a respondent's date of birth is less than 15 years prior to the interview date, then that date of birth is adjusted to be exactly 15 years prior to the interview date.

- c. If a sibling's date of death is later than the interview date, then that date of death is adjusted back to the interview date
- 5. Checks for missing dates of birth. If any, these records are removed from the data and a warning is issued indicating the number of records that were deleted.
- 6. Transforms the data such that there is one row per person per five-year period of each person's life.
- 7. Returns the person-period data frame with the following columns:

```

mother: an integer that uniquely identifies each sibling group in
the data
v001: cluster number
v005: sample weight
v008: respondent date of interview in century-month-day code
mm1: sex (male=1; female=2)
mm9: pregnancy status at death for women who died between ages 15
and 49 (various coding schema across surveys)
mmidx: integer that uniquely identifies sibling within sibship
(respondent = 0)
dob: date of birth in century-month-day code
dod: date of death in century-month-day code
age5: integer code associated with five-year age period (0 = 0-4
years; 1= 5-9 years; 2 = 10-14 years...)
age: factor description of five-year age period ("0-4", "5-9",
"10-14"...)
```

In the above example, the output data frame is called "df". Users may choose an alternative name for the data frame if so desired. The person-period data frame output by the PrepForSiblingDHS function is ready to be used in the estimation of adult and maternal mortality, described in the following section. A screen shot showing several lines of the data frame output by the PrepForSiblingsDHS function using the DHS6 model individual recode file is shown in figure 1.

**Figure 1. Sample showing several lines of data frame output produced by the PrepForSiblingDHS Function**

	mother	v001	v005	v008	mm1	mm9	mmidx	dob	dod	age5	age
190	6	1	1057703	1386.935	1	NA	2	840.0269	NA	7	35-39
191	6	1	1057703	1386.935	1	NA	2	840.0269	NA	8	40-44
192	6	1	1057703	1386.935	1	NA	2	840.0269	NA	9	45-49
193	6	1	1057703	1386.935	1	NA	3	871.6715	1356.6715	0	0-4
194	6	1	1057703	1386.935	1	NA	3	871.6715	1356.6715	1	5-9
195	6	1	1057703	1386.935	1	NA	3	871.6715	1356.6715	2	10-14
196	6	1	1057703	1386.935	1	NA	3	871.6715	1356.6715	3	15-19
197	6	1	1057703	1386.935	1	NA	3	871.6715	1356.6715	4	20-24
198	6	1	1057703	1386.935	1	NA	3	871.6715	1356.6715	5	25-29
199	6	1	1057703	1386.935	1	NA	3	871.6715	1356.6715	6	30-34
200	6	1	1057703	1386.935	1	NA	3	871.6715	1356.6715	7	35-39
201	6	1	1057703	1386.935	1	NA	3	871.6715	1356.6715	8	40-44
202	6	1	1057703	1386.935	2	NA	4	1020.8580	NA	0	0-4
203	6	1	1057703	1386.935	2	NA	4	1020.8580	NA	1	5-9
204	6	1	1057703	1386.935	2	NA	4	1020.8580	NA	2	10-14

Showing 201 to 218 of 248,247 entries, 11 total columns

## 6. ESTIMATING ADULT AND MATERNAL MORTALITY FROM THE PERSON-PERIOD DATA FILE

After transforming the microdata from the DHS sibling histories on the individual recode file into a person-period file, the second step is to carry out the estimation of adult and maternal mortality indicators. There are several decisions that must be made for this estimation, including with respect to the handling of missing data, the definition of the reference period for the estimates, and whether rates should be standardized to the age distribution of the survey reference population. A second R function, called `SiblingSurvival`, is developed to produce estimates of adult and maternal mortality according to users' specification on these decisions, among others. As for the `PrepForSiblingsDHS` function discussed in the previous section, the `SiblingSurvival` function can be loaded into the R environment using the `source()` command:

```
source("FUNCTION.SiblingSurvival.r")
```

To execute the `SiblingSurvival` function and generate estimates of adult and maternal mortality, users must provide several pieces of information to specify the treatment of the data and the types of output desired.

```
out <- SiblingSurvival(PreppedSibData = df,
                      include.Respondent = FALSE,
                      impute.Sex = FALSE,
                      PregRelCode = c(2, 3, 6),
                      recall.back = 7,
                      recall.end = 0,
                      stderror = FALSE,
                      age.standardize = TRUE,
                      pr.recode = "ZZPR62FL.DTA")
```

The above example takes as input the person-period file output by the `PrepForSiblingsDHS` function (through the parameter `PreppedSibData = df`). The remaining parameters operate as follows:

`include.Respondent` is a logical object (TRUE/FALSE) indicating whether the respondents' exposures should be included in the denominators for the estimation of mortality rates. Usually the exposures of respondents should be excluded from the calculation and are, by definition, excluded in the estimation of male mortality since all of the respondents are female. The potential biases associated with the exclusion of respondents as living siblings have been investigated and debated in the literature (Trussell and Rodriguez (1990), Gakidou and King (2006), Masquellier (2013)). While the `SiblingSurvival` function presented here defaults to excluding the exposures of respondents (via `include.Respondent = FALSE`), by changing this parameter to TRUE, users can explore the implications of this exclusion and make an alternative decision.

`impute.Sex` is a logical object (TRUE/FALSE) that instructs the function on how to handle records where information on the sex of the person is missing. The default selection is FALSE, which has the effect of excluding these records from the estimation. If changed to TRUE, records with missing sex will be imputed via a random binomial and a warning will be written to indicate for how many persons sex was imputed.

`PregRelCode` is a vector of numeric codes that identify whether the death of a woman aged 15-49 years was related to pregnancy. Most DHS of at least the last decade have used the code 2 to identify deaths that occurred while a woman was pregnant, the code 3 to identify deaths that occurred during

childbirth and the code 6 to identify deaths that occurred within a short period of time following the end of a pregnancy (typically, either two months or 42 days). The default setting for this parameter is thus `c(2, 3, 6)`.

The parameters `recall.back` and `recall.end` are used to define the reference period to be used in the estimation. They are expressed as years related to each respondent's interview date. Many DHS reports evaluate adult and maternal mortality over the seven years prior to the survey and the default parameters are defined thus, as `recall.back = 7` and `recall.end = 0`. While a seven-year lookback is typical for the reference period, users may wish to explore shorter reference time windows, in particular, for surveys with larger samples or higher mortality levels.

The `std.error` parameter is a logical object (TRUE/FALSE) that instructs the function on whether to calculate standard errors and include them in the output. If true, the function executes a `jackknife` procedure to calculate standard errors, re-estimating each of the mortality indicators while excluding from the data one cluster at a time and then describing the deviance of those estimates from the average. This approach is consistent with that used for other indicators estimated and reported by the DHS Program (Measure DHS, 1996).

The `age.standardize` parameter is a logical object (TRUE/FALSE) that instructs the function on whether to standardize the estimated mortality rates to the age distribution of the survey reference population. This may be desirable if the age distribution of reported siblings differs meaningfully from that of the survey reference population. This decision impacts the estimates only for the broad aggregate age group 15-49 years. Estimates for five-year age groups are not affected. If `age.standardize = TRUE`, then the user must also identify the location of the DHS household member recode file, which is used to describe the age distribution of the survey reference population.

`pr.recode` is a character field that identifies the DHS household member recode file. As with the individual recode file called by the `PrepForSiblingsDHS` function, this can be the name of a file stored in the working directory, or a file path that leads to a different location.

The `SiblingSurvival` function carries out the following steps:

1. Reads the data frame output by the `PrepForSiblingsDHS` function
2. If `impute.sex = TRUE`, looks for records with missing sex and imputes sex via a random binomial
3. For each person-period, identifies the deaths and years of exposure that fall within the recall period defined by `recall.back` and `recall.end`.
4. If `include.Respondent = FALSE`, removes records with `mmidx = 0` from the data.
5. If `age.standardize = TRUE`, reads the Stata format household member recode file into the R environment and calculates the proportional distribution of males and females aged 15-49 years by five-year age group.
6. For each five-year age group in the age range 15-49 years, the function estimates the following indicators:
  - a. the number of deaths by sex, weighted according to sampling weights ("death")
  - b. the number of years of exposure by sex, weighted according to sampling weights ("exposure")
  - c. the number of years of exposure by sex, unweighted ("exposure.unw")
  - d. the number of pregnancy related deaths to women, weighted according to sampling weights ("death.mat")
  - e. the number of years of exposure to maternal deaths, weighted according to sampling weights ("exposure.mat"). This value may

- differ from "exposure" for women because it excludes records for which the pregnancy-related death code is missing.
- f. the mortality rate  

$$m_x = \text{death/exposure}$$
  - g. the probability of dying  

$$q_x = (10*m_x)/(2+m_x) \quad (\text{from Croft (2011) pg. 27})$$
  - h. the maternal mortality rate  

$$m_x.\text{mat} = \text{death.mat} / \text{exposure.mat}$$
  - i. the proportion maternal among deaths to females  

$$\text{pmdf} = m_x.\text{mat} / m_x * 100$$
  - j. the natural log of the pmdf

For the aggregate age range 15-49 years, the function sums the deaths and exposures over five-year age groups and estimates the mortality rates and probabilities of dying (35q15) by sex, as well as the proportion maternal among deaths to females aged 15-49 years. If `age-standardize = TRUE`, the  $m_x$ ,  $m_x.\text{mat}$ , and `pmdf` for the 15-49 years age range are standardized to the age distribution of the survey reference population calculated in step 5 above.

The output of the `SiblingSurvival` function is compiled in a single data frame with the following columns:

`sex`: 1=male; 2=female

`Age`: the opening age of the age group (15, 20, 25, ..., 45)

`AgeInt`: the width of the age group (5 = five-year age group; for `Age = 15` and `AgeInt = 35`, refers to the full age range 15-49 years)

`Indicator`: the mortality indicator that is estimated (identified in a through j of step 6 described above)

`Estimate`: the value of the estimate corresponding to each sex, `Age`, `AgeInt` and `Indicator`

`SE`: the standard error associated with the estimated rates, probabilities and proportions. If `std.error = FALSE`, this column is omitted from the output data frame.

A screen shot showing several lines of the data frame output by the `SiblingSurvival` function using the DHS6 model individual and household member recode files is shown in figure 2.

**Figure 2. Sample showing several lines of data frame output produced by the SiblingSurvival Function**

	sex	Age	AgeInt	Indicator	Estimate	SE
57	1	15	5	mx	3.866703e-03	6.343953e-04
58	1	20	5	mx	4.920091e-03	7.103987e-04
59	1	25	5	mx	3.845495e-03	5.931677e-04
60	1	30	5	mx	5.665562e-03	8.220757e-04
61	1	35	5	mx	4.429736e-03	9.241472e-04
62	1	40	5	mx	8.929681e-03	1.662174e-03
63	1	45	5	mx	7.040469e-03	2.009225e-03
64	1	15	35	mx	5.164978e-03	3.793303e-04
65	2	15	5	mx	5.439424e-03	8.843954e-04
66	2	20	5	mx	6.718303e-03	7.423259e-04
67	2	25	5	mx	4.523053e-03	8.461382e-04
68	2	30	5	mx	5.177044e-03	7.289971e-04
69	2	35	5	mx	5.154927e-03	1.166011e-03
70	2	40	5	mx	6.111632e-03	1.268889e-03
71	2	45	5	mx	6.109199e-03	1.587687e-03

Showing 66 to 83 of 104 entries, 6 total columns

## 7. RECOMMENDATIONS FOR USE

The R scripts described in this paper are developed to facilitate estimation of adult and maternal mortality using the microdata from the sibling survival histories collected from adult female responses to DHS. They enable standardized processing of the more than 100 DHS that have collected the maternal mortality module across countries and over time. They also allow users, including those with limited proficiency in data manipulation or R programming, to produce adult and maternal mortality estimates and assess the impact of changing certain parameters such as, the time period covered by the estimates, the width of the reference window, the treatment of missing data, and whether to standardize to the age distribution of the survey reference population. In addition, if so desired, the scripts estimate standard errors associated with each of the estimated rates, probabilities and proportions. This information can then be used to assess the appropriateness of the reference period width, for example, or the degree of uncertainty surrounding each estimate.

The mortality estimates produced using these scripts provide valuable information on the risks of dying between the ages of 15 and 49 years in countries for which reliable vital statistics are not available. As such, they are an important input towards the assessment of mortality levels and age patterns for the *World Population Prospects*. Moreover, the Population Division is a member of the Maternal Mortality Estimation Inter-Agency Group (MMEIG), which is responsible for producing regularly updated estimates of maternal mortality risks around the world.<sup>1</sup> These scripts facilitate the computation of the proportion

<sup>1</sup> Trends in maternal mortality 2000 to 2017: Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO.



maternal among deaths to females, which are the key inputs to the models used by the MMEIG to estimate global, regional and country-level trends in maternal mortality.

#### APPENDIX: R SCRIPTS

1. `FUNCTION.PrepForSiblingsDHS.r`
2. `FUNCTION.SiblingSurvival.r`

## REFERENCES

- Croft, Trevor (2011). Mortality estimates from major sample surveys: Towards the design of a database for the monitoring of mortality levels and trends. United Nations Population Division, Technical Paper No. 2011/2.
- Gakidou, E., and G. King (2006). Death by survey: Estimating adult mortality without selection bias from sibling survival data. *Demography*, vol. 43, No. 3, pp. 569-585.
- Masquelier, B. (2013). Adult mortality from sibling survival data: A reappraisal of selection biases. *Demography*, vol. 50, No. 1, pp. 207-228.
- Measure DHS (1996). Sampling manual, DHS-III basic documentation No 6. Calverton: Macro International. Available from [https://dhsprogram.com/pubs/pdf/AISM5/DHS\\_III\\_Sampling\\_Manual.pdf](https://dhsprogram.com/pubs/pdf/AISM5/DHS_III_Sampling_Manual.pdf)
- Rutstein, S. O., and J. G. Rojas (2006). Guide to DHS Statistics. Calverton, Maryland: ORC Macro. Available from [http://www.measuredhs.com/pubs/pdf/DHSG1/Guide\\_DHS\\_Statistics.pdf](http://www.measuredhs.com/pubs/pdf/DHSG1/Guide_DHS_Statistics.pdf)
- Timaeus, I. (2013). Estimation of adult mortality from sibling histories, International Union for the Scientific Study of Population. Available from <http://demographicestimation.iussp.org/content/estimation-adult-mortality-sibling-histories>.
- Trussell J., and G. Rodriguez (1990). A Note on the sisterhood estimator of maternal mortality. *Studies in Family Planning*, vol. 21, No. 6, pp. 344-346. (Nov. - Dec., 1990).
- United Nations (1983). Manual X: Indirect Techniques for Demographic Estimation. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/81. Available from [www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html](http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html).