

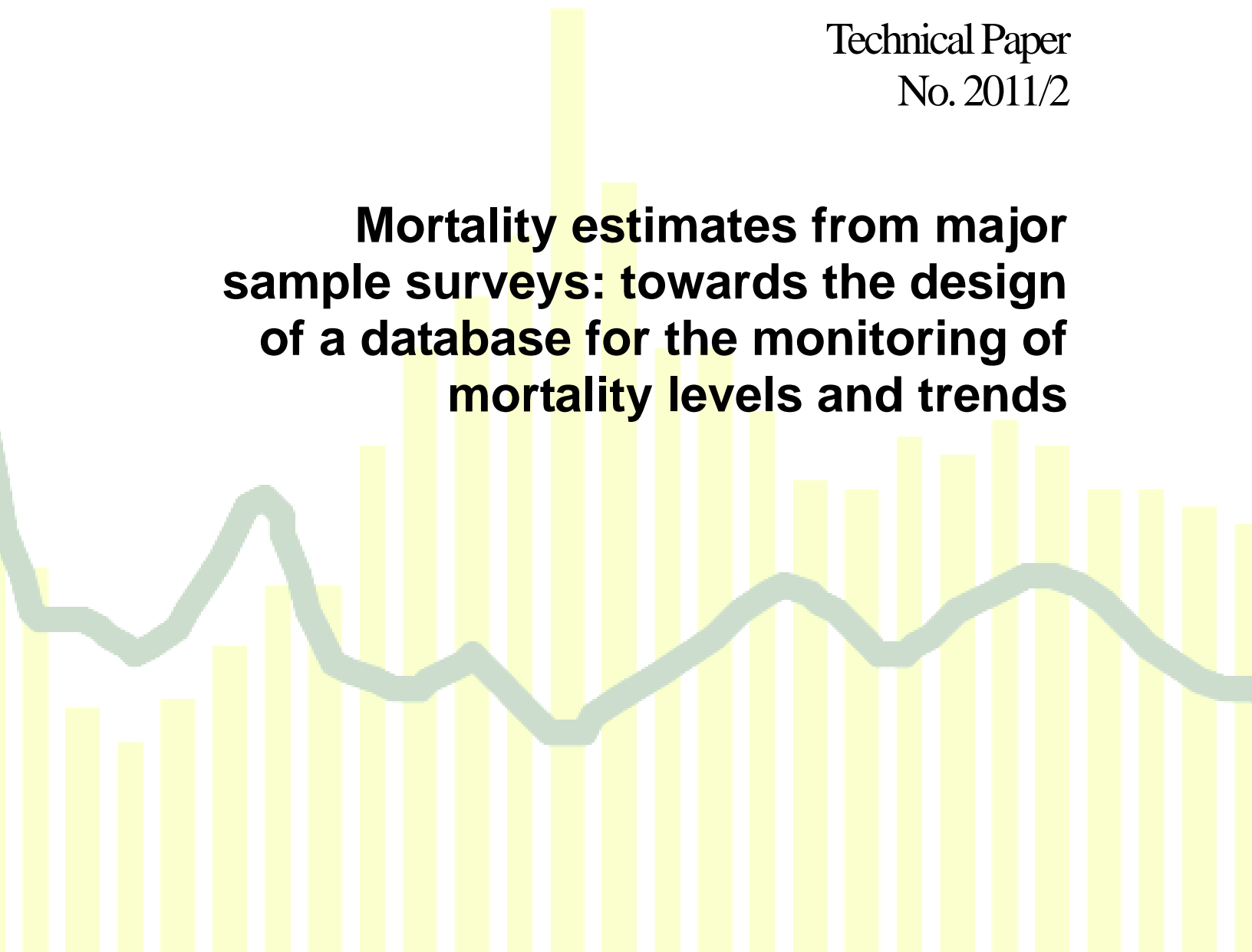
United Nations  
Department of Economic and Social Affairs



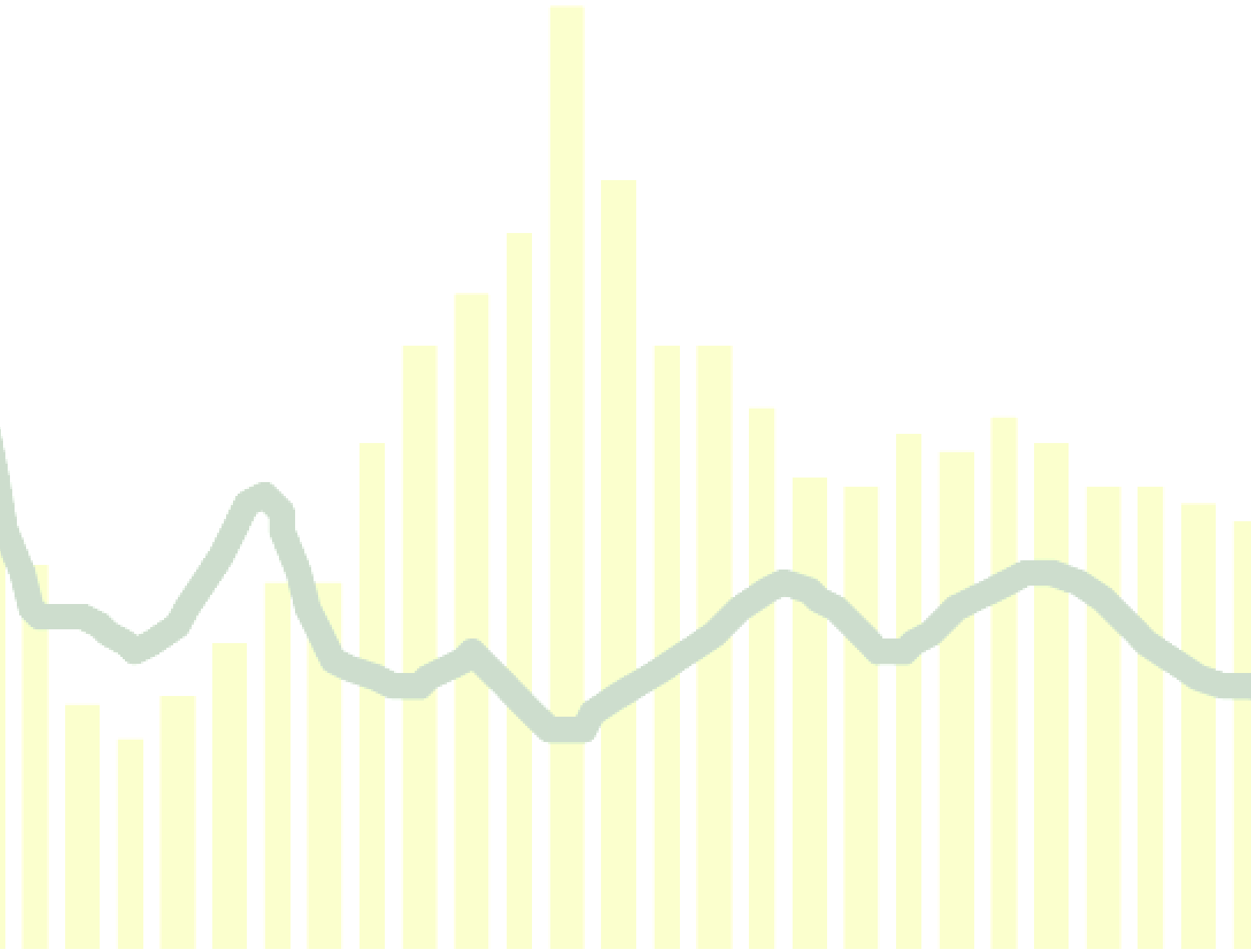
Population Division

Technical Paper  
No. 2011/2

**Mortality estimates from major  
sample surveys: towards the design  
of a database for the monitoring of  
mortality levels and trends**



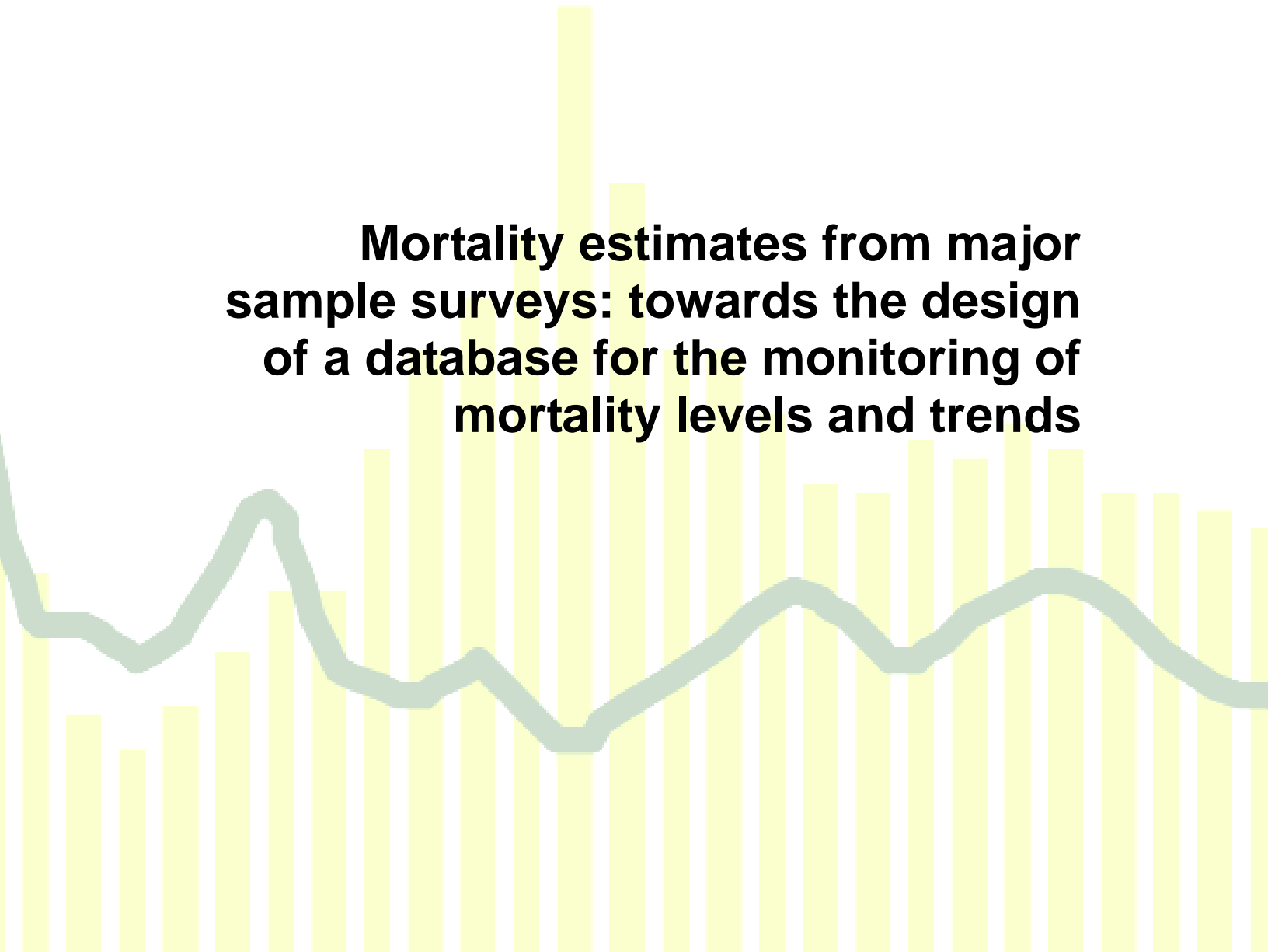
This page is intentionally left blank



Population Division

Technical Paper

No. 2011/2



**Mortality estimates from major  
sample surveys: towards the design  
of a database for the monitoring of  
mortality levels and trends**

## NOTE

The views expressed in this paper do not imply the expression of any opinion on the part of the United Nations Secretariat.

The designations employed and the presentation of material in this paper do not imply the expression of any opinion whatsoever on the part of the United Nations Secretariat 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.

## PREFACE

The Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat is responsible for providing the international community with up-to-date and scientifically objective information on population and development. The Population Division provides guidance on population and development issues to the United Nations General Assembly, the Economic and Social Council and the Commission on Population and Development and undertakes regular studies on population estimates and projections, fertility, mortality, migration, reproductive health, population policies and population and development interrelationships.

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.

Amongst its different responsibilities, the Population Division is also in charge of estimating levels and trends of mortality for all countries of the world. In this regard, there is a need to develop and populate a database containing relevant data for the estimation of mortality and the analysis of mortality differentials. Facilitating access to such data, both to Population Division staff and to academia, will contribute to foster work on the analysis of mortality and improve available mortality estimates.

As part of this endeavour, this paper provides a comprehensive and up-to-date overview about data that emanate from surveys and that are used to estimate levels and trends of mortality in childhood and adulthood, especially for developing countries. It defines the type of data that should be extracted from major survey programmes and suggests examples of appropriate tabulations that facilitate the best usage of these data. The paper was written by Mr. Trevor Croft of Blancroft Research International and was edited by staff members of the Population Division. The Population Division is grateful to Mr. Trevor Croft for having contributed to this paper.

The *Technical Paper* series as well as other population information may 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 (212) 963-3179, fax (212) 963-2147.

This page is intentionally left blank

## Table of contents

	<i>Page</i>
Preface .....	iii
Explanatory notes .....	viii
Introduction .....	1
Household survey programmes .....	1
Indicators .....	2
Childhood mortality estimation methods .....	3
DIRECT ESTIMATION .....	3
Data needs .....	3
Direct estimation variants .....	4
Considerations in using the direct estimation approaches .....	4
Calculation algorithm – the theory .....	7
Calculation algorithm – the practice .....	8
Alternative algorithm .....	10
Cohort estimates of mortality .....	13
Time periods .....	13
Recommendations for data storage .....	14
INDIRECT ESTIMATION .....	16
Data needs .....	17
Calculation algorithm .....	18
Assumptions .....	20
Comparison of the three variants of the indirect method .....	20
Recommendations for the storage of indirect mortality estimates .....	21
Men’s surveys .....	24
Adult mortality estimation methods .....	24
DIRECT ESTIMATION – SIBLING HISTORY .....	24
Data needs .....	24
Considerations in the sibling history .....	25
Calculation algorithm .....	26
Time periods .....	28
Recommendations for data storage .....	28
DIRECT ESTIMATION – RECENT HOUSEHOLD DEATHS .....	30
Data needs .....	30
Considerations in using the recent household deaths approach .....	31
Calculation algorithm .....	32
Recommendations for data storage .....	32
INDIRECT ESTIMATION .....	34
Data needs .....	35
Calculation algorithm – maternal orphanhood .....	35
Limitations of the maternal orphanhood approach .....	37
Recommendations for data storage .....	38
Calculation algorithm – sibling survivorship .....	38
Limitations of the sibling survivorship approach .....	39
Recommendations for data storage .....	39
Data Analysis Issues .....	40
WEIGHTING OF DATA .....	41
SAMPLE SIZE AND SAMPLING ERRORS .....	42
MISSING DATA AND SPECIAL VALUES .....	44
Dates of key events .....	44
Other types of data .....	45
EVER-MARRIED SAMPLES .....	45
DISAGGREGATION .....	46

Database contents .....	48
AGGREGATION RULES .....	50
PUBLISHED DATA VS. RE-ANALYZED DATA .....	50
Survey sources .....	51
DEMOGRAPHIC AND HEALTH SURVEYS .....	51
Data availability .....	51
Data content and limitations .....	51
Other survey-specific issues .....	52
REPRODUCTIVE HEALTH SURVEYS .....	52
Data availability .....	52
Data content and limitations .....	53
Other survey-specific issues .....	53
MULTIPLE INDICATOR CLUSTER SURVEYS .....	53
Data availability .....	53
Data content and limitations .....	53
Other survey specific issues .....	54
WORLD HEALTH SURVEYS .....	54
Data availability .....	54
Data content and limitations .....	54
Other survey specific issues .....	55
LIVING STANDARD MEASUREMENT SURVEYS .....	55
Data availability .....	55
Data content and limitations .....	55
Other survey specific issues .....	55
PAN ARAB PROGRAMME ON FAMILY WELFARE .....	56
Data availability .....	56
Data content and limitations .....	56
Other survey specific issues .....	56
GULF FAMILY HEALTH SURVEYS .....	56
Data availability .....	57
Data content and limitations .....	57
Other survey specific issues .....	57
CONTRACEPTIVE PREVALENCE SURVEYS .....	57
Data availability .....	57
Data content and limitations .....	57
Other survey specific issues .....	58
WORLD FERTILITY SURVEYS .....	58
Data availability .....	58
Data content and limitations .....	58
Other survey specific issues .....	59

#### LIST OF TABLES

Table 1	Numerators, denominators, probabilities and childhood mortality rates from direct estimation, Senegal DHS 2005 .....	10
Table 2	Numerators, denominators, probabilities and childhood mortality rates using an alternative algorithm, Ecuador RHS 2004 .....	12
Table 3	Childhood mortality rates by period using an alternative algorithm, Ecuador RHS 2004 .....	13
Table 4	Childhood mortality rates, Armenia DHS 2005 .....	15
Table 5	Weighted numerators and weighted and unweighted denominators, Armenia DHS 2005 .....	16
Table 6	Value of exact age x, by age or duration group, for three variants of indirect estimation of child mortality .....	19
Table 7	Child mortality rates using CEB/CS method, Coale-Demeny Models (Trussell equation), Jamaica MICS 2005 .....	22



Table 8	Child mortality rates using CEB/CS method, UN Models, Jamaica MICS 2005.....	23
Table 9	Weighted numerators and weighted and unweighted denominators, Jamaica MICS 2005.....	23
Table 10	Adult mortality rates by direct estimation from sibling history (0-6 years preceding the survey), Cameroon DHS 2004 .....	29
Table 11	Weighted numerators and weighted and unweighted denominators, Cameroon DHS 2004.....	30
Table 12	Adult mortality rates by direct estimation from recent household deaths (two years preceding the survey), India NFHS 1998/99.....	33
Table 13	Mean age at childbearing, based on births in the year preceding the survey, Burundi DHS 1987.....	36
Table 14	Adult mortality rates using maternal survivorship method, Burundi DHS 1987.....	37
Table 15	Adult mortality rates using sibling survivorship method, Cameroon DHS 2004.....	39
Table 16	MICS3 standard coding system for missing and special responses.....	43
Table 17	Partial example of data file of numerators and denominators by single year period, age group, sex and background variable for input into database .....	49
Table 18	Transformed data file of numerators and denominators by single year period, age group, sex and background variable for input into relational database.....	49

## FIGURE

Figure I	Lexis diagramme showing cohorts exposed to mortality at ages $a_1$ to $a_u$ during the period $t_x$ to $t_y$ .....	7
----------	--	---

## ANNEXES

Annex I	Direct Estimation Questionnaire Modules for Childhood Mortality .....	64
Annex II	Indirect Estimation Questionnaire Modules for Childhood Mortality.....	66
Annex III	Direct Estimation Questionnaire Modules for Adult Mortality .....	69
Annex IV	Indirect Estimation Questionnaire Modules for Adult Mortality .....	71
Annex V	Key variables from sample survey programmes (childhood mortality) .....	73
Annex VI	Key variables from sample survey programmes (adult mortality) .....	75

## Explanatory notes

### The following abbreviations have been used:

AGFUND	Arab Gulf Program for Development
AIDS	Acquired immune deficiency syndrome
AIS	AIDS Indicator Survey
BR	Birth recode
CDC	Centers for Disease Control
CEBCS	Children ever born and children surviving
CMC	Century month code
CMR	Child mortality rate <sup>1</sup>
CPS	Contraceptive Prevalence Survey
ISSA/CSPPro	Information System Security Association/Census and Survey Processing System
DHS	Demographic and Health Survey
DOM	Duration of marriage
DRH	Division of Reproductive Health
ESCWA	Economic and Social Commission for Western Asia
GHS	Gulf Health Survey
HIV	Human immunodeficiency virus
IMR	Infant mortality rate <sup>1</sup>
IOMS	Islamic Organization for Medical Sciences
IPPF	International Planned Parenthood Federation
IR	Individual recode
KIS	Key Indicator Survey
LAS	League of Arab States
LSMS	Living Standards Measurement Survey
MIS	Malaria Indicator Survey
MICS	Multiple Indicator Cluster Survey
MR	Men's recode
NFHS	National Family Health Survey
NN	Neonatal mortality rate <sup>1</sup>
ODA	Overseas Development Administration
OPEC	Organization of the Petroleum Exporting Countries
PAPFAM	Pan Arab Programme on Family Health
PAPCHILD	Pan Arab Programme on Child Welfare
PNN	Postneonatal mortality rate
RHS	Reproductive Health Survey
SPA	Service Provision Assessment
TSFB	Time since first birth
U5MR	Under-five mortality rate <sup>1</sup>
UNFPA	United Nations Population Fund
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
WFS	World Fertility Survey
WHO	World Health Organization
WHS	World Health Survey

---

<sup>1</sup> Throughout this paper, the term "rate" is used to signify either a rate or a probability. Among other cases, it should be noted that the CMR (or 4q1), IMR, NN and U5MR all refer to probabilities, not rates.

## **Introduction**

This paper describes the data requirements for a database on child<sup>1</sup> and adult mortality estimates based on household surveys, with particular focus on the major large-scale demographic household survey programmes. It describes the basic child and adult mortality indicators that can be derived from these surveys and the different approaches possible to calculate estimates of these indicators. It then describes the data that must be extracted from the survey datasets in order to calculate the estimates and provides details of the calculations.

To guide the development of database applications, the paper provides a set of recommendations for data tabulation and storage that will enable maximum flexibility in the analysis of mortality for different ages and time periods. It then discusses the disaggregation of mortality estimates for the analysis of mortality differentials by demographic or socio-economic characteristics, as well as issues related to sample size requirements, sampling errors around the estimates, and the use of sample weights. Particular attention is paid throughout the paper to the handling of cases in which data that are necessary for the estimation of mortality, such as the date of birth of an individual, are missing from the data record.

The paper does not describe in detail the theory behind the estimation methods nor does it extensively discuss the benefits and limitations of each of the methods, although it does discuss key issues pertaining to each and provide references to sources containing more thorough discussion. Rather, focus is placed on the mechanics of the calculation of the estimates and the issues to be addressed in computing the estimates using data from the major household survey programmes. Much of the material in the paper is compiled from other sources, including the demographic literature and the websites and technical reports of the various household survey programmes.

## **Household survey programmes**

This paper will review the infant and child mortality estimation processes and the adult mortality estimation processes used by major demographic and health survey programmes. In particular the paper looks at the following major household survey programmes:

- Demographic and Health Surveys (DHS);
- Reproductive Health Surveys (RHS);
- Multiple Indicator Cluster Surveys (MICS)
- World Health Survey (WHS);
- Living Standards Measurement Studies (LSMS);
- Pan Arab Programme on Family Health (PAPFAM/PAPCHILD);
- Gulf Health Surveys (GHS);
- Contraceptive Prevalence Surveys (CPS);
- World Fertility Surveys (WFS).

Details concerning the history and content of each of these household survey programmes are provided later in the paper.

## Indicators

This paper will discuss the computation of five indicators on childhood<sup>1</sup> mortality and two indicators on adult mortality from the major household survey sources. The indicators considered are as follow:

Childhood mortality indicators:

- Under-five mortality (U5MR or 5q0): the probability of dying between birth and exact age 5 years;
- Child mortality (CMR or 4q1): the probability of dying between exact ages 1 and 5 years;
- Infant mortality (IMR or 1q0): the probability of dying between birth and exact age 1 year;
- Neonatal mortality (NN): the probability of dying between birth and exact age 1 month;
- Post neonatal mortality (PNN): the difference between infant and neonatal mortality.

Of these indicators, three are considered particularly useful for monitoring mortality levels in childhood, as well as trends and differentials. The under-five mortality rate has long been used as a key measure of childhood mortality, measuring the probability of dying from all causes before the age of 5 years among children. The infant mortality rate has been used historically as an important measure of childhood mortality, although its popularity has waned somewhat in recent years, and it is more commonly available from routine reporting systems than the under-five mortality rate. As levels of childhood mortality decline, an increasing proportion of child deaths occur in the first month of life and so the neonatal mortality rate has gained increased importance.

The other two indicators of childhood mortality are produced as by-products from the calculations of the three indicators above, but are useful in their own right. The child mortality rate provides an indication of mortality levels of children after the first year of life. Deaths that occur during that period are primarily due to causes that are preventable if the right services and interventions are in place. The post-neonatal mortality “rate” is computed as the arithmetic difference between the infant mortality rate and the neonatal mortality rate and provides an indication of the incidence of death in children between exact age 1 month and exact age 1 year. While not strictly a rate, the arithmetic difference is used by convention for simplicity because in practice there is almost no difference between a rate calculated as the probability of dying between age 1 month and age 1 year and the arithmetic difference between the infant mortality rate and the neonatal mortality rate.

Adult mortality indicators:

- Adult mortality to age 60 ( ${}_{45}q_{15}$ ): the probability of dying between exact ages 15 and 60 years;
- Adult mortality to age 50 ( ${}_{35}q_{15}$ ): the probability of dying between exact ages 15 and 50 years.

The age of 15 years is generally used as the starting point for adult mortality estimates as it is roughly the age at which declining childhood mortality risks transition to increasing adult mortality risks (Hill, 2003). The choice for the upper age limit is usually 60 years because the age range from 15 to 60 avoids the difficulties related to the measurement of old-age mortality. In some cases, age 50 has been used as the upper limit to produce estimates of  ${}_{35}q_{15}$ , particularly in the case of the sibling history approach, to be presented below. Alternative measures such as life expectancy at age 5 ( $e_5$ ), 10 ( $e_{10}$ ) or 15 ( $e_{15}$ ) have also occasionally been used.

## Childhood mortality estimation methods

Two principal types of mortality estimation are commonly in use for the estimation of childhood mortality—direct estimation methods and indirect estimation methods. Direct estimation methods use observed or reported data on survivorship of individual children, including data on their date of birth, survival status and date of death or age at death for those who have died. Indirect estimation methods apply modelling techniques to cumulative data on the number of children ever born and children surviving or dead classified by age group of the mother, by duration of marriage duration of the mother or by the time since her first birth.

### DIRECT ESTIMATION

Procedures for the direct estimation of child mortality from household survey data have been well established and documented, particularly in WFS (Rutstein, 1983; Rutstein, 1984) and DHS publications (Sullivan et al., 1990; Sullivan et al., 1994; Bicego and Ahmad, 1996; Mahy, 2003a; Rutstein and Rojas, 2006), and the description here is adapted from the cited texts. The approaches to direct estimation that were developed in these two survey programmes have been adopted extensively by other household survey programmes.

#### *Data needs*

Direct estimation methods based on household surveys typically use retrospective birth or pregnancy histories to collect the data needed for the computation of mortality indicators. Birth or pregnancy histories include information for each birth or pregnancy that the respondent has ever had and usually include at a minimum:

- Month and year of birth of each child;
- Sex of each child;
- Survival status of each child (i.e., alive or dead);
- Age of each surviving child;
- Age at death of each deceased (or date of death);
- In the case of pregnancy histories, information on the outcome of each pregnancy (i.e., live birth, still birth, miscarriage, or induced abortion).

Birth or pregnancy histories may be collected in either full or truncated forms. Full birth or pregnancy histories cover all live births or pregnancies that the woman being interviewed had by the date of the survey. This information is usually collected in chronological order starting with the first birth or pregnancy, but sometimes it is collected in reverse chronological order. Truncated birth or pregnancy histories typically cover only births or pregnancies occurring during a fixed period of time, such as the five years preceding the survey. The information in truncated histories is usually collected in reverse chronological order starting from the date of the survey. Truncated histories facilitate fieldwork by limiting the amount of information that needs to be obtained, but they provide limited observations of mortality for older ages (over 2 years of age), and do not provide information on trends in mortality. Moreover, truncated histories have in practice been less effective than full birth histories at producing good quality data, principally due to omission and misplacement of births, particularly those of children who have died.

Birth history data are typically organized into data files where each record describes one birth. For DHS, these are known as Birth Recode (BR) Files, while in other survey programmes different

names are used for the files, including Birth History (MICS, WHS), Child (RHS), or Fertility files (LSMS).

In addition to the above data, two other necessary variables are the date (month and year) of interview, and the sample weight (see section on weighting for more details).

Annex I contains examples of the questions used to collect information on birth histories in the DHS, MICS, LSMS and RHS surveys.

#### *Direct estimation variants*

There are three main variants of direct methods for estimating childhood mortality rates, described in the *DHS Guide to Statistics* as follows:

1. A vital statistics approach in which the numbers of deaths to children under age 12 months in a particular period are divided by the numbers of births in the same period. What is estimated is a rate of mortality but not a probability; a variation in the number of births with time will change the rate without changes in the underlying probabilities. To correct this, separation factors would need to be used, which would have to come from the other variants.
2. A true cohort life table approach in which deaths to children under age 12 months of a specific cohort of births are divided by the number of births in that cohort to give an estimate of infant mortality. This procedure gives true probabilities of death, but has the drawback that all children in the cohort must have been born at least 12 months before the survey to be fully exposed to mortality, thus not taking into account the most recent experience. This requirement of full exposure becomes more limiting the higher the age segment of interest: for under-five mortality rates, only the information on children born five or more years before the survey can be utilized. Another drawback is that true cohort probabilities are not specific to a particular period at death, but instead relate to the date of birth of the cohort. Therefore the effects of events that affect several cohorts at the same time, for example, a famine, appears to be spread out over time.
3. A synthetic cohort life table approach in which mortality probabilities for small age segments based on real cohort mortality experience are combined into the more common age segments. This approach allows full use of the most recent data and is also specific for time periods (Rutstein and Rojas, 2006, page 70).

The synthetic cohort life table approach is the method that most household survey programmes have adopted for direct estimation of mortality, although the specifics of the approach may vary from one programme to another. Accordingly, the synthetic cohort life table approach and the calculations needed for that approach will be presented first. The true cohort life table approach and the corresponding calculations will also be discussed.

#### *Considerations in using the direct estimation approaches*

In the synthetic cohort life table approach, there are a number of decisions that need to be made regarding the handling of incomplete or incorrect information on the date of birth or the age at death of the child. In some cases, this information is missing entirely. In others, responses are believed to be “heaped” on a rounded age such as 12 months or 1 year, affecting the accuracy of mortality rates that are bounded by such ages.

It should be noted that date of death is rarely asked for, or recorded, in household surveys as the quality of reporting on dates of birth and death has not been sufficiently good. In most household surveys, dates of birth are only recorded to the month, not to the day, because respondents in many parts of the world are unable to provide accurate reporting of the day of birth. Even when asking for the month and year of birth there is only partial reporting of date of birth in many surveys. In order to properly identify neonatal deaths (those occurring at less than one month of age), the age at death is asked rather than the date of death. In the WFS, DHS and most other survey programmes, age at death is recorded in days if the death was within one month of birth, recorded in months if the death was within two years of birth and recorded in years thereafter.

Much of the early work on the calculation of mortality rates from household surveys was done as part of the World Fertility Survey (WFS) programme which ran from 1972 to 1984, and many of the decisions on how to handle deficient data date from that time.

To handle incomplete date of birth information, there are two alternative approaches that could be employed:

1. Drop cases where the date of birth is not fully known. Where the number of cases with incomplete date of birth information is very small and omission of the date is equally likely to occur for children who have died as for those who have survived, using this option will have little effect on the mortality rates. However, in reality this is very rarely the case. In most surveys the number of cases with incomplete date of birth information is not negligible and dropping these cases may bias the results. Even when the number is relatively small, it is much more common for children who have died to have incomplete dates of birth than children who have survived, and thus dropping these cases would bias the mortality rates downwards.
2. Impute complete date of birth for all cases from incomplete data. In practice most household survey programmes impute the incomplete dates of birth of each child to produce a complete date of birth. A number of methods exist for imputation of information, and a full description of the recommended imputation process is beyond the scope of this paper. The WFS and DHS approaches use the following basic steps (described more fully in Croft, 1991):
  - a. Create initial logical ranges in months in which the birth may have taken place based on existing date information.
  - b. Constrain the logical ranges based on isolated constraints (constraints that only affect the date of a single event), e.g. age of child or age at death if the child died.
  - c. Further constrain the logical ranges based on neighbouring constraints (constraints that may affect dates of several events), e.g. minimum intervals between births, minimum age at first birth, duration of amenorrhea, or abstinence after a birth and before the next birth.
  - d. Randomly impute a final date from the constrained logical range.

In practice, in most survey programmes the imputation of dates of key events has already been done in the production of the publicly available data file, so the decision has already been made by the data producer. However, the original data is often available and it is possible to re-impute the data if desired. In the case of DHS, the original data is not provided, but a flag variable is included in the dataset indicating the data on which the imputation was based, and from this it would be possible to reconstruct the original data if it were desirable to re-impute the data. In general, however, it is recommended that the already imputed values be used for consistency of results.

Incomplete or missing ages at death also require special handling. Clearly the mortality estimates would be biased downwards if children with missing ages at death were dropped from the calculations. Thus some form of imputation is used to produce an estimated age at death for all children that have died. In the DHS surveys, the proportion of children with incomplete ages at death is relatively small and imputation is performed using a “hot deck” process based on the age at death for the preceding child of the same parity and form of reporting (day, month or year, if known) found in the data file. This quasi-random technique preserves the variance of responses in the data set. Alternative procedures are also available such as random imputation of age at death based on the distribution of reported ages at death. In the DHS, imputed ages at death are produced in the process of preparing the distributed data, but for many survey programmes some handling of incomplete age at death information must be performed by the user prior to the mortality analysis.

It was noted above that reported age at death is notoriously heaped on certain values. In particular, age at death is frequently reported as 12 months or as one year. In DHS and most other survey programmes, reporting of one year for age at death is not strictly permitted and interviewers are asked to probe for the age at death in months. Even so, in most surveys there are still a few cases reported as one year. Furthermore, even after probing, the age at death is frequently reported as 12 months and this response is much more common than any of the neighbouring responses, clearly showing the heaping of responses. This heaping of responses is important since, if the response was rounded to 12 months then it causes the transfer of the death across the one-year boundary and thus may lead to an under estimation of infant mortality. There are various possibilities as to how to treat the heaping, including:

1. Redistributing deaths evenly as infant (less than 12 months) or child (12 months or more) deaths, or by some other fixed proportion, e.g. one quarter or one third.
2. Redistributing the deaths according to the age distribution of death between, say, 8 and 16 months, i.e. 8-11 and 13-16 months.
3. Using the data without any correction for heaping.

The decision as to which of the redistribution techniques should be used is arbitrary. Moreover, it is difficult to quantify the number of cases where age at death is rounded up and the extent of rounding up probably varies from one survey to another. For these reasons DHS and most other survey programmes do not adjust age at death for heaping and use the data as reported.

The same issue is likely to occur with reported deaths at the one month mark, where the age at death may have been rounded up from a value of less than 30 days and may thus underestimate neonatal mortality. However, it is even more difficult to quantify the amount of rounding up to the age at death of one month, so the data are almost always used without any adjustment for heaping on age at death of one month.

A decision that is partially affected by data quality is the length of age segments to be used in tabulating the data for the synthetic cohort approach (see description in the following pages). The length of the individual age segments for the synthetic cohort could theoretically be as short as the units in which the data are reported. Because date of birth is only reported in months and years, the age of living children can only be measured to the month. This means that the smallest unit that can be used is one month in length. However, age at death is reported in years for deaths at age two or higher, so the length of the age segment should be at least one year at these ages.

One approach would be to use single month segments up to two years of age and single year segments from that point on. However, because of age heaping and potentially small sample sizes in



some cells, the WFS and DHS have used the following age groups for the individual segments in months: 0, 1-2, 3-5, 6-11, 12-23, 24-35, 36-47, and 48-59.

An alternative approach used in the RHS is to use individual age segments of one month at all ages. However, to do so it is necessary to deal with ages at death of two years and higher, which are recorded in years. To work around this problem, the RHS randomly impute ages at death in months based on the reported age at death in years, using a uniform distribution. In other words, if a child died at  $x$  years, the age at death in months is randomly imputed in the range  $[x*12: x*12+11]$ . The RHS mortality rate calculations also have some other differences from the WFS/DHS approach and these are described following the description of the WFS/DHS approach below.

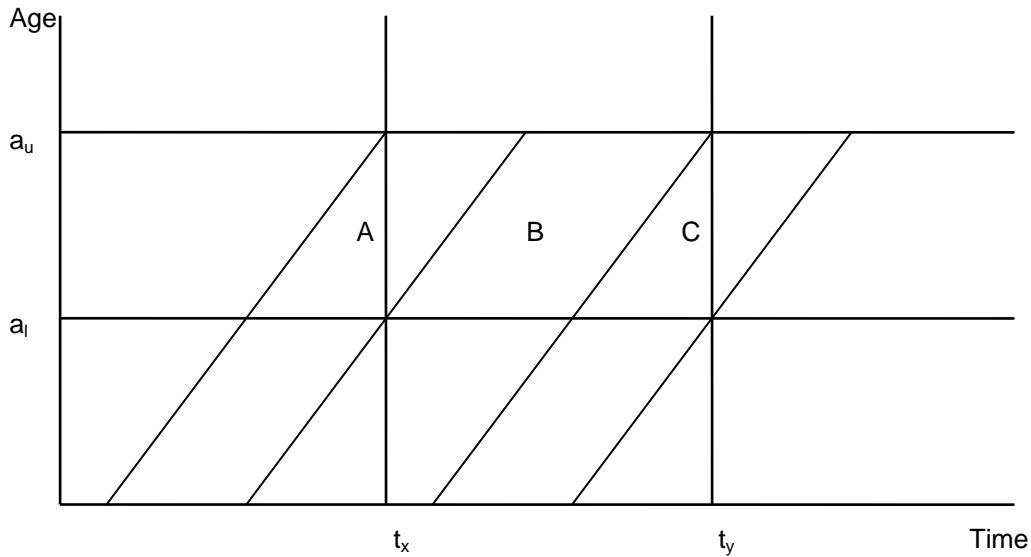
*Calculation algorithm - the theory*

The following description of the calculation used to obtain child mortality estimates through the synthetic cohort approach is adapted from the DHS Guide to Statistics (Rutstein and Rojas, 2006, page 71).

The first step is to calculate the component death probabilities that will ultimately yield the mortality estimates. The component probabilities are calculated for age segments 0, 1–2, 3–5, 6–11, 12–23, 24–35, 36–47, and 48–59 months of completed age. Each component is defined by a time period and an age interval.

Within these two parameters, three birth cohorts of children are included, as indicated in figure I. One cohort of children is completely included and two are partially included. If the lower and upper limits of the age interval are given by  $a_l$  and  $a_u$ , respectively, and the lower and upper limits of the time period are given by  $t_x$  and  $t_y$ , respectively, then the three cohorts are defined as children born between dates  $t_x - a_u$  and  $t_x - a_l$  (cohort A),  $t_x - a_l$  and  $t_y - a_u$  (cohort B) and  $t_y - a_u$  and  $t_y - a_l$  (cohort C).

**Figure I. Lexis diagramme showing cohorts exposed to mortality at ages  $a_l$  to  $a_u$  during the period  $t_x$  to  $t_y$ .**



Cohorts A and C are only partially exposed to mortality between ages  $a_l$  and  $a_u$  during time period  $t_x$  to  $t_y$  and account needs to be taken of this partial exposure. Because the age intervals of the component probabilities are small, the assumption is made that the exposure to mortality and deaths of birth cohorts A and C are well represented by taking one-half of the total exposure (i.e., one-half of the survivors at age  $a_l$  of the children of the given cohort) and one-half of the deaths (with the special exception noted below).

The numerators of the component death probabilities are equal to the sum of one-half of the deaths between ages  $a_l$  and  $a_u$  to children of cohort A, plus all of deaths between ages  $a_l$  and  $a_u$  to children of cohort B, plus one-half of the deaths between ages  $a_l$  and  $a_u$  to children of cohort C. The denominator is equal to the sum of one-half of the survivors at age  $a_l$  of children of cohort A, plus all of the survivors at age  $a_l$  of children of cohort B, plus one-half of the survivors at age  $a_l$  of children of cohort C. The component death probabilities are calculated by dividing the numerator for each age range and time period by the denominator for that range and period.

Another procedure is used for the time period that ends with the date of the survey. For this time period, numerators are calculated as the sum of one-half of the deaths between ages  $a_l$  and  $a_u$  to children of cohort A, plus all of the deaths between ages  $a_l$  and  $a_u$  to children of cohort B, plus all of the deaths between ages  $a_l$  and  $a_u$  to children of cohort C. This change is necessary because all the deaths reported in the survey for cohort C for this time period represent one-half of the deaths that will have occurred to the cohort between ages  $a_l$  and  $a_u$ .

#### *Calculation algorithm – the practice*

The calculation algorithm commonly used is consistent with the theoretical calculation described above but it is implemented in a different manner. Using the information in the birth history file on the duration of survival of each child, the algorithm usually tabulates mortality rates for multiple time periods at the same time. One of the benefits of the direct estimation method is the ability to produce trends in child mortality from the retrospective birth history in this manner. The algorithm first tabulates the numerator (the death of a child if the child died), and then tabulates the denominator (the exposure for the child, whether the child is alive or dead).

In tabulating the death the algorithm identifies the lower bound of the age group ( $a_l$ ) that the child was in when he or she died (i.e.  $a_l \leq a_i < a_u$  where  $a_i$  is the age of the child at death), and the time period ( $t_x - t_y$ ) in which the death occurred (i.e.  $t_x \leq t_j < t_y$ ) where  $t_j$  is the month in which the child reached the lower bound of the age group  $a_l$ . If the upper limit of the age group ( $a_u$ ) is in the same time period ( $t_j + a_u - a_l < t_y$ ), this is equivalent to the child being in cohort B, and the whole death is counted for the age group and the time period. However, if the upper age limit is in the following time period ( $t_j + a_u - a_l \geq t_y$ ), this is equivalent to the child being in cohort C. In this case, one half of the death is counted for the age group and the time period, and the other half is counted for the same age group, but for the next time period (equivalent to cohort A, but for the next time period).

The algorithm then tabulates the denominator (exposure) experienced by the child during its life. It follows the individual child from birth to the child's death (if he or she died before age 5 years), to age 5 years (if the child is 5 years or older or died at age 5 or older) or to the date of interview (if the child is younger than 5 years). For each age group in which the child was alive, it identifies the time period in which that age group was reached. Then, as in the tabulation of the deaths above, the algorithm identifies whether the upper limit of the age group is in the same time period or in the following time period. If it is in the same time period (equivalent to cohort B) the case is counted in the denominator for the age group and time period. If the upper limit of the age group is in the next time period (equivalent to cohort C) then one half is counted in the denominator for the age group and

time period, and then one half is counted for the same age group, but for the following time period (equivalent to cohort A, but for the next time period).

A few details of the algorithm bear highlighting. First, once an age group has been identified, the calculation of whether the child is in cohort B or cohort C is based on the lower and upper limits of the age group, and not on the actual age or age at death of the child. Second, the month of interview is excluded from all calculations as this is a partial month and does not contribute a full month of deaths or exposure. Third, for the most recent time period ending with the date of interview, deaths of children falling in cohort C are counted as one whole death rather than half a death, following the special exception noted above. Lastly, the length of the time periods used can not be less than one year, with the possible exception of the most recent time period terminating with the interview.

Once the numerators and denominators have been tabulated by age group and time period, the component probabilities ( $P_x$ ) of dying in each age group are calculated for each time period by dividing the numerators by the denominators. Finally, the various mortality rates for the time periods are computed as follows:

Neonatal mortality rate:

$$NN = P_0$$

where  $P_0$  is the probability of dying in month 0 (the first month of life).

Infant mortality rate ( ${}_1q_0$ ):

$$IMR = 1 - \prod_x (1 - q_x)$$

where x is age groups 0, 1-2, 3-5, and 6-11

Post-neonatal mortality “rate”<sup>2</sup>:

$$PNN = IMR - NN$$

Child mortality rate ( ${}_4q_1$ ):

$$CMR = 1 - \prod_x (1 - q_x)$$

where x is age groups 12-23, 24-35, 36-47, and 48-59

Under-five mortality rate ( ${}_5q_0$ ):

$$U5MR = 1 - \prod_x (1 - q_x)$$

where x is age groups 0, 1-2, 3-5, 6-11, 12-23, 24-35, 36-47, and 48-59

An example of the numerators, denominators, probabilities of dying and mortality rates calculated from a DHS survey is provided in table 1. The probabilities of dying at each age group are calculated by dividing the age at death in months by the exposure for each age group. The mortality rates are then calculated from the probabilities of dying using the algorithm above.

TABLE 1. NUMERATORS, DENOMINATORS, PROBABILITIES AND CHILDHOOD MORTALITY RATES FROM DIRECT ESTIMATION, TOTAL SAMPLE, SENEGAL DHS 2005

	<i>Years preceding the survey</i>				
	<i>0-4</i>	<i>5-9</i>	<i>10-14</i>	<i>15-19</i>	<i>20-24</i>
Age at death in months (numerator)					
0	367.6	451.5	310.8	224.3	178.5
1-2	74.7	86.3	63.1	30.4	46.5
3-5	54.6	94.0	44.7	39.6	41.5
6-11	142.2	145.0	106.3	63.9	53.1
12-23	179.0	259.6	157.7	131.5	76.6
24-35	223.7	171.9	138.8	120.4	72.9
36-47	106.5	100.3	83.3	51.5	26.4
48-59	83.9	53.9	54.3	33.6	12.2
Exposure to age groups in months (denominator)					
0	10 600.8	9 527.3	7 605.2	5 540.8	3 388.3
1-2	10 116.3	9 073.7	7 256.7	5 254.3	3 140.9
3-5	9 857.2	8 890.1	7 154.2	5 116.4	3 014.3
6-11	9 708.3	8 673.5	6 983.5	4 888.6	2 821.7
12-23	9 456.2	8 227.8	6 674.6	4 476.1	2 503.4
24-35	9 109.9	7 563.9	6 132.6	3 943.8	2 115.3
36-47	8 739.6	7 099.2	5 533.2	3 407.1	1 752.8
48-59	8 325.7	6 758.0	5 066.7	2 985.9	1 438.8
Probability of dying for age groups in months					
0	0.0347	0.0474	0.0409	0.0405	0.0527
1-2	0.0074	0.0095	0.0087	0.0058	0.0148
3-5	0.0055	0.0106	0.0062	0.0077	0.0138
6-11	0.0146	0.0167	0.0152	0.0131	0.0188
12-23	0.0189	0.0316	0.0236	0.0294	0.0306
24-35	0.0246	0.0227	0.0226	0.0305	0.0345
36-47	0.0122	0.0141	0.0151	0.0151	0.0151
48-59	0.0101	0.008	0.0107	0.0112	0.0085
Mortality rates (deaths per thousand live births)					
Neonatal mortality (NN)	34.7	47.4	40.9	40.5	52.7
Post-neonatal mortality (PNN)	26.4	34.6	28.7	25.3	44.2
Infant mortality (1q0)	61.1	82.0	69.5	65.8	96.9
Child mortality (4q1)	64.2	74.4	70.2	83.6	85.9
Under-five mortality (5q0)	121.3	150.3	134.8	143.9	174.5

The WFS/DHS approach to the direct estimation of mortality rates has also been used by other household survey programmes. In particular, the MICS surveys use the same approach in surveys in which birth histories are included. The PAPFAM/PAPCHILD surveys also use the same approach.

#### *Alternative algorithm*

An alternative algorithm is used in some household survey programmes, including the RHS, to calculate the desired mortality estimates. This method is somewhat simpler than the approach described above. Similarly to the approach used in WFS/DHS, above, it is also a synthetic cohort life table approach; however, the key difference is that the length of the individual age segments is defined as one month in length for all ages. As mentioned above, to avoid the problem of using age at death reported in years for children dying after two years of age, the RHS approach randomly imputes the age at death in months within the year using a uniform distribution.

The RHS essentially use a simple life table approach and build up a 60 month matrix of exposure and deaths for each month that the child was alive in the first five years of life. For each child, the algorithm executes the following steps:

1. Calculate the date of birth of the child as a century month code (CMC)<sup>3</sup>, denoted *bcmc*.
2. Calculate the date of death of the child as a century month code (*dcmc*). In this process an age at death in months is calculated for all deaths, imputing the age in months from the age in years, if necessary, using random imputation with a uniform distribution.
3. Repeat the following steps a-c for age in months *x*, where *x* is 1-60<sup>4</sup>
  - a. Calculate the CMC corresponding to age *x* (*agecmc*) as the CMC date of birth (*bcmc*), plus the age in months *x*, minus 1.
  - b. If *agecmc* is within the period of interest and does not exceed the date of death (*dcmc*) tally the case to the exposure for month *x* in the period. Note that a child born prior to the period of interest may contribute to the tally of exposure from the point at which the child reached the age corresponding to the start of the period. For example, a child born 6 months prior to the start of the period would contribute exposure to the period from the seventh month.
  - c. If *agecmc* is within the period of interest and equals the date of death (*dcmc*) tally the case to the deaths for month *x* in the period.
4. After tallying the deaths and exposure for each month, calculate the probability of dying at exact age *x* ( $q_x$ ) by dividing the deaths at age *x* by the exposure at age *x* for each month 1-60.
5. Calculate the probability of surviving to age *x* as  $l_x = \prod_{i=1-x} (1 - q_i)$ ,  
or more simply as  $l_x = l_{x-1} * (1 - q_x)$ , with  $l_0 = 1$ .
6. Calculate the probability of dying by age *x* as  $1 - l_x$ .

An example table of the data produced using this algorithm with RHS data is presented in table 2 for the most recent five-year period preceding the survey, with the resulting mortality estimates for five five-year periods presented in table 3.

This algorithm makes some assumptions that differ from those used in the WFS/DHS approach, but that simplify the algorithm. First, the algorithm essentially assumes that all births take place on the first day of the month and that a death in the first month of life takes place in the same month as the birth. In the DHS surveys, the general assumption is that the birth takes place on average in the middle of the month and a death in the first month of life may take place either in the latter half of that month or the first half of the following month. The difference between these approaches is that the RHS approach tallies a whole case in the first month of exposure and similarly for the death, while DHS would tally half a case in the month of birth, and similarly the death would be tallied as half a case at month *x* and half at month *x*+1 if the age at death is reported as *x* months.

TABLE 2. NUMERATORS, DENOMINATORS, PROBABILITIES AND CHILDHOOD MORTALITY RATES USING AN ALTERNATIVE ALGORITHM, 0-4 YEARS PRECEDING THE SURVEY, ECUADOR RHS 2004

Months	Deaths	Exposure	$q_x$	$l_x$	${}_xq_0$
			(per thousand)		
1	127.2	7 305.7	17.408	982.6	17.4
2	31.9	7 235.6	4.404	978.3	21.7
3	13.4	7 228.5	1.849	976.5	23.5
4	8.9	7 195.6	1.242	975.2	24.8
5	10.0	7 212.9	1.392	973.9	26.1
6	3.0	7 186.5	0.412	973.5	26.5
7	5.5	7 160.1	0.761	972.7	27.3
8	1.6	7 130.9	0.228	972.5	27.5
9	7.0	7 143.3	0.985	971.6	28.4
10	3.8	7 125.8	0.531	971.0	29.0
11	4.8	7 097.2	0.682	970.4	29.6
12	1.1	7 065.8	0.158	970.2	29.8
13	0.0	7 055.0	0.000	970.2	29.8
14	0.9	7 061.5	0.134	970.1	29.9
15	1.9	7 035.3	0.264	969.8	30.2
16	2.0	7 030.8	0.291	969.6	30.4
17	5.2	7 029.8	0.733	968.9	31.1
18	0.3	7 014.0	0.047	968.8	31.2
19	0.8	7 028.2	0.111	968.7	31.3
20	0.0	6 985.3	0.000	968.7	31.3
21	0.7	6 970.0	0.098	968.6	31.4
22	1.9	6 983.0	0.278	968.3	31.7
23	0.0	6 996.6	0.000	968.3	31.7
24	0.8	6 986.1	0.111	968.2	31.8
25	0.0	6 975.6	0.000	968.2	31.8
26	0.0	6 988.0	0.000	968.2	31.8
27	1.6	6 984.4	0.228	968.0	32.0
28	0.9	6 953.8	0.136	967.9	32.1
29	2.7	6 956.9	0.383	967.5	32.5
30	1.4	6 941.1	0.203	967.3	32.7
31	1.1	6 949.1	0.161	967.2	32.8
32	1.5	6 903.8	0.212	966.9	33.1
33	1.7	6 867.1	0.250	966.7	33.3
34	0.0	6 889.1	0.000	966.7	33.3
35	0.8	6 849.4	0.114	966.6	33.4
36	1.9	6 803.3	0.285	966.3	33.7
37	0.9	6 794.4	0.139	966.2	33.8
38	0.0	6 774.3	0.000	966.2	33.8
39	1.5	6 764.9	0.217	966.0	34.0
40	0.8	6 792.7	0.115	965.9	34.1
41	0.0	6 768.1	0.000	965.9	34.1
42	0.0	6 739.6	0.000	965.9	34.1
43	0.0	6 737.8	0.000	965.9	34.1
44	0.0	6 735.0	0.000	965.9	34.1
45	0.0	6 720.0	0.000	965.9	34.1
46	0.0	6 720.3	0.000	965.9	34.1
47	0.0	6 744.4	0.000	965.9	34.1
48	1.8	6 749.6	0.273	965.6	34.4
49	1.4	6 738.7	0.214	965.4	34.6
50	0.0	6 735.2	0.000	965.4	34.6
51	0.0	6 705.6	0.000	965.4	34.6
52	0.0	6 691.6	0.000	965.4	34.6
53	0.9	6 706.9	0.135	965.3	34.7
54	0.0	6 689.3	0.000	965.3	34.7
55	0.0	6 680.2	0.000	965.3	34.7
56	0.0	6 666.7	0.000	965.3	34.7
57	0.0	6 664.7	0.000	965.3	34.7
58	0.0	6 638.8	0.000	965.3	34.7
59	0.0	6 643.6	0.000	965.3	34.7
60	0.0	6 622.2	0.000	965.3	34.7

TABLE 3. CHILDHOOD MORTALITY RATES BY PERIOD USING AN ALTERNATIVE ALGORITHM, TOTAL SAMPLE, ECUADOR RHS, 2004

<i>Mortality rates (deaths per thousand live births)</i>	<i>Years preceding the survey</i>				
	<i>0-4</i>	<i>5-9</i>	<i>10-14</i>	<i>15-19</i>	<i>20-24</i>
Neonatal mortality (NN)	17.4	20.6	16.3	24.8	22.0
Post-neonatal mortality (PNN)	12.4	13.0	14.1	17.0	22.2
Infant mortality (1q0)	29.8	33.6	30.4	41.8	44.2
Child mortality (4q1)	5.1	5.9	9.8	11.5	22.7
Under 5 mortality (5q0)	34.7	39.3	39.9	52.8	65.9

### *Cohort estimates of mortality*

True cohort estimates of mortality are estimates produced based on a particular cohort of births. For example, births 1-4 years preceding the survey can be used to produce estimates of infant mortality rates. With small changes to either the DHS or the RHS approach to the direct mortality estimates, true cohort estimates of mortality can be produced. Rather than using a fixed period and tallying deaths and exposure according to the period in which the child attained a particular age, the deaths and exposure are tallied according to the cohort in which the birth of the child occurred.

The main drawback of cohort estimates of mortality rather than period estimates is that the mortality estimates can only be properly calculated for children who have had the opportunity to experience the full duration covered by the rate. In other words children must have been born at least 12 months prior to the survey to permit the calculation of the infant mortality rate and at least 60 months before the survey to permit the calculation of the under-five ( ${}_5q_0$ ) or child ( ${}_4q_1$ ) mortality rates. Additionally, the probability of dying is often affected by period-related events such as famines, epidemics or war, which affect several cohorts simultaneously, rather than by cohort-related factors.

### *Time periods*

Retrospective estimates of childhood mortality from household surveys are usually presented for a number of time periods to permit analysis of trends over time. Direct estimation procedures offer flexibility in the choice of time periods for which estimates can be produced. Typically results are presented in most survey programmes for a five-year period preceding the survey. It is possible to calculate mortality rates for smaller time periods such as single years, but because of sample size limitations and the fact that childhood mortality is a relatively rare event (measured per thousand instead of per cent as for many other indicators), the resulting sampling errors around the estimates become very large.

In the DHS, national level estimates of mortality are presented for five-year periods preceding the survey, while disaggregated estimates for a variety of characteristics are presented only for a 10-year period. Examples of sampling errors in the estimates can be found in DHS reports.

The exact boundaries of the time period preceding the survey vary between the survey programmes. In the DHS, MICS, and PAPFAM programmes, the rates have typically been based on a five-year period preceding the survey, including months 1-60 before the survey, where month 0 is the month of interview. In this case, the period covered actually varies from respondent to respondent, but overall covers 0-4 years before the survey. In the RHS surveys, the period chosen is a fixed period, starting in the month before the first interview and extending back five years before that point (for

example, July 1999-June 2004). Alternatively, calendar year periods could be selected, and in fact this was common practice in phase 1 of DHS.

The choice of the number of periods for which to present results also varies from programme to programme. DHS typically presents national level trend data for three five-year periods preceding the survey in the final reports; however, data are presented in the DHS STATcompiler (<http://www.statcompiler.com/>) for five five-year periods. In the RHS, mortality rates are typically presented only for the most recent five-year period.

Some caution should be exercised when using retrospective estimates that refer to periods more distant from the survey. The birth history questions are generally asked to women of reproductive age, typically 15-49. As periods further back in time are used, there is increasing age censorship in terms of the respondent. For example, if a period 15 years before the survey is considered, then only births to women aged under 35 are used in the estimation. If there are differentials in childhood mortality by age of the mother, then this could bias trends further in the past that exclude older mothers. Also, estimates for the more distant past are based on a smaller portion of the sample, giving rise to large confidence intervals around the estimates. Additionally, the further back in time estimates are produced, the greater is the risk of underestimating the mortality rates.

The discussion here has focused mostly on the use of five-or ten-year periods, but there are occasions when periods of other length are beneficial. In some recent DHS surveys, the mortality rates presented by five-year periods show a slight increase in the period 5-9 years before the survey compared with 10-14 years before the survey, and then a dramatic decline in the most recent period (0-4 years preceding the survey). However, this appears to be an artefact of the interview process where some omission of births (particularly for children who have died) and some shifting of births outside of the most recent five-year period seem to have occurred. Interviewers may omit or shift these births in order to avoid asking a long series of child health questions. Using seven year periods rather than five-year periods appears to resolve the issue of shifting of births out of the five- year period and, in most cases, eliminates the apparent increase in mortality and suggests a steadier decline in mortality over time. While this is not a surprising result, as using a longer period is effectively smoothing the data, it has proven useful in assessing data in situations where no reason is known for an increase in mortality followed by a dramatic decline.

#### *Recommendations for data storage*

The majority of the household survey programmes follow the first approach described to produce direct estimates of child mortality, including WFS, DHS, MICS, and PAPFAM/PAPCHILD. The recommendation would be to use this approach to provide the best compatibility with existing published data. For those surveys using different approaches the differences in the estimates will be small and possibly non-existent when data is rounded to estimates per thousand (without decimal places).

Direct mortality estimates are typically published for five-year periods preceding the survey, and it is recommended that mortality estimates for each of the five childhood mortality indicators be calculated for five-year periods. It is also recommended that the estimates be stored in the database with the maximum precision possible, but be presented from the database without decimal places. Furthermore, it is recommended that these indicators be calculated for up to five periods of five years preceding the survey, but with caution for the two earliest periods due to potential biases described above.

A sample of the estimates to be stored is presented in table 4.



TABLE 4. CHILDHOOD MORTALITY RATES, ARMENIA DHS 2005

<i>Mortality rates(deaths per thousand live births)</i>	<i>Years preceding the survey</i>				
	<i>0-4</i>	<i>5-9</i>	<i>10-14</i>	<i>15-19</i>	<i>20-24</i>
Neonatal mortality (NN)	11.2	19.5	14.5	12.6	15.0
Post-neonatal mortality (PNN)	9.8	10.7	23.0	17.0	40.6
Infant mortality (1q0)	21.0	30.2	37.5	29.6	55.5
Child mortality (4q1)	0.0	5.4	5.8	9.3	2.6
Under-five mortality (5q0)	21.0	35.4	43.1	38.6	58.0

In addition to storing the final estimates, it is recommended that the aggregated numerators and denominators used in calculating the estimates be stored in the database. This would mean storing the following:

- Weighted number of deaths in age group x in period i;
- Weighted number of children exposed in age group x in period i;
- Unweighted number of children exposed in age group x in period i.

The unweighted number of children exposed is included because decisions on minimum sample size for reliable estimates are generally made based on unweighted numbers of cases contributing in the denominator (see section on weighting for more details).

It is also recommended that the data above be stored by single year periods preceding the survey. This provides the flexibility to calculate mortality rates for periods of any multiple of years preceding the survey by calculating the probabilities of dying at each age group for the combined period and then producing the desired mortality rates for that period, following the calculation of each indicator as described earlier. Table 5 provides an example of the recommended table of numerators and denominators. The table shows only five years of data, but it is recommended to produce the numerators and denominators for 25 years to permit calculation of five five-year estimates.

In the example below, the numerators and denominators are displayed with four decimal places. The recommendation would be to store the numerators and denominators as floating point numbers in the database to the maximum precision possible, but to display numerators, denominators and rates without decimal places. Keeping the maximum precision possible in the numerators and denominators will avoid problems of rounding error in the calculated mortality estimates.

The database should also store the rules for aggregating the numerators and denominators so that any of the rates can be produced for periods of any length.

Additionally (or alternatively) the numerators and denominators may be stored by calendar year if it is desired to calculate mortality rates for calendar year periods. The same may also be done for cohort rates, storing the numerators and denominators for single calendar year cohorts.

TABLE 5. WEIGHTED NUMERATORS AND WEIGHTED AND UNWEIGHTED DENOMINATORS,  
TOTAL SAMPLE, ARMENIA DHS 2005

	Years preceding the survey				
	0	1	2	3	4 ...
<b>Age at death in months (numerator)</b>					
0	2.6067	7.6732	3.9999	8.2292	3.1640
1-2	0.8780	1.7692	0.4674	0.0000	1.2015
3-5	1.1574	0.6306	0.2731	0.0000	0.0000
6-11	0.1938	0.1938	0.0000	2.9539	3.1210
12-23	1.2927	0.0000	0.9144	0.9144	0.5734
24-35	0.9660	0.0000	0.7137	0.7137	0.0000
36-47	0.0000	0.0000	0.0000	0.0000	0.0000
48-59	0.0000	0.0000	0.0000	0.0000	0.0000
<b>Exposure to age groups in months (denominator)</b>					
0	337.2608	328.9470	305.0733	283.3391	266.1409
1-2	321.3618	334.2479	290.4720	270.0623	265.5658
3-5	303.2680	346.1689	285.4277	257.4052	263.3702
6-11	307.7135	326.3123	282.5960	254.7796	258.4057
12-23	310.0322	290.6645	267.3610	257.2948	273.7671
24-35	290.6645	266.4465	256.3804	273.1937	309.0927
36-47	266.4465	255.6667	272.4801	309.0927	362.1094
48-59	255.6667	272.4801	309.0927	362.1094	349.2150
<b>Exposure to age groups in months (denominator) – unweighted</b>					
0	315.5000	299.5000	296.0000	269.5000	264.5000
1-2	297.5000	304.0000	288.0000	257.5000	266.0000
3-5	277.0000	310.0000	286.5000	246.5000	264.5000
6-11	280.5000	301.5000	277.5000	250.5000	264.5000
12-23	290.0000	279.0000	262.5000	263.5000	271.0000
24-35	279.0000	262.0000	263.0000	270.5000	299.0000
36-47	262.0000	262.5000	270.0000	299.0000	350.5000
48-59	262.5000	270.0000	299.0000	350.5000	344.5000

#### INDIRECT ESTIMATION

Indirect estimation methods use indirect information from reported numbers of children ever born and children surviving or dying, or the proportion of children dead. The indirect estimation methods were first developed in the 1960s, most notably by Brass (1964). These methods are described in a number of publications, particularly in Chapter III of *Manual X: Indirect Techniques for Demographic Estimation* (United Nations, 1983). They are sometimes referred to as Children Ever Born and Children Surviving (CEBCS). In this paper we review the data needs, issues and limitations of three variations of the indirect estimation of childhood mortality:

1. Estimation of childhood mortality classified by age of the mother (AGE);
2. Estimation of childhood mortality classified by duration of marriage (DOM);
3. Estimation of childhood mortality classified by time since first birth (TSFB).

The first two of these variations have been widely used for many years (and are detailed in United Nations, 1983), while the third variation—classified by time since first birth—is newer (Hill and Figueroa, 1999; Croft and Hill, 2008).

It should be noted that these methods provide estimates for infant mortality ( ${}_1q_0$ ), under-five mortality ( ${}_5q_0$ ) and child mortality ( ${}_4q_1$ ), but not for neonatal or post-neonatal mortality.

### *Data needs*

The data needed for the indirect estimation methods are much simpler than those needed for the direct estimation methods. Summary data are needed for the following:

- a. Children ever born
  - i. Total number of children ever born, or
  - ii. Mean number of children ever born.
  
- b. Children surviving
  - i. Total number of children surviving, or
  - ii. Mean number of children surviving, or
  - iii. Proportion of children surviving, or
  - iv. Total number of children who died, or
  - v. Mean number of children who died, or
  - vi. Proportion of children who died.
  
- c. Total number of women
  - i. Total number of all women (including never married women) (AGE)
  - ii. Total number of ever-married women (DOM)
  - iii. Total number of all women who have given birth (TSFB).

Only one piece of information is required from each list for groups (a) and (b) above, from which the others in each list can be derived. For the third group, the total number of all women, ever-married women, or women who have given birth is needed depending on the variation to be used. For each variation, the data should be disaggregated as follows:

1. Age of the woman in five-year age groups (AGE) for all women;
2. Duration of marriage in five-year groups (DOM) for ever-married women;
3. Time since first birth in five-year groups (TSFB) for women who have ever given birth.

Additionally, it is recommended that the data on children ever born and children surviving are disaggregated by sex to permit sex-specific mortality estimation.

The basic data above may be collected with just two questions to ascertain the number of children ever born and the number of children surviving, but are typically collected in a longer set of questions designed to probe to ensure better completeness of the data, and to allow disaggregation of the data by sex. The questions are typically asked in the following sequence:

1. Have you ever given birth?
2. Do you have any sons or daughters to whom you have given birth who are now living with you?
3. How many sons live with you?
4. How many daughters live with you?
5. Do you have any sons or daughters to whom you have given birth who are alive but do not live with you?
6. How many sons are alive but do not live with you?
7. How many daughters are alive but do not live with you?
8. Have you ever given birth to a boy or girl who was born alive but later died?
9. How many boys have died?

10. How many girls have died?
11. Just to make sure that I have this right, you have had in total \_\_\_ births during your life. Is this correct?

These questions are often known as the Brass questions (after Professor William Brass, who first developed the methods for deriving mortality estimates from these data). While the exact questions may vary slightly in different survey programmes, they are designed to elicit the same information. Example sets of questions from several survey programmes are given in Annex II. From this set of questions, the responses can be aggregated to produce the summary tabulations needed for carrying out the indirect estimation.

### *Calculation algorithm*

The basic steps in the calculation of the indirect estimates are:

1. Aggregate the total number of children ever born and children surviving by either age group, duration of marriage or time since first birth, as desired. Also aggregate the total number of respondents by the same grouping.
2. Calculate the average parity per woman (mean number of children ever born), by group:
 
$$P_i = \text{CEB}_i / N_i$$
 where  $P_i$  is the parity for group  $i$ ,  $\text{CEB}_i$  is the number of children ever born for group  $i$ , and  $N_i$  is the number of women in group  $i$ .
3. Calculate the proportion of children dead, by group:
 
$$D_i = (\text{CEB}_i - \text{CS}_i) / N_i$$
 where  $D_i$  is the proportion dead for group  $i$  and  $\text{CS}_i$  is the children surviving for group  $i$ .
4. Calculate the multipliers to use, as follows:
 
$$k_i = a_i + b_i * P_1/P_2 + c_i * P_2/P_3$$
 where  $a_i$ ,  $b_i$  and  $c_i$  are taken from tables of coefficients by group and model life table (United Nations, 1983; Hill and Figueroa, 1999; Croft and Hill, 2008). Alternatively, if the United Nations model life tables are to be used, calculate the multipliers as follows:
 
$$k_i = a_i + b_i * P_1/P_2 + c_i * P_2/P_3 + d_i * M$$
 where  $a_i$ ,  $b_i$ ,  $c_i$ , and  $d_i$  are taken from tables of coefficients by group and model life table (United Nations, 1990), and  $M$  is the mean age at maternity<sup>5</sup>.
5. Calculate the probabilities of dying, as follows:
 
$$q_x = k_i * D_i$$
 where  $q_x$  is the probability of dying by exact age  $x$

Depending on the variation of the method used, the value of  $x$  in the last step will vary, as follows in table 6.

TABLE 6. VALUE OF EXACT AGE X, BY AGE OR DURATION GROUP, FOR THREE VARIANTS OF INDIRECT ESTIMATION OF CHILD MORTALITY

Index <i>i</i>	Age		Duration of marriage		Time since first birth	
	Group	Age <i>x</i>	Group	Age <i>x</i>	Group	Age <i>x</i>
1	15-19	1	0-4	2	0-4	2
2	20-24	2	5-9	3	5-9	3* / 5**
3	25-29	3	10-14	5	10-14	5
4	30-34	5	15-19	10	15-19	10* / 5**
5	35-39	10	20-24	15	20-24	15
6	40-44	15	25-29	20		
7	45-49	20	30-34	25		

NOTES: \* As originally used in Hill and Figueroa, 1999; \*\* As proposed in Croft and Hill, 2008.

6. Calculate the reference period for the estimates, as follows:

$$t_x = e_i + f_i * P_1/P_2 + g_i * P_2/P_3$$

where  $e_i$ ,  $f_i$  and  $g_i$  are taken from tables of coefficients by group and model life table (United Nations, 1983; United Nations, 1990; Hill and Figueroa, 1999; Croft and Hill, 2008).

7. Finally, transform the  $q_x$  values from each group into  ${}_1q_0$ ,  ${}_4q_1$  and  ${}_5q_0$  estimates for each reference point. The transformation is done using linear interpolation between model life table levels as follows:

- Select the model life table to use, and select the table for males, females or both sexes as appropriate.
- Identify the correct column ( $q_x$  or  $l_x$ ) of the life table.
- Identify the mortality levels between which the  $q_x$  value lies. If the life table provides probabilities of surviving ( $l_x$ ), use  $1 - q_x$  to identify the mortality levels. If the  $q_x$  (or  $1 - q_x$ ) value is lower than the lowest value or higher than the highest value in the life table, then the transformation can not be performed. By convention,  $q_x$  values higher than the highest levels of mortality are transformed to .999 and  $q_x$  values lower than the lowest level of mortality are transformed to 0.
- Calculate the fraction the  $q_x$  value is from the smaller of the two bounding values in the life table as  $\Theta = (q_x - q_{xi+1}) / (q_{xi} - q_{xi+1})$ , where  $i$  and  $i+1$  are the mortality levels.
- Identify the column of the life table for the IMR ( ${}_1q_0$ ) and U5MR ( ${}_5q_0$ ) rates ( $x = 1$  for IMR, and  $x = 5$  for U5MR). Calculate the IMR and U5MR from the corresponding columns by locating the same mortality levels in each table and interpolating between the levels using  $\Theta$ , as follows:

$$q_1 = q_{i+1} + \Theta * (q_{i1} - q_{i+1}),$$

and

$$q_5 = q_{5i+1} + \Theta * (q_{5i} - q_{5i+1}).$$

If the tables provide  $l_x$  values instead of  $q_x$  values, use  $l_x = 1 - q_x$  and perform a similar interpolation.

- f. Calculate CMR from U5MR and IMR as follows:  $CMR = (U5MR - IMR) / (1 - IMR)$ .

### *Assumptions*

The indirect estimation models are based on a number of assumptions. The most important assumptions are:

- Accurate survivorship data – namely, that respondents accurately report data on CEB and CS;
- Knowledge of the shape of fertility and mortality schedules – namely, that the schedules prevailing in a particular population follow a pattern, by age or marriage duration, similar to those embedded in the models;
- Stationary demographic conditions – namely, that fertility and mortality levels have been constant for the past 15 or 20 years; and
- Homogenous mortality conditions – namely, that the children born to women in different age, marriage duration or time since first birth categories are exposed to the same risks of mortality.

In many cases, however, most of the assumptions, if not all, are violated:

- The quality of reporting on CEB and CS is somewhat variable
- The model life table patterns were based mostly on European historical data, but the mortality patterns seen in many countries, particularly those in parts of Africa do not match the model life tables so well.
- Mortality levels have changed significantly over the past decade or two in many countries.
- Children born to very young and very old women as well as first births have higher risks of mortality than other children.

Even with these caveats, the indirect methods have been demonstrated to produce estimates of mortality that are reasonably comparable to those from direct methods and have proven useful for monitoring mortality during childhood, particularly in the absence of complete birth history information. Albeit, the results from the respective methods may differ considerably in some countries.

### *Comparison of the three variants of the indirect method*

Each of the three variations of the indirect method has both benefits and drawbacks. The variant based on women's age is the most commonly used, but it has been shown to have problems with estimates produced for the youngest age groups, 15-19 and 20-24, typically over-estimating the childhood mortality for these groups. In part this is a real effect of higher mortality among children born to younger mothers, but it is also influenced by a selection effect whereby women of lower socioeconomic groups tend to start childbearing at younger ages and their children have higher risks of mortality (Hill and Figueroa, 1999). Additionally, the sample of births and deaths to the youngest women is relatively small and so the random errors are greater in these groups. These problems are exacerbated in surveys with relatively small sample sizes and in settings with low fertility and relatively low mortality. Unfortunately these age groups provide the estimates for the most recent points in time. It is common practice to ignore the estimate based on the 15-19 age group, and often for the 20-24 age group.

An additional problem occurs with the age-based variant when used with surveys that include only ever-married women (see section below on ever-married samples). In this case, in order to obtain correct parity distributions inflation factors are needed to inflate the denominator in each age group

based on the inverse of the proportion of women ever married in each age group. The numerators are unchanged based on the assumption that there are no births outside of marriage.

The duration of marriage variant has some advantage over the age variant in that it tends to provide better estimates of recent mortality as it is less affected by selection bias and sample size issues among young women. The duration of marriage variant can be subject to another form of selection bias, however. In countries where childbirth outside of marriage is not unusual, the risk of mortality for children born out of wedlock is not captured until the mother is married (if at all). Additionally, when the experience of these children is captured it enters at the wrong duration of exposure. For example, if a child was born 6 years ago, when the mother was 16, and the mother married in the year preceding the survey at age 21, the exposure proxy (duration of marriage) would reflect no more than one year of exposure when in reality it was much longer.

The duration of marriage variant is particularly useful for surveys with samples of ever-married women as no adjustment for the never-married women is necessary (on the assumption that they have had no children) in contrast to the age variant which requires such an adjustment.

The variant based on time since first birth was proposed more recently by Hill and Figueroa (1999). It avoids certain problems of the other two variants. Like the duration of marriage variant, it has an advantage over the age variant in that it is not highly affected by socioeconomic bias or small sample sizes for the most recent estimates. The time since first birth variant also avoids the problems related to substantial numbers of births occurring outside of marriage. It does require date of first birth to be collected and usually this is only done in surveys that collect full birth histories. One exception to this is the MICS which ask date of first birth in all surveys together with the Brass questions. The time since first birth variant can also be easily used with ever-married samples, similarly to the duration of marriage variant.

#### *Recommendations for the storage of indirect mortality estimates*

Overall the recommendation for the mortality database would be to produce all three variants where possible and store the under-five, infant and child mortality rates produced by each variant for one selected model life table. Examples of the computed rates are shown in table 7 (Coale-Demeny models) and table 8 (United Nations models) for the age variants.

TABLE 7. PROBABILITIES OF DYING BEFORE AGE 1 USING CEB/CS METHOD  
 COALE-DEMENY MODELS (TRUSSELL EQUATION)  
 JAMAICA MICS 2005

		<i>North</i>		<i>South</i>		<i>East</i>		<i>West</i>	
<i>Age</i>	<i>Age i</i>	<i>q<sub>i</sub></i>	<i>t<sub>i</sub></i>	<i>q<sub>i</sub></i>	<i>t<sub>i</sub></i>	<i>q<sub>i</sub></i>	<i>t<sub>i</sub></i>	<i>q<sub>i</sub></i>	<i>t<sub>i</sub></i>
15-19	1	0.0230	1.0178	0.0222	1.0084	0.0237	1.0265	0.0235	1.0217
20-24	2	0.0167	2.3248	0.0174	2.3367	0.0176	2.4073	0.0175	2.3841
25-29	3	0.0320	4.2323	0.0342	4.3474	0.0340	4.4713	0.0337	4.4150
30-34	5	0.0265	6.5053	0.0276	6.7762	0.0273	6.9434	0.0272	6.8304
35-39	10	0.0321	9.0447	0.0320	9.4928	0.0317	9.7081	0.0316	9.4838
40-44	15	0.0436	11.7666	0.0431	12.4093	0.0429	12.6945	0.0428	12.2788
45-49	20	0.0440	14.6619	0.0437	15.5522	0.0436	15.9249	0.0436	15.2046
		<i>Reference IMR date</i>		<i>Reference IMR date</i>		<i>Reference IMR date</i>		<i>Reference IMR date</i>	
15-19	1	0.0230	2004.9	0.0000	2004.9	0.0237	2004.8	0.0235	2004.9
20-24	2	0.0158	2003.6	0.0000	2003.5	0.0169	2003.5	0.0165	2003.5
25-29	3	0.0275	2001.6	0.0314	2001.5	0.0311	2001.4	0.0295	2001.5
30-34	5	0.0216	1999.4	0.0252	1999.1	0.0244	1998.9	0.0230	1999.0
35-39	10	0.0234	1996.8	0.0283	1996.4	0.0268	1996.2	0.0245	1996.4
40-44	15	0.0283	1994.1	0.0360	1993.5	0.0340	1993.2	0.0305	1993.6
45-49	20	0.0255	1991.2	0.0355	1990.3	0.0320	1990.0	0.0283	1990.7
		<i>Reference U5MR date</i>		<i>Reference U5MR date</i>		<i>Reference U5MR date</i>		<i>Reference U5MR date</i>	
15-19	1	0.0284	2004.9	0.0000	2004.9	0.0264	2004.8	0.0278	2004.9
20-24	2	0.0187	2003.6	0.0000	2003.5	0.0184	2003.5	0.0189	2003.5
25-29	3	0.0349	2001.6	0.0352	2001.5	0.0352	2001.4	0.0356	2001.5
30-34	5	0.0265	1999.4	0.0276	1999.1	0.0273	1998.9	0.0272	1999.0
35-39	10	0.0290	1996.8	0.0314	1996.4	0.0302	1996.2	0.0291	1996.4
40-44	15	0.0361	1994.1	0.0411	1993.5	0.0388	1993.2	0.0370	1993.6
45-49	20	0.0321	1991.2	0.0405	1990.3	0.0364	1990.0	0.0341	1990.7



TABLE 8. PROBABILITIES OF DYING BEFORE AGE I USING CEB/CS METHOD,  
UNITED NATIONS MODELS, JAMAICA MICS 2005

Age	Age i	Latin American		Chilean		South Asian		Far Eastern		General	
		$q_i$	$t_i$	$q_i$	$t_i$	$q_i$	$t_i$	$q_i$	$t_i$	$q_i$	$t_i$
15-19	1	0.0222	1.0626	0.0245	1.2334	0.0222	1.0523	0.0224	1.1890	0.0223	1.1127
20-24	2	0.0177	2.3030	0.0181	2.4516	0.0178	2.3539	0.0175	2.4209	0.0176	2.3344
25-29	3	0.0344	4.0566	0.0349	4.3007	0.0347	4.1432	0.0340	4.1677	0.0342	4.0823
30-34	5	0.0277	6.3130	0.0277	6.6268	0.0279	6.4316	0.0272	6.4100	0.0274	6.3496
35-39	10	0.0322	8.9076	0.0315	9.2994	0.0323	9.0966	0.0315	8.9565	0.0320	8.9388
40-44	15	0.0423	11.8487	0.0427	12.2769	0.0435	12.1966	0.0423	11.7203	0.0424	11.8356
45-49	20	0.0438	15.4495	0.0436	15.8453	0.0440	16.0852	0.0435	14.9645	0.0438	15.3439
		<i>IMR</i>	<i>Reference date</i>	<i>IMR</i>	<i>Reference date</i>	<i>IMR</i>	<i>Reference date</i>	<i>IMR</i>	<i>Reference date</i>	<i>IMR</i>	<i>Reference date</i>
15-19	1	0.0000	2004.8	0.0000	2004.6	0.0000	2004.8	0.0224	2004.7	0.0000	2004.8
20-24	2	0.0000	2003.6	0.0000	2003.4	0.0000	2003.5	0.0162	2003.5	0.0000	2003.5
25-29	3	0.0285	2001.8	0.0322	2001.6	0.0000	2001.7	0.0292	2001.7	0.0291	2001.8
30-34	5	0.0225	1999.6	0.0253	1999.2	0.0234	1999.4	0.0227	1999.5	0.0227	1999.5
35-39	10	0.0000	1997.0	0.0000	1996.6	0.0000	1996.8	0.0246	1996.9	0.0248	1996.9
40-44	15	0.0300	1994.0	0.0360	1993.6	0.0323	1993.7	0.0304	1994.2	0.0307	1994.0
45-49	20	0.0297	1990.4	0.0353	1990.0	0.0319	1989.8	0.0289	1990.9	0.0301	1990.5
		<i>U5MR</i>	<i>Reference date</i>	<i>U5MR</i>	<i>Reference date</i>	<i>U5MR</i>	<i>Reference date</i>	<i>U5MR</i>	<i>Reference date</i>	<i>U5MR</i>	<i>Reference date</i>
15-19	1	0.0000	2004.8	0.0000	2004.6	0.0000	2004.8	0.0268	2004.7	0.0000	2004.8
20-24	2	0.0000	2003.6	0.0000	2003.4	0.0000	2003.5	0.0189	2003.5	0.0000	2003.5
25-29	3	0.0365	2001.8	0.0359	2001.6	0.0000	2001.7	0.0360	2001.7	0.0361	2001.8
30-34	5	0.0277	1999.6	0.0277	1999.2	0.0279	1999.4	0.0272	1999.5	0.0274	1999.5
35-39	10	0.0000	1997.0	0.0000	1996.6	0.0000	1996.8	0.0298	1996.9	0.0303	1996.9
40-44	15	0.0385	1994.0	0.0404	1993.6	0.0407	1993.7	0.0377	1994.2	0.0383	1994.0
45-49	20	0.0381	1990.4	0.0396	1990.0	0.0401	1989.8	0.0356	1990.9	0.0375	1990.5

In addition, it is also recommended to store the tables of numerators and denominators needed to calculate each of the variants, including unweighted denominators to permit a review of sample sizes to ensure an adequate sample for the estimates. An example is shown in table 9.

TABLE 9. WEIGHTED NUMERATORS AND WEIGHTED AND UNWEIGHTED DENOMINATORS  
TOTAL SAMPLE, JAMAICA MICS 2005

Age group	Mean number of children ever born	Mean number children surviving	Proportion dead	Number of women	Unweighted number of women
15-19	0.137	0.134	0.022	665.0	664
20-24	0.820	0.806	0.017	558.3	557
25-29	1.632	1.576	0.034	479.9	478
30-34	2.379	2.314	0.027	515.8	531
35-39	2.868	2.779	0.031	538.6	535
40-44	3.214	3.077	0.043	503.3	485
45-49	3.551	3.396	0.044	386.1	397
<b>Total</b>	<b>1.945</b>	<b>1.877</b>	<b>0.035</b>	<b>3 647.0</b>	<b>3 647</b>

## *Men's surveys*

Most men's surveys conducted under the DHS programme also include the set of reproduction summary (Brass) questions and could potentially be used for mortality analysis; however, no efforts to use such data are documented in the literature.

### **Adult mortality estimation methods**

As for child mortality, both direct and indirect methods exist for the estimation of adult mortality. However, the use of these methods and the quality of the estimates produced by them has been much more limited. As noted by Hill and Choi (2004), "Estimates of adult mortality for the developing world are less satisfactory than estimates of child mortality for two main reasons: no equivalent of the birth history for estimating child mortality from household surveys has been developed for adult mortality, and indirect estimation techniques do not seem to be as robust as indirect estimates of childhood mortality based on summary birth histories."

Timæus and Graham (1989) and Hill (2001) describe a number of different methods for adult mortality estimation. In the sections below, four of these methods will be discussed in detail. Two are direct estimation methods, the first based on detailed sibling histories and the second based on data on recent deaths in the household. The other two are indirect estimation methods based on summary responses on the survivorship of the respondent's mother or siblings.

#### DIRECT ESTIMATION - SIBLING HISTORY

##### *Data needs*

The sibling history method has become the main direct estimation method for adult mortality based on household surveys. The primary motivation for collecting the sibling history, especially in the DHS, is to estimate maternal mortality, but it can be used to produce estimates of adult mortality more generally. The sibling history is similar to the birth history used for child mortality estimation—in effect it is the birth history of the mother of the respondent. The sibling histories collect information for each birth or pregnancy that the respondent's mother ever had, and usually include at a minimum:

- Sex of the sibling;
- Survival status of the sibling – alive or dead;
- Age of sibling for surviving siblings;
- Age at death for siblings who died;
- Years since death or year of death for siblings who died.

Because sibling histories are designed to collect information for the estimation of maternal mortality, they often include information for female siblings as to whether the sibling was pregnant at the time of death, death was during childbirth, or within 42 days (or 6 weeks) of delivery.

In addition, the module is usually preceded by two questions, the first to ascertain the number of children that the mother of the respondent had, and the second to ascertain how many of the siblings were born before the respondent was born. This last question is used to place the respondent in order within the sibling history as part of the imputation process prior to analysis. Note that the sibling history usually does not include the respondent in the list of siblings.

Sibling history data are sometimes found in sibling-based data files, however, for DHS the data are found in the Individual Recode (IR) files, with all siblings listed in the same case. For these files, the data either need to be processed with software that can read repeating elements within a case, such as CPro, or be reformatted into a sibling history file to facilitate the processing. In some surveys (e.g. DHS surveys) the data are collected from both women and men. In these cases, the results presented in the DHS reports are usually based on the combined data for women and men aged 15-49. However, the data reported by the men and women are stored separately in the Individual Recode (IR) (Women) and the Men's Recode (MR) files and need to be pulled together for the analysis. In addition to the above data from the sibling module, two other variables are typically needed to produce mortality estimates – month and year of interview, and sample weight (see section on weighting for more details).

See Annex III for examples of sibling histories from the DHS and WHS surveys.

### *Considerations in the sibling history*

The major source of sibling history data is the DHS. The DHS Guide to Statistics (Rutstein and Rojas, 2006, page 130) describes the general calculation methods for producing mortality estimates from these data. The discussion below will clarify a few details of the calculation algorithm that differ from those described in the DHS guide.

In general the quality of data collected in the sibling history is significantly worse than that collected in the birth history. The information collected is also collected with less precision than in the birth history. All ages, ages at death and years since death in the sibling history are reported in completed years. Similar to the specifications for child mortality and the direct approach using the birth history, there are certain decisions concerning the calculations that need to be made, principally related to:

- a. The handling of incomplete information on date of birth.
- b. The handling of incomplete information on date of death.
- c. The inclusion or exclusion of information about the respondent.
- d. The length of period to cover.
- e. The size of age groups.

Many of the decisions concerning this approach were made in the late 1980s and early 1990s in DHS (Ruttenberg and Sullivan, 1991).

- a. For incomplete date of birth information, one may decide to drop the cases or to impute the dates of birth. Dropping cases with incomplete dates of birth will almost certainly create biases in the estimates, so this is not recommended. DHS imputes a CMC date of birth for all siblings, based on the reported age of the sibling (for living siblings) or the reported age at death and either the year of death or the number of years since death (for dead siblings), and based on the order of siblings and minimum intervals between siblings for those without age or age at death and years since death. The imputation process for the sibling histories is quite complicated and a full description is beyond the scope of this paper, but in essence it mirrors the approach used for the imputation of dates of events in the birth history. As the accuracy of most information is only to the year, the imputation process randomly imputes months based on a range of possible dates from the information provided, usually within a 12 month or shorter range.

- b. For incomplete date of death information, it is clearly inadvisable to drop cases as this will significantly affect the adult mortality rates. Thus it is necessary to impute the date of death. As is done for the date of birth, DHS imputes a CMC date of death for each sibling who has died, based on the age at death, years since death and imputed date of birth of the sibling. The exact month of death is imputed for all deaths within a range of possible dates.

Note that in DHS the imputed CMC date of birth and date of death are provided in the individual recode (IR) data files.

- c. There has been debate over whether to include or exclude the information about the respondent as a living sibling in the calculation of the adult mortality rates, and one's first reaction would be to include their information to avoid biasing the estimates. There are, however, other potential biases in the method in that people with no siblings are not included as there is no one else to report on them, and sibships in which all siblings have died are excluded from the analysis. However, Trussell and Rodriguez (1990) showed that these potential biases cancel each other out if the assumption is made that the adult mortality rates are independent of sibship size. This assumption has generally been accepted, so the information about the respondent is not included in the calculations of the adult mortality rates.
- d. Because in DHS the sibling history has mainly been used for estimating maternal mortality, a period of seven years typically has been chosen to provide sufficient maternal deaths to produce a reasonable estimate. The choice of a seven-year period is a compromise between a more common five-year period and a longer 10-year period that would preclude calculating the rates for an earlier period of equal length. For adult mortality, provided the sample size were sufficient, the period used could be shorter than seven years, but by convention seven year periods are generally used. The periods are usually defined as periods preceding the interview (excluding the month of interview as this provides only a partial month of exposure), but could alternatively be defined as calendar years.
- e. In most surveys using the sibling history approach, such as DHS, with sample sizes averaging around 7000 households, the total number of adult deaths reported is relatively small. For this reason, five-year age groups are usually used; however, it is also possible to tabulate the data for single year age groups.

#### *Calculation algorithm*

The calculation algorithm used by DHS is somewhat similar to that used for child mortality, except that age-specific death rates (based on person-years of exposure during each age interval for the given time period) are calculated from the dataset rather than component death probabilities (based on survivors to the beginning of each age interval during the given time period). The age-specific death rates must be converted to death probabilities. Most surveys using the sibling history approach are based on samples of individuals aged 15-49, the age groups used in the mortality estimates are also ages 15-49, and thus for a summary measure of adult mortality probability it is most advisable to produce an estimate of  ${}_{35}q_{15}$  from the sibling history approach rather than  ${}_{45}q_{15}$ . The  ${}_{35}q_{15}$  estimate can be transformed into a  ${}_{45}q_{15}$  estimate by using an appropriate model life table and level. Note that in the published DHS reports, an age-adjusted general mortality rate for ages 15-49 is presented as the summary measure, rather than  ${}_{35}q_{15}$ .

Each age-specific mortality rate is defined by a time period and an age group. The age groups used are 15-19, 20-24, ..., 45-49 (or alternatively by single year 15-49). For each sibling the numerators and denominators are tallied as follows:

- Numerators: If the sibling died within the time period, tally the case for the age group in which the sibling died.
- Denominators: Tally the months of exposure experienced by the sibling in each age group within the period. Depending on the size of the time period, the sibling may have been exposed in multiple age groups. For example, if five-year age groups are being used with a seven year time period, then a sibling who is now 26 years and 4 months at interview (based on the difference between the CMC date of interview and the CMC date of birth) contributes 15 months of exposure (ignoring the month of interview) to the 25-29 age group, 60 months of exposure to the 20-24 age group and 9 months of exposure to the 15-19 age group over the last seven years. Similarly, if the sibling died two years and seven months before the interview (based on the difference between the CMC date of interview and the CMC date of birth) and was aged 31 years and eight months at death (based on the difference between the CMC date of death and the CMC date of birth), then the sibling contributes 20 months of exposure (ignoring the month of interview) to age group 30-34 and 34 months of exposure to age group 25-29. After tallying the denominators, they are divided by 12 to convert into person-years of exposure.

Age-specific mortality rates ( ${}_5m_x$ ) are calculated by dividing the numerator for each age group and time period by the denominator for that group and period. To compute  ${}_{35}q_{15}$ , the age-specific mortality rates must first be converted into death probabilities using the formula:

$${}_5q_x = (10 * {}_5m_x) / (2 + {}_5m_x)$$

Death probabilities for each five-year age group can then be combined to produce estimates of  ${}_{35}q_{15}$  as follows:

$${}_{35}q_{15} = 1 - \prod_x (1 - q_x)$$

where x is age groups 15-19, 20-24, ..., 45-49.

While the sibling history includes siblings up to and beyond the age of 60, it is not recommended that the data be used for estimating age specific mortality rates beyond age group 45-49 as (i) the data are not representative of all siblings of that age, (ii) the siblings in these age ranges are only older siblings of the respondents and may experience different mortality risks than younger siblings, and (iii) the sample of siblings of these ages rapidly reduces with increasing age. For these reasons, it is suggested that if an estimate of  ${}_{45}q_{15}$  is desired, it should be derived from the estimate of  ${}_{35}q_{15}$  by interpolation between the mortality levels from an appropriate model life table. However, care should be taken in certain populations, particularly those most affected by HIV, as the patterns of mortality may differ greatly from those of the model life tables.

Also note that the age distribution of siblings aged 15-49 is different from the age distribution of respondents. For example, siblings within the age range 15-49 that only have siblings outside of this age range will never be reported. Thus the distribution provides minimums at the ends of the age range and a maximum at the mid point of the range (age group 30-35).

### *Time periods*

It is possible to produce adult mortality rates for several periods preceding the survey. However, estimates referring to the more distant periods are subject to a number of difficulties. The further back in time the reference period is, the younger the respondents would have been at that time. For example, when using the period 7-13 years preceding the survey, there are no respondents older than 42 years old, and thus increasing biases are likely to be introduced in the age specific mortality rates for the older ages.

Additionally, adult mortality estimates from sibling histories are known to be under-reported in most instances. Again, the further back in time events are reported, the worse the reporting problems appear to be. It is quite common when using the sibling history that adult mortality estimates for earlier time periods are lower than estimates for the more recent period due to reporting problems. This should be borne in mind when deciding on the use of earlier time periods.

The majority of direct adult mortality estimates are provided for either five-, seven- or ten-year periods preceding the survey. Seven-year estimates are recommended as a balance between avoiding large sampling error for shorter periods due to small numbers of events, while still providing relatively recent estimates and avoiding the worsening reporting for events that occurred further back in time. It is recommended that estimates be calculated for two seven year periods preceding the survey, but with caution concerning the earlier period. If it is desirable to disaggregate mortality estimates by any socioeconomic characteristic, it is recommended that this be done for 10-year time periods due to sample size considerations.

### *Recommendations for data storage*

It is recommended to store the age-specific mortality rates for seven-year periods calculated using the approach described above. The adult mortality data should always be presented disaggregated by sex as the mortality risks may be quite different between males and females. Estimates should be stored in the database with the maximum precision possible along with the aggregated deaths and person-years of exposure used in calculating them, though final results should be presented per thousand without using decimals.

In addition to storing the final estimates, it is recommended that the aggregated numerators and denominators used in calculating the estimates be stored in the database. This would mean storing the following:

- Weighted number of deaths in age group  $x$  in period  $i$ ;
- Weighted number of person-years of exposure in age group  $x$  in period  $i$ ;
- Unweighted number of person-years of exposure in age group  $x$  in period  $i$ .

The unweighted number of person-years of exposure is included because decisions on minimum sample size for reliable estimates are generally made based on unweighted numbers of cases contributing in the denominator (see section on weighting for more details).

An example table of estimates is presented in table 10.

TABLE 10. ADULT MORTALITY RATES BY DIRECT ESTIMATION FROM SIBLING HISTORY  
TOTAL SAMPLE (0-6 YEARS PRECEDING THE SURVEY)  
CAMEROON DHS 2004

<i>Age</i>	<i>Deaths</i>	<i>Exposure</i>	<i>Age specific mortality rates (per thousand)</i>	<i>(x-15)Q15 (per thousand)</i>
<i>Both sexes</i>				
15-19	246.9	82,355.0	2.998	14.9
20-24	391.6	83,944.1	4.666	37.7
25-29	492.3	70,216.3	7.011	70.9
30-34	427.4	53,490.6	7.990	107.5
35-39	347.0	38,032.4	9.124	147.4
40-44	262.1	24,157.0	10.851	192.7
45-49	161.3	13,755.9	11.728	238.9
45Q15				321.3
<i>Male</i>				
15-19	125.0	41,529.4	3.010	15.0
20-24	172.9	42,281.4	4.088	34.9
25-29	219.4	35,455.1	6.187	64.4
30-34	212.5	27,020.2	7.865	100.6
35-39	208.2	19,399.8	10.734	147.9
40-44	162.0	12,363.0	13.107	202.3
45-49	92.9	6,986.1	13.291	253.9
45Q15				343.5
<i>Female</i>				
15-19	121.9	40,825.6	2.985	14.8
20-24	218.8	41,662.7	5.251	40.4
25-29	272.9	34,761.2	7.851	77.5
30-34	214.9	26,470.4	8.117	114.4
35-39	138.8	18,632.6	7.448	146.8
40-44	100.1	11,794.1	8.486	182.4
45-49	68.5	6,769.9	10.115	222.0
45Q15				<b>297.8</b>

It is recommended that the data above also be stored by single year periods preceding the survey. This provides the flexibility to calculate mortality rates for any desired span of years preceding the survey by summing the numerators and denominators across that number of years for each age group, calculating the mortality rates for each age group for the combined period and then producing the desired summary indicator for that period. In addition, the numerators and denominators could be stored by single years of age, and the mortality indicators calculated directly from these based on death rates by single year of age. However, the reported ages of siblings tend to be heaped on certain responses, so grouped data may in fact perform better. Experience using DHS data has shown that the results differ little whether calculated for single year age groups or five-year groups and there does not seem to be significant advantages of storing single year age groups.

Table 11 provides an example of the recommended table of numerators and denominators for single year periods. The table shows only five years of data, but it is recommended to produce the numerators and denominators for at least 15 years to permit calculation of two seven-year estimates, three five-year estimates or a ten-year estimate. The numerators and denominators are displayed to three decimal places in the example below, but should be stored to the maximum precision possible to avoid rounding errors in later calculations.

TABLE 11. WEIGHTED NUMERATORS AND WEIGHTED AND UNWEIGHTED DENOMINATORS  
CAMEROON DHS 2004

	<i>Years preceding the survey</i>				
	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4 ...</i>
<b>Age at death in years (numerator)</b>					
15-19	27.590	26.040	35.125	38.612	36.540
20-24	48.989	43.053	62.348	54.018	73.607
25-29	64.723	62.361	88.006	85.227	73.250
30-34	61.397	54.590	81.617	59.486	56.433
35-39	52.209	48.454	62.750	53.193	51.748
40-44	45.784	40.600	51.670	50.037	32.067
45-49	25.894	32.769	26.500	28.688	19.830
<b>Exposure to age groups in years (denominator)</b>					
15-19	10 702.003	11 191.235	11 596.143	11 857.167	12 105.473
20-24	12 161.390	12 252.635	12 259.974	12 257.030	12 007.679
25-29	11 348.898	10 954.878	10 438.403	9 974.181	9 617.006
30-34	8 776.634	8 368.200	8 102.714	7 665.204	7 258.606
35-39	6 576.029	6 195.782	5 817.719	5 456.998	5 037.603
40-44	4 422.374	4 063.953	3 767.004	3 417.149	3 126.606
45-49	2 659.340	2 402.689	2 157.156	1 938.743	1 765.135
<b>Exposure to age groups in years (denominator) – unweighted</b>					
15-19	10 662.833	11 173.001	11 573.167	11 829.001	12 046.916
20-24	12 074.500	12 174.667	12 203.000	12 198.584	11 970.167
25-29	11 315.333	10 922.917	10 440.083	10 007.583	9 640.250
30-34	8 809.167	8 440.749	8 170.416	7 757.084	7 406.167
35-39	6 718.250	6 347.917	5 979.417	5 623.582	5 193.417
40-44	4 547.584	4 178.833	3 885.500	3 529.667	3 250.000
45-49	2 772.916	2 504.583	2 252.251	2 027.000	1 830.083

#### DIRECT ESTIMATION RECENT HOUSEHOLD DEATHS

Questions on deaths that have occurred in the household in the recent past have been widely employed in censuses in developing countries and have also been used in a number of sample household surveys (e.g., WFS, India NFHS 1991/2 and 1998/9). The method is simple in that it just asks about deaths that have occurred in the past *x* years, where *x* is typically one year in censuses, but has been two years in many of the surveys that have used the approach. Generally for each death in the recent time period, the sex of the deceased and the age at death are asked. The total living population from the survey provides the denominator for the mortality rates. Note that questions on recent household deaths are also the only survey questions that potentially provide data on mortality over the entire age span.

#### *Data needs*

For the recent household deaths approach, the data requirements are quite simple. For each person that died in the time period, the following information is elicited:

- Sex of the person;
- Age at death;
- Date of death.



Additionally in surveys that collect data on both *de facto* and *de jure*<sup>6</sup> residents, the residence status (resident or visitor) of the person who died is also collected. In these cases, typically results are presented for *de jure* residents.

Household deaths data are often found in separate data files with one record per death. However, for DHS the data are found in the Household Recode (HR) files, with all deaths listed in the same case. For these files, the data either needs to be processed with software that can read repeating elements within a case, such as CSPro, or be reformatted into a separate household deaths file to facilitate the processing.

As for other direct methods, the date of interview and the sample weight are also needed.

See Annex III for an example of the recent household deaths module as used in the India NFHS.

#### *Considerations in using the recent household deaths approach*

The choice of the period of analysis using the recent household deaths approach is typically made prior to data collection and will define whether the data is collected covering a one, two or three year period. Note that for simplicity of data collection, the period used in collecting the date may be defined as all deaths occurring on or after a fixed date. For analysis, however, it common to define a slightly narrower period preceding the interview, and exclude the month of interview from the period selected. For example, if a survey took place between March and May of 2007, the interviewers may ask for all deaths in the household on or after January 1, 2005. While analyzing the data, the period of interview may then be restricted to deaths in months 1-24 preceding the interview (where month 0 is the month of interview). Note that with longer periods for the reporting of deaths, the quality of the reporting is poorer for deaths further back in time. On the other hand, a longer time period increases the sample of deaths and exposure and provides estimates with smaller sampling errors.

When using the recent household deaths approach, there are a number of ways in which the denominator for the age specific mortality rates may be produced. Typically the denominator is based on the household population as reported in the household schedule (or listing); however, this provides a distribution of the population at the time of the survey by age (or age group) and does not reflect the person-years of exposure in each age group at the time of exposure to the risk of mortality. If the time period is very short, for example one year, there is only a small difference. However, if a longer time period, such as 3 years, is used then adjustments should be made. Various approaches have been used to make this adjustment.

For example, in the India NFHS 1998/99, which used a two year recall period, the intercensal population growth rate was applied to the age distribution to project the population distribution backwards to produce a midpoint estimate of the population. The number in each age group was doubled to approximate the person-years of exposure within the two year period of analysis.

Alternatively it could be assumed that a person aged  $x$  at the time of interview contributed on average one half of a person-year of exposure at age  $x$  and one half at age  $x-1$  in the preceding one year period. For a two-year period, this would be one half of a person-year at age  $x$ , one whole person-year at age  $x-1$  and one half at age  $x-2$ .

### Calculation algorithm

The algorithms for calculating adult mortality rates from data on recent household deaths are similar to those for other direct estimation approaches.

First tabulate the deaths and exposure needed to compute age-specific mortality rates. Rates may be calculated for all age groups starting from ages 0, 1-4, 5-9, ..., however only the age groups 15-19, 20-24, ..., 55-59 are needed for adult mortality estimation.

To compute the numerator of the age-specific mortality rate tally any death that occurred within the period of analysis according to the age group in which the person died. To obtain the denominator, from the household schedule (or listing) of all household members first select the correct population of household members (typically *de jure* household members) and then tally the person years of exposure in each age group during the period. See the discussion above concerning the calculation of the person-years of exposure. Calculate the age-specific mortality rate by dividing the numerator for each age group by the denominator for that group.

If it is desired to compute a summary measure such as  ${}_{35}q_{15}$ , the age-specific mortality rates must first be converted into death probabilities using the formula:

$${}_5q_x = (10 * sm_x) / (2 + 5 * sm_x)$$

Then, age-specific probabilities of death can be combined to produce estimates of  ${}_{45}q_{15}$  as follows:

$${}_{45}q_{15} = 1 - \prod_x (1 - q_x)$$

where x is age groups 15-19, 20-24, ..., 55-59

### Recommendations for data storage

While the use of the recent deaths approach is more common in censuses, the approach has been used in a few household surveys and this section provides recommendations for the storage of estimates and underlying data from surveys. Typically, for surveys, a two year period is used in the analysis of household deaths data. It is recommended to store the final estimates for all five-year age groups, by sex, as well as any desired summary indicator such as  ${}_{35}q_{15}$  or  ${}_{45}q_{15}$ . An example of estimates to be stored is presented in table 12.

As sampling error can be an issue with this method and it is only generally used with surveys that have large sample sizes, it is recommended that the estimates be calculated with five-year age groups, although the data may be analyzed using single years of age. As can be seen in table 12, the numerators can be relatively small in some cells, even in a survey of over 90,000 households.

TABLE 12. ADULT MORTALITY RATES BY DIRECT ESTIMATION FROM RECENT HOUSEHOLD DEATHS  
DURING THE TWO YEARS PRECEDING THE SURVEY  
INDIA NFHS 1998/1999

Age	Deaths	Exposure	Age specific mortality rates (per thousand)	(x-15)q <sub>15</sub> (per thousand)
Both sexes				
0	1 344.4	21 574.2	62.3	--
1-4	597.2	88 103.1	6.8	--
5-9	247.1	123 665.1	2.0	--
10-14	134.6	117 273.9	1.1	--
15-19	211.8	101 792.4	2.1	10.4
20-24	275.0	87 459.7	3.1	25.8
25-29	242.3	80 809.9	3.0	40.3
30-34	224.9	67 467.3	3.3	56.2
35-39	226.5	63 894.2	3.5	72.8
40-44	234.6	48 375.4	4.9	95.1
45-49	320.1	42 406.4	7.5	128.8
50-54	400.8	31 757.5	12.6	182.4
55-59	383.5	28 842.0	13.3	235.3
60-64	753.7	29 127.8	25.9	--
65-69	650.2	19 597.5	33.2	--
70-74	884.5	15 213.7	58.1	--
75-79	494.7	6 386.8	77.5	--
80+	1 618.3	7 960.1	203.3	--
Male				
0	742.7	11 182.0	66.4	--
1-4	245.8	45 222.3	5.4	--
5-9	118.5	64 532.0	1.8	--
10-14	60.6	61 000.3	1.0	--
15-19	89.5	52 584.5	1.7	8.5
20-24	113.7	42 889.7	2.7	21.6
25-29	114.2	39 289.8	2.9	35.7
30-34	137.9	33 612.5	4.1	55.3
35-39	124.9	33 495.5	3.7	72.8
40-44	155.0	25 737.7	6.0	100.4
45-49	199.9	22 565.0	8.9	139.5
50-54	242.8	17 102.5	14.2	198.9
55-59	215.3	13 279.8	16.2	261.8
60-64	401.6	14 666.1	27.4	--
65-69	389.2	10 063.6	38.7	--
70-74	508.8	8 628.4	59.0	--
75-79	284.4	3 567.4	79.7	--
80+	809.2	4 191.9	193.0	--
Female				
0	601.6	10 392.3	57.9	--
1-4	351.5	42 880.7	8.2	--
5-9	128.6	59 133.1	2.2	--
10-14	74.0	56 273.5	1.3	--
15-19	122.3	49 207.8	2.5	12.4
20-24	161.3	44 570.0	3.6	30.1
25-29	128.1	41 520.1	3.1	45.0
30-34	87.0	33 854.8	2.6	57.2
35-39	101.6	30 398.6	3.3	72.8
40-44	79.6	22 637.7	3.5	89.0
45-49	120.3	19 841.4	6.1	116.3
50-54	158.0	14 655.0	10.8	162.9
55-59	168.2	15 562.3	10.8	207.2
60-64	352.2	14 461.7	24.4	--
65-69	261.1	9 533.9	27.4	--
70-74	375.8	6 585.4	57.1	--
75-79	210.3	2 819.4	74.6	--
80+	809.1	3 768.2	214.7	--

In addition to storing the final estimates, it is recommended that the aggregated numerators and denominators used in calculating the estimates be stored in the database. This would mean storing the following:

- Weighted number of deaths in age group  $x$  in the period;
- Weighted number of person-years of exposure in age group  $x$  in the period;
- Unweighted number of person-years of exposure in age group  $x$  in the period (not shown in the example above).

The unweighted number of person-years of exposure should be included as decisions on minimum sample size for reliable estimates are generally made based on unweighted numbers of cases contributing in the denominator (see section on weighting for more details).

#### INDIRECT ESTIMATION

The main methods of indirect estimation of adult mortality from household survey data are based around the concept of survivorship of relatives – orphanhood<sup>7</sup> (or survivorship of parents), widowhood (or survivorship of spouse), and sibling survivorship. These methods are described in detail in several publications, including: Chapter IV of *Manual X: Indirect Techniques for Demographic Estimation* (United Nations, 1983); *Methods for Estimating Adult Mortality* (United Nations, 2002); and *Methods for Measuring Adult Mortality in Developing Countries: A Comparative Review* (Hill, 2001). Only the maternal orphanhood and sibling survivorship methods will be discussed in detail here.

The principle behind these methods is similar to that for the CEBCS methods for child mortality, namely that the target person was alive at some known point in the past and information exists about the length of exposure to the risk of mortality – since the birth of the respondent for maternal orphanhood; on average since the birth of the respondent for sibling survival – and the age at which exposure began.

The standard indicators produced from these approaches differ from those produced by the direct methods described earlier. The indicators produced are:

- Maternal orphanhood –  $l_{(25+x)}/l_{25}$ , the conditional probability of surviving from age 25 to age  $25+x$ .
- Sibling survivorship –  $l_x/l_{15}$ , the conditional probability of surviving from age 15 to age  $x$ , from which  ${}_{35}q_{15}$  is often calculated as for the sibling history direct approach.

For both methods, these indicators can then be transformed into a more familiar summary indicator of mortality such as  ${}_{45}q_{15}$  estimates using model life tables, on the assumption that the mortality patterns correspond to those in the model life table selected.

In addition to these methods, there are several lesser-used variations of indirect methods including paternal orphanhood, widowhood (or spouse survival), maternal orphanhood before and since marriage, and paternal orphanhood before and since marriage. The paternal orphanhood method is similar to the maternal orphanhood method, but makes more assumptions about the mortality patterns of men. Additionally, it requires an estimate of  $M$ , the mean age of fathers at conception, which is more difficult to produce than the mean age at childbearing for women as the data needed for

the calculation of M are rarely collected. Alternative methods of calculating M make further assumptions that may not hold in practice.

The widowhood method is rarely used and the data needed for this are rarely collected in household surveys. The United Nations *World Mortality Report 2007* (United Nations, 2007) records only two countries for which the data were collected for this method since 1990.

The maternal and paternal orphanhood before and since marriage methods (Timæus, 1991b) were attempts to produce methods that provided more recent estimates than the original maternal and paternal orphanhood methods. The questions needed for these approaches are rarely included in most household surveys.

#### *Data needs*

The questions needed for the maternal orphanhood and sibling survivorship methods are relatively simple:

- Maternal orphanhood:
  - “Is your mother alive?” (Yes, No, Unknown)
  - An estimate of the mean age at childbearing
- Sibling survivorship:
  - Number of brothers ever born
  - Number of brothers surviving to age 15
  - Number of brothers surviving
  - Number of sisters ever born
  - Number of sisters surviving to age 15
  - Number of sisters surviving

In addition to these, the age of the respondent, date of interview and sample weight are also needed.

The data for each of these variants are typically found in a woman-based data file, such as the Individual Recode (IR) file for DHS, or the Female file in other surveys. The sibling survivorship method may be applied to data aggregated from the sibling history. The method may also be applied to data gathered in a summary form in the questionnaire. The MICS, for example, collects data in summary form in an optional maternal mortality module.

#### *Calculation algorithm - maternal orphanhood*

The original maternal and paternal orphanhood methods proposed by Brass and Hill (1973) used an equation of the form:

$$l_{25+n}/l_{25} = W_n * S_{n-5} + (1-W_n) * S_n$$

where  $W_n$  is the weighting factor for the age group with upper limit  $n$  and  $S_{n-5}$  is the proportion of mothers alive for respondents aged  $n-5$  to  $n-1$ . Note that the weighting factors vary depending on the value of  $M$ , the mean age at childbearing (see below). The weighting factors are given in Manual X, table 86 (United Nations, 1983).

In 1977 Hill and Trussell proposed a regression equation for maternal orphanhood:

$$l_{25+n}/l_{25} = a_n + b_n * M + c_n * S_{n-5}$$

In the steps described below, the regression equation approach will be used. The basic steps in the calculation of the indirect estimates are:

1. Calculate the mean age at childbearing, M, as

$$M = \frac{\sum_1^7 a_i B_i}{\sum_1^7 B_i}$$

where  $i=1$  represents age group 15-19,  $i=2$  is 20-24, ...,  $i=7$  is 45-49,  $B_i$  is the number of births in a fixed period (usually one year preceding the survey) and  $a_i$  is the midpoint of age group  $i$  during the period used. Note that if one year preceding the survey is the period used, the midpoint is six months younger than the current midpoint of the age group, i.e. the midpoint for age group 15-19 would be 17, not 17.5. If a five-year period is used, the midpoint for each group is 2.5 years younger than the current midpoint of the group, i.e. for the 15-19 age group during the five-year period the midpoint is age 15. An example is given in table 13.

TABLE 13. MEAN AGE AT CHILDBEARING  
BASED ON BIRTHS IN THE YEAR PRECEDING THE SURVEY, BURUNDI DHS 1987

<i>Age</i>	<i>Births</i>
15-19	17.4
20-24	203.6
25-29	288.3
30-34	203.5
35-39	128.2
40-44	49.0
45-49	13.8
<b>Total</b>	<b>903.9</b>
<b>Mean age at childbearing</b>	<b>29.3</b>

2. Aggregate the total number of respondents with mother alive (or dead) by five-year age group of the respondent, the total number of respondents by five-year age group, and the total number of respondents who did not know or state the survival status of their mother.
3. For each age group, compute the proportion of mothers alive by dividing the number of mothers alive by the number of respondents less the number of respondents for whom the survival status of the mother was not known.
4. Compute the conditional probability of surviving using the regression formula, as follows:

$$l_{25+n}/l_{25} = a_n + b_n * M + c_n * S_{n-5}$$

The coefficients  $a_n$ ,  $b_n$ , and  $c_n$  are given in *Manual X*, table 93 (United Nations, 1983).

5. Compute the reference date for the estimates using the formula:

$$t_n = (n-2.5)*(1-u_n)/2$$

where

$$u_n = 0.3333*\ln(S_{n-5}) + Z(M+n-2.5) + 0.0037(27-M)$$

and  $Z$  is a standard function of age whose values are found in *Manual X*, table 88 (United Nations, 1983).

6. Finally transform the conditional probabilities of surviving from age 25 to 25+n ( $l_{25+n}/l_{25}$ ) for each age group into the required adult mortality rate  ${}_{45}q_{15}$  associated with each age group. This is done in several steps:
- First, identify the level of mortality from the adult survivorship ratios for Coale-Demeny model life tables for use with adult mortality estimates from orphanhood (*Manual X*, Annex VI, tables 207-210) for the selected life table, by interpolating between levels for the correct age group.
  - Find the probability of surviving from birth to age 15 ( $l_{15}$ ) and the probability of surviving from birth to age 60 ( $l_{60}$ ) from the Coale-Demeny model life tables for the same model, interpolating between levels based on the level identified in step “a”.
  - Compute the adult mortality rate for each age group as

$${}_{45}q_{15} = 1 - (l_{60} / l_{15})$$

Table 14 provides an example of the indirect estimation of adult mortality using the maternal survivorship approach.

TABLE 14. ADULT MORTALITY RATES USING MATERNAL SURVIVORSHIP METHOD, BURUNDI DHS 1987

Age	Maternal survivors	Maternal survivorship unknown	Number of respondents	Number of respondents (unweighted)	Interview date	Proportion of mothers alive	$P_{25+n} / P_{25}$	Reference date	${}_{45}q_{15}$
15-19	660.2	0.0	731.1	740	1,987.4	0.9030	0.9085	1,979.7	0.2523
20-24	645.0	5.7	779.1	785	1,987.4	0.8340	0.8478	1,977.8	0.2949
25-29	628.0	4.8	803.4	818	1,987.4	0.7865	0.8128	1,976.4	0.2743
30-34	446.4	1.1	622.1	608	1,987.4	0.7189	0.7592	1,975.3	0.2648
35-39	269.0	0.0	486.1	469	1,987.4	0.5534	0.6005	1,973.6	0.3403
40-44	129.8	0.2	269.1	273	1,987.4	0.4829	0.5366	1,973.4	0.2935
45-49	102.4	1.1	279.1	277	1,987.4	0.3683	0.4116	–	0.2720

Mean age at childbearing,  $M = 29.3$ . West model life table.

### *Limitations of the maternal orphanhood approach*

There are several limitations with the maternal orphanhood method. First, the respondents may not know or be willing to recognize that they were adopted and thus would be reporting on the survival status of their adoptive mother rather than their birth mother, who may have passed away. This is known as the “adoption effect” and it may generate a downward bias in the mortality estimates,

especially for younger respondents. Second, the selected model life tables may not adequately describe the pattern of adult mortality in a given region; this is particularly the case for many African countries. Additionally, the mean age at childbearing,  $M$ , based on current fertility may not reflect the mean age of childbearing of the mothers' generation. Timæus pointed out several limitations in the method and proposed variations of the method based on orphanhood since age 20 (Timæus, 1991a), and later, the orphanhood since marriage method (Timæus, 1991b). Both of these are improvements over the earlier method, but there is a lack of data for all of these methods. The data for the maternal orphanhood method are available for several DHS and WFS surveys as well as a variety of other national surveys, mainly from the 1970s and 1980s. The data are also available from a number of censuses, including in more recent years. The data for the paternal orphanhood method are also available from most of the same surveys; however, the estimates produced from the paternal orphanhood method have greater potential problems than the maternal orphanhood method.

#### *Recommendations for data storage*

If the maternal orphanhood method is to be used to provide historical estimates of levels of adult mortality, it is recommended that all of the data shown in table 14 be stored in the database, together with the final estimates in the last column.

#### *Calculation algorithm – sibling survivorship*

The basic steps in the sibling survivorship method are similar to the maternal orphanhood method and use the following regression equation:

$$l_n/l_{15} = a_n + b_n * S_{n-5}$$

where  $S_{n-5}$  is the proportion of brothers (or sisters) who, having survived to age 15, are still alive for respondents aged  $n-5$  to  $n-1$ .

The steps in the computation are as follows:

1. Aggregate the total number of siblings over the age of 15 currently alive, the total number of siblings who reached the age of 15 and the total number of respondents by five-year age group of the respondent.
2. For each age group, compute the proportion of siblings surviving among those who had reached age 15. In order to do so, divide the number of siblings currently alive by the number of siblings who reached age 15.
3. Compute the conditional probability of surviving using the regression formula, as follows:

$$l_n/l_{15} = a_n + b_n * S_{n-5}$$

The coefficients  $a_n$  and  $b_n$ , (as well as  $c_n$  and  $d_n$  below) are given in United Nations (2002), table V.1.

4. Compute the reference date for the estimates using the formula:

$$t_n = c_n + d_n * \ln(S_{n-5})$$



5. Transform the conditional probabilities of surviving ( $l_n/l_{15}$ ) into the adult mortality estimates  ${}_{35}q_{15}$  using the table of “Adult mortality estimates from survival of siblings: Conditional survival probabilities  $l_n/l_{15}$  for model life table translation” in table V.2 in United Nations 2002.

Results for Cameroon are shown in table 15.

TABLE 15. ADULT MORTALITY RATES USING SIBLING SURVIVORSHIP METHOD,  
CAMEROON DHS 2004

Age	Sibling survivors	Siblings survived to age 15	Number of respondents	Number of respondents (unweighted)	Interview date	Proportion of siblings alive	$l_n/l_{15}$	Reference date	${}_{35}q_{15}$	${}_{45}q_{15}$
15-19	7 632.9	7 936.2	2 684.5	2 680	2 004.4	0.9618	NA	NA	NA	NA
20-24	8 689.5	9 100.4	2 251.9	2 215	2 004.4	0.9548	0.9556	2 001.2	0.2072	0.2858
25-29	7 718.4	8 162.4	1 747.1	1 738	2 004.4	0.9456	0.9385	1 999.1	0.1840	0.2595
30-34	6 053.4	6 499.9	1 349.9	1 352	2 004.4	0.9313	0.9214	1 997.1	0.1708	0.2442
35-39	4 941.3	5 431.8	1 080.1	1 092	2 004.4	0.9097	0.8988	1 995.4	0.1669	0.2398
40-44	3 434.5	3 864.8	831.8	846	2 004.4	0.8887	0.8785	1 994.0	0.1563	0.2275
45-49	2 708.0	3 194.0	710.9	733	2 004.4	0.8478	0.8364	1 993.0	0.1636	0.2359

Note that, as for the direct estimation approach using the sibling history, the preferred summary estimate of mortality is  ${}_{35}q_{15}$  rather than  ${}_{45}q_{15}$  as most household surveys interview samples of women aged 15-49. The estimate of  ${}_{35}q_{15}$  can be transformed into an estimate of  ${}_{45}q_{15}$  using the appropriate model life table.

#### *Limitations of the sibling survivorship approach*

United Nations (2002) provides an overview of the sibling survivorship method, its assumptions as well as some of its limitations. Stanton and others (1997), in an assessment of the quality of maternal mortality estimates produced by sibling methods, provide some comments on the adult mortality estimates derived from the sibling survivorship approach; they suggest that the levels of mortality are likely to be under-estimated. The sibling survivorship approach is likely to suffer from some of the same problems as the direct sibling history approach, particularly omission of siblings who have died.

#### *Recommendations for data storage*

The recommendation for the sibling survivorship approach would be to produce and store in the data base the information provided in table 13 above. The data should be tabulated and stored disaggregated by sex as well as for both sexes combined. While it may be possible to disaggregate by other characteristics, it is not generally recommended. In particular disaggregating by residence may be misleading as the siblings of the respondents may not live or have lived in the same region or type of place of residence as the respondent.

When sibling data have been collected in both the women’s and men’s questionnaires, it is recommended that the data from both questionnaires be combined rather than just using the data from the women’s questionnaire. This exercise is possible for a number of DHS surveys.

## Data Analysis Issues

Various data analysis issues have been touched upon in the preceding sections. In this section some of these issues will be reviewed in more detail.

### WEIGHTING OF DATA

The vast majority of household surveys use probability samples with different probabilities of selection within different sample domains or sampling units. Sample designs frequently use unequal probabilities of selection of households to increase the number of cases available for analysis and reduce the sample variability for certain areas or subgroups. Because of the unequal probability of selection of households, sample weights need to be used when tabulating cases to produce the proper representation of the population.

The sample weights are essentially the inverse of the probability of selection, but are often adjusted for household non-response within the sampling domain. The sampling weights are then normalized by multiplying them by a constant factor so that the resulting weighted number of complete household interviews matches the unweighted number of complete household interviews. Once the household weights are calculated, individual weights are calculated by multiplying the household weight by the inverse of the non-response at the individual level within the sampling domain and then normalized so the total number of individual women's complete interviews matches the unweighted number.

In some surveys the normalization process is not performed and instead the sample weights raise the weighted number of households and individual women to national estimates of the total number of households and women.

The details of the calculation of the weights are dependent on the sample design and will have been performed by a sampling statistician prior to the initial analysis of the data. The sample weights are traditionally appended to the data file as variables in each case.

Note that in some survey programmes the household and individual women's weights are stored in the data files without decimal places, but may have a number of implied decimal places. In DHS survey, six implied decimal places are used, so it is necessary to divide the weight variable by 1,000,000 to produce the correct weight. Similarly, in WFS surveys the weight variables have 3 implied decimal places and so the weight variable is divided by 1000 to produce the correct weight. In most other survey programmes the weights are stored with the correct decimal placement. Occasionally the sample weights are not stored in the dataset but are provided in separate documentation or in the survey report.

There are very few surveys that do not require the use of sample weights to produce representative results. Such samples are known as self-weighting samples. One country that has frequently used self-weighting samples is Ghana, where the population is sufficiently evenly distributed across the sample domains to permit the use of a self-weighting sample. This, however, is the exception, rather than the rule. Even in cases of self-weighted samples, a weight variable set to 1 is typically included in the data file to simplify the use of the data and avoid the need to treat the self-weighted sample differently from non-self weighted samples.

To weight the data, most statistical packages use a statement to indicate that all results should be weighted. For example, using a DHS dataset in SPSS, the following commands would be included:

```
COMPUTE RWEIGHT = V005/1000000.  
WEIGHT BY RWEIGHT.
```

The sample weights should be used in all analyses to produce representative results.

#### SAMPLE SIZE AND SAMPLING ERRORS

Infant and child mortality are relatively rare events, measured per thousand rather than per cent as for many other indicators. To produce accurate estimates of child mortality requires a large sample size. However, most household surveys are not designed to produce highly accurate estimates of child mortality, but rather aim for high accuracy of a number of other indicators.

The Multiple Indicator Cluster Survey Manual 2005 (UNICEF, 2006) states that the infant mortality rate is not mentioned as a key indicator on which decisions on sample size are based because the sample size that would be necessary to measure child mortality indicators with the same precision, as is recommended for other indicators, would be too large – in the tens of thousands – and would be impractical. Instead indicators such as immunization coverage are typically chosen as key indicators for decisions on sample size.

Because the sample sizes are relatively small, the sampling errors around child mortality estimates are quite large. DHS reports provide sampling errors for child mortality rates in the appendices of each final report. To give an example, in the Zambia DHS 2001/02 the 95 per cent confidence interval around the under-five mortality rate of 168 per thousand for the five years preceding the survey was 159-181 (a relative error of 3.7%), based on a sample size of 7,658 women. For the infant mortality rate of 95 per thousand, the confidence interval was 86-104 (a relative error of 4.6%)

When these data are disaggregated into urban and rural areas, DHS presents child mortality estimates only for the 10-year period preceding the survey. Even with this doubling of the time period, the confidence interval around the under-five mortality estimate of 140 for this period increases to 123-157 (a relative error of 6.0%). Similarly the infant mortality rate for the 10-year period for urban areas was 77 per thousand with a confidence interval of 65-88 (a relative error of 7.4%).

Disaggregating the results into smaller groups, such as regions increases the sampling errors further. For example, the under-five mortality from the same survey for Copperbelt province for the 10-year period was 134 per thousand with a confidence interval of 108-159 (a relative error of 9.4%), and for the infant mortality rate of 68 the confidence interval was 52-83 (a relative error of 11.6%).

The example from Zambia shows that the sampling errors quickly increase as the sample size is reduced or the sample is disaggregated. The figures above are taken from a survey with a relatively high mortality rate and a moderate size sample. Taking an example now from the Armenia DHS 2000, a slightly smaller survey with lower child mortality rates, the infant mortality rate was estimated at 36 per thousand for the five years preceding the survey, but with a confidence interval of 25-47 (a relative error of 15.0%). The under-five mortality rate shows similar relative error at a very slightly higher level of mortality as the majority of deaths under age five in Armenia take place in infancy. Looking at the neonatal mortality rate, the problem worsens and the confidence interval around the neonatal mortality rate of 19 is 12-27 (a relative error of 20.2%).

Disaggregated into urban and rural areas, the child mortality rates in Armenia were only presented for a 10-year period and the confidence interval for the urban areas around the estimate of infant mortality of 36 per thousand was 26-48 (a relative error of 15.2%). This confidence interval is almost identical to that for the national sample for the five-year period, but is based on roughly half the sample, for twice the time period.

As can be seen from these examples, the average DHS survey provides a sample size that is reasonable for national estimates of child mortality for a five-year period, but the sampling errors quickly grow as the sample size is reduced or the data are disaggregated.

It should also be pointed out that the sampling errors are the one source of error that can be reasonably measured in surveys. Non-sampling errors are probably at least as great as the sampling errors, if not greater. The assumption has generally been made that the non-sampling errors are generally random in nature and are likely to cancel out. However, there is clear evidence from some surveys that this is not the case and that the survey estimates are sometimes biased in one direction, typically downward, underestimating mortality.

Some preliminary work has been done on the estimation of sampling errors from indirect estimates of child mortality. With the Zambia DHS 2001/02, the indirect estimate of under-five mortality using North model was 176 from the 25-29 age group with a reference date of mid 1997. Based on this preliminary work the standard error around this estimate was 8.4 per thousand for a confidence interval of 159-193 (a relative error of 4.8%). This is a slightly higher sampling error than the one from the direct estimation, and it seems to be the case that the error in the indirect estimates is somewhat greater than in the direct estimates.

The adult mortality rates also suffer from the same issues of sampling error as the child mortality rates. While adult mortality between 15 and 60 is not so rare, adult mortality within specific age groups is rare – in the younger adult age groups, rarer than child mortality. The age-specific mortality rates thus will have relatively large sampling errors around each rate, and consequently the  ${}_{45}q_{15}$  compound rate will be affected by these sampling errors.

It can be assumed that the indirect adult mortality estimates would have large confidence intervals similar to those of the indirect child mortality estimates. The author is unaware of any work on estimating the size of these errors, but the same approach used to estimate sampling errors for the indirect child mortality estimates could in principle be applied to the indirect adult mortality methods.

Certain rules have been used in some survey programmes to set minimum sample sizes for displaying mortality rates. As described in the DHS Phase II Tabulation Plan (Measure DHS, 2007, p. 106), DHS uses the following rule concerning the presentation of mortality estimates: “Rates based on 250 to 499 unweighted exposed persons should be shown in parentheses. Rates based on fewer than 250 unweighted exposed persons should not be shown (\*) and appropriate footnotes should be added to the tables if either or both of these cases exist. Rates are presented without any decimal points because of the relatively high sampling errors for the estimates.” It should be noted that these are absolute minimums for the presentation of results and that any estimates based on as few as 500 unweighted exposed persons would be likely to have extremely large sampling errors.

Such considerations underline the need for both weighted and unweighted denominators in the mortality database. Judgements concerning adequate sample size must be made based on the unweighted cases. For each type of estimate, it is important to correctly define the denominator to which the rule on sample size refers. For direct estimates of child mortality, it refers to the unweighted

denominator of the youngest age group contributing to the mortality rate. For indirect estimates of child mortality, it is the unweighted number of children ever born at each age or duration group (the denominator for the proportion surviving). In the case of direct adult mortality estimates, the rule refers to the unweighted denominator for each age-specific mortality rate. Note that for the recent household deaths approach, if the exposure period is greater than one year, for example a three-year period, then the denominator of the age-specific mortality rates is person-years of exposure, but the unweighted denominator for decisions concerning sample size is persons, i.e. roughly one-third of the actual denominator. For indirect adult mortality methods the denominator under consideration is the unweighted number of persons in the age group (the denominator for the proportion surviving).

The unweighted denominators should be stored in the database together with the weighted numerators and denominators.

#### MISSING DATA AND SPECIAL VALUES

In earlier sections on the direct estimation approaches for child mortality and adult mortality there has been some discussion of the handling of missing data. In this section, issues pertaining to missing data and other special responses are explored further.

All variables in survey datasets may potentially contain missing or other special values. Such values may be in one of several forms, including “don’t know” responses, refusals to answer, or items correctly or incorrectly skipped by the interviewers. It is important to make a distinction between missing data that is correctly missing because the question was properly skipped during the interview and values that are missing because the interviewer failed to record a response to a question that should have been asked.

For example, the MICS3 standard coding system for missing and special responses is shown in table 16.

TABLE 16. MICS3 STANDARD CODING SYSTEM FOR MISSING AND SPECIAL RESPONSES

<i>Response</i>	<i>Variable length</i>				
	<i>Alphabetic</i>	<i>1 character</i>	<i>2 characters</i>	<i>3 characters</i>	<i>4 characters</i>
Other	X	6	96	996	9996
No/None	Y	na	na	Na	na
Inconsistent	na	7	97	997	9997
Doesn't know	Z	8	98	998	9998
Missing	?	9	99	999	9999
Not applicable	Blank	Blank	Blank	Blank	Blank

The DHS programme uses the same coding system, and other programs use variations on this. In this system, “not applicable” is the term used for missing values that occur when questions are correctly skipped during interview. “Missing” is the term used for missing values that are incorrectly left empty during the fieldwork. In the discussion below, the “not applicable” cases will not be considered as “missing” values (although in many statistical software their values are called missing values).

Proper handling of missing values in analysis is crucial to obtaining accurate results. In mortality analysis some of the key variables tend to contain higher proportions of missing or special values than most other variables. In particular, the proportion of “missing” and “doesn’t know” responses is often high in the date and age variables that are central to the direct estimation approaches for child and adult mortality.

Decisions concerning the handling of missing data and special values essentially fall into two categories:

- Exclude cases with missing data or other special values from analysis;
- Replace the missing data or other special values with an in-range value, based on a particular set of rules.

The first approach is the easier to implement, but can lead to biases in the results. For example, if cases with missing values for the age at death are dropped from the numerator for a mortality rate, then the mortality rate would be underestimated.

The second approach is more complicated to implement as a set of rules needs to be developed as to how to replace or impute a new value for the values with missing or other special responses, and the rules will vary depending on the kind of data to be imputed.

#### *Dates of key events*

Because variables concerning the dates of key events and the ages of respondents and their children are central to the analysis of many surveys, most survey programmes have developed rules for the imputation of missing data for certain key dates and the related age variables. In the DHS, WFS, MICS and some other survey programmes, dates of the following key events are imputed and are already available in the datasets:

- Date of birth of the respondent;
- Date of first marriage;
- Date of birth of each child;
- Age at death for each child that died.

The date of interview is also required.

From these variables, the following variables are then imputed:

- Age of the respondent;
- Age of the respondent at first marriage;
- Age of the respondent at each birth;
- Age of each child or age at death for each child that died.

The imputation process for the dates of key events is briefly described in the section above concerning considerations in using direct estimation for child mortality. A full description of the imputation process can be found in “DHS Data Editing and Imputation” (Croft, 1991).

A similar imputation process is used by DHS for the sibling history and the imputation of dates of birth and death of each sibling.

In surveys that have not already used an imputation process to impute these information it will be necessary to carry out such a process. Alternatively if the number of cases with missing data is very small, dropping these cases is likely to only have a small effect on the results.

#### *Other types of data*

The data user will also be required to make decisions regarding the handling of special responses for other variables used in mortality analysis. For example, in surveys that do not collect birth or pregnancy histories and use the CEB/CS approach to indirect estimation, missing data on the numbers of children born, surviving or dead need to be handled. It is sometimes possible to impute one piece of information from another. For example, if a full set of reproduction summary questions have been asked, including the total number of children ever born, the number of children alive and the number of children dead, then it is possible to compute one of these three pieces of information from the other two if only one piece is missing. If multiple pieces of information are missing, then the options are more limited:

- Drop the case from the analysis;
- Impute a midpoint value (median) for the item;
- Impute according to a set of rules.

In most cases where the number of cases with missing data is small, the first option tends to be the most commonly used, and the bias in the estimates is usually small. The decision on the approach to be used should only be made after a thorough review of the data to evaluate the completeness of the reporting.

#### EVER-MARRIED SAMPLES

Certain surveys have used “ever-married samples” rather than “all women samples”. An “ever-married sample” is a sample in which only ever-married women are selected as women eligible for interview within a household. Women in consensual unions or separated from consensual unions are considered as ever-married. Such samples have typically been used in cases where it has been felt inappropriate to interview single women on some of the issues covered in the survey and where sexual behaviour and childbirth out of marriage are considered extremely rare. These samples have principally been used in Middle Eastern countries, but were more widely used during the WFS survey programme.

In these surveys, any statistic that is based on all women needs to take into account the ever married sample. Direct estimates of child mortality are unaffected, on the assumption that there are no births outside of marriage, as are the indirect estimation methods of child mortality using the duration of marriage or the time since first birth variants. However the indirect estimation method based on age of the respondent requires a sample of all women. To adjust for this, “all-women factors” need to be applied.

All-women factors are multipliers to adjust ever-married sample to statistics based on all women and are essentially the inverse of the proportion of women who are ever married or in union based on the household data. All-women factors are specific for individual years of age and the particular subgroup to be estimated (e.g., separate factors exist for urban and for rural subgroups). The all-women factors are used to inflate the denominators used in the age variant of the indirect approach, but not the numerators, assuming no children are born to never married women.

For the estimation of adult mortality, the ever-married samples can be used without adjustment under the assumption that age at marriage and survivorship of parents or siblings is not related.

In DHS surveys, the all-women factors have been calculated for a few subgroups (total, urban/rural, region, and education groups) and are available in the individual recode data files. In most other surveys the all-women factors will have to be calculated from the household data prior to analysis of the mortality data.

## DISAGGREGATION

Disaggregation of data permits the estimation of the levels of mortality for subpopulations and the study of differentials in mortality between different groups within the population. One dimension of disaggregation, age, is inherently built into the indicators under discussion. Each mortality rate is described as the probability of dying between age  $x$  and age  $x+n$  where  $x$  and  $n$  are dependent on the indicator in question (e.g.,  $5q_0$  or  $35q_{15}$ ). Additional age breakdowns of the mortality rates are possible; however, the indicators that have been discussed in this paper cover the principal indicators used in the study of mortality.

A second key dimension, sex, has also been mentioned previously. Mortality risks are known to differ by sex both early in life and late in life and it is recommended that all estimates of mortality be produced separately for each sex as well as for both sexes together.

Several other dimensions can be considered as potential variables for the disaggregation of mortality rates, including socio-economic characteristics such as urban/rural residence, region of residence, level of education, household income, household wealth quintiles, ethnicity, language, religion or occupation. Additionally, certain demographic characteristics are useful for the study of child mortality, such as age of mother at birth, birth order, birth size, or previous birth interval.

One of the problems with the analysis of mortality from survey data is that socio-economic characteristics are available for the respondent to the interview, but not for the deceased or surviving relatives for whom the mortality rates are calculated. For example, for adult mortality rates, residence, level of education, household wealth, and household income are not ascertained for the siblings or parents of the respondent, only for the respondent.

Even for child mortality, the information available is not about the child himself or herself, but about the respondent and the household in which the respondent currently lives. However, for child mortality it is generally assumed that the information obtained from the respondent is applicable to the children, particularly those under the age of five.

For adult relatives, this assumption is much less likely to be valid. Particularly with the level of urban migration seen in many countries, either the respondent or one or more of their relatives may have migrated and thus the mortality measured may be for a different location than the location in which the respondent is living.

Only the recent household deaths approach to adult mortality estimation avoids this problem. Using this method the location of the relatives is known and thus these data may be associated with particular locations, and with other characteristics of the household.



For the demographic characteristics, the only known characteristics for adult relatives are typically the sex and in some cases the age (or age at death) of the relative. Other demographic characteristics are not known. For child mortality a few other characteristics are known and may be used with direct estimation of child mortality.

In general, care should be taken in disaggregating the mortality estimates as the sampling errors around estimates increase quickly as the sample size is reduced. It is recommended to disaggregate the data into the fewest groups possible. The DHS programme produces only 10-year estimates for disaggregated data. Similarly the RHS surveys present only 10-year estimates in their reports. Disaggregated estimates for 5-year periods and for indirect methods are possible, but care should be taken to ensure a sufficiently large sample size. Preferably, the groups into which the data are disaggregated should be dichotomous. The MICS produce disaggregated estimates of child mortality using indirect methods, but generally only with dichotomous variables.

Taking each of the socio-economic characteristics mentioned above in turn:

- Urban/rural residence – it is recommended that this breakdown be used for child mortality. For adult mortality this has been used with the recent household deaths approach, but because of the issues of location of the relatives it is not typically used for the other methods.
- Region of residence – most household survey samples are not sufficiently large to present results by individual regions and it is frequently necessary to group regions to produce sufficiently large samples when producing mortality estimates. Even using grouped regions, most surveys publish only 10-year estimates of child mortality. For adult mortality, region may be used with the recent household deaths method, but the issue of location makes it unusable for other methods.
- Level of education – the mother’s level of education is often used for disaggregating rates, and is a powerful predictor of differences in child mortality rates. However it is difficult to recommend a consistent point of disaggregation due to varying levels of education in different countries. In some countries with high levels of mortality, education can be dichotomized into those with no education and those with some education. In countries with a medium level of mortality, the most useful dichotomy is often between those with some secondary education versus those with no secondary education, and in countries with lower levels of mortality one might compare respondents with only compulsory education (or less) to those with higher levels of education. Level of education of the mother is only applicable for child mortality. Level of education of the father could also be used, but is less directly correlated with mortality levels and less frequently collected in household surveys.
- Household income – the measurement of household income is quite complicated and, among the survey programmes discussed here, is only done in LSMS surveys. Due to the relatively few surveys collecting household income, it is not recommended that this variable be used for disaggregation within the database.
- Household wealth quintiles have been used as a proxy for household income in a number of survey programmes in recent years, including DHS and MICS. The household wealth quintiles are based on an asset index that is derived using principal component analysis from data on household assets. DHS uses the household wealth quintiles directly for disaggregation of child mortality in 10-year rates. For 5-year rates and for indirect estimates, it is recommended that the quintiles be grouped into a dichotomous variable of the poorest 60 per cent and the richest 40 per cent. This dichotomous disaggregation could also be used with the recent household deaths approach for adult mortality, but would not be applicable for other approaches to adult mortality.

- Ethnicity, language and religion – these variables have been used in various surveys to disaggregate mortality rates. However, their use raises a number of potential difficulties. Firstly, there is considerable sensitivity concerning the issues of ethnicity, language and religion in many countries. Secondly, these variables are probably best used dichotomized into the major group and all other groups, due to sample size issues, but this can also add to the sensitivities concerning the results. These variables are not collected in all surveys, but have been collected in a few DHS surveys and a number of the MICS3 surveys. Due to the limited number of surveys collecting these data, and the issues around using this data, the recommendation would be not to use ethnicity, language or religion variables for disaggregation in the database.
- Occupation – variables related to occupation were collected throughout the WFS surveys and have been collected in most DHS surveys. However, they have rarely been used for disaggregating mortality rates principally because they do not discriminate well between groups. Dichotomized versions of occupation variables have occasionally been used, with the groups being agricultural occupations and non-agricultural occupations, but again these have not proven very useful. The recommendation would be not to use occupation for disaggregation of child mortality.

Consideration could be made as to whether to produce disaggregated estimates of mortality broken down by sex within each group. While this would increase the sampling error problems, sex disaggregated mortality rates for subgroups could potentially be useful. It should be noted, though, that because of the sampling errors any differences between sexes are unlikely to be statistically significant.

### **Database contents**

The database should be designed to maintain the key indicators described in earlier sections, for the most common time periods as presented in the discussion. Estimates should be disaggregated to the extent possible.

To provide the most flexibility in data analysis, it is also recommended that the database maintain the numerators, denominators and unweighted denominators used in each method, as well as the resulting mortality rates. For the direct estimation methods, it is recommended that these be stored by single year preceding the survey. The reasons for doing this are twofold:

1. Providing the numerators and denominators permits the calculation of indicators other than those selected for inclusion if the database. For example, for direct child mortality estimation it would also permit the calculation of  ${}_2q_0$  or  ${}_3q_0$ , indicators of childhood mortality that are occasionally used, but are not common.
2. Providing the numerator and denominators by single year periods permits the flexibility to aggregate the data across different periods and produce, say, 5-year, 7-year and 10-year rates from the same data.

An example of the data to be stored in the database is shown in table 17.

TABLE 17. PARTIAL EXAMPLE OF DATA FILE OF NUMERATORS AND DENOMINATORS BY SINGLE YEAR PERIOD, AGE GROUP, SEX AND BACKGROUND VARIABLE FOR INPUT INTO DATABASE

<i>Survey name</i>	<i>Backvar</i>	<i>Backval</i>	<i>Sex</i>	<i>Age group</i>	<i>Period</i>	<i>Deaths</i>	<i>Deaths (unweighted)</i>	<i>Exposure</i>	<i>Exposure (unweighted)</i>
Armenia DHS 2005	Total	0	3	0	0	2.61	3	337.26	315.5
Armenia DHS 2005	Total	0	3	1	0	0.88	1	321.36	297.5
Armenia DHS 2005	Total	0	3	2	0	1.16	1	303.27	277
Armenia DHS 2005	Total	0	3	3	0	0.19	0.5	307.71	280.5
Armenia DHS 2005	Total	0	3	4	0	1.29	1	310.03	290
Armenia DHS 2005	Total	0	3	5	0	0.97	1	290.66	279
Armenia DHS 2005	Total	0	3	6	0	0	0	266.45	262
Armenia DHS 2005	Total	0	3	7	0	0	0	255.67	262.5
Armenia DHS 2005	Total	0	3	0	1	7.67	4.5	328.95	299.5
Armenia DHS 2005	Total	0	3	1	1	1.77	2.5	334.25	304
Armenia DHS 2005	Total	0	3	2	1	0.63	1	346.17	310
Armenia DHS 2005	Total	0	3	3	1	0.19	0.5	326.31	301.5
Armenia DHS 2005	Total	0	3	4	1	0	0	290.66	279
Armenia DHS 2005	Total	0	3	5	1	0	0	266.45	262
Armenia DHS 2005	Total	0	3	6	1	0	0	255.67	262.5
Armenia DHS 2005	Total	0	3	7	1	0	0	272.48	270
...	...	...	...	...	...	...	...	...	...

<sup>a</sup> Backvar refers to background variable and backval refers to background value.

This table can also be easily transformed into a format such as the following that is more amenable for importing directly into a relational database as shown in table 18.

TABLE 18. TRANSFORMED DATA FILE OF NUMERATORS AND DENOMINATORS BY SINGLE YEAR PERIOD, AGE GROUP, SEX AND BACKGROUND VARIABLE FOR INPUT INTO RELATIONAL DATABASE

<i>Survey name</i>	<i>Backvar</i>	<i>Backval</i>	<i>Sex</i>	<i>Age group</i>	<i>Period</i>	<i>Variable</i>	<i>Value</i>
Armenia DHS 2005	Total	0	3	0	0	Deaths	2.61
Armenia DHS 2005	Total	0	3	0	0	exposure	337.26
Armenia DHS 2005	Total	0	3	1	0	Deaths	0.88
Armenia DHS 2005	Total	0	3	1	0	exposure	321.36
Armenia DHS 2005	Total	0	3	2	0	Deaths	1.16
Armenia DHS 2005	Total	0	3	2	0	exposure	303.27
Armenia DHS 2005	Total	0	3	3	0	Deaths	0.19
Armenia DHS 2005	Total	0	3	3	0	exposure	307.71
Armenia DHS 2005	Total	0	3	4	0	Deaths	1.29
Armenia DHS 2005	Total	0	3	4	0	exposure	310.03
Armenia DHS 2005	Total	0	3	5	0	Deaths	0.97
Armenia DHS 2005	Total	0	3	5	0	exposure	290.66
Armenia DHS 2005	Total	0	3	6	0	Deaths	0
Armenia DHS 2005	Total	0	3	6	0	exposure	266.45
Armenia DHS 2005	Total	0	3	7	0	Deaths	0
Armenia DHS 2005	Total	0	3	7	0	exposure	255.67
...	...	...	...	...	...	...	...

There are other forms that should also be considered for storing the numerators and denominators. For example, numerators and denominators could be stored by single calendar year as well as by single year preceding the survey to permit the calculation of mortality rates for particular

calendar year periods. Also, for the direct estimation of child mortality the numerators and denominators could be stored by single month age group for ages under 2 years to permit the grouping of the data in different ways than that used by DHS. In theory this could also be applied to smaller time periods in the first month to permit the calculation of mortality rates, say, in the first seven days of life.

For indirect estimation methods, the numerators and denominators should be stored in order to permit the re-calculation of rates according to different life tables than the life table selected for the estimates included in the database. One of the reasons for doing this is that the judgments as to the most applicable model life table may change over time as new data become available.

#### AGGREGATION RULES

For the numerators and denominators described above to be truly useful, the database would also need to store the rules for the aggregation of the data to produce the final mortality rates. For the direct estimation methods this would mean storing the rules for compounding the age specific mortality rates to produce the final rates desired. For the indirect methods, this would mean implementing each of the indirect methods for child and adult mortality, based on the aggregate numerators and denominators.

The database would also need to be capable of aggregating the single year numerators and denominators of the direct approaches into periods desired for the rates. For example, for direct estimates of child mortality, assuming the numerators and denominators have been stored by single year period (see table 5), then the aggregation rules needed would be:

1. Sum the numerators and denominators over the time period desired (typically five or ten years).
2. Calculate the probability of dying at each age group by dividing the numerators by the denominators for each age group.
3. Calculate the mortality rates as one minus the product of the survival probabilities (the probability of surviving is one minus the probability of dying) at each age group for the rates of interest (see pages 8-9).

Considerable care would have to be taken, though, to ensure that mortality rates were calculated on sufficiently large sample sizes to produce reliable rates. A good rule of thumb for the direct estimation of child mortality is a minimum of 500 unweighted cases of exposure in age group 0.

#### PUBLISHED DATA VS. RE-ANALYZED DATA

If it is not possible to re-analyze the original survey data to produce the recommended data for storage in the database, then it may be necessary to limit the database contents to previously published estimates. In many cases only the final mortality estimates are published, in which case these could be stored in the database. However, it is not unusual for the numerators and denominators to be published along with the mortality rates and these could also be stored in the database. By including the aggregation rules described above in the database design, the database would then permit the calculation of a variety of different indicators from those numerators and denominators, such as estimates based on a different life table, providing greater information than in the published material.

## Survey sources

The following major survey programmes have been reviewed as sources of data for the estimation of child and adult mortality:

- Demographic and Health Surveys (DHS)
- Reproductive Health Surveys (RHS)
- Multiple Indicator Cluster Surveys (MICS)
- World Health Survey (WHS)
- Living Standards Measurement Studies (LSMS)
- Pan Arab Programme on Family Welfare (PAPFAM/PAPCHILD)
- Gulf Health Surveys (GHS)
- Contraceptive Prevalence Surveys (CPS)
- World Fertility Surveys (WFS)

### DEMOGRAPHIC AND HEALTH SURVEYS

The Demographic and Health Surveys (DHS) are managed by Macro International Inc. under contract with the U.S. Agency for International Development. Since 1984, Measure DHS (also known as the Demographic and Health Surveys) has carried out more than 200 surveys in over 75 countries. The DHS surveys usually have sample sizes of between 5,000 and 30,000 households per country and typically include interviews of all women aged 15-49 years and men aged 15-54 years in selected households. These surveys measure indicators such as fertility, contraceptive use, infant and child mortality, child health, maternal morbidity and mortality, and knowledge and attitudes about HIV/AIDS. Surveys are typically conducted roughly every five years in most countries to allow comparisons over time. Macro International also provides a range of survey products other than the standard DHS, that includes Interim Surveys, Key Indicator Surveys (KIS), AIDS Indicator Surveys (AIS), Malaria Indicator Surveys (MIS) and Service Provisions Assessments (SPA).

#### *Data availability*

Data are made available through the Measure DHS website ([www.measuredhs.com](http://www.measuredhs.com)) and are generally freely available online to bona fide researchers. Applications for access to data maybe made online and are approved on a country by country basis. Applications require a brief description (or abstract) of the project or the analysis to be carried out. Restrictions exist in access to data for certain surveys and these require specific approval from the executing agency in the country that conducted the survey. The DHS data archive provides assistance in requesting permission for access to restricted surveys.

Datasets are available in several formats, including their native format ISSA/CSPRO hierarchical ASCII files, as well as in SPSS, Stata and SAS formats.

#### *Data content and limitations*

All standard DHS surveys include full birth histories<sup>8</sup> as well as the reproduction summary (or Brass questions) that permit the calculation of both direct and indirect estimates of child mortality (using age, duration of marriage and time since first birth methods). DHS final reports present only direct child mortality estimates and indirect estimates are not produced. National level estimates are

produced for several five-year periods preceding the survey, but disaggregated estimates are only presented for a single ten-year period.

The maternal mortality module from which adult mortality can be calculated using the sibling history is an optional module that has been used in a number of DHS surveys. Indirect estimates using the sibling survivorship method can also be calculated from these datasets. When the sibling histories are used, the DHS final reports usually provide direct sex- and age-specific adult mortality estimates for five-year age groups for one or two seven-year periods, as well as an age-standardized general mortality rate for ages 15-49 combined.

A number of surveys in the first phase of DHS (DHS I – 1984-1989) included the questions for the maternal and paternal orphanhood method for adult mortality, and most of these also included questions on orphanhood before and since marriage. Adult mortality estimates from these questions were not produced in the DHS final reports, however.

#### *Other survey-specific issues*

DHS uses *de facto* samples of individual women and selects all *de facto* women in each selected household for interview. Household listings include both *de facto* and *de jure* household members, and require selection of the correct household members for analysis. Dates of key events are imputed as necessary and the imputed dates are available in the datasets. All-woman factors exist in the datasets for surveys with ever-married samples.

### REPRODUCTIVE HEALTH SURVEYS

Reproductive Health Surveys (RHS) are managed through an agreement between the Centers for Disease Control (CDC)/Division of Reproductive Health (DRH) and the U.S. Agency for International Development (USAID). Since the mid-1990s, the RHS programme has conducted over 30 nationally-representative household surveys primarily in Latin America and the Caribbean, Central and Eastern Europe and the former Soviet Union. Typical sample sizes range between 5000 and 7000 households, and include interviews with one selected woman aged 15-49 years per household. Separate surveys of men are also carried out in some countries. The surveys of women measure indicators such as fertility, contraceptive use, infant and child mortality, child health, maternal morbidity and mortality, and knowledge and attitudes about HIV/AIDS and sexually-transmitted infections. Four types of surveys are now available: the International/Reproductive Health Surveys (RHS) of reproductive-age women, Male Reproductive Health Surveys, Young Adult Reproductive Health Surveys and Reproductive Health Programme Evaluation Surveys.

#### *Data availability*

The datasets from the Reproductive Health Surveys fall into two main groups – surveys from Latin America and those from Eastern Europe and the Commonwealth of Independent States. The surveys from Latin America are generally freely available upon request, but the surveys from Eastern Europe require written permission from the executing agency in the country that conducted the survey.

Data files are available as SAS datasets, e.g.:

cc_yyyy_creadas.sas7bdat	-	Women's data with household data attached
cc_yyyy_hijos.sas7bdat	-	Children's data from birth or pregnancy history

RHS reports and datasets may be ordered from the CDC website at:  
<http://www.cdc.gov/reproductivehealth/Surveys/SurveyCountries.htm>

#### *Data content and limitations*

Most Reproductive Health Surveys include a full birth or pregnancy history from which direct estimates of childhood mortality can be produced. Additionally, most of these surveys also include the reproduction summary questions from which indirect estimate of child mortality can be produced. RHS final reports typically include direct estimates of child mortality for a single ten-year period only, and do not present indirect estimates.

#### *Other survey-specific issues*

In the RHS surveys a single woman is selected per household for interview. When men's surveys are collected, these are collected as separate samples in separate households.

### MULTIPLE INDICATOR CLUSTER SURVEYS

Multiple Indicator Cluster Surveys (MICS) are household surveys developed by UNICEF to assist countries in filling data gaps for monitoring the situation of children and women. MICS was originally developed in response to the World Summit for Children to measure progress towards an internationally agreed set of mid-decade goals. The first round of MICS (MICS1) was conducted around 1995 in more than 60 countries. A second round of surveys (MICS2) was conducted in 2000 (around 65 surveys). The third round (MICS3) focused on providing a monitoring tool for the World Fit for Children, the Millennium Development Goals (MDGs), as well as for other major international commitments and was conducted in 2005/6 in around 50 countries. Survey questionnaires are modular tools and consist of 3 questionnaires: a household questionnaire, a questionnaire for women aged 15-49, and a questionnaire for children under the age of 5. Survey content includes child mortality, nutrition, child health, reproductive health, water and sanitation, early child development, education, child protection and HIV.

#### *Data availability*

Datasets are available online for the majority of the MICS2 and MICS3 surveys (<http://www.childinfo.org/mics.html>). Applications for a username and password may be made online and will be sent to the researcher within a week. Executing agencies from MICS2 and MICS3 from certain countries have either not provided the data to UNICEF New York or not provided permission to distribute the data. No datasets are available at UNICEF New York for the MICS1 surveys.

Data files are available as four SPSS datasets, including:

HH.SAV	Household data
HL.SAV	Household listing data
WM.SAV	Women's data
CH.SAV	Children under five's data
BH.SAV	Birth history data (where applicable).

### *Data content and limitations*

The standard MICS surveys include the reproduction summary questions plus questions on the date of first and last birth but do not include birth histories. Thus the surveys permit the calculation of indirect estimates of child mortality using either the age or time since first birth method. A small number of MICS3 surveys, including those in Iraq, Malawi, Somalia, and Sudan, also collected full birth histories permitting direct estimation of child mortality.

MICS surveys also include an optional maternal mortality module that can also be used for the calculation of female adult mortality using the sibling survivorship indirect method.

### *Other survey specific issues*

MICS2 & 3 surveys use *de jure* samples and selects all women aged 15-49 and all children under five for the individual interviews.

## WORLD HEALTH SURVEYS

World Health Surveys (WHS) were developed and implemented by the World Health Organization and have been completed in 70 countries since 2002, the year of the first round. The WHS offers a choice of modules covering a) health states of populations, measuring health in multiple domains; b) risk factors (e.g. tobacco, alcohol, pollution) and their association with health states; c) responsiveness of health systems (whether health systems meet the legitimate expectations of people); d) coverage, access and utilization of key health services; and e) health care expenditures. The WHS recommended sample size is 5,000 people per country with one eligible person randomly selected per household – a male or female 18 years of age or older). However, sample sizes have varied between 1,000 and 10,000 people for each country survey.

### *Data availability*

WHS survey data are available online for download (<http://www.who.int/healthinfo/survey/whsresults/en/index.html>). All datasets are freely available to bona fide researchers. To download the datasets, researchers must first submit a WHS data use agreement to WHO in order to receive a username and password. The following files are included:

countryname-ID.dta	Household identification and weighting
WHS_countryname_F2.dta	Rest of the household data
WHS_countryname_F3.dta	Household roster
WHS_countryname_F4.dta	Household occupation
WHS_countryname_F5.dta	Rest of the individual data
WHS_countryname_F6.dta	Birth history data
WHS_countryname_F7.dta	Sibling data

Data files are available as Stata datasets.

### *Data content and limitations*

Most WHS surveys include a maternity history and a sibling history, permitting analysis of child mortality and adult mortality via direct estimation. However, WHO researchers have questioned



the quality of the reporting in the maternity histories and have not produced child mortality estimates from these data. Preliminary analysis of some of the surveys suggests considerable under-reporting of child deaths. Likewise, WHO has not published estimates of adult mortality based on the WHS sibling history data.

#### *Other survey specific issues*

WHS surveys use a *de jure* sample and interview one female or male respondent over the age of 18 in each household.

Sex of children in the maternity histories is coded 1=Female, 2=Male, contrary to the norm in other household survey programmes. Dates of events are stored as strings and require parsing to extract the month and year of the event. It appears that the string is empty if the year of the event is unknown. If the year is known but the month is not known then the month is set to 01. Unfortunately it appears impossible to differentiate between events that took place in January and events for which the month was not known.

### LIVING STANDARD MEASUREMENT SURVEYS

Living Standard Measurement Surveys (LSMS) are managed by the Development Economics Research Group (DECRG) of the World Bank. The LSMS started as a research project in the early 1980s in response to a perceived need for policy relevant data to allow policy makers to move beyond measuring rates of unemployment, poverty, and health care use, for example, to understanding the determinants of these observed social sector outcomes. Since 1985, over 80 LSMS surveys have been carried out in 32 countries. Sample sizes range from 1,500 to 11,280 households per survey. The LSMS surveys collect household data that can be used to assess household welfare, to understand household behaviour, and to evaluate the effect of various government policies on the living conditions of the population. Accordingly, LSMS surveys collect data on many dimensions of household well-being, including consumption, income, savings, employment, health, education, fertility, nutrition, housing and migration.

#### *Data availability*

The LSMS survey documentation and datasets are available online; however, there are different levels of access to certain datasets. Some are freely available online after completing a data user agreement, while others require permission to be obtained from the country before the World Bank can provide the data. Older datasets may be provided on CD-ROM. To ascertain the access policy and procedures for a given survey, select the survey from the LSMS Dataset Finder (<http://iresearch.worldbank.org/lsmssurveyFinder.htm>).

#### *Data content and limitations*

Not all LSMS collect information relevant to mortality estimation. The online Dataset Finder allows users to search surveys by topic. Some LSMS provide maternity histories, allowing direct estimates of child mortality, while others only include the summary reproduction questions that can be used for indirect estimates. LSMS surveys do not include information on adult mortality.

### *Other survey specific issues*

LSMS surveys typically used roster based household questionnaires and interview *de jure* residents of the household. All *de jure* women aged 15-49 are interviewed for the fertility section.

#### PAN ARAB PROGRAMME ON FAMILY WELFARE

The Pan Arab Programme on Family Health (PAPFAM) and its predecessor programme, the Pan Arab Programme on Child Welfare (PAPCHILD) are executed by the League of Arab States (LAS) and are supported by a number of donors, including AGFUND, UNFPA, OPEC Fund, WHO, IOMS, IPPF, UNICEF and ESCWA. PAPFAM is designed to make available a set of social, economic, demographic and health indicators needed to formulate, design and monitor development and health programs and policies at the country level; to build the regional family health and reproductive health data bank; and publish in-depth analytical and comparative studies that will assist in using the findings for policy formulation and programme monitoring and evaluation at the regional level. Eight surveys (Algeria 2002, Djibouti 2002, Lebanon 2004, Morocco 2003-2004, Occupied Palestinian Territory 2006, Sudan 2006, Syria 2001, and Tunisia 2001) have been completed since 2001. Under the earlier PAPCHILD project, 10 surveys were carried out between 1991 and 1997: Algeria 1992, Egypt 1991, Lebanon 1996, Libya 1995, Mauritania 1990, Morocco 1997, Sudan 1992-1993, Syria 1993, Tunisia 1994-1995, Yemen 1991-1992.

### *Data availability*

Datasets are not available online. According to PAPFAM, to access the micro data, they should be requested officially from the participant countries. The request should include the objectives of the research, ensuring that a copy of the output of the analysis will be shared with the country before finalising. This can be done through PAPFAM and PAPFAM will take the necessary actions with the countries. Contact information for PAPFAM may be found at: <http://www.papfam.org/pap/English/engindex.html>.

### *Data content and limitations*

The questionnaire content for the PAPFAM surveys is similar to that of the DHS surveys and typically includes a full birth history. PAPFAM reports include the child mortality estimates for the five-year period preceding the survey. The methodology used by PAPFAM for mortality estimation is the same as that of DHS. Some surveys also include the reproduction summary questions and permit the calculation of indirect estimates.

PAPFAM surveys do not include questions for the calculation of adult mortality.

### *Other survey specific issues*

Most PAPFAM surveys use ever-married samples, so any indirect estimation of child mortality based on age would require the use of all-women factors to adjust the data.

#### GULF FAMILY HEALTH SURVEYS

Gulf Family Health Surveys (GFHS) and its predecessor programme, the Gulf Child Health Surveys (GCHS) were household survey programmes conducted from 1995-1998 and 1987-1989,

respectively. The earlier programme focused on child health, while the latter added information on the health status of adults. The GFHS comprised four questionnaires including a household questionnaire collecting socio-economic and environmental data about the households; a household members questionnaire concerning the health status of each member; an individual questionnaire concerning the reproductive health of women aged 15-49; and a questionnaire for children under five concerning their health status.

The surveys were carried out in six Gulf States in each programme: Bahrain 1989, Kuwait 1987, Oman 1988-89, Qatar 1987, Saudi Arabia 1987, and the United Arab Emirates 1987 in the GCHS; Bahrain 1995, Kuwait 1996, Oman 1995, Qatar 1998, Saudi Arabia 1996, and the United Arab Emirates 1995 in the GFHS.

#### *Data availability*

Data from the Gulf Family Health Survey data and the Gulf Child Health Survey datasets are not known to be available for further analysis. Users may wish to contact the individual institutions that executed the surveys to request access to the datasets.

#### *Data content and limitations*

The GFHS surveys all included birth histories and would permit the estimation of child mortality using direct estimation methods, while the majority of the GCHS surveys only included the reproduction summary questions. The surveys did not include modules for adult mortality.

#### *Other survey specific issues*

The GFHS and the GCHS surveys interviewed only national populations and excluded non-nationals from the survey.

### CONTRACEPTIVE PREVALENCE SURVEYS

Contraceptive Prevalence Surveys (CPS) were carried out between 1977 and 1985 to study fertility and family planning in developing countries. Most CPS were conducted by a division of Westinghouse Public Applied Systems (which became part of Macro International in 1987 and is now responsible for the DHS) and were funded by the United States Agency for International Development (USAID). In parallel, the CDC Division of Reproductive Health also carried out several CPS surveys. The CPS and WFS programmes were the immediate predecessor programmes that led to the DHS programme.

#### *Data availability*

Raw data files for the CPS conducted by Westinghouse can be requested by contacting the DHS data archive. Most of these CPS surveys are available, although the datasets are in inconsistent formats. For the CPS that were executed by CDC, that agency must be contacted to request access to the datasets.

#### *Data content and limitations*

The CPS surveys typically included the reproduction summary questions that permit the calculation of indirect estimates of child mortality. A small number of surveys included a birth history

and permit direct estimation of child mortality. The CPS surveys did not include questions related to adult mortality.

#### *Other survey specific issues*

The survey data from the CPS are not well archived or documented and the datasets are not in any kind of standardized format. Many of the datasets require documenting and further editing of the data to make the survey data files usable, and therefore producing mortality estimates from these datasets would entail a considerable amount of work with each survey.

### WORLD FERTILITY SURVEYS

The World Fertility Surveys (WFS) were carried out between 1972 and 1984 under the auspices of the International Statistical Institute with funding from UNFPA, USAID and the United Kingdom Overseas Development Administration (ODA). Forty two surveys were conducted in developing countries with twenty more surveys conducted separately by national governments in developed countries. The developing country surveys covered topics including fertility, contraception, family planning, marriage, and occupation. The programme ended more than 20 years ago, but set the standard for all household surveys carried out since then.

#### *Data availability*

The main standard recode datasets for the 42 WFS conducted in developing countries are available online through the Office of Population Research at Princeton University at <http://opr.princeton.edu/archive/wfs/>. Copies of these files and of a number of other WFS files, including some of the household data, are also available from the author. A CD was produced by the Dynamic Data Base (DDB) at the International Statistical Institute in 1993, but this CD contains less than half of the developing country surveys together with a number of developed country datasets.

The main datasets available are:

ccSRnn.dat	-	Standard recode data from the women's questionnaire
ccSSnn.dat	-	Supplemental recode data from the women's questionnaire.

Additionally, household and household member data files are available for some surveys.

#### *Data content and limitations*

The 42 surveys in developing countries were based on standard questionnaires and are available as standard recode files, using the same variable names across the files. The 20 surveys in developed countries did not use standardized questionnaires and as a result the datasets are less uniform.

Each of the developing country surveys included a full birth history together with the reproduction summary questions, allowing for both direct and indirect estimation of mortality.

The surveys in Cameroon, Jordan, Lesotho, Mauritania, Mexico, Morocco, Peru, Philippines, Republic of Korea, Sudan (North), and Syria also included questions for indirect estimation of adult mortality.

### Other survey specific issues

The surveys used a *de facto* sample of women aged 15-49 in most surveys. Several surveys used ever-married samples rather than all-women samples and thus require the use of all-women factors for indirect estimation of child mortality.

---

#### NOTES

<sup>1</sup> Throughout this paper, the terms “child” and “childhood” can be used interchangeably to refer to the period from birth to age five. More specifically, when referring to the CMR or 4q1, the reference period is between exact ages one and five.

<sup>2</sup> The post-neonatal mortality “rate” is not strictly a rate; by convention, it is the arithmetic difference between the infant and neonatal mortality.

<sup>3</sup> The century month code (CMC) as used in most surveys is defined as the number of months from the start of the 20th century, with January 1900 = CMC 1, February 1900 = CMC 2, ..., January 2000 = CMC 1201. The general formula is  $CMC = (YYYY - 1900) * 12 + MM$ , where YYYY is the year of the event and MM is the month of the event within that year. Note that in some RHS surveys the CMC has been calculated with a different base as  $CMC = YYYY * 12 + MM$ , with CMC 0 = January 0000, and January 2000 = CMC 24001.

<sup>4</sup> This is actually the upper bound of the age interval in months for each age. The bounds are actually  $(x-1, x)$ .

<sup>5</sup> The mean age at maternity (M) may be calculated based on all births in a fixed period, say the last year or last 5 years. However in place of the actual mean age, a value of 27 years is commonly used. See Annex III of Manual X (United Nations, 1983) for details concerning the calculation of the mean age at childbearing.

<sup>6</sup> *De jure* residents are the usual residents of the household. *De facto* residents are persons that slept in the household the night before the interview. Using both *de facto* and *de jure* residents would potentially permit double counting of residents as they may be counted in more than one household. To avoid this problem, analysis is performed using either the *de facto* or the *de jure* residents, but not the combination of the two.

<sup>7</sup> Note that the term orphanhood is used to mean the loss of one or both parents, not solely both parents.

<sup>8</sup> The two earliest DHS surveys in El Salvador (1986) and Liberia (1986) used a truncated 5 year birth history, but since that time all have included full birth or pregnancy histories.

## REFERENCES

- Bicego, G., and O. B. Ahmad (1996). Infant and Child Mortality. *DHS Comparative Studies*, No. 20. Calverton, Maryland: Macro International Inc. Available from [http://www.measuredhs.com/pubs/pub\\_details.cfm?ID=28](http://www.measuredhs.com/pubs/pub_details.cfm?ID=28).
- Brass, W. (1964). Uses of Census and Survey Data for the Estimation of Vital Rates. Paper prepared for the African Seminar on Vital Statistics, Addis Ababa, 14-19 December 1964. E/CN.14/CAS.4/V57.
- Brass, W. and K. Hill (1973). Estimating adult mortality from orphanhood. *Proceedings of the International Population Conference*, Liège, vol. 3, pp. 111-123. Liège: International Union for the Scientific Study of Population.
- Croft, T. (1991). DHS Data Editing and Imputation. *Demographic and Health Surveys World Conference Proceedings*, II, pp. 1337-1356, Columbia, Maryland: IRD/ORC Macro. Available from [http://www.measuredhs.com/pdfs/DHS\\_Data\\_Editing\\_Tcroft.pdf](http://www.measuredhs.com/pdfs/DHS_Data_Editing_Tcroft.pdf).
- Croft, T., and K. Hill (2008). Indirect Estimation of Child Mortality: A Reappraisal. (Unpublished).
- Hill, K., and J. Trussell (1977). Further Developments in Indirect Mortality Estimation. *Population Studies*, vol. 31, No. 2 (July), pp. 313-334.
- Hill, K., and M-E. Figueroa (1999). Child Mortality Estimation by Time Since First Birth. Johns Hopkins Population Center Papers on Population WP 99-05. Baltimore, Maryland: Johns Hopkins Population Center. Available from <http://www.jhsph.edu/popcenter/publications/pdf/WP99-05.pdf>.  
Also in Brass Tacks: Essays in Medical Demography, edited by B. Zaba and J. Blacker. London: The Athlone Press.
- Hill, K. (2001). Methods for Measuring Adult Mortality in Developing Countries: A Comparative Review. The Global Burden of Disease 2000 in Aging Populations, Research Paper No. 01.13. Cambridge, MA: Harvard Burden of Disease Unit, Center for Population and Development Studies. Available from <http://www.hsph.harvard.edu/burdenofdisease/publications/papers/Methods%20for%20Measuring%20Adult%20Mortality.pdf>.
- \_\_\_\_\_ (2003). Adult Mortality in the Developing World; What we Know and How we Know it. Training Workshop on HIV/AIDS and Adult Mortality in Developing Countries, UN/POP/MORT/2003/1, New York. September 8-13, 2003. Available from [http://www.un.org/esa/population/publications/adultmort/HILL\\_Paper1.pdf](http://www.un.org/esa/population/publications/adultmort/HILL_Paper1.pdf).
- Hill, K., and Y. Choi (2004). The Adult Mortality in Developing Countries Project: Substantive Findings. Paper prepared for Adult Mortality in Developing Countries Workshop, July 2004. Available from [http://www.ceda.berkeley.edu/events/AMDC\\_Papers/Hill\\_Choi\\_Summary-amdc.pdf](http://www.ceda.berkeley.edu/events/AMDC_Papers/Hill_Choi_Summary-amdc.pdf).
- Mahy, M. (2003). Childhood Mortality in the Developing World: A Review of Evidence from the Demographic and Health Surveys. *DHS Comparative Reports*, No. 4. Calverton, Maryland: ORC Macro. Available from [http://www.measuredhs.com/pubs/pub\\_details.cfm?ID=450](http://www.measuredhs.com/pubs/pub_details.cfm?ID=450).
- Measure DHS (2007). *Guidelines for the MEASURE DHS Phase II Main Survey Report*. Calverton, Maryland: Macro International Inc. Available from [http://www.measuredhs.com/pubs/pub\\_details.cfm?ID=717](http://www.measuredhs.com/pubs/pub_details.cfm?ID=717).
- Trussell, J., and G. Rodriguez (1990). A note on the sisterhood estimator of maternal mortality. *Studies in Family Planning*, No. 21, pp. 344-346.
- Rutstein, S. O. (1983). Infant and child mortality: Levels, trends, and demographic differentials. *World Fertility Survey Comparative Studies*, No 24. Voorburg, Netherlands: International Statistical Institute.

- \_\_\_\_\_ (1984). Infant and child mortality: Levels, trends, and demographic differentials. *World Fertility Survey Comparative Studies*, No. 43. Voorburg, Netherlands: International Statistical Institute.
- 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)
- Rutenberg, N., and J. Sullivan (1991). Direct and indirect estimates of maternal mortality from the sisterhood method. In *Proceedings of the Demographic and Health Surveys World Conference*, Washington DC, 1991. vol. 3, pp. 1669-1696. Columbia, Maryland: IRD/Macro International Inc.
- Stanton, C., N. Abderrahim, and K. Hill (1997). DHS Maternal Mortality Indicators: An Assessment of Data Quality and Implications for Data Use. *Demographic and Health Surveys Analytical Report*, No. 4. Calverton, Maryland: Macro International Inc.
- Sullivan, J. M., G. T. Bicego, and S. O. Rutstein (1990). Assessment of the quality of data used for the direct estimation of infant and child mortality in the Demographic and Health Surveys. In *Assessment of DHS-I data quality*. DHS Methodological Report No. 1. Columbia, Maryland: IRD/Macro Systems, Inc. Available from [http://www.measuredhs.com/pubs/pub\\_details.cfm?ID=142](http://www.measuredhs.com/pubs/pub_details.cfm?ID=142).
- Sullivan, J. M., S. O. Rutstein, and G. T. Bicego (1994). Infant and child mortality. *DHS Comparative Studies*, No. 15. Calverton, Maryland: Macro International Inc. Available from [http://www.measuredhs.com/pubs/pub\\_details.cfm?ID=23](http://www.measuredhs.com/pubs/pub_details.cfm?ID=23)
- Timæus, I., and W. Graham (1989). Measuring adult mortality in developing countries: a review and assessment of methods. *Policy Research Working Paper Series*, No. 155. Washington DC: The World Bank. Available from [http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/1989/04/01/000009265\\_3960927213856/Rendere d/PDF/multi0page.pdf](http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/1989/04/01/000009265_3960927213856/Rendere d/PDF/multi0page.pdf)
- Timæus, I. (1991a). Estimation of Mortality from Orphanhood in Adulthood. *Demography*, Vol. 28, No. 2, pp. 213-227.
- \_\_\_\_\_ (1991b). Estimation of Adult Mortality from Orphanhood Before and Since Marriage. *Population Studies*, Vol. 45, No. 3 (November), pp. 455-472.
- UNICEF (2006). *Multiple Indicator Cluster Survey Manual 2005*. New York.
- United Nations (1982). *Model Life Tables for Developing Countries*. United Nations publication, Sales No. E.81.XIII.7.
- \_\_\_\_\_ (1983). *Manual X: Indirect Techniques for Demographic Estimation*. Population Studies, No. 81, ST/ESA/SER.A/81 (United Nations publication, Sales No. E.83.XIII.2). Available from [http://www.un.org/esa/population/publications/Manual\\_X/Manual\\_X.htm](http://www.un.org/esa/population/publications/Manual_X/Manual_X.htm)
- \_\_\_\_\_ (1990). *Step-by-Step Guide to the Estimation of Child Mortality*. United Nations publication, Sales No. E.89.XIII.9.
- \_\_\_\_\_ (2002). *Methods for Estimating Adult Mortality*. Working Paper ESA/WP/WP.175. Available from <http://www.un.org/esa/population/publications/adultmort/Complete.pdf>
- \_\_\_\_\_ ((2007). *World Mortality Report 2007* (POP/DB/MORT/2007).

This page is intentionally left blank



## **Annexes**

**Annex I. Direct Estimation Questionnaire Modules  
for Childhood Mortality**

*DHS*

211 Now I would like to record the names of all your births, whether still alive or not, starting with the first one you had. RECORD NAMES OF ALL THE BIRTHS IN 212. RECORD TWINS AND TRIPLETS ON SEPARATE LINES.									
212  What name was given to your (first/next) baby?  (NAME)	213  Were any of these births twins?	214  Is (NAME) a boy or a girl?	215  In what month and year was (NAME) born?  PROBE: What is his/her birthday?	216  Is (NAME) still alive?	217 IF ALIVE:  How old was (NAME) at his/her last birthday?  RECORD AGE IN COMPLETED YEARS.	218 IF ALIVE  Is (NAME) living with you?	219 IF ALIVE:  RECORD HOUSEHOLD LINE NUMBER OF CHILD (RECORD '00' IF CHILD NOT LISTED IN HOUSEHOLD)	220 IF DEAD:  How old was (NAME) when he/she died?  IF '1 YR', PROBE: How many months old was (NAME)? RECORD DAYS IF LESS THAN 1 MONTH; MONTHS IF LESS THAN TWO YEARS; OR YEARS.	221  Were there any other live births between (NAME OF PREVIOUS BIRTH) and (NAME)?
01	SING... 1 MULT.. 2	BOY ..1 GIRL..2	MONTH <input type="text"/> <input type="text"/> YEAR <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	YES .. 1 NO ... 2 <input type="checkbox"/> 220	AGE IN YEARS <input type="text"/> <input type="text"/>	YES ... 1 NO .... 2	LINE NUMBER <input type="text"/> <input type="text"/> <input type="checkbox"/> (NEXT BIRTH)	DAYS . 1 <input type="text"/> <input type="text"/> MONTHS 2 <input type="text"/> <input type="text"/> YEARS 3 <input type="text"/> <input type="text"/>	
02	SING... 1 MULT.. 2	BOY ..1 GIRL..2	MONTH <input type="text"/> <input type="text"/> YEAR <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	YES .. 1 NO ... 2 <input type="checkbox"/> 220	AGE IN YEARS <input type="text"/> <input type="text"/>	YES ... 1 NO .... 2	LINE NUMBER <input type="text"/> <input type="text"/> <input type="checkbox"/> (NEXT BIRTH)	DAYS . 1 <input type="text"/> <input type="text"/> MONTHS 2 <input type="text"/> <input type="text"/> YEARS 3 <input type="text"/> <input type="text"/>	YES..... 1 NO ..... 2

NOTE: The module is shown here truncated after two births, but the DHS questionnaire contains a full birth history.

MICS

BIRTH HISTORY MODULE										BH
NOW I WOULD LIKE TO RECORD THE NAMES OF ALL YOUR BIRTHS, WHETHER STILL ALIVE OR NOT, STARTING WITH THE FIRST ONE YOU HAD.										
Record names of all the births in BH1. Record twins and triplets on separate lines.										
BH1	BH2	BH3	BH4	BH5	BH6	BH7	BH8	BH9	BH10	
					<i>If alive</i>			<i>If dead</i>		
WHAT NAME WAS GIVEN TO YOUR (First/next) BABY? (name)	WERE ANY OF THESE BIRTHS TWINS?	IS (name) A BOY OR A GIRL?	IN WHAT MONTH AND YEAR WAS (name) BORN? <i>Probe: What is his/her birthday?</i>	IS (name) STILL ALIVE?	HOW OLD WAS (name) AT HIS/HER LAST BIRTHDAY? <i>Record age in completed years</i>	IS (name) LIVING WITH YOU?	<i>Record household line number of child (record '00' if child not listed in household)</i>	HOW OLD WAS (name) WHEN HE/SHE DIED? <i>If "Y"; probe: How many months old was (Name)?</i> <i>Record days if less than 1 month; months if less than two years; or years.</i>	WHERE THERE ANY OTHER LIVE BIRTHS BETWEEN (name of previous birth) AND (name)	
01	Sing .... 1 Mult..... 2	Boy ..... 1 Girl..... 2	Month ..... ____ Yr _____	Yes... 1 No .... 2⇒BH9	____	Yes. 1 No .. 2	____ ⇒ next line	Days.....1 ____ Months .....2 ____ Years.....3 ____		
02	Sing .... 1 Mult..... 2	Boy ..... 1 Girl..... 2	Month ..... ____ Yr _____	Yes... 1 No .... 2⇒BH9	____	Yes. 1 No .. 2	____ ⇒BH10	Days.....1 ____ Months .....2 ____ Years.....3 ____	Yes .... 1 No..... 2	

NOTE: The module is shown here truncated after two births, but the MICS questionnaire contains a full birth history.

LSMS Example –Tajikistan LSMS 1999

PLEASE MAKE A COMPLETE LIST, STARTING WITH THE FIRST, OF ALL THE CHILDREN YOU HAVE GIVEN BIRTH TO DURING YOUR LIFE, AND FILL OUT THE OTHER QUESTIONS IN THE TABLE.  
PLEASE LIST ALL CHILDREN, EVEN THOSE WHO ONLY LIVED FOR A SHORT TIME.

CHILD ID	7	8		9	10	11			12	13
	What is the child's name?	Month	Year	Was it a boy or a girl?	Is this child still alive surviving?	How many months or years did the child live?	DAYS	MONTHS	YEARS	THIS COLUMN IS FOR INTERVIEWER= S USE  IF CHILD STILL IN HOUSEHOLD, INTERVIEWER WRITE CHILD ID NUMBER
	FIRST NAME (If applicable)			BOY.....1 GIRL.....2	YES.....1 <QUESTION 12 NO.....2					
1										
2										
3										
4										
5										
6										
7										
8										
9										

**Annex II. Indirect Estimation Questionnaire Modules  
for Childhood Mortality**

*DHS*

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP								
201	Now I would like to ask about all the births you have had during your life. Have you ever given birth?	YES ..... 1 NO ..... 2	<206								
202	Do you have any sons or daughters to whom you have given birth who are now living with you?	YES ..... 1 NO ..... 2	<204								
203	How many sons live with you? And how many daughters live with you? IF NONE, RECORD '00'.	SONS AT HOME ..... <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr><tr><td> </td><td> </td></tr></table> DAUGHTERS AT HOME ..... <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr><tr><td> </td><td> </td></tr></table>									
204	Do you have any sons or daughters to whom you have given birth who are alive but do not live with you?	YES ..... 1 NO ..... 2	<206								
205	How many sons are alive but do not live with you? And how many daughters are alive but do not live with you? IF NONE, RECORD '00'.	SONS ELSEWHERE ..... <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr><tr><td> </td><td> </td></tr></table> DAUGHTERS ELSEWHERE ... <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr><tr><td> </td><td> </td></tr></table>									
206	Have you ever given birth to a boy or girl who was born alive but later died? IF NO, PROBE: Any baby who cried or showed signs of life but did not survive?	YES ..... 1 NO ..... 2	<208								
207	How many boys have died? And how many girls have died? IF NONE, RECORD '00'.	BOYS DEAD ..... <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr><tr><td> </td><td> </td></tr></table> GIRLS DEAD ..... <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr><tr><td> </td><td> </td></tr></table>									
208	SUM ANSWERS TO 203, 205, AND 207, AND ENTER TOTAL. IF NONE, RECORD '00'.	TOTAL ..... <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>									
209	CHECK 208: Just to make sure that I have this right: you have had in TOTAL _____ births during your life. Is that correct?  YES <input type="checkbox"/> NO <input type="checkbox"/> PROBE AND CORRECT 201-208 AS NECESSARY.										
210	CHECK 208: ONE OR MORE BIRTHS <input type="checkbox"/> NO BIRTHS <input type="checkbox"/>		<226								

MICS

CHILD MORTALITY MODULE		CM
<p><i>This module is to be administered to all women age 15-49.</i></p> <p><i>All questions refer only to LIVE births.</i></p>		
<p>CM1. NOW I WOULD LIKE TO ASK ABOUT ALL THE BIRTHS YOU HAVE HAD DURING YOUR LIFE. HAVE YOU EVER GIVEN BIRTH?</p> <p><i>If "No" probe by asking:</i> I MEAN, TO A CHILD WHO EVER BREATHED OR CRIED OR SHOWED OTHER SIGNS OF LIFE – EVEN IF HE OR SHE LIVED ONLY A FEW MINUTES OR HOURS?</p>	<p>Yes ..... 1</p> <p>No ..... 2</p>	<p>2⇒ NEXT MODULE</p>
<p>CM2A. WHAT WAS THE DATE OF YOUR FIRST BIRTH?</p> <p>I MEAN THE VERY FIRST TIME YOU GAVE BIRTH, EVEN IF THE CHILD IS NO LONGER LIVING, OR WHOSE FATHER IS NOT YOUR CURRENT PARTNER.</p> <p><i>Skip to CM3 only if year of first birth is given. Otherwise, continue with CM2B.</i></p>	<p>Date of first birth</p> <p>Day ..... __ __</p> <p>DK day ..... 98</p> <p>Month ..... __ __</p> <p>DK month ..... 98</p> <p>Year ..... __ __ __ __</p> <p>DK year ..... 9998</p>	<p>⇒CM3 ⇓CM2B</p>
<p>CM2B. HOW MANY YEARS AGO DID YOU HAVE YOUR FIRST BIRTH?</p>	<p>Completed years since first birth ..... __ __</p>	
<p>CM3. DO YOU HAVE ANY SONS OR DAUGHTERS TO WHOM YOU HAVE GIVEN BIRTH WHO ARE NOW LIVING WITH YOU?</p>	<p>Yes ..... 1</p> <p>No ..... 2</p>	<p>2⇒CM5</p>
<p>CM4. HOW MANY SONS LIVE WITH YOU?</p> <p>HOW MANY DAUGHTERS LIVE WITH YOU?</p>	<p>Sons at home ..... __ __</p> <p>Daughters at home ..... __ __</p>	
<p>CM5. DO YOU HAVE ANY SONS OR DAUGHTERS TO WHOM YOU HAVE GIVEN BIRTH WHO ARE ALIVE BUT DO NOT LIVE WITH YOU?</p>	<p>Yes ..... 1</p> <p>No ..... 2</p>	<p>2⇒CM7</p>
<p>CM6. HOW MANY SONS ARE ALIVE BUT DO NOT LIVE WITH YOU?</p> <p>HOW MANY DAUGHTERS ARE ALIVE BUT DO NOT LIVE WITH YOU?</p>	<p>Sons elsewhere ..... __ __</p> <p>Daughters elsewhere ..... __ __</p>	
<p>CM7. HAVE YOU EVER GIVEN BIRTH TO A BOY OR GIRL WHO WAS BORN ALIVE BUT LATER DIED?</p>	<p>Yes ..... 1</p> <p>No ..... 2</p>	<p>2⇒CM9</p>
<p>CM8. HOW MANY BOYS HAVE DIED?</p> <p>HOW MANY GIRLS HAVE DIED?</p>	<p>Boys dead ..... __ __</p> <p>Girls dead ..... __ __</p>	
<p>CM9. <i>Sum answers to CM4, CM6, and CM8.</i></p>	<p>Sum ..... __ __</p>	

*LSMS Example –Tajikistan LSMS 1999*

1. Have you ever been pregnant, even if you had a pregnancy that lasted only a few weeks?

YES ... 1

NO ... 2 <IF NO, PLEASE GO TO QUESTION 28

2. How many children have you given birth to. Please include births where the child only lived a few short hours or died later?

NUMBER \_\_\_\_\_

IF NONE, PLEASE WRITE ZERO AND GO TO QUESTION 22

3. Some times it happens that children die. It may be painful to think about such memories and I am sorry to ask you about them.

However it is important to get the right information.

In all, how many of your children have not survived?

NUMBER \_\_\_\_\_

**Annex III. Direct Estimation Questionnaire Modules  
for Adult Mortality**

*DHS Sibling History*

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES					SKIP
901	Now I would like to ask you some questions about your brothers and sisters, that is, all of the children born to your natural mother, including those who are living with you, those living elsewhere and those who have died. How many children did your mother give birth to, including you?	NUMBER OF BIRTHS TO NATURAL MOTHER ..... <input type="text"/>					
902	CHECK 901:  TWO OR MORE BIRTHS <input type="checkbox"/> ? ONLY ONE BIRTH (RESPONDENT ONLY) <input type="checkbox"/>						←914
903	How many of these births did your mother have before you were born?	NUMBER OF PRECEDING BIRTHS ..... <input type="text"/>					
904	What was the name given to your oldest (next oldest) brother or sister?	[1] _____	[2] _____	[3] _____	[4] _____	[5] _____	[6] _____
905	Is (NAME) male or female?	MALE .....1 FEMALE .....2	MALE ..... 1 FEMALE ..... 2	MALE ..... 1 FEMALE ..... 2	MALE .....1 FEMALE.....2	MALE..... 1 FEMALE ..... 2	MALE .....1 FEMALE .....2
906	Is (NAME) still alive?	YES.....1 NO .....2 ↳GO TO 908 DK.....8 ↳GO TO [2]	YES ..... 1 NO ..... 2 ↳GO TO 908 DK ..... 8 ↳GO TO [3]	YES ..... 1 NO ..... 2 ↳GO TO 908 DK ..... 8 ↳GO TO [4]	YES.....1 NO.....2 ↳GO TO 908 DK.....8 ↳GO TO [5]	YES ..... 1 NO ..... 2 ↳GO TO 908 DK ..... 8 ↳GO TO [6]	YES .....1 NO .....2 ↳GO TO 908 DK.....8 ↳GO TO [7]
907	How old is (NAME)?	<input type="text"/> GO TO [2]	<input type="text"/> GO TO [3]	<input type="text"/> GO TO [4]	<input type="text"/> GO TO [5]	<input type="text"/> GO TO [6]	<input type="text"/> GO TO [7]
908	How many years ago did (NAME) die?	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
909	How old was (NAME) when he/she died?	<input type="text"/> IF MALE OR DIED BEFORE 12 YEARS OF AGE GO TO [2]	<input type="text"/> IF MALE OR DIED BEFORE 12 YEARS OF AGE GO TO [3]	<input type="text"/> IF MALE OR DIED BEFORE 12 YEARS OF AGE GO TO [4]	<input type="text"/> IF MALE OR DIED BEFORE 12 YEARS OF AGE GO TO [5]	<input type="text"/> IF MALE OR DIED BEFORE 12 YEARS OF AGE GO TO [6]	<input type="text"/> IF MALE OR DIED BEFORE 12 YEARS OF AGE GO TO [7]
910 - 914	Questions related to maternal mortality – not shown	...	...	...	...	...	...

India NFHS 1998/1999 Recent Household Deaths Module

NO. _____		QUESTIONS AND FILTERS		CODING CATEGORIES		SNIP			
51		Did any usual resident of this household die since January 1997		YES.....1	NO.....2	63			
52		How many persons died?		TOTAL DEATHS.....					
53	54	55	56	57	58	59	60	61	62
What (was/were) the name(s) of the person(s) who died?	Was (NAME) a male or a female?	How old was he/she when he/she died? RECORD DAYS IF LESS THAN ONE MONTH, MONTHS IF LESS THAN TWO YEARS, OR YEARS	In what month and year did (NAME) die?	What did (NAME) die of?	CHECK 54 AND 55: DECEASED WAS FEMALE AGED 15-49 AT THE TIME OF DEATH	Was (NAME) pregnant when she died?	Did (NAME) die during childbirth?	Did (NAME) die within two months after the end of a pregnancy or childbirth?	Was the death of (NAME) due to a complication of the pregnancy or childbirth?
01] _____ (NAME)	MALE.....1 FEMALE...2	DAYS....1 MONTHS..2 YEARS...3	MONTH.. YEAR...		YES.....1 NO.....2 (GO TO NEXT DEATH) <	YES.....1 NO.....2 (GO TO 62) <	YES.....1 NO.....2 (GO TO NEXT DEATH) <	YES.....1 NO.....2 (GO TO NEXT DEATH) <	YES.....1 NO.....2 (GO TO NEXT DEATH) <
02] _____ (NAME)	MALE.....1 FEMALE...2	DAYS....1 MONTHS..2 YEARS...3	MONTH.. YEAR...		YES.....1 NO.....2 (GO TO NEXT DEATH) <	YES.....1 NO.....2 (GO TO 62) <	YES.....1 NO.....2 (GO TO NEXT DEATH) <	YES.....1 NO.....2 (GO TO NEXT DEATH) <	YES.....1 NO.....2 (GO TO NEXT DEATH) <
03] _____ (NAME)	MALE.....1 FEMALE...2	DAYS....1 MONTHS..2 YEARS...3	MONTH.. YEAR...		YES.....1 NO.....2 (GO TO NEXT DEATH) <	YES.....1 NO.....2 (GO TO 62) <	YES.....1 NO.....2 (GO TO NEXT DEATH) <	YES.....1 NO.....2 (GO TO NEXT DEATH) <	YES.....1 NO.....2 (GO TO NEXT DEATH) <
63		RECORD THE TIME.		HOUR.....		MINUTES.....			



**Annex IV. Indirect Estimation Questionnaire Modules  
for Adult Mortality**

*Parental Orphanhood Questions – example from Egypt DHS 1988*

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP TO
611	<p>Now I would like to ask some questions about your parents and the parents of your (first) husband.</p> <p>Is your mother still alive?</p> <p>Is your father still alive?</p> <p>Is your (first) husband's mother still alive?</p> <p>Is your (first) husband's father still alive?</p>	<p align="right">YES NO DK</p> <p>RESPONDENT'S MOTHER..1 2 8</p> <p>RESPONDENT'S FATHER..1 2 8</p> <p>FIRST HUSBAND'S MOTHER.....1 2 8</p> <p>FIRST HUSBAND'S FATHER.....1 2 8</p>	
612	<p>CHECK 611: AT LEAST ONE PARENT NOT LIVING OR DON'T KNOW <input type="checkbox"/></p> <p>ALL PARENTS ALIVE <input type="checkbox"/></p>		→614
613	<p>Was (MENTION PARENTS NOT ALIVE NOW OR DK) alive at the time you began living together with your (first) husband?</p>	<p align="right">YES NO DK</p> <p>RESPONDENT'S MOTHER..1 2 8</p> <p>RESPONDENT'S FATHER..1 2 8</p> <p>FIRST HUSBAND'S MOTHER.....1 2 8</p> <p>FIRST HUSBAND'S FATHER.....1 2 8</p>	

**MICS3 Maternal mortality module**

This module may be used for the analysis of adult female mortality using the sibling survival method.

MATERNAL MORTALITY MODULE										MM
Administer to each adult household member. Copy name and line number of each adult (age 15 or over) in the household. If one of these adults is not at home, another adult may respond for him/her. Indicate this by placing a '1' in MM3, and insert the number of proxy respondent in MM4. For household members below age 15, leave rows blank										
MM1. Line no.	MM2. Name	MM3. IS THIS A PROXY REPORT? 1 YES ⇐ MM4 2 NO ⇐ MM5	MM4. Line no. of proxy respondent (from household listing HL1)	MM5. HOW MANY SISTERS (BORN TO THE SAME MOTHER) HAVE YOU EVER HAD?	MM6. HOW MANY OF THESE SISTERS EVER REACHED AGE 15?	MM7. HOW MANY OF THESE SISTERS (WHO ARE AT LEAST 15 YEARS OLD) ARE ALIVE NOW?	MM8. HOW MANY OF THESE SISTERS WHO REACHED AGE 15 OR MORE HAVE DIED?	MM9. HOW MANY OF THESE DEAD SISTERS DIED WHILE PREGNANT, OR DURING CHILD BIRTH, OR DURING THE SIX WEEKS AFTER THE END OF PREGNANCY? 98= DON'T KNOW		
LINE	NAME	Y N	LINE		98= DON'T KNOW	98= DON'T KNOW	98= DON'T KNOW	98= DON'T KNOW		
01		1 2								
02		1 2								
03		1 2								
04		1 2								
05		1 2								
06		1 2								
07		1 2								
08		1 2								
09		1 2								
10		1 2								
11		1 2								
12		1 2								
13		1 2								
14		1 2								
15		1 2								

**Annex V. Key variables from sample survey programmes  
(childhood mortality)**

<b>Variable</b>	<b>DHS</b>	<b>WFS</b>	<b>MICS</b>	<b>RHS<sup>1</sup></b>	<b>LSMS<sup>2</sup></b>	<b>WHS</b>
Naming convention	Standard across all surveys	Standard across all surveys	Usually standard but with some variations	Mostly standard for Latin American surveys, non-standard for others	Survey specific (below are examples only)	Standard across surveys
Sample weight	V005	V006	wmweight	Pesomef	Self weighted	pweight (in file _ID)
Date of interview (CMC)	V008	V007	Cmcdoiw	Entcmc <sup>3</sup>	Date_yr * 12 + Date_mth	q0008 (in file _F2 – date in day, month, year form)
<b>Direct</b>						
Data file name	Birth recode (ccBRvv)	Standard recode (ccSRvv <sup>4</sup> )	Birth history (BH)	Children’s file (cc_yyyy_hijos)	Children’s file (SSec11CH + SSec1)	Children’s file (WHS_Country_F6.dta)
Date of birth of child (CMC)	B3	BDAT (or Bxx2)	BH4M, BH4Y	Hijocmc	S11q8_2 * 12 + S11q8_1	q5002 (date in month and year form)
Sex (1 = Male, 2 = Female)	B4	BSEX (or Bxx3)	BH3	Sexh	S11q9	q5003 (1=FEMALE 2=MALE)
Survival status (Alive, Dead)	B5 (1,0)	BADM (98, <> 98)	BH5 (1,2)	Vivo (1,2)	S11q10 (1,2)	q5004 <sup>5</sup> (1=Alive 5=Dead)
Age at death	B7 <sup>6</sup> B6 <sup>7</sup>	BADM (or Bxx4), BADY (or Bxx5)	BH9U/A, BH9N/B <sup>8</sup>	Edmur <sup>7</sup>	S11q11_1, S11q11_2, S11q11_3 <sup>9</sup>	q5008a, q5008b, q5008c <sup>10</sup> q5007 (date of death – month and year)

Variable	DHS	WFS	MICS	RHS <sup>1</sup>	LSMS <sup>2</sup>	WHS
<b>Indirect</b>						
Data file name	Individual recode (ccIRvv)	Standard recode (ccSRvv)	Women's file (WM)	Women's file (cc_yyyy_cread as)	Women's file (SSec11A + SSec1)	Children's file (WHS_Country_F6.dta)
Children ever born	V201	V208	CM9 or CEB	Hijocasa + Hijofuer + Hijomuer	S11q5	Sum of children
Children surviving	V202 + V203 + V204 + V205	V213	CM4A + CM4B + CM6A + CM6B or Surviv	Hijocasa + Hijofuer	S11q5 – S11q6	Sum of children with q5004 = 1
Age group	V013	V011	Wage	Edad	(S1q6 / 5) – 2	(q1002 / 5) – 2 In file _F2
Duration of marriage group	V513	V117	(Cmcdoiw – Wdom) / 60	(Entcmc – (Anounion*12 + Mesunion))/60	Not collected	Not collected
Time since first birth group	(V008- V211) /60	(V007- BDAT(1)) /60	(Cmcdoiw – dofb) <sup>11</sup> /60	(Entcmc – (Ano1er*12 + Mes1er)) /60	((Date_yr - S11q8_2) * 12 + (Date_mth – S11q8_1)) / 60	(cmc <sup>12</sup> of q0008 (in file _F2) – cmc <sup>13</sup> of q5002 for 1 <sup>st</sup> child (in file _F6)) / 60

<sup>1</sup> Based on Ecuador RHS 2004.

<sup>2</sup> Based on Tajikistan LSMS 1999.

<sup>3</sup> Note that the century month is calculated here as YYYY \* 12 + MM, rather than the more standard (YYYY-1900) \* 12 + MM.

<sup>4</sup> Women based file with all births listed in the same case.

<sup>5</sup> Note that the codes for sex are reversed for the WHS surveys compared with the other survey programmes.

<sup>6</sup> Imputed age at death in months.

<sup>7</sup> Reported age at death as one variable combining units (1=Days, 2=Months, 3=Years) and number of those units in the form UNN.

<sup>8</sup> Reported as units and number of those units as for DHS, RHS. Variable names are sometimes BH9U and BH9N, and sometimes BH9A and BH9B depending on the survey.

<sup>9</sup> Age at death reported as days, months and/or years, and includes some cases reported as 1.5 years or months.

<sup>10</sup> Age at death reported as days, months and/or years.

<sup>11</sup> Date of first birth (dofb) must be constructed from CM2AM, CM2AY, and CM2B.

<sup>12</sup> Century month code (cmc) must be constructed from the date of interview as a string in q0008 in the form dd/mm/yyyy, e.g. cmcq0008 = (yyyy-1900)+mm.

<sup>13</sup> Century month code (cmc) must be constructed from the date of birth of the first child (childno = 1) as a string in q5002 in the form mm/yyyy e.g. cmcq5002 = (yyyy-1900)+mm.

**Annex VI. Key variables from sample survey programmes  
(adult mortality)**

<b>Variable</b>	<b>DHS</b>	<b>WFS</b>	<b>MICS</b>	<b>RHS<sup>1</sup></b>	<b>WHS</b>
Naming convention	Standard across all surveys	Semi-standard	Usually standard but with some variations	Survey specific (below are examples only)	Standard across surveys
<b>Direct</b>					
<b>Sibling History</b>					
Data file name	Individual recode [Male recode] (ccIRvv) [ccMRvv]			Women's file (cc_yyyy_creadas) <sup>2</sup>	Sibling's file (WHS_Country_F7.dta)
Sample weight	V005 [MV005]			Pesomef	pweight (in file _ID)
Date of interview (CMC)	V008 [MV008]			Entcmc <sup>3</sup>	q0008 (in file _F2)
Age of respondent	V013 [MV013]			Edad	(q1002 / 5) – 2 In file _F2
Date of birth of sibling (CMC)	MM4 [SMM4]			<i>Living sister:</i> P1201x (age of each surviving sister) <i>Dead sister:</i> P1204_n (years since death), P1205_n (age at death)	q5106 (date in month and year form)
Sex (1 = Male, 2 = Female)	MM1 [SMM1]			2 – Females only	q5105 <sup>4</sup> (1=Female 2=Male)
Survival status (Alive, Dead)	MM2 [SMM2] (1,0)			<i>Separate group of variables for surviving sisters and for dead sisters</i>	q5108 (1=Alive 5=Dead)
Date of death (CMC)	MM8 <sup>5</sup> [SMM8]			P1204_n (years since death), P1205_n (age at death)	q5111 (age at death), q5112 (years since death) Combine to produce CMC date of death

Variable	DHS	WFS	MICS	RHS <sup>1</sup>	WHS
<b>Recent household deaths</b>					
Data file name	Individual recode [Male recode] (ccIRvv) [ccMRvv]	Household members data (ccHMvv or ccHMdata)			
Sample weight	HV005	I206			
Date of interview (CMC)	HV008	I302/ I303 (Month and year of interview)			
Sex	HV104 for living, SH54 for deceased <sup>6</sup>	M101 for living, H606 for deceased			
Age	HV105 for living, SH55C for deceased <sup>6</sup>	M102 for living, H607, H608/ U608 for deceased			
Date of death	SH56C <sup>6</sup>	H609/ H610 (Month and year of death)			
<b>Indirect</b>					
<b>Maternal/Paternal orphanhood</b>					
Data file name	Individual recode ccIRvv	Household members data (ccHMvv or ccHMdata)			
Sample weight	V005	I206			
Date of interview (CMC)	V008	I302/ I303 (Month and year of interview)			
Mother, Father alive	V514 (1,0), V515 (1,0)	M401 (1,2) M404 (1,2)			

Variable	DHS	WFS	MICS	RHS <sup>1</sup>	WHS
<b>Sibling survival</b>					
Data file name	Individual recode (cciRvv)		Household members data (HL) <sup>7</sup>	Women's file (cc_yyyy_creadas) <sup>2</sup>	Sibling's file (WHS_Country_F7.dta)
Sample weight	V005		hhweight	Pesomef	pweight (in file _ID)
Date of interview (CMC)	V008		Doic = (HH5Y-1900)*12 + HH5M	Entcmc <sup>8</sup>	q0008 (in file _F2)
Age group	V013		Wage	Edad	(q1002 / 5) - 2 In file _F2
Siblings alive	Count (MM2 = 1)		MM7	P1200HER	Count (q5108 = 1)
Siblings surviving to age 15	Count (MM2 = 1 & (V008 - MM4) >= 180) + Count (MM2 = 0 & (MM8-MM4) >= 180)		MM6	Count (p1201x >= 15) + Count (P1205_n >= 15)	Count (q5108 = 1 & (q0008 - q5106) >= 180) + Count (q5108 = 5 & q5111 >= 15)

<sup>1</sup> Based on Ecuador RHS 2004.

<sup>2</sup> The Reproductive Health Surveys support the calculation of maternal mortality estimates and adult female mortality estimates, but not estimates for males or both sexes combined.

<sup>3</sup> Note that the century month is calculated here as YYYY \* 12 + MM, rather than the more standard (YYYY-1900) \* 12 + MM.

<sup>4</sup> Note that the codes for sex are reversed for the WHS surveys compared with the other survey programmes.

<sup>5</sup> Imputed date of death in CMC.

<sup>6</sup> Survey specific variable – example from India NFHS 1998/1999.

<sup>7</sup> The Multiple Indicator Cluster Surveys (MICS) support the calculation of maternal mortality estimates and adult female mortality estimates, but not estimates for males or both sexes combined.

<sup>8</sup> Note that the century month is calculated here as YYYY \* 12 + MM, rather than the more standard (YYYY-1900) \* 12 + MM.