Theme: Strengthening vital statistics and cause-of-death data

Mortality statistics: a tool to improve understanding and quality

Carla AbouZahr

Lene Mikkelsen

Rasika Rampatige

Alan Lopez

Working Paper Series • Number 13 • November 2010 • WORKING PAPER

Health Information Systems Knowledge Hub

School of Population Health
University of Queensland

Strengthening health systems in Asia and the Pacific through better evidence and practice

For the PDF version of this paper and other related documents, visit

www.uq.edu.au/hishub

About this series

The Health Information Systems Knowledge Hub's Working Paper Series is the principal means to disseminate the knowledge products developed by the hub as easily accessible resources that collectively form a lasting repository of the research findings and knowledge generated by the hub's activities. Working papers are intended to stimulate debate and promote the adoption of best practice for health information systems in the region. The series focuses on a range of knowledge gaps, including new tools, methods and approaches, and raises and debates fundamental issues around the orientation, purpose and functioning of health information systems. Generally, working papers contain more detailed information than a journal article, are written in less-academic language, and are intended to inform health system dialogue in and between countries and a range of development partners.

Many working papers have accompanying products such as summaries, key points and action guides. The full range of documents, as well as other resources and tools, is available on the Health Information Systems Knowledge Hub website at www.uq.edu.au/hishub/publication-tools.

The opinions or conclusions expressed in the Working Paper Series are those of the authors and do not necessarily reflect the views of institutions or governments.

Acknowledgments

We gratefully acknowledge the assistance of Rafael Lozano in preparing the table and graphs for this guide.

Author details

Carla AbouZahr, Lene Mikkelsen, Rasika Rampatige and Alan Lopez, Health Information Systems Knowledge Hub, University of Queensland

Published by the Health Information Systems Knowledge Hub,
School of Population Health
The University of Queensland
Public Health Building,
Herston Rd, Herston Qld 4006, Australia

Please contact us for additional copies of this publication, or send us feedback:

Email: hishub@sph.uq.edu.au
Tel: +61 7 3346 4732
Fax: +61 7 3365 5442
www.uq.edu.au/hishub

Design by Biotext, Canberra, Australia

Contents

Acronyms and abbreviations

ASMR	age-specific mortality rate
CDR	crude death rate
ICD-10	International Classification of Diseases, 10th revision
IMR	infant mortality rate
NNMR	neonatal mortality rate
PNNMR	postneonatal mortality rate
U5MR	under-five mortality rate
WHO	World Health Organization

Objectives of this guide

This guide is intended to help build analytical capacity to assess the quality of mortality statistics that are currently being collected in order to improve their value in informing health policies and programs.

Countries routinely invest significant resources into collecting mortality data from a variety of sources, including civil registration systems, health care facilities, ongoing longitudinal demographic and health surveillance, and from other data sources such as censuses or household surveys. The primary purpose is to generate critical information to guide public health decision-making. However, data cannot be used appropriately or with any confidence if insufficient attention is paid to the quality. In the absence of systematic data quality assessment, and adjustment where necessary, the data that have been collected—often at great expense—cannot be used to their full potential to guide decision-making.

To assist countries in validating and correcting their mortality data, the World Health Organization (WHO), in partnership with the Health Information Systems Knowledge Hub at the University of Queensland (UQ), Brisbane, has developed this mortality statistics assessment guide and toolkit. The guide describes relatively simple ways of analysing the internal validity and coherence of mortality data, and shows how comparisons with other external sources of mortality data can be used to assess data consistency and plausibility. By carrying out these simple checks, data collectors and practitioners will be able to diagnose weaknesses in their data. If this information is used in conjunction with an assessment of the functioning of the civil registration and vital statistics systems using the WHO/UQ guide (WHO and UQ 2010), country decision-makers will have all the tools necessary to develop and target strategies for improving the availability and quality of mortality data. The checks will also assist users in the interpretation of the data so that they can better understand prevailing levels, trends and patterns of mortality in their populations.

This guide describes simple ways of analysing the internal validity and coherence of mortality data, and shows how comparisons with other external sources of mortality data can be used to assess data consistency and plausibility.

Ten simple steps

We describe a 10 step process for assessing the quality of mortality data. The 10 steps can be applied to datasets from different sources, but steps 6–10 are not relevant for survey and census data, as these sources do not generate cause-of-death information using *International Statistical Classification of Diseases and Related Health Problems*, 10th revision (ICD-10) standards (WHO 2007).

At each step, users are led through a process of checking for errors, calculating key indicators, interpreting the public health significance of the indicator values and reflecting on how to use the information to diagnose possible weaknesses in their mortality data systems. The 10 steps are:

Step 1 Prepare basic tabulations of deaths by age, sex and cause-of-death.

Step 2 Review crude death rates.

Step 3 Review age and sex-specific death rates.

Step 4 Review the age distribution of deaths.

Step 5 Review child mortality rates.

Step 6 Review the distribution of major causes -of-death.

Step 7 Review age patterns of major causes-of-death.

Step 8 Review leading causes-of-death.

Step 9 Review ratio of noncommunicable disease deaths to communicable disease deaths.

Step 10 Review ill-defined causes- of-death.

Using the electronic mortality data quality assessment tool

Applying the 10 steps

This 10 step process can be applied to any mortality dataset. In many settings, mortality data will be the product of the national civil registration and vital statistics systems that routinely collect and compile information to produce statistics on births, deaths and causes-of-death. Data on mortality by age and sex (but not cause) can also be collected through the decennial census. Mortality data, including information on causes-of-death, are also generated through longitudinal demographic surveillance in specific sites. In some settings, the most regular source of data on mortality for a population is the routine health information system that records deaths occurring in hospitals. Although these data cannot be considered nationally representative (because they are biased towards deaths occurring in health care facilities and usually confined to the public sector), they can, nonetheless, provide useful information on patterns of hospital mortality and may be of considerable value for understanding mortality patterns in specific sectors of the population. This is especially true in urban areas, where a high proportion of deaths are likely to occur in a health care setting.

To automate the data quality assessment process described step by step in this guide, an easy-to-use electronic tool is available[1] that will perform the calculations needed for the data quality review and automatically generate the associated figures and tables. To use the tool, it is helpful to have basic computer skills and familiarity with software packages such as Microsoft Excel and Access. However, the tool does not require either advanced expertise in software packages, or advanced statistical or computing skills.

The tool aggregates and presents mortality data in a format that makes them easier to analyse. It automatically:

- verifies and checks for gross data errors (eg maternal deaths ascribed to males)
- generates information on the reliability of certification and coding practices (eg identifying invalid underlying causes-of-death)
- carries out basic calculations of health indicators and generates figures, such as the distribution of broad causes-of-death by age group, and age, sex and cause-specific death rates
- summarises the data in formats that facilitate data sharing and presentation.

Users of this guide are strongly recommended to use the accompanying electronic tool to facilitate the computations and analyses of data described in the following pages.

Following up the results of the review

The purpose of conducting a data quality assessment as outlined in the 10 steps is to diagnose possible problems with the mortality data collection system(s) and to take action to address them. It is important to stress that the review should not be seen as a fault-finding exercise, designed to identify errors and apportion blame. Rather, the purpose is to engage with all those producing and using mortality data—at all levels—to identify weaknesses in the data with a view to correcting problems in the systems that generate them. Ongoing efforts are needed to assure data quality and the regular assessment of the quality of mortality data should become an integral activity of the health information system.

In situations where mortality statistics being reviewed emanate from a civil registration or vital statistics system with information on the causes-of-deaths, it is strongly

1 This tool can be accessed at www.uq.edu.au/hishub.

recommended to thoroughly assess the functioning of the civil registration system using the WHO/UQ comprehensive assessment tool (WHO and UQ 2010). This tool not only provides a detailed framework and road map to identify deficiencies with the mortality data collection system, but also provides detailed guidance about prioritising actions and interventions to improve specific functions.

Step 1 Basic tabulations of deaths by age, sex and cause-of-death

The first step is to aggregate the individual death records and tabulate the available data on deaths by age, sex and causes (using ICD-10 codes).

As a minimum, the tabulations should include:

- numbers of deaths for a specified year
- by sex (ie for males and females separately)
- by age at death using the following age groupings
  - within the first 28 days after birth
  - between completed months 1 and 11
  - between completed years 1 and 4
  - completed years 5–9
  - completed years 10–14 and so on, by 5-year age groups, up to completed years 80–84
  - completed years 85 and over
- by ICD-10 short list of causes.

In addition, the tabulations should include the mid-year population for the same year, sex and age group. Population estimates are generally available from the decennial census and intercensal projections produced by the National Statistics Office. These data will be used for the calculation of rates and ratios that will be explained in the subsequent steps. A standard template for tabulating the mortality and cause-of-death data is shown as an example in Table 1. It is strongly recommended that countries adhere to the age detail shown in the table. Mortality statistics should always be tabulated and analysed separately for males and females.

It is important that age at death be recorded with precision. A death occurring to a child aged 4 years and 11 months should be classified in the 1–4 years age group. Only when the child has completed the 4th year of age (ie had their 5th birthday) should the death be counted in the 5–9 years age group. It is usual practice to use five-year age groups except for deaths occurring in children under 5, which are subdivided into those occurring within the first month of life (28 completed days), those occurring between the ages of 1 and 11 months, and those occurring between the ages of 1 and 4 years. Precision is also important at older ages, which should continue to be grouped into five-year categories at least up to the age of 85 years.

It is poor practice to only tabulate age of death to some relatively low terminal age such as 55+ or 65+. Increasingly, more and more deaths are occurring in populations after about age 50, and it is extremely important for preventive efforts to distinguish between a death at age 80–84 years and an adult death at a much younger age, like 60–64 or 65–69 years old. The use of these standard five-year age groupings is important because the same age groups are used to compile census data on population size and distribution that are used as denominators for the calculation of rates and ratios.

Ideally, causes-of-death should be shown by the ICD-10 three-digit or four-digit codes. However, many countries have only higher order grouping, such as the ICD-10 short or condensed list of major causes. Although not as informative as the more detailed codes, these groupings can still provide useful information for analysis of data quality and hence for use in policy debates.

The WHO/UQ electronic tool that accompanies these guidelines provides alternative data entry formats for use depending on the degree of detail in the available data. Some datasets will include more detail by age, such as deaths within the first 24 hours after birth, deaths by single completed year of age at death, and causes-of-death using the detailed ICD-10 four-digit classification. The electronic tool can handle a variety of formats and levels of detail.

The purpose of this initial tabulation is to identify gross errors in the dataset. The WHO mortality tool will automatically identify clearly incorrect causes-of-death such as male maternal deaths, suicides among young children or prostate cancer deaths in females. It will also draw the attention of users to invalid use of certain codes as underlying causes-of-death, thus alerting data managers to potential quality problems in coding causes-of-death from death certificates or in the certification process (eg implausible sequence of morbid conditions reported on the death certificate).

Once the data have been entered according to the format recommended in Table 1, and the gross errors identified and corrected, the tool will automatically calculate totals and distributions of deaths by sex, age group and cause. Steps 2–10 involve the calculation and analysis of key indicators that can alert users to possible weaknesses in their mortality dataset.

Step 2 Crude death rates

The second step in assessing the quality of a set of mortality data is to review the calculated level of the crude death rate (CDR). This is done for two reasons. First, the CDR is the simplest measure of mortality that can provide insights into the health status of a population over time. Second, the CDR provides a useful indicator of possible problems with the completeness of mortality data.

The objectives of Step 2 are to enable users to:

- define and calculate the CDR
- understand the public health relevance of the CDR
- interpret the CDR and judge its limitations
- use the CDR as an approximate indicator of completeness of death registration
- use the CDR as the first step to analyse the quality of mortality data.

Table 1 Recommended data tabulation format

Year	Numbers of deaths by age group									
	<28 days	1–11 months	1–4 years	5–9 years	10–14 years	15–19 years	20–24 years	25–29 years	80–84 years	85+ years
ICD-10 code or short list										
Males										
Females										
Both sexes										
ICD-10 code or short list										
Males										
Females										
Both sexes										
ICD-10 code or short list										
Males										
Females										
Both sexes										
(Repeat this for each individual ICD-10 cause-of-death for which data are available)										
Total mid-year population by age group										
Males										
Females										
Both sexes										

Definition and calculation of the crude death rate

The CDR is a measure of the number of deaths in a population, relative to the size of that population during a given period of time. The CDR is typically expressed in units of deaths per 1000 individuals per year; thus, a crude death rate of 9.5/1000 in a population of 500 000 indicates there were 4750 deaths per year in the total population (9.5/1000 × 500 000).

The CDR is defined and calculated as follows:

$$\text{CDR} = \frac{\text{Number of deaths in the usual resident population in a given year}}{\text{Size of the mid-year resident population in that year}} \times 1000$$

Because mortality rates for males and females differ across all ages, it is useful to calculate the CDR separately for both sexes.

$$\text{CDR females} = \frac{\text{Number of deaths among females in the usual resident population in a given year}}{\text{Size of the mid-year resident female population in that year}} \times 1000$$

$$\text{CDR males} = \frac{\text{Number of deaths among males in the usual resident population in a given year}}{\text{Size of the mid-year resident male population in that year}} \times 1000$$

It is important that both numerator and denominator refer to the same population in terms of geography and time. It is standard practice to take the size of the population at mid-year as the denominator because population size may vary during the year (due to migration, births and deaths) and the mid-year population serves as an estimate of the average population exposed to the risk of dying over the course of the year.

Interpreting the crude death rate

The CDR is called a 'crude' rate because it does not take into consideration the age and sex structure of the population. In practice, the risk of death in a given population group varies according to age and sex as well as patterns of socioeconomic status, and environmental and other factors. For example, populations with a large proportion of young children or a high proportion of elderly people will, other things being equal, have relatively higher CDRs. This is because mortality risks are highest at youngest and the oldest ages. In general, mortality rates are higher among males than females. Therefore, when comparing populations across countries, geographic areas or over time, it is important to use age and sex-specific mortality rates alongside the CDR (see Step 3). This controls for differences in a population's age and sex structure across the populations being compared.

Crude death rate and population structures

In order to interpret the CDR, it is helpful to refer to the population age–sex pyramid, a graphical illustration of the distribution of the population by standard age groups (usually 5-year groups). The population pyramid typically consists of two back-to-back bar graphs, with age groups on the vertical axis and population size in each age group on the horizontal axis. Males are conventionally shown on the left and females on the right. The bars can represent either the absolute numbers (more common) or percentages of the total (male or female) population in each 5-year age group.

In most developed countries, the age–sex pyramid is constructed on the basis of annual birth and death data from the civil registration system and censuses every 10 years. In countries where civil registration systems are weak, age–sex population pyramids can only be reliably estimated from the census. Intercensal estimates of population size by age and sex generally need to be estimated from mortality rates derived from model life tables, which are inherently uncertain. The United Nations Population Division generates regular updates on national population sex and age structures, which should be used where there is doubt about the reliability of country population data.

The use of age–sex pyramids in helping to interpret CDRs is illustrated in Figure 1. The CDR for Sudan in 2005 is estimated at 13 per 1000 population compared with 9 per 1000 population in Japan. This difference reflects the fact that Sudan has a high proportion of children aged below 4 years and this is precisely the age group where mortality rates are highest. By contrast, Japan has a much smaller percentage of population in this age group, although it has a large proportion of older people aged 60+, when death rates are also high. However, this is insufficient to counteract the effect of a large population of children in Sudan, among whom death rates are comparatively high.

Lower limits for the crude death rate

Based on many decades of experience in calculating CDRs, demographers have demonstrated that there is generally a lower limit for the CDR of around 5 per 1000. For example, during the past 20–30 years, Japan has consistently registered the lowest age-specific mortality rates in the world. Yet throughout this period, the CDR in Japan never fell below 5 per 1000.

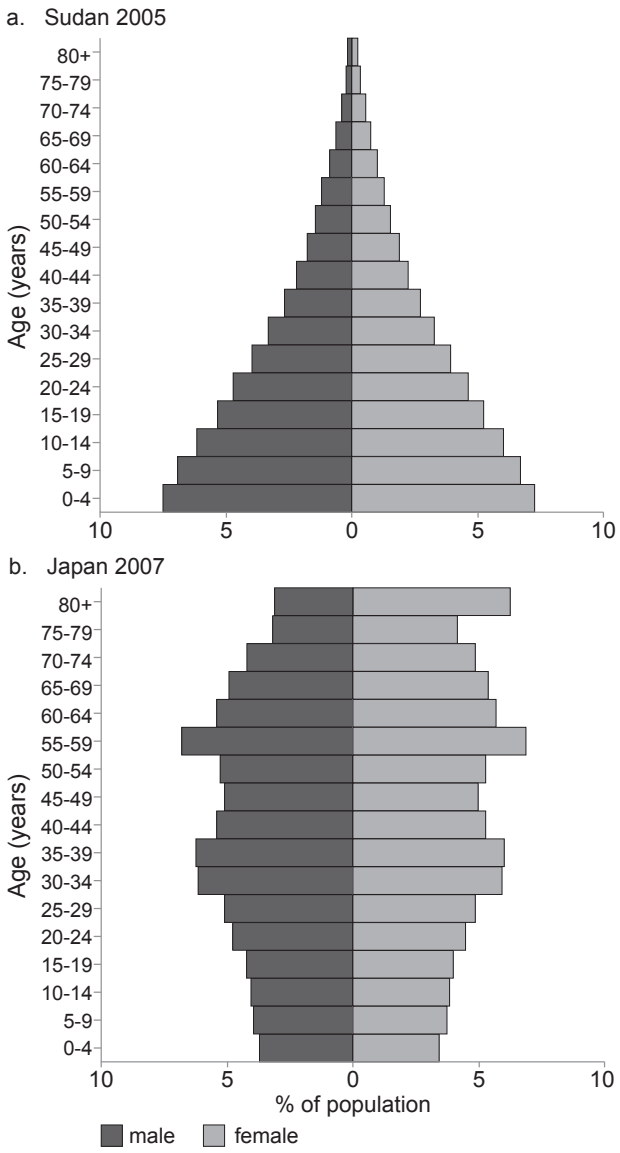
Table 2 shows the combinations of life expectancy and population growth rates that are associated with different levels of the CDR. In many parts of the developing world, population growth rates are typically around 2 per cent each year. In such populations, the CDR can never get below 5 per 1000, and even for the CDR to fall below 7 per 1000, life expectancy would need to be 75 years or more. This is relatively uncommon in developing countries and hence low CDRs should be treated with great suspicion.

Any CDR under 5 per 1000 should be treated with extreme caution, as such a figure is strongly suggestive of incomplete death registration.

However there are exceptional populations that have both high growth rates—due to natural increase (excess of births over deaths), immigration or both—and low age-specific mortality rates, including low child death rates, implying a comparatively high life expectancy at birth. Several of the Gulf States[2] do in fact have a CDR below five because of this particular demographic configuration. In the vast majority of countries, however, this does not apply and low CDRs below 5 per 1000 are typically a sign of underreporting of deaths.

Trends in crude death rates

An analysis of CDR trends over time can help to improve understanding of the evolution of mortality in a given setting. Moreover, looking at mortality trends over time is a useful way of identifying possible problems with data quality. For example, sudden fluctuations in registered deaths indicate data quality problems because in the absence of severe epidemics, wars or natural disasters, mortality levels change only very marginally from one year to another. This is shown clearly in Figure 2, which shows CDR trends in Japan from 1950 to 2005. Trends in the CDR emerge over time, although it is important to note that there typically are only small fluctuations from



Source: Calculated from UN Population Division estimates (http://esa.un.org/unpp/index.asp)

Figure 1 Population age–sex pyramids for Sudan (2005) and Japan (2007)

2 Arab states of the Persian Gulf

year to year. Large fluctuations may arise due to changes in death registration practices (such as legislation to facilitate delayed registration of deaths that occurred several years earlier). These factors need to be taken into account when interpreting trends in the CDR.

To better understand trends in the CDR, it is useful to compare the CDR with trends in other related indicators, such as under-five mortality rates, life expectancy and the proportion of the population aged 65 years and older. This comparison is shown in Figure 3.

Putting these data together on one graph highlights the nature of the temporal relationship between them in a country with good vital statistics on deaths. In particular, the striking decline in the CDR in Japan between 1950 and 1980—from more than 10 per 1000 to about 5.5 per 1000 (right-hand scale)—coincided with a large decline in mortality in children under 5 years old and is reflected in growing life expectancy during the period, as one would expect (left-hand scale).

Since the 1980s, the CDR in Japan has started to rise, coinciding with a gradual increase in the percentage of population aged 65 years and over (left-hand scale). This ageing of the population in Japan is due to the fact that an increasing number of children and adults are surviving to reach old age. By 2005, the CDR had increased to 8 per 1000, reflecting rising mortality in the growing cohort of older people. Note that despite this increase in CDR, under-five mortality continued to decline and life expectancy continued to increase. A rise

Table 2 Crude death rates at different levels of life expectancy and population growth

Life expectancy (years)	Annual rate of population growth (%)									
	5.0	3.0	2.5	2.0	1.5	1.0	0.5	0.0	–0.5	–1.0
40	27.4	24.1	23.6	23.4	23.6	24.1	24.1	25.0	26.2	27.8
45	21.6	19.5	19.3	19.4	19.6	20.2	21.1	22.2	23.7	25.6
50	16.8	15.7	15.8	16.1	16.7	17.5	18.6	20.0	21.8	23.9
55	12.7	12.5	12.9	13.4	14.2	15.2	16.5	18.2	20.2	22.5
60	9.4	9.9	10.4	11.1	12.1	13.3	14.8	16.7	18.8	21.3
65	6.6	7.7	8.4	9.2	10.3	11.7	13.4	14.8	16.7	19.5
70	4.3	5.8	6.6	7.6	8.8	10.4	12.2	14.3	16.7	19.5
75	2.6	4.4	5.2	6.3	7.6	9.2	11.1	13.3	10.9	8.8
80	1.5	3.4	4.2	5.3	6.7	8.3	10.2	12.5	15.1	18.1

Note: Cell values are crude death rate estimates for given values of life expectancy and population growth rates. They have been estimated from the Coale–Demeny 'west' family regional model life tables for females (Coale and Demeny 1966).

Figure 2 Crude death rate trends in Japan, females (1950–2007)

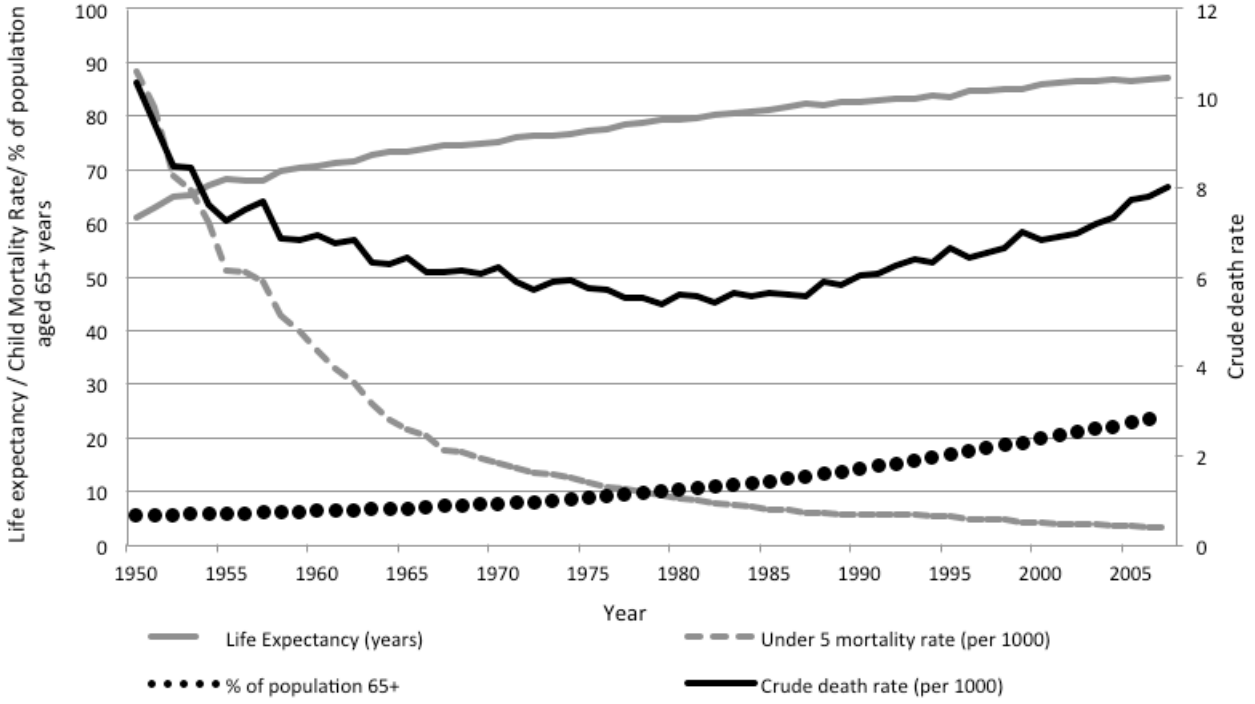


Figure 3 Major demographic trends in Japan, females (1950–2007)

in the CDR after a long period of mortality decline is to be expected since it reflects the postponement of death to older ages. As seen for Japan, the CDR started to rise when life expectancy reached about 80 years and the proportion of the elderly (people aged 65 and over) in the population reached about 10 per cent. A key issue to note is that even in a population such as Japan, with very high levels of life expectancy overall, the CDR always exceeded 5 per 1000.

In most countries, estimates of life expectancy, the child mortality rate and percentage of population aged 65 years and older are published in annual official statistics. This enables a similar analysis to be undertaken to compare trends in these indicators with trends in the CDR. On this basis, countries can judge whether their CDR appears plausible, and hence whether or not their reporting of deaths has been reasonably complete.

Summary of Step 2

- Calculate the CDR. A level less than 5 per 1000 is strongly indicative of incomplete registration of deaths.
- Compare the CDR with data on population age and sex structure by calculating a population age–sex pyramid for your country. If the proportion of young children in the population is high, you should expect the CDR to be relatively high. The same is true when the proportion of older people in the population rises.
- Examine the CDR for males and females separately. You should generally expect the CDR for males to be higher than for females. Deviations from this pattern could indicate that women and girls face severe disadvantages in terms of health and nutrition. Alternatively, there may be problems with data completeness and quality with systematic underreporting of female deaths.
- Examine CDR trends over time and compare them with trends in other measures, such as mortality in children under 5 years old, percentage of population aged 65 years and older, and life expectancy at birth. Any rapid fluctuations from year to year indicate possible data problems. You should see a similar trend pattern over time for these indicators as that shown for Japan.

Step 3 Age and sex-specific death rates

In Step 2, we analysed a mortality dataset by calculating the CDR for the population. However, the CDR is a 'crude' rate because it does not take into consideration the age and sex structure of the population. As we saw from Step 2, populations with a large proportion of young children or a high proportion of elderly people will, other things being equal, have relatively higher CDRs because mortality risks are highest at the youngest and oldest ages. Moreover, mortality rates are generally higher among males than females across all age groups. Therefore, when comparing the mortality of populations across countries, geographic areas or over time, it is important to use both age-specific and sex-specific mortality rates alongside the CDR, and to examine these detailed age and sex-specific rates for possible age misreporting of deaths.

The objectives of Step 3 are to enable users to:

- define and calculate the mortality rate specific to a population age group (usually a five-year grouping), known as the age-specific mortality rate (ASMR)
- understand the public health relevance of the ASMR
- interpret the ASMR and understand its limitations
- use the ASMR to assess the quality of mortality data.

Definition and calculation of age-specific mortality rates

The ASMR is calculated as the total number of deaths, occurring at a specified age or in a specified age group, in a defined geographic area (eg country, state, county) divided by the mid-year population of the same age in the same geographic area. By contrast to the CDR, which is expressed per 1000 population, the ASMR is generally expressed as a rate per 100 000 population. This is because there are many fewer deaths within each age group compared with the numbers occurring in the total population. The standard demographic practice is to calculate the ASMR for 5-year age groups, namely < 1, 1–4, 5–9, 10–14 ... 80–84 and 85+. The ASMR is calculated as follows:

$$\text{ASMR} = \frac{\text{Deaths in a specific age group in a population during a specified time period}}{\text{Total mid-year population in the same age group, population and time period}} \times 100\,000$$

Disaggregation of age-specific mortality rates by sex

As noted in Step 2, there are important differences in patterns and levels of mortality between males and females across all age groups. Therefore, it is standard practice to calculate ASMRs separately for males and females within each age group.

Dealing with fluctuations

In countries and settings with small population numbers, the annual number of deaths at specific ages may be very small. As a result, the ASMR would tend to fluctuate and be too unstable for analysis. In order to overcome this problem, it is usual to calculate the ASMR during a 3–5 year period to average out annual fluctuations. This is illustrated in Figure 4 for a small Pacific Island population, which shows the large fluctuations in annual ASMR and the smoothed trend data produced by using a 3-year moving average. Alternatively, it is possible to expand the age group or area to be studied, thus increasing the numbers of deaths in the calculation of ASMR.

Interpreting age and sex patterns of mortality

Age-specific mortality rates

Once the ASMR has been calculated for each age group and sex, the next step is to examine the pattern of the data by age to assess plausibility. In order to do this, it is important to have an independent source of comparative data on ASMR—for example, the census. If there is no independent source within a country, it is possible to compare the ASMR with figures from similar countries and settings. The following examples can help in improving the understanding and interpretation of age and sex patterns of mortality in a given country. They also show how this analysis can assist in determining the quality and completeness of the mortality data within specific age groups.

As a general rule, in all settings, mortality rates are high during infancy and early childhood and fall to their lowest levels between the ages of 5 and 14 years. Subsequently, mortality rates start to rise with increasing age and increase exponentially beyond age 35 or so. Figure 5 shows patterns of mortality across age for Australia, where death registration is complete, compared to Russia and South Africa, where death registration is less complete or essential information

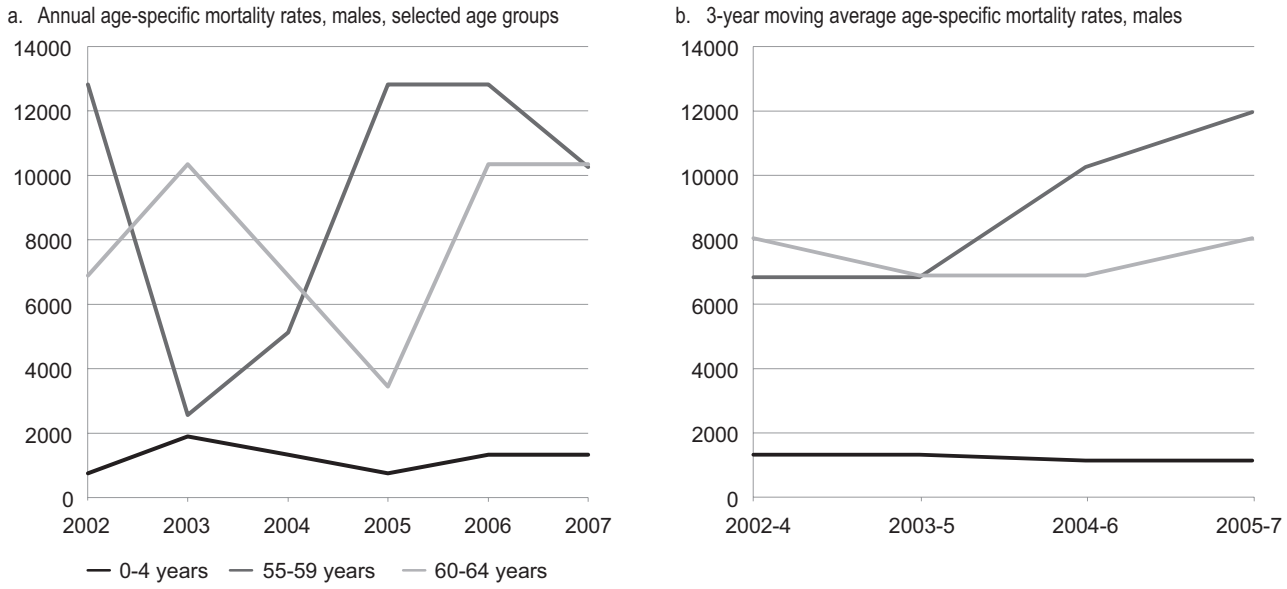


Figure 4 Annual age-specific mortality rates for selected age groups (males) and smoothed trends using a 3-year moving average

about the death is missing (eg unknown age or sex). In Australia, mortality rates are very low up to the age of about 15 years old, and although there is a small increase for males during the ages of 15–34 years due to accidents and other injuries, death rates only really begin to rise sharply after about age 55 years. This pattern is typical of most low-mortality populations.

In Russia and South Africa, mortality in infants is relatively high (this is particularly marked in South Africa) but declines during childhood. In South Africa, there is a 'bump' in mortality during reproductive ages in both sexes, reflecting premature mortality due to AIDS-related illnesses. A similar bump may occur in females of reproductive ages in settings where maternal mortality is very high.

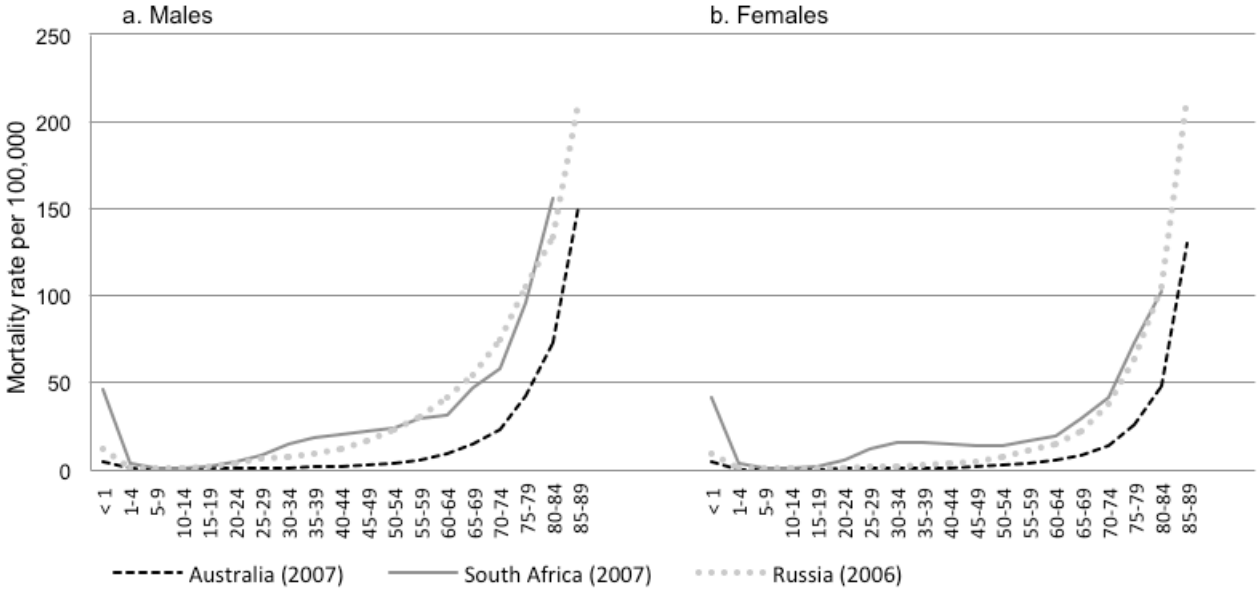


Figure 5 Age-specific mortality rates for Australia, Russia and South Africa, males and females

Comparing your data with this pattern can provide a simple check on the quality of the mortality data and indicate possible underregistration of deaths at certain ages. It is not the level of mortality that matters in this comparison but the relative age pattern of the ASMR among different age groups.

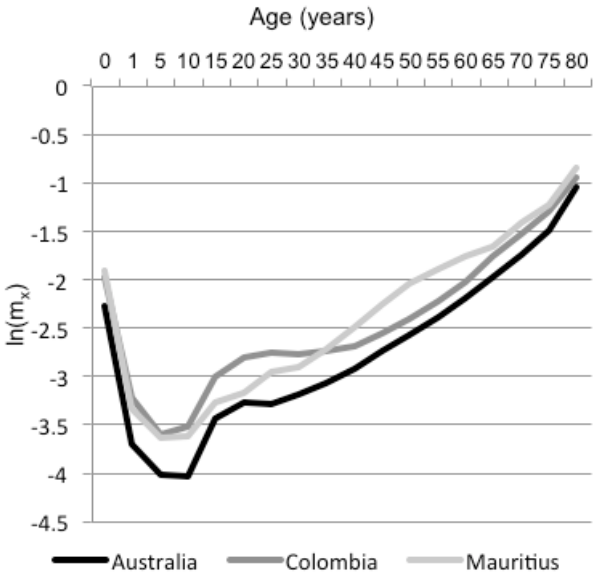
As noted above, beyond about 35 years of age, death rates rise exponentially with age. Therefore, the natural logarithm of the age-specific death rate (mx), written as ln(mx), should be a straight line as age (x) increases.[3] Figure 6 shows examples of ln(mx) for three countries—Australia, Colombia and Mauritius—with very different patterns of mortality and variable quality of mortality data.

The primary purpose of preparing a graph of the log of the death rate at each age is to examine the data for irregular or implausible changes in ln(mx) from age to age. In countries with high maternal or injury mortality in young adults (especially males), death rates will rise steeply (ie ln(mx) will rise) around age 15 years, peak at age 25, and decline to a new low at about age 35 years old. Subsequently, the ASMR will rise linearly with age. Any other departure from this linear pattern in adult death rates suggests that deaths are being selectively (by age) underreported or that there is misreporting of the correct age of death. This is particularly common at older ages.

With this in mind, we can make the following observations from Figure 6 showing age-specific death rates for males:

- Australia—All deaths are registered and hence the ln(mx) increases smoothly in a straight line with increasing age (x), as would be expected. Note the slight bump around ages 15–25 years old, indicating an excess in injury-related deaths in this age group.
- Mauritius—Notice that in this case the ln(mx) does not increase linearly with age after about age 65, suggesting underreporting of deaths, particularly at the oldest ages.
- Colombia—Note the large bump in mortality at ages 15–34 years old due to accidents and other violent deaths. One would expect to see a similar large bump in the ln(mx) graph at these ages in countries with high AIDS-related mortality.

3 'mx' is the standard demographic notation to indicate the level of the ASMR (written as 'm') in any age group 'x'



Source: Institute for Health Metrics and Evaluation database

Figure 6 Log of male age-specific death rates for Australia, Mauritius and Colombia

Thus, plotting the ln(mx) will help to identify if there are any age groups where deaths are being selectively underreported (eg older ages in Mauritius). In addition, by comparing the graph of ln(mx) for your population with a neighbouring country with good quality mortality data, it will be possible to assess whether, and to what extent, deaths are being systematically underreported at all ages. This will be the case if the graph for ln(mx) for your population is systematically lower than the graph for a neighbouring population.

Ratio of male to female mortality rates

As already observed, mortality rates tend to be higher at all ages for males than females. To better understand these male–female differences, it is useful to calculate the ratio of male:female mortality rates by age group. If the ASMR was the same for both sexes, the ratio would be 1 (ie a straight line) for all ages. In practice, the male:female ASMR ratio shows considerable variation over different age groups and at different period of time. Figure 7 shows typical patterns of the male:female ratio in settings with different overall mortality levels, as reflected by levels of infant mortality.

Male death rates are higher than female death rates everywhere except in societies with very low female

status. As the status of women in society improves and discriminatory practices against females disappear, female death rates should be lower than male rates at all ages (Waldron 1982). As Figure 7 shows, in settings with high levels of infant mortality (>100 per 1000 live births), the male mortality excess is relatively small because of high female mortality in reproductive ages. As overall mortality declines, this pattern changes and male mortality is higher than female mortality across all age groups. As already noted, death rates among males aged 15–29 years old tend to be higher largely due to accidents and other external causes. A secondary peak in the male:female ratio of mortality rates typically occurs around ages 55–64 years because males tend to die at higher rates from chronic disease than females, due primarily to increased risk factors such as tobacco, poor diet and being overweight or obese.

Users should prepare a similar chart showing the male:female ratio of age-specific death rates based on the latest available mortality data and compare your pattern with one of the curves shown in Figure 7. If the pattern of male:female ratio of age-specific death rates is very different from what would be expected given your level of infant mortality, there are good reasons for questioning the quality—that is, the completeness of death registration—of the reported data, particularly for females.

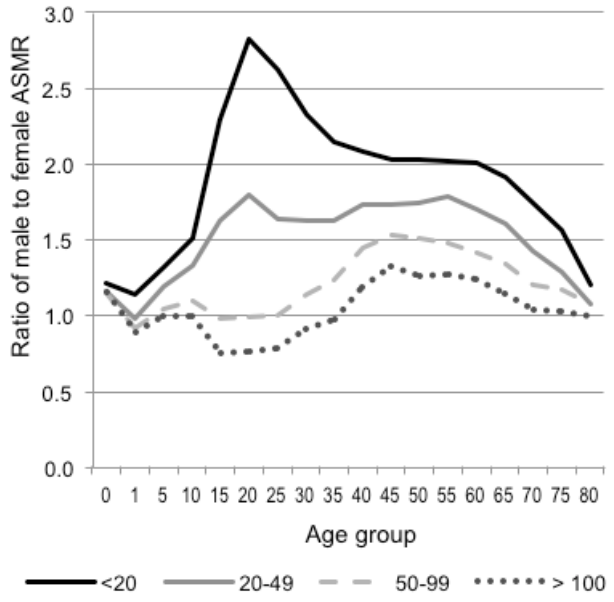


Figure 7 Male:female age-specific mortality rate (ASMR) ratios at different levels of infant mortality

Note that in comparing your age patterns of the sex mortality ratio to one of those from Figure 7, it is important to use an independent value of the infant mortality rate derived from censuses or surveys, or estimated by the United Nations, WHO or other sources. Do not use the value from your vital registration data, which could be underestimated.

Summary of step 3

- Calculate age and sex-specific mortality rates.
- Examine the ASMR across all age groups for each sex separately. You should find a pattern of relatively high mortality in the 0–4 years age group, very low mortality in the age groups 5–14 and an exponentially increasing mortality rate after the age of about 35.
- Plot the logarithm of the death rate at each age. It should increase smoothly and linearly with age after about 35 years old.
- Examine the ratio of male:female ASMRs across all ages. In general, you would expect male mortality rates to be higher than for females, especially in the age groups 15–35 years old, as young males are more likely to die as a result of violence, road traffic accidents and other external causes. High mortality rates in young adults may also be due to AIDS-related illnesses. In some cases, female deaths are less likely to be recorded than male deaths, leading to higher than expected ratios of male:female death rates.

Step 4 Age distribution of deaths

In Step 3 we looked at the age and sex-specific mortality rates, and at how these vary at different levels of overall mortality. The objective of Step 4 is to examine the age distribution of reported deaths. This age distribution should look quite different depending on the overall level of mortality in a population. The basic tabulations of data prepared in Step 1 can be used to prepare a chart showing the distribution of deaths by age group. You should use that same broad age group as shown in Figure 5 to tabulate your mortality data for this exercise. Your calculated distribution of deaths should then be compared with one of the expected distributions shown in Figure 8 that most closely resembles the level of mortality in your population, as reflected in the infant mortality rate.

To determine which of the four models is most relevant to your situation, use an independent estimate (derived from censuses or surveys, or estimated by the United Nations, WHO or other sources) of the infant mortality rate as follows:

- If your infant mortality rate is less than 20 per 1000, the age distribution of your reported deaths should be similar to that shown in panel A in Figure 8.
- If infant mortality is between 20 and 50 per 1000, the age distribution of your reported deaths should be similar to that shown in panel B in Figure 8.
- If infant mortality is between 50 and 100 per 1000, the age distribution of your reported deaths should be similar to that shown in panel C in Figure 8.
- If infant mortality is over 100 per 1000, the age distribution of your reported deaths should be similar to that shown in panel D in Figure 8.

Significant departures from these model age distributions of deaths suggest that the reporting of deaths by age is selectively biased. One reason for such bias may be the way age at death is reported. For example, people tend to have a strong preference to report age at death as a number ending in 0 or 5 (eg 45, 50, 55). This is commonly known as digit preference or age heaping. In other instances, the age of the deceased person may be misreported; it is common for families to report that the deceased person was older than they actually were. This highlights the importance of checking the plausibility of age patterns of mortality, and to test for underreporting of deaths in certain age groups by plotting the graph of ln(mx) versus age (x), as described above.

An example of the application of this check on data quality is shown in Figure 9, which gives the reported age distributions of deaths calculated from civil registration data for Sri Lanka, and from the Sample Registration System (SRS) for India. Sri Lanka has an estimated infant mortality rate of 8 per 1000 (hence panel A should be used as the comparator) while the infant mortality rate for India is closer to 60 per 1000 (hence panel C is chosen). This comparison shows that the age distribution of deaths in Sri Lanka is very similar to what was expected (panel A), but in India, the SRS appears to have more deaths at ages 60–74 years and fewer deaths at ages 75+ than expected from a comparison with panel C. This may or may not reflect problems with misreporting of the age at death for older adults, and should be investigated further.

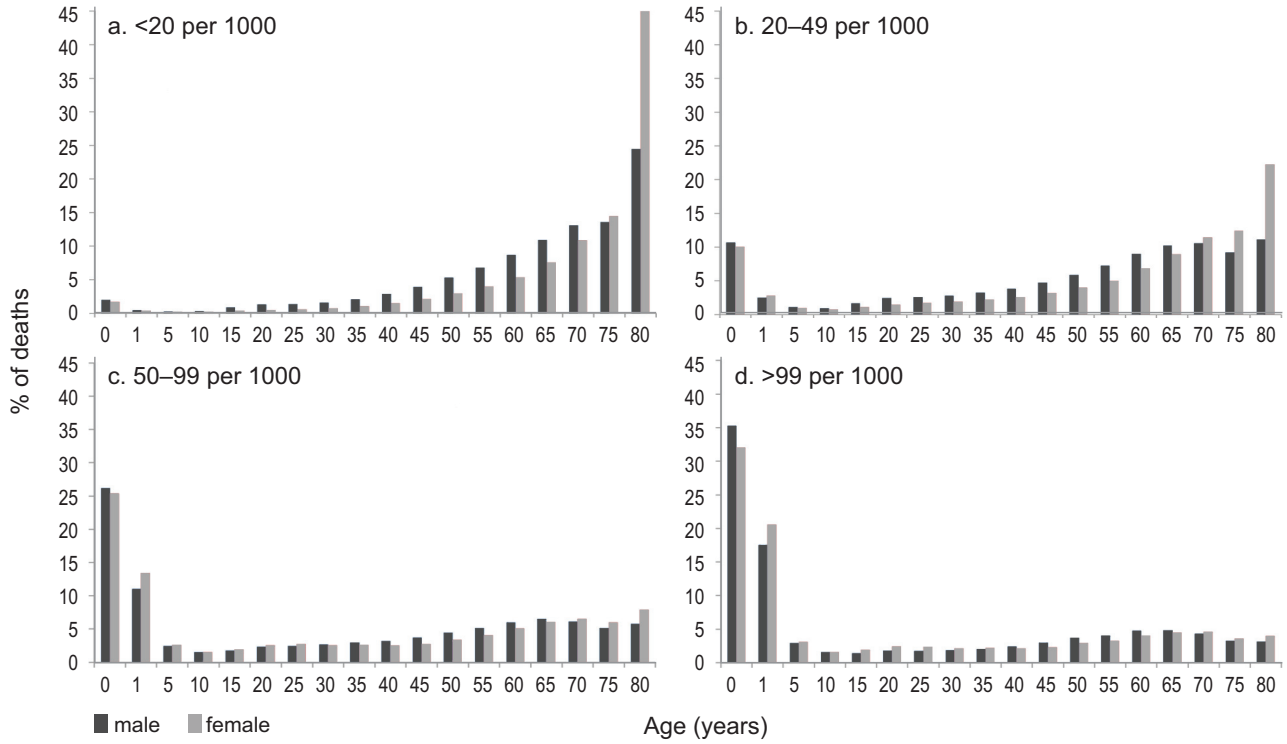


Figure 8 Typical age distributions of reported deaths at different levels of infant mortality

Draw a chart showing the distribution of deaths by age (for each sex separately) and compare the pattern you see with that which would be expected given your level of infant mortality. Distortions in mortality patterns may be due to poor recording of age at death and should be investigated.

Summary of step 4

- Compare the age and sex distribution of your reported deaths with expected age–sex distributions based on your estimated level of infant mortality as shown in Figure 8. Departures from these expected patterns can be indicative of underreporting of deaths at certain ages for males or females. If, for example, you have very low infant and child mortality rates and also low adult mortality rates, you should suspect problems with the registration of adult deaths.

Step 5 Child mortality rates

Mortality among children under five years old, more than any other age group, reflects a range of economic, social and health conditions that all affect population health. Child mortality is therefore a key indicator for public health monitoring. Mortality in children under five can be divided into several components:

- neonatal mortality—mortality among infants aged less than 28 days old
- postneonatal mortality—mortality in infants older than 28 days but less than 1 year old
- infant mortality—mortality among infants aged less than one year (neonatal and postneonatal deaths)
- under-five mortality—mortality among children aged less than 5 years old.[4]

The objectives of Step 5 are to enable users to:

- define and calculate indicators of under-five mortality
- understand the public health relevance of measures of under-five mortality
- interpret the indicators of under-five mortality and understand their limitations
- use under-five mortality indicators from various sources to analyse the quality of mortality data.

Definition and calculation of under-five mortality indicators

Under-five mortality rate

The under-five mortality rate (U5MR) is defined as deaths in children aged 0–4 years in a given population over a specified time period divided by the total number of live births in that population over the same period.

$$\text{U5MR} = \frac{\text{Number of deaths in children aged less than five in a specified time period}}{\text{Number of live births in the same time period}} \times 1000$$

However, because of the very different age pattern of mortality risks among children, it is usual statistical practice to transform the mortality rate in children under five into a probability of dying before age five, assuming that children would be subject to the ASMRs of that period. Thus, the U5MR is, strictly speaking, not

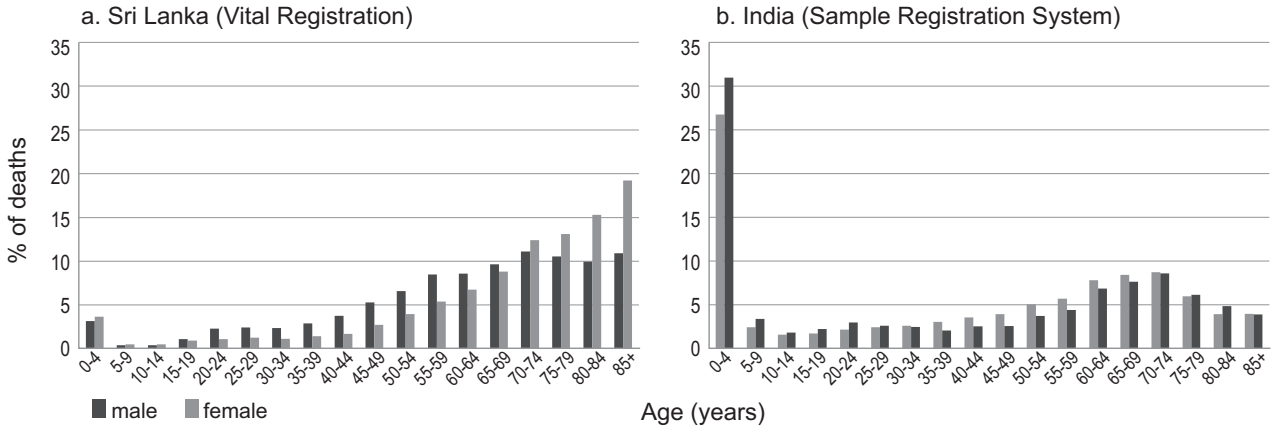


Figure 9 Age distribution of reported deaths in Sri Lanka and India

4 Mortality in children aged between 1 and >5 years is commonly referred to as child mortality.

a rate (ie the number of deaths divided by the number of population at risk during a certain period of time) but a probability of death, expressed as a rate per 1000 live births.[5]

Infant mortality rate

The calculation of the infant mortality rate (IMR) is the same as for the U5MR with the exception that the numerator is the number of deaths in children aged less than one year old (ie died before their first birthday).

$$\text{IMR} = \frac{\text{Number of deaths in infants aged less than one year old in a specified time period}}{\text{Number of live births in the same time period}} \times 1000$$

Neonatal mortality rate

The calculation of the neonatal mortality rate (NNMR) is the same as for the IMR with the exception that the numerator only includes deaths in children less than one month (28 days) old.

$$\text{NNMR} = \frac{\text{Number of deaths in infants aged less than 28 days in a specified time period}}{\text{Number of live births in the same time period}} \times 1000$$

Neonatal deaths may be subdivided into early neonatal deaths, occurring during the first seven days of life, and late neonatal deaths, occurring after the seventh day but before 28 completed days of life.

Postneonatal mortality rate

The calculation of the postneonatal mortality rate (PNNMR) is the same as for the NNMR with the exception that the numerator only includes deaths in infants aged from 28 days to one year old.

$$\text{PNNMR} = \frac{\text{Number of deaths in infants aged between 28 days and one year old in a specified time period}}{\text{Number of live births in the same time period}} \times 1000$$

5 There is a well-defined method for calculating the probability of a child dying between birth and age 5 years (written as ${}_5q_0$) from data on the ASMR at age 0 (defined as deaths at age 0 divided by mid-year population at age 0, and written ${}_1m_0$) and at age 1–4 years (defined as deaths at age 1–4 years divided by mid-year population at ages 1–4 years, written as ${}_4m_1$). Specifically,
${}_5q_0 = 1 - (1 - {}_1q_0)(1 - {}_4q_1)$
where ${}_1q_0 = {}_1m_0/(1 + (0.7)\,{}_1m_0)$ and ${}_4q_1 = ((4)\,{}_4m_1)/(1 + (2.4)\,{}_4m_1)$
where ${}_1q_0$ is the probability of an infant dying between birth and their first birthday, and ${}_4q_1$ is the probability of an infant who survives until their first birthday dying before age 5 years. These calculations are performed automatically in the accompanying electronic tool.

Definitions

The reliability of under-five, infant and neonatal mortality estimates depends on the accuracy and completeness of reporting and recording births and deaths. It is essential to apply standard international terminologies and definitions to ensure comparability over time, and across areas or countries. These have been defined in the WHO ICD-10 (WHO 2007). Differences in IMRs, and especially NNMRs, can be greatly affected by the failure to apply the standard definition of live birth.[6] In practice, underreporting and misclassification of under-five deaths are common, especially for deaths occurring very early in life, many of which are misclassified as stillbirths. In such cases, countries often do not record both the early neonatal death and the live birth. This is poor public health practice, as data on both events are critical to improve maternal and child health services. An example of the calculation of the U5MR, IMR and NNMR based on birth registration and death data is given below.

Table 3 Child deaths by age and calculation of mortality indicators

	Male	Female	Total
Neonatal deaths registered	1 563	895	2 458
Infant deaths registered	2 075	1 677	3 752
Under-five deaths registered	3 980	3 456	7 436
Live births registered	191 263	182 275	373 538

Neonatal mortality rate (both sexes combined) = (2458/373 538)*1000 = 6.6 per 1000
Infant mortality rate (both sexes combined) = (3752/373 538)*1000 = 10.0 per 1000
Under-five mortality rate (both sexes combined) = (7436/373 538)*1000 = 19.9 per 1000
Note: The U5MR would then need to be converted into the probability of dying before age 5 years (${}_5q_0$) in order to use it to assess the completeness of recording of child deaths in the vital registration system.

Sources of data on under-five mortality

In principle, the civil registration system can generate annual data on under-five mortality at both national

6 Live birth: The complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of the pregnancy, which, after such separation, breathes or shows any other evidence of life such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles, whether or not the umbilical cord has been cut or the placenta is attached (ICD-10).

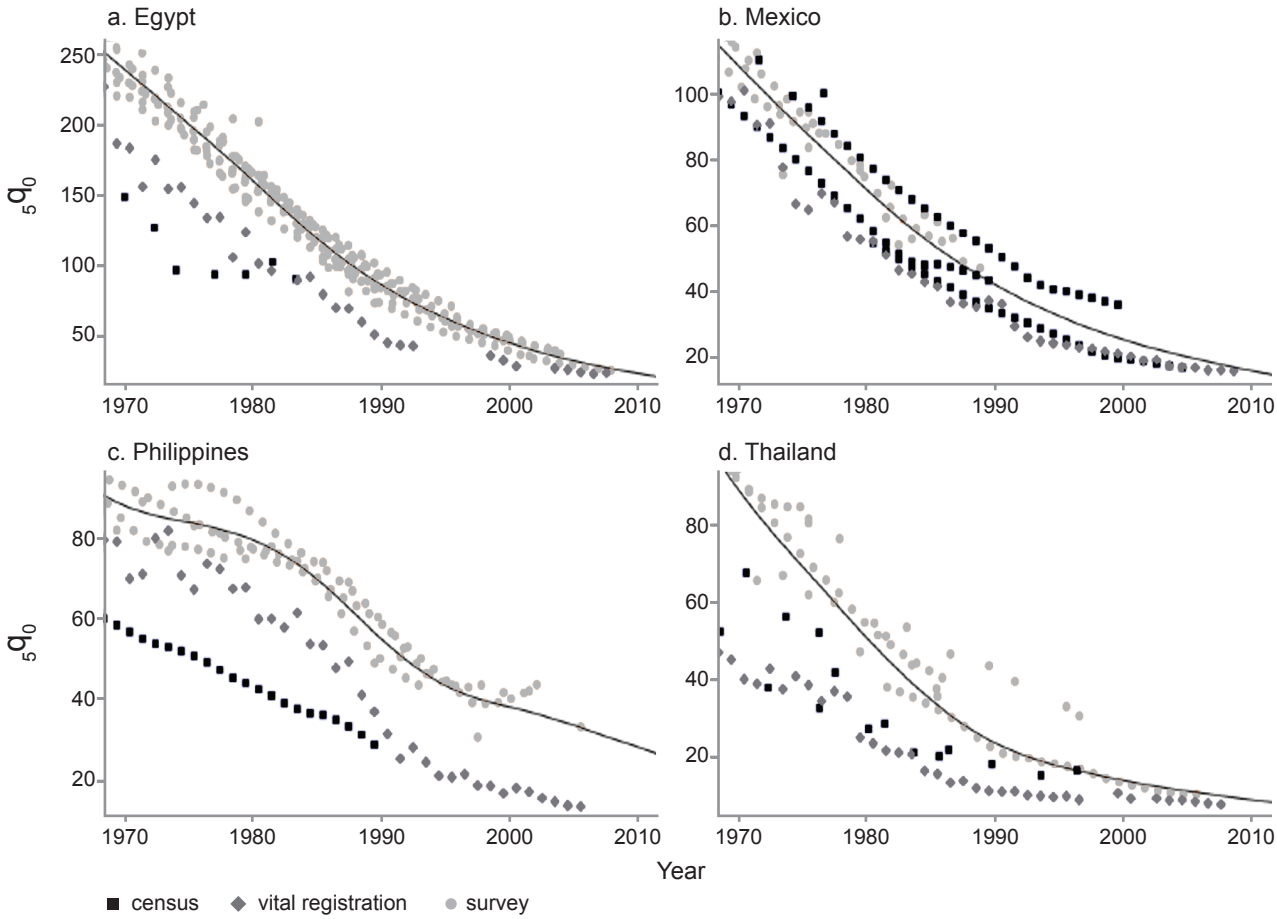


Figure 10 Observed (from vital registration) and estimated levels of the under-five mortality rate, selected countries (1960 - 2000)

and subnational levels, and on a continuous basis. Where civil registration systems are complete, ASMRs among children and infants can be calculated directly from the number of deaths by age and number of births registered. However, the coverage and quality of civil registration systems is often questionable in developing countries, and the resulting vital statistics may be incomplete and biased.

There are particular reasons why deaths occurring in young children are less likely to be registered than deaths in adulthood. In settings where civil registration is not universal, deaths are generally only registered when there are some benefits attached to doing so; for example to claim land ownership and inheritance, or to claim compensation by the dependants. Registering the death of a child is not usually linked to such a benefit and as a result many such deaths remain unregistered. In such settings, data on infant and child mortality estimated from censuses and surveys tend to be more reliable.

In countries with incomplete registration systems, census done every 10 years can be used to generate estimates of child mortality using direct or indirect techniques (UNPD 2001). The direct method involves questions to respondents about deaths in the household during a specified period of time. More commonly, an indirect method is used based on questions to female respondents on children ever born and children that are still alive. Brass-type methods and model life tables are then used to obtain an estimate of under-five mortality (UNPD 1983). However, the census is, by definition, an infrequent occurrence (ie only every 10 years), so it is not a good source of data for ongoing monitoring. It does, however, serve a very useful function of providing an alternate source that can be used to validate data from vital registration on the number of child deaths registered and hence the level of child mortality.

In most developing countries, household surveys provide the most common source of data on child mortality using both direct and indirect methods. The indirect method asks questions about children ever born and children still alive, as for the census. The direct method

involves taking a detailed birth history for each birth that a woman has had during her lifetime. These births histories are then converted to rates of child mortality corresponding to a particular period in time.

Interpreting different estimates of the under-five mortality rate

Most countries have data on child mortality from multiple sources, including the civil registration system, censuses, household surveys and the routine health information system. In this section, we show how information from reliable censuses and surveys can be used to assess the completeness of child mortality reporting by the civil registration system.

In order to compare the data reported from civil registration with estimates from other sources (eg census), household surveys or estimates developed by United Nations agencies, the numbers of deaths and population for age groups 0 years and 1–4 years are used to calculate age-specific death rates, which are then converted into an age-specific probability of dying. Large differences between U5MRs calculated from the reported data, and the levels estimated from censuses and surveys by international agencies are likely to be due to underreporting of child deaths in the country.

Figure 10 shows U5MRs for Egypt, Mexico, the Philippines and Thailand. The data are derived from various sources, including censuses, surveys and the civil registration system.

This visual display of data from different sources clearly shows the extent to which the U5MRs derived from civil registration appear to be systematically lower than those derived from the census or household surveys, especially during the earlier periods. This is indicative of substantial underreporting of deaths of children under five in the civil registration system. By comparing the line of best fit for the estimated U5MR derived from censuses and surveys with observed values calculated from the civil registration system for the same year(s) (symbolised by diamonds in Figure 10 for each country), it is possible to estimate the completeness of civil registration of child deaths by comparing the distance of the vital registration estimate from the solid line, year by year.

This analysis concluded that under-five deaths in Thailand were grossly underreported in the national civil registration system during the 1970s and 1980s. However, levels of reporting appear to have improved dramatically in the most recent decade (the trend in the vital registrations estimate for Thailand is getting closer and closer to the solid line of best fit for the true level of the child mortality rate). Similarly, the registration system in the Philippines appears to have significantly underestimated the U5MR, especially in the earlier period. Underreporting of under-five mortality in Egypt and Mexico appears to have diminished significantly in recent years.

Users should produce similar figures for their country or populations with death registration, bringing together on one graph estimates of under-five mortality derived from difference sources, including civil registration, to help interpret the multiple data points and diagnose possible incompleteness levels in death registration. To facilitate this, users can refer to the Child Mortality Estimation database (WHO and United Nations Children's Fund), which brings together available datasets from different sources on a country-by-country basis, and presents the information in tables and figures.[7] Plots of child mortality are also available from the Institute for Health Metrics and Evaluation, University of Washington, which also maintains a database of child mortality data (Rajaratnam et al 2010).

Direct measures of incompleteness of death reporting

Special studies can also be carried out to determine the extent of underreporting of deaths. The most widely used of these so-called direct methods are 'capture–recapture' studies where deaths reported in the civil registration system for a sample of the population are compared (on a case-by-case basis) with deaths 'captured' in an independent survey of the same population.[8]

This capture–recapture methodology (more formally known as the Chandrasekar–Deming method) can be used to estimate underreporting of deaths in any routine mortality surveillance system (Sekara and Deming 1949). Table 4 shows the results of a capture–recapture study of deaths reported in the Chinese national

7 www.childmortality.org/cmeMain.html

8 'Independence' as applied to capture–recapture studies means that the probability of a death not being reported under the civil registration system is not related to (ie is independent of) the probability that the same death will not be reported in another system or survey. In practice, this is very difficult to achieve.

disease surveillance points system in the late 1990s. This confirmed the higher rate of underreporting of death among children compared with adults and among females compared with males at all ages (Rao et al 2005).

Table 4 Underreporting of deaths by age and sex (per cent), Disease Surveillance Points system, China (1996–98)

Sex	<5 years	5–29 years	30–59 years	>60 years	Total
Male	19.8	12.6	10.7	12.6	12.4
Female	23.6	18.6	14.1	13.2	14.1
Total	**21.6**	**14.7**	**12.0**	**12.9**	**13.1**

Table 5 shows the results of a study in Thailand that estimated the percentage of underreporting of deaths by age group in the civil registration system (Popakkam et al 2010). Again, underreporting of deaths was found to be much higher in the 0–4 years age group, probably due to the reasons described earlier in this section.

Table 5 Underreporting of deaths by age, Thailand (2005)

	Age groups				
	0–4	5–49	50–74	75+	All ages
Percentage undercount in the civil registration system	42.8	14.8	7.7	5.9	8.7

Although not all countries will have the technical and financial resources to carry out capture–recapture studies, we have illustrated their application here to highlight the fact that underreporting of deaths is likely to be much higher among children than adults, and hence special attention should be paid to evaluating probable levels of underreporting of child deaths using the methods proposed in this section.

Summary of Step 5

- Calculate under-five, infant, neonatal and postneonatal mortality rates, and convert the U5MR to a probability of dying before age five years.
- Bring together, in one chart, estimates during the past 20–30 years of the probability of dying before age 5 ($_5q_0$) from different sources, including civil registration, the census, household surveys and other studies, as shown in Figure 10. Use the results to estimate the likely degree of underreporting of deaths in children less than five years old in the civil registration system by comparing levels with those estimated from censuses or surveys.

Steps 6–10 Cause-of-death

Steps 6–10 focus on simple steps to assess the plausibility of data on causes-of-death.

Information on the levels and patterns of mortality among different population groups is essential for public health authorities and for the effective allocation of resources to health care. However, a fully functioning civil registration and vital statistics system should not only register deaths by age and sex, but should also have mechanisms for assigning the cause-of-death according to international standards as expressed in the ICD-10. Only a medically qualified doctor should determine the cause-of-death. A coding expert trained in the ICD-10 rules and principles should determine the underlying cause-of-death, from a death certificate properly filled out by a physician, as defined in the ICD-10. Note that this coding expert should not be a medical doctor as this is not the best use of their time.

The objectives of steps 6–10 are to enable users to:

- calculate broad patterns of causes-of-death using available data on mortality by age, sex and cause
- critically analyse and interpret cause-of-death data
- assess the plausibility of the cause-of-death patterns emerging from the data.

Definition of the underlying cause-of-death

The quality of cause-of-death data depends on the reliability of death certification and the accuracy of coding. These are two separate, but related, functions. Death certification, which should only be done by a qualified medical practitioner, involves correctly completing an international form (medical certificate of death). This information is then translated into a code (alpha-numeric digital code) from among the approximately 3000 underlying causes-of-death in the ICD-10 by a qualified and trained coder (not the physician who certified the death, as they are unlikely to have been formally trained in the coding of information given on a death certificate).

There are well-established rules for assigning the cause-of-death. It is essential that deaths be classified not by the immediate cause-of-death but by the underlying cause; that is, the cause that initiated the sequence of events leading to the death. It is the underlying cause-of-death that generates information that is useful for public health purposes. The underlying cause-of-death, as defined by WHO, is the disease or injury that initiated the train of events leading directly to death, or the circumstances of the accident or violence that produced the fatal injury. Under international rules for selecting (ie coding) the underlying cause from the reported conditions, every death is attributed to one (and only one) underlying cause based on information reported on the death certificate. The International Form of Medical Certificate of Cause-of-Death was specially designed to facilitate the selection of the underlying cause-of-death when two or more causes are recorded on the death certificate. This certificate is shown in Box 1 and should be filled in only by a trained medical practitioner. Moreover, all countries are strongly urged to use this certificate to certify death, and not some other adaptation of it, which will be of limited public health value.

Currently, only about 70 WHO Member Countries produce good-quality cause-of-death data from their civil registration and vital statistics systems (Mathers et al 2005). Although a further 50 countries produce some cause-of-death data, the quality of the information is problematic because of poor certification and coding practices. In these settings, deaths that occur outside health care facilities and hospitals are rarely medically certified and consequently many of these deaths are assigned to nonspecific or ill-defined causes.

Even where medical certification of the cause-of-death is common practice, it does not necessarily mean that the correct cause-of-death is written on the death certificate in the correct way. Most doctors certify death infrequently, and their medical school training may have been forgotten or be out of date. Lack of diagnostic

Box 1 International form of medical certificate of cause of death

INTERNATIONAL FORM OF MEDICAL CERTIFICATE OF CAUSE OF DEATH

Cause of death		Approximate interval between onset and death
I Disease or condition directly leading to death*	(a)	.....
	due to (or as a consequence of)	
Antecedent causes Morbid conditions, if any, giving rise to the above cause, stating the underlying condition last	(b)	.....
	due to (or as a consequence of)	
	(c)	.....
	due to (or as a consequence of)	
	(d)	.....
II Other significant conditions contributing to the death, but not related to the disease or condition causing it	.....	.....
	.....	.....

**This does not mean the mode of dying, e.g. heart failure, respiratory failure. It means the disease, injury, or complication that caused death.*

facilities and awareness of the importance of cause-of-death data, combined with inexperience and human error, contribute to poor diagnostic accuracy. In addition, there may be financial or social consequences for the family that deter the doctor from reporting the true cause-of-death.

For all these reasons, any dataset with information on causes-of-death by age and sex should be carefully reviewed and assessed to identify and correct potential quality problems. Unless this is done as a matter of course, public health authorities using the data risk diverting resources away from those conditions that are causing the most serious problems of suffering and death in their communities.

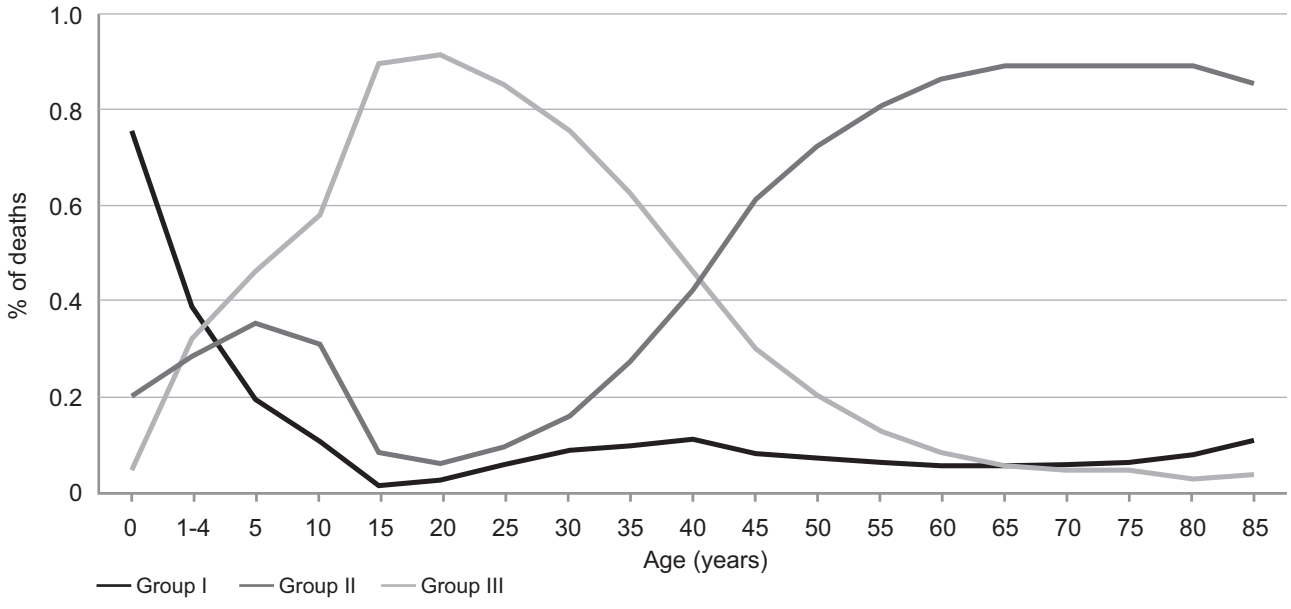
Step 6 Distribution of major causes-of-death

A first step in any quality assessment of cause-of-death data is to calculate the percentage of death distribution by broad disease groups and compare the results with what would be expected given the level of life expectancy for the population. These expected patterns have been developed by demographers and epidemiologists on the basis of many years of data and observations on patterns of causes-of-death in different settings. Any significant deviation from the expected pattern that cannot be explained by some local, external factor should be viewed as a potential problem with the quality of the cause-of-death data.

The ICD-10 contains over 3000 possible causes-of-death. All of these causes can be further condensed into three very broad groups of causes-of-death:

Group I[9] Infectious and parasitic diseases (eg tuberculosis, pneumonia, diarrhoea, malaria, measles)
Maternal/perinatal causes (eg maternal haemorrhage, birth trauma)
Malnutrition

9 ICD-10: A00-B99, G00-G04, N70-N73, J00-J06, J10-J18, J20-J22, H65-H66, O00-O99, P00-P96, E00-E02, E40-E46, E50, D50-D53, D64.9, E51-64



Redistributed unknown age and ill-defined diseases proportionately

Figure 11 Distribution of broad causes-of-death (groups I, II and III) by age (males, Venezuela 2007)

Group II[10] Noncommunicable diseases (eg cancer, diabetes, heart disease, stroke)
Mental health conditions (eg schizophrenia)

Group III[11] Injuries (eg accidents, homicide, suicide).

The expected percentage distribution of causes-of-death into these three broad groups varies in different countries according to where they stand in relation to the 'health transition'—an interrelated set of changes in demographic structures, patterns of disease and risk factors. Demographic changes include lower mortality rates among children under five years old and declining fertility rates, which result in an ageing population. Epidemiological changes include a shift in the main causes-of-death and disease away from infectious diseases, such as diarrhoea and pneumonia (diseases traditionally associated with poorer countries), towards noncommunicable diseases such as cardiovascular disease, stroke and cancers. Changes in patterns of risk include declines in risk factors for infectious diseases (eg undernutrition, unsafe water and poor sanitation) and increases in risk factors for chronic diseases (eg being overweight, and using alcohol and tobacco). Thus, a simple but effective way of checking the plausibility of mortality data is to compare the observed patterns of causes-of-death with what would be expected given the local levels of life expectancy. As a general rule, countries with low life expectancy are characterised by high levels of mortality due to infectious and parasitic diseases especially in childhood, along with high maternal mortality (ie Group I causes). As life expectancy rises, the pattern of mortality changes, with more deaths occurring in older age groups due to noncommunicable conditions such as cardiovascular diseases and cancers (ie Group II causes).

Table 6 shows how the percentage of deaths assigned to various causes in each of groups I, II and III is expected to change as life expectancy increases. Thus, a country with an average life expectancy of 55 years would typically have about 22 per cent of deaths due to Group I causes-of-death and 65 per cent due to Group II causes. A country with higher life expectancy of 65 years would typically have a smaller percentage of deaths due to Group I conditions (around 13 per cent) and correspondingly more deaths due to Group II conditions (74 per cent).

10 ICD-10: C00-C97, D00-D48, D55-D64 (minus D 64.9) D65-D89, E03-E07, E10-E16, E20-E34, E65-E88, F01-F99, G06-G98, H00-H61, H68-H93, I00—I99, J3—J98, K00-K92, N00-N64, N75-N98, L00-L98, M00-M99, Q00-Q99

11 ICD-10: V01-Y89

Table 6 Expected distribution of cause-of-death according to life expectancy by broad groups

Life expectancy	55 years	60 years	65 years	70 years
Group I cause-of-death	22%	16%	13%	11%
Group II cause-of-death	65%	70%	74%	78%
Group III cause-of-death	13%	14%	13%	11%

Note that these are model-based percentage distributions derived from WHO's large database on causes-of-death and mortality rates. It is unlikely that any country would fit exactly these proportions, but significant departures from them suggest potential problems with the certification or coding of causes-of-death.

Users should review their most recent available data on causes-of-death data and calculate the distribution by broad groups of causes (note that ill-defined causes, such as symptoms and cause-of-death unknown, should be excluded from the calculation of percentage of death assigned to groups I, II and III). The findings can then be compared with the expected distribution in Table 6 according to the average life expectancy in the country. However, in doing this comparison, it is important to use an independent source of life expectancy data (eg WHO, the United Nations or from your census), not the life expectancy calculated from the civil registration data, as this may be unreliable if the system is incomplete.

Summary of step 6

- Use a simple spreadsheet to tabulate your data on cause-of-death by age, sex and broad causes-of-death (groups I, II and III).
- Calculate the percentage distribution of deaths by broad cause groups (groups I, II and III). Do not include ill-defined causes. Compare the distribution with the expected distribution for a country with the same level of average life expectancy as your country, as shown in Table 6. Use an independent estimate of life expectancy for this comparison (eg from your country's census). Do not use life expectancy from the vital registration data unless they are known to be complete.

Step 7 Age pattern of broad groups of causes-of-death

All leading causes-of-death in a population follow a predictable age pattern that has been identified from decades of epidemiological research. The next step is to check whether the age pattern of deaths from broad causes is consistent with what one would expect from epidemiological research and modelling. These age patterns do not change very much with increasing life expectancy (although the percentage of deaths in each cause group will—see Table 6). Figure 11 shows a typical distribution of deaths across groups I, II and III at different ages for a country (Venezuela) with a life expectancy of around 70 years.[12] At each age, the graph shows the expected proportion (fraction) of deaths at that age that are likely to occur on average. At any age, the three fractions will add up to 100 per cent.

Figure 11 shows a commonly found pattern of distribution of causes-of-death by age in settings with relatively high life expectancy. Ill-defined causes-of-death have been omitted.

The proportion of deaths due to Group I causes (infectious, parasitic and maternal/perinatal causes) is high among children, but declines thereafter to very low levels, although it may rise again at older ages (above approximately 80 years old) due to pneumonia.

The proportion of deaths due to Group II causes is relatively high in children (eg due to some cancers), declines in adulthood, but rises significantly at older ages due to the increasing incidence of cancers, cardiovascular diseases and stroke.

The proportion of deaths due to Group III causes (ie external causes-of-death including accidents and violence) is generally highest in young adulthood. This pattern is especially marked among males.

This is a typical cause-of-death pattern by age and would not be replicated exactly in every country. However, significant departures from this pattern should be closely investigated as they are suggestive of problems such as poor death certification and coding practices, and age-specific misreporting of deaths.

In general, the charts for males and females should be broadly similar, although there is often higher mortality due to external causes among young males, while young women may have high mortality due to maternal causes (which would increase the fraction from Group I causes).

The principal reason for carrying out this step is to identify serious biases in the data. Depending on the data source, there are strong tendencies to avoid coding deaths to infectious diseases (or to overcode them) or to ignore injury deaths (Group III). This check will help to identify the extent of these biases in your data.

Summary of Step 7

- Plot the cause-of-death patterns by sex and age group, and compare your findings with the typical patterns for groups I, II and III shown in Figure 11.

Step 8 Leading causes-of-death

An analysis of leading causes-of-death can also indicate the reliability of cause-of-death data and is another way to check reporting in the civil registration system. Figure 12 shows the percentage distribution of leading causes (by specific disease groups) globally, and in low-income, middle-income and high-income countries (using definitions from the World Bank). These charts can assist countries to ascertain divergences in their reported leading causes-of-death compared with leading causes-of-death estimated by WHO and other researchers. These global estimates refer to the average experience of all countries in each of the country groups; hence, it is unlikely that the percentage distribution of deaths in any one country would match them exactly. However, significant departures from these average rankings of leading causes-of-death are suggestive of problems with the quality of cause-of-death data.

Note that these comparative distributions of leading causes-of-death do not include ill-defined causes. However, countries should include this category in their rankings in order to see how frequently these causes are coded. In many cases, ill-defined causes may be in the top three or four leading causes-of-death. This suggests serious problems with certification or coding in the country. These ill-defined causes—unfortunately, commonly reported—are of absolutely no value for informing public health policies and debates in countries.

12 WHO mortality database

Summary of Step 8

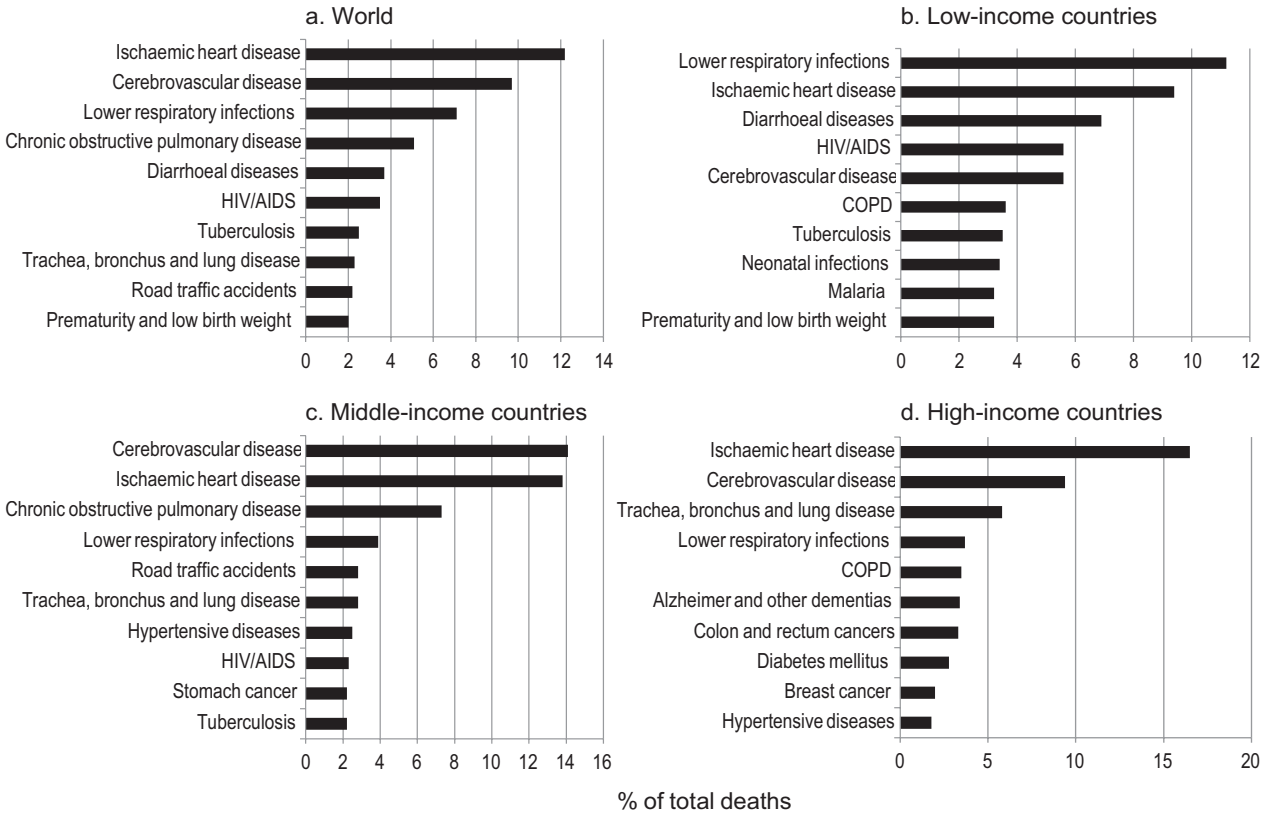
- Calculate the leading causes-of-death from your data and compare the findings with the typical patterns for all ages (both sexes) shown in Figure 12.

Step 9 Ratio of noncommunicable to communicable causes-of-death

As countries develop their health systems, communicable disease such as diarrhoea and pneumonia, as well as maternal, perinatal and nutritional risks will be increasingly brought under control. As a result, more people will survive to adulthood, where chronic diseases such as ischaemic heart disease, stroke, cancer and chronic obstructive pulmonary diseases claim more lives. Hence, the simple ratio of Group II:I deaths should progressively increase as a country moves through the epidemiological transition (ie as life expectancy increases).

This is illustrated in Figure 13, which shows the ratio of deaths from noncommunicable diseases (Group II) to communicable diseases (Group 1) in selected World Bank income groupings (both sexes combined) (WHO 2008). If there were the same numbers of deaths in each broad disease group, the ratio would be 1.

Figure 13 shows that, globally, there are more than twice as many deaths due to Group II causes as Group I causes. In high-income countries, noncommunicable diseases account for nearly 14 times as many deaths as communicable diseases. By contrast, in low-income countries, there are roughly the same numbers of deaths due to communicable as noncommunicable diseases, so the ratio is nearly 1. In middle-income countries, there are about five times as many deaths due to noncommunicable diseases compared with



HIV/AIDS = human immunodeficiency virus or acquired immunodeficiency syndrome

Figure 12 Leading causes of death globally, and in low, middle and high-income countries (2005)

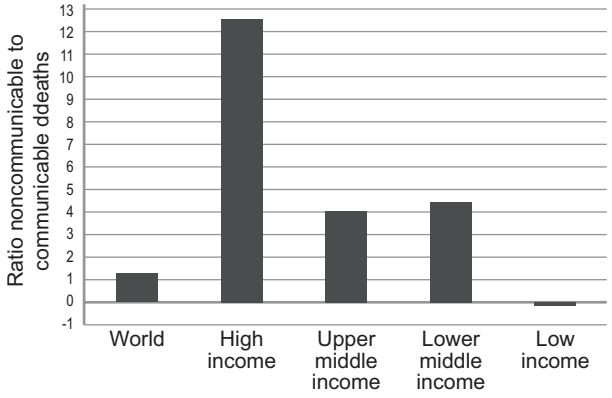


Figure 13 Ratio of noncommunicable to communicable diseases by country income groupings (2004)

communicable diseases. This reflects the fact that in high and middle-income countries, most deaths occur later in life, due to chronic conditions such as cancers and cardiovascular diseases. In low-income countries, by contrast, most deaths occur in childhood, due to infectious diseases conditions such pneumonia, diarrhoea and vaccine-preventable conditions, as well as perinatal causes.

Over time, as child mortality decreases and life expectancy increases, the pattern in low-income countries will start to look more like that observed in middle and high-income countries. This is illustrated in Figure 14, which shows estimated trends in the ratio of noncommunicable to communicable conditions in China, India and Latin America. In India in 1990, there were more deaths due to communicable diseases than to noncommunicable diseases; hence, the ratio is less

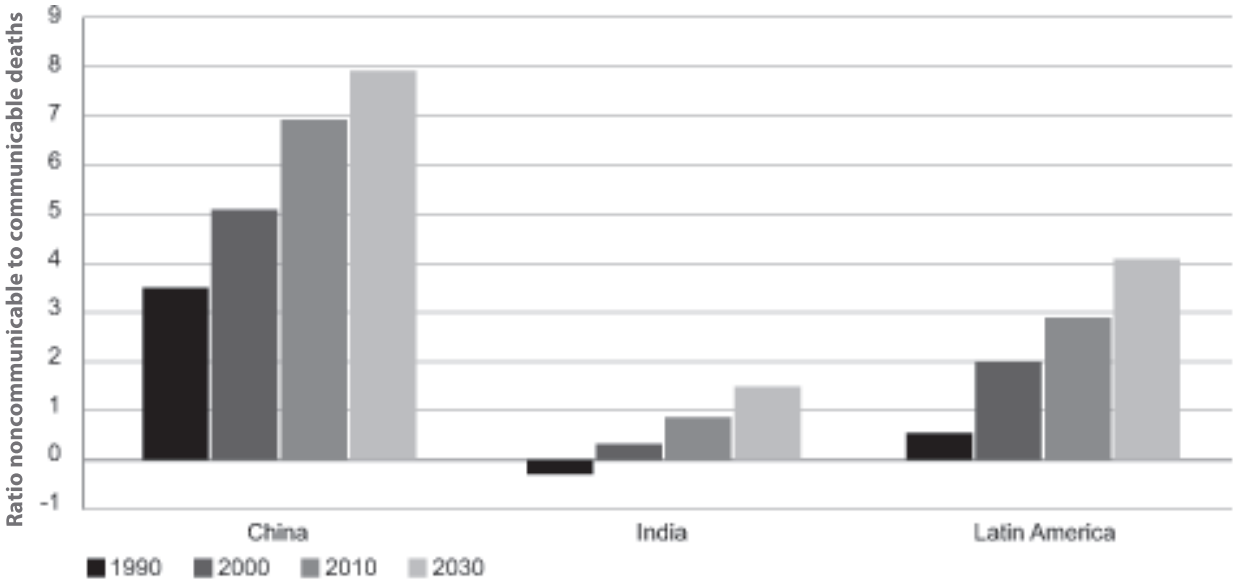


Figure 14 Estimated trends in ratio of noncommunicable to communicable deaths, selected regions (1990–2030)

than 1. Since 2000, however, deaths due to noncommunicable diseases have exceeded those due to communicable diseases.

Departures from this overall picture are suggestive of errors in cause-of-death data.

Summary of step 9

- Calculate the ratio of deaths from noncommunicable diseases to communicable diseases (Group II to Group I deaths) and compare your findings to those of the most appropriate comparator group as shown in Figures 13 and 14.

Step 10 Ill-defined causes-of-death

As noted in Step 6, when a death occurs and is medically certified, every effort should be made to correctly ascertain the underlying cause-of-death in order to be able to draw conclusions about the leading causes and about the need for priority public health interventions. Classification of deaths to ill-defined conditions does not generate information of public health value. Where a high proportion of all deaths is classified as being due to ill-defined causes, the cause-of-death distribution will be biased and unreliable.

At the end of this section, users should be able to:

- define and calculate the proportion of deaths attributed to ill-defined causes-of-death
- understand the implications for the overall quality of mortality statistics of a high proportion of ill-defined causes-of-death
- understand the definition and calculation of ill-defined categories in cause-of-death data.

Ill-defined causes are vague diagnoses often described as 'symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified' that the ICD-10 advises should not be used as the underlying cause-of-death.

These ill-defined codes arise from two sources:

i. Deaths classified as ill-defined (Chapter XVIII of ICD-10).

ii. Deaths classified to any one of the following vague or unspecific diagnoses:

- I46.1 (sudden cardiac death, so described)
- I46.9 (cardiac arrest, unspecified)
- I95.9 (hypotension, unspecified)
- I99 (other and unspecified disorders of the circulatory system)
- J96.0 (acute respiratory failure)
- J96.9 (respiratory failure, unspecified)
- P28.5 (respiratory failure of newborn)
- C76, C80, C97 (ill-defined cancer sites)
- Y10-Y34, Y872 (injury not specified, as accidentally or purposefully inflicted).

Deaths classified to either of these two categories of ill-defined diagnoses are insufficiently detailed to be of value for public health purposes, although in the majority of cases they help to describe the overall mortality due to broad disease (eg cardiovascular or respiratory disease) or injury groups. Separately identifying their frequency in cause-of-death tabulations is essential to decide upon remedial action to reduce their use. This could involve interventions to improve certification practices or coding practices, or both.

Although there will always be individual cases where it is not possible to classify the cause to a specific ICD-10 category due to lack of appropriate information, such cases should be relatively infrequent. As a general principle, the proportion of ill-defined deaths coded to either category i or ii (above) should collectively not exceed 10 per cent for deaths at ages 65 years and older, and should be less than 5 per cent for deaths at ages below 65 years.

When reviewing a data series of cause-of-death information, it is important to study how the proportion of ill-defined causes-of- death has changed over time. Large fluctuations may be indicative of changes in coding practices rather than real changes in patterns of mortality.

Table 7 provides a hypothetical example of how to assess the extent of ill-defined causes-of-death. Out of 12 341 deaths that occurred in this population in a given year, 2052 were assigned to either a category i (1021) or category ii (1031) diagnosis. Thus, the total proportion of deaths assigned to ill-defined causes is 2051/12 341 × 100 = 16.6 per cent, higher than what is considered desirable.

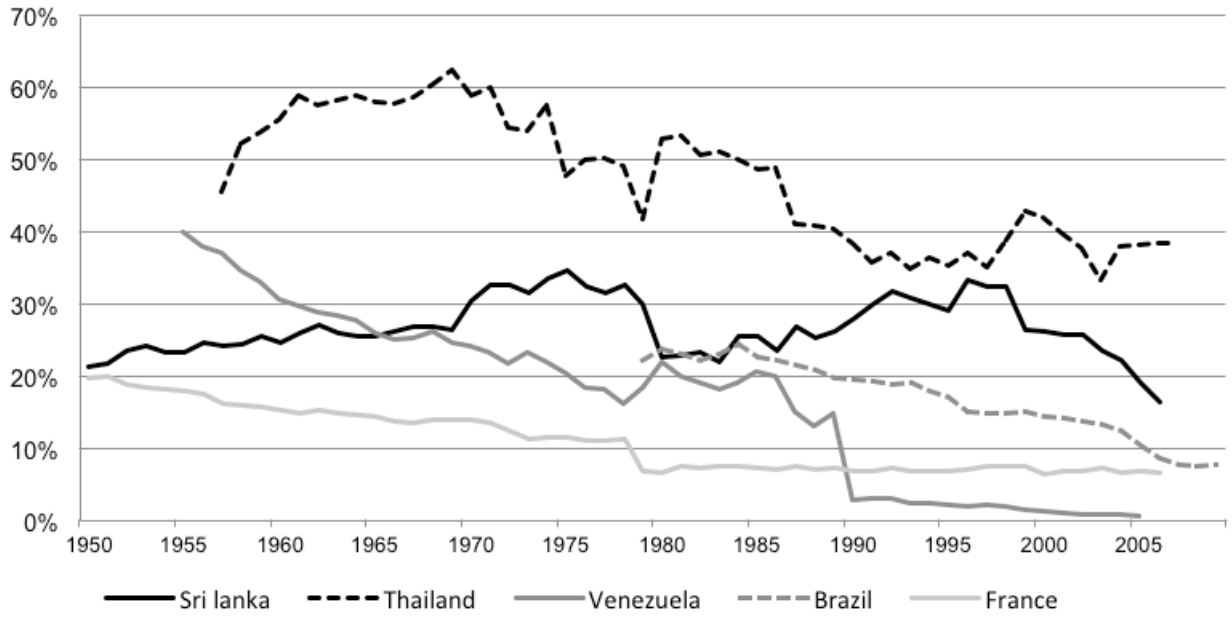


Figure 15 Trends in percentage of deaths assigned to ill-defined codes, selected countries (1950–2008)

Table 7 Calculating the percentage of deaths assigned to ill-defined causes

ICD-10 code	Number of deaths
I46.1	146
I46.9	203
I95.9	102
I99	174
J96.0	147
J96.9	161
P28.5	98
R codes	1 021
Total deaths attributed to ill-defined causes	2 052
Total deaths in population	**12 341**

Figure 15 shows the trend in the percentage of deaths assigned to ill-defined codes in selected countries for 1950–2000. Developed countries tend to have a lower percentage of deaths assigned to ill-defined categories than developing countries because of better developed cause-of-death reporting systems where all deaths are certified by a medical practitioner, which is often not the case in developing countries where a significant proportion of deaths occur outside hospitals.

Brazil has achieved significant reductions in the percentage of deaths assigned to ill-defined causes, with a decrease of more than 50 per cent between 1980 and 2008. In Thailand, ill-defined categories accounted for more than 40 per cent of all deaths in 2008. In Sri Lanka, the proportion of ill-defined causes-of-death remains unacceptably high despite some improvements in recent years. The overuse of ill-defined causes-of-death is not only an issue for developing countries. For example, in France in 1950, 20 per cent of all deaths were assigned as ill defined; however, by the early 1980s, the percentage had declined to less than 10 per cent. Both Brazil and Venezuela have achieved significant improvements in recent years, particularly Venezuela.

Conclusions

The proportion of deaths assigned to ill-defined causes tends to be higher for deaths occurring at older ages. There are many possible explanations, including the fact that many such deaths occur outside health care facilities and also because of the existence of multiple comorbidities that renders such deaths harder to correctly diagnose. Nonetheless, with good certification and coding practices, it should be possible to reduce this proportion to less that 10 per cent of deaths among the elderly.

Summary of step 10

- Calculate the proportion of category i and ii ill-defined causes in your cause-of-death data for ages <65, 65+ and all ages. The total should not exceed 5 per cent of deaths at ages below 65 and 10 per cent of deaths at age 65+.
- Calculate the trend in the proportion of ill-defined deaths (all ages) and use this information to interpret trends in specific causes-of-deaths.
- If the proportion increases or decreases over time, it is likely that real changes in disease-specific mortality will be correspondingly lower or higher than your data indicates. For example, if the proportion of ill-defined deaths has declined substantially, increases in the percentage of deaths observed for specific causes may largely be spurious, arising due to better certification and coding practices.

Improving the quality of vital statistics will be of inestimable value to public health decision-makers. It will greatly increase confidence in the data.

This guide and the accompanying electronic tool provide guidance on simple actions that can and should be taken to assess the quality of mortality data, particularly vital statistics on deaths and causes-of-death. The aim of conducting such a review of data quality is to diagnose problems and identify potential solutions. Solutions may include:

- extending civil registration to remote and underserved areas
- introducing incentives to encourage accurate reporting of all births and deaths
- improving the training of medical doctors in death certification
- improving the skills of coders to correctly assign underlying causes-of-death
- improving the quality and completeness of medical records so that doctors have all the information they need to correctly certify causes-of-death.

More specific guidance on interventions to improve data quality can also be gained by applying the full WHO/UQ Comprehensive Vital Statistics Assessment Tool.[13]

The guide places emphasis on three particular aspects of data quality:

- The completeness of the data. (Are all deaths registered?)
- The age pattern of reported deaths. (Is there serious age-specific misreporting or underreporting?)
- The plausibility of cause-of-death data using a series of comparisons and internal consistency checks.

Although these are essential, other dimensions of data quality might be considered as well, particularly timeliness. Cause-of-death data that are 5–10 years out of date are of reduced value for good health policy and planning. We have tried to write this guide so that the operation and rationale for the basic 10 steps are readily interpretable. Continuous data quality improvement

13 WHO and UQ 2010 www.uq.edu.au/hishub

requires continuous assessment. It is not intended that these steps be applied once or infrequently. They should form an integral part of the health information system.

Improving the quality of vital statistics will be of inestimable value to public health decision-makers. It will greatly increase confidence in the data, and thereby facilitate and promote the use of mortality and cause-of-death statistics to ensure that resource allocation is evidence informed, and focuses on interventions most needed to improve overall population health levels.

References

Coale AJ and Demeny P (1966). *Regional Model Life Tables and Stable Populations*, Princeton University Press, Princeton, NJ.

Mathers C, Fat DM, Inoue M et al (2005). Counting the dead and what they died from: an assessment of the global status of cause-of-death data. *Bulletin of the World Health Organization* 83(3):171–177.

Porapakkam Y, Rao C, Pattaraarchachai J et al (2010). Estimated causes-of-death in Thailand, 2005: implications for health policy. *Population Health Metrics* 8:14.

Rao C, Lopez AD, Yang G et al.(2005). Evaluating national cause-of-death statistics: principles and application to the case of China. *Bulletin of the World Health Organization* 83:618–625.

Sekara C and Deming W (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association* 44(245):101–115.

UNPD (United Nations Population Division) (2001). *Principles and Recommendations for a Vital Statistics System*, revision 2, Department of Economic and Social Affairs, Statistics Division, United Nations, New York. http://unstats.un.org/unsd/publication/SeriesM/SeriesM_19rev2E.pdf

UNPD (United Nations Population Division) (1983). *Manual X: Indirect Demographic Estimation*, United Nations, New York.

Waldron D (1982). The sex differential in mortality rates: Demographic and behavioral factors. *American Journal of Epidemiology* 15(2).

WHO (World Health Organization) (2007). *International Classification of Diseases and Related Health Problems*, 10th revision, WHO, Geneva.

WHO (World Health Organization) (2008). *Projections of Mortality and Burden of Disease, 2004–2030*, WHO, Geneva. www.who.int/healthinfo/global_burden_disease/projections/en/index.html

WHO (World Health Organization) and UQ (University of Queensland) (2010). *Improving the Quality and Use of Birth, Death and Cause-of-Death Information: Guidance for a Standards-Based Review of Country Practices*, WHO, Geneva. www.uq.edu.au/hishub//?page=123572 (Accessed 17 November 2010).

The Knowledge Hubs for Health Initiative

The Health Information Systems Knowledge Hub is one of four hubs established by AusAID in 2008 as part of the Australian Government's commitment to meeting the Millennium Development Goals and improving health in the Asia and Pacific regions. All four hubs share the common goal of expanding the expertise and knowledge base to help inform and guide health policy.

The Knowledge Hubs are funded by AusAID's Strategic Partnership for Health Initiative.

Health Information Systems Knowledge Hub

The University of Queensland

Aims to facilitate the development and integration of health information systems into the broader health system strengthening agenda, and increase local capacity to ensure that cost-effective, timely, reliable and relevant information is available. The Health Information Systems Knowledge Hub also aims to better inform health information systems policies across Asia and the Pacific.
www.uq.edu.au/hishub

Human Resources for Health Knowledge Hub

The University of New South Wales

Aims to contribute to the quality and effectiveness of Australia's engagement in the health sector in the Asia–Pacific region by developing innovative policy options for strengthening human resources for health systems. The hub supports regional, national and international partners to develop effective evidence-informed national policy-making in the field of human resources for health.
www.hrhhub.unsw.edu.au

Health Finance and Health Policy Knowledge Hub

The Nossal Institute for Global Health (University of Melbourne)

Aims to support regional, national and international partners to develop effective evidence-informed national policy-making, particularly in the field of health finance and health systems. Key thematic areas for this hub include comparative analysis of health finance interventions and health system outcomes; the role of non-state providers of health care; and health policy development in the Pacific.
www.ni.unimelb.edu.au

Compass: Women's and Children's Health Knowledge Hub

Compass is a partnership between the Centre for International Child Health, The University of Melbourne, Menzies School of Health Research and Burnet Institute's Centre for International Health.

Aims to enhance the quality and effectiveness of women's and children's health interventions and focuses on supporting the Millennium Development Goals 4 and 5—improved maternal and child health, and universal access to reproductive health. Key thematic areas for this hub include regional strategies for child survival; strengthening health systems for maternal and newborn health; adolescent reproductive health; and nutrition.
www.wchknowledgehub.com.au

A strategic partnerships initiative funded by the Australian Agency for International Development

The University Of Queensland AUSTRALIA

The University of Melbourne

UNSW THE UNIVERSITY OF NEW SOUTH WALES

HUMAN RESOURCES FOR HEALTH KNOWLEDGE HUB

The University of Melbourne

The Nossal Institute for Global Health

menzies school of health research

Burnet Institute
Medical Research. Practical Action.