



Bloomberg  
Philanthropies



DATA FOR  
HEALTH INITIATIVE

CRVS DEVELOPMENT SERIES

# Improving vital statistics for informed policy: The importance of data quality

March 2018



*Transforming data into information for action*



## Resources available from the University of Melbourne, Bloomberg Philanthropies Data for Health Initiative

### *CRVS development series*

The CRVS development series, generated through the Initiative, form a lasting archive of concise and easily accessible evidence and knowledge on strengthening CRVS systems. The content is based on a combination of technical knowledge and country experiences, as well as the scientific literature. The series is intended to stimulate debate and ideas for in-country CRVS policy, planning and capacity building, and promote the adoption of best practice to strengthen CRVS systems worldwide.

### *CRVS technical outcome series*

This series focuses on filling a range of scientific knowledge gaps and offering new tools, methods, findings and approaches for CRVS systems and data improvement. The series has a strong empirical focus. It reports on works in progress, particularly for large or complex technical initiatives, and on specific components of projects that may be of more immediate relevance to stakeholders.

### *CRVS resources and tools*

Capacity-building resources and tools are designed to influence and align CRVS processes with established international or best-practice standards and to help countries improve their systems. These resources, which are used extensively in the Initiative's training courses, aim to change practice and ensure countries benefit from such changes by developing critical CRVS capacity among technical officers and ministries.

### *CRVS country perspectives*

CRVS country perspectives describe the capacity-building experiences and successes of strengthening CRVS systems in partner countries, including fellowship reports. The series describes the state of CRVS systems improvement in partner countries and lessons learnt, and provides a baseline for comparison over time and between countries.

### *CRVS action guides and summaries*

Many papers from the development and technical outcome series have accompanying action guides or summaries, which provide a succinct overview of key points and, in the case of action guides, a suggested way forward for countries.

Published by the University of Melbourne, Civil Registration and Vital Statistics Improvement, Bloomberg Philanthropies Data for Health Initiative

Melbourne School of Population and Global Health  
Building 379  
207 Bouverie Street  
Carlton  
VIC 3053, Australia

+61 3 9035 6560

CRVS-info@unimelb.edu.au

[www.mspgh.unimelb.edu.au/dataforhealth](http://www.mspgh.unimelb.edu.au/dataforhealth)

**Made possible through funding from Bloomberg Philanthropies**  
[www.bloomberg.org](http://www.bloomberg.org)

---

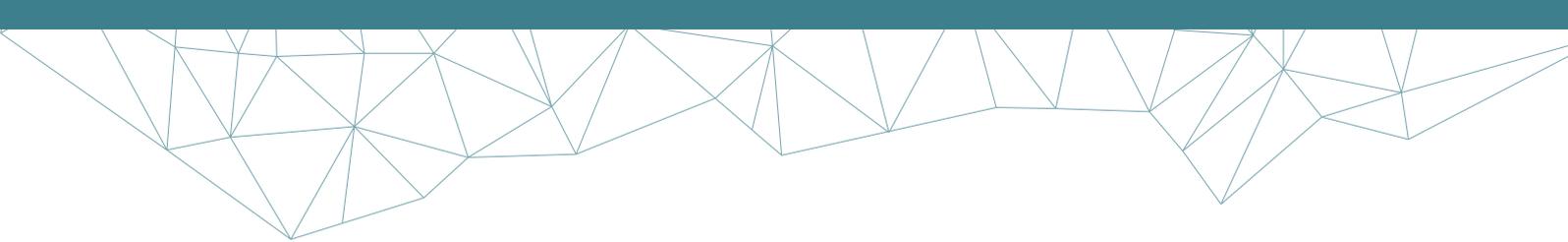
### Acknowledgments

Many people were involved in the drafting, reviewing and finalisation of this paper, including Nicola Richards, Renee Sorchik, Claire Brolan, Lene Mikkelsen, and Tim Adair, the University of Melbourne.

---

### Suggested citation

University of Melbourne. *Improving vital statistics for informed policy: The importance of data quality*. CRVS development series. Melbourne, Australia: University of Melbourne, Civil Registration and Vital Statistics Improvement, Bloomberg Philanthropies Data for Health Initiative; 2018.



# Contents

<b>Abbreviations</b> .....	4
<b>Key terms</b> .....	4
<b>Key points</b> .....	5
<b>The importance of data quality</b> .....	6
<b>Coverage and completeness: key aspects of data quality</b> .....	8
<b>Ensuring data quality during notification and registration</b> .....	8
Strategy 1: Strengthen capacity in local registration and health offices to check the micro and compiled data before they are transferred further .....	9
Strategy 2: Use information technology to expand capacity and reduce errors .....	10
<b>Data quality checks for aggregated birth and death data</b> .....	12
Data quality checks for tabulated birth data.....	12
Data quality checks for tabulated mortality data .....	14
<b>Timeliness</b> .....	16
<b>Summary</b> .....	18
<b>Related resources and products</b> .....	19
University of Melbourne, BD4H Initiative, CRVS Knowledge Gateway: Library .....	19
University of Melbourne, BD4H Initiative, CRVS Knowledge Gateway: Learning Centre .....	19
University of Melbourne, BD4H Initiative, CRVS Knowledge Gateway: Courses .....	19
Further reading .....	19

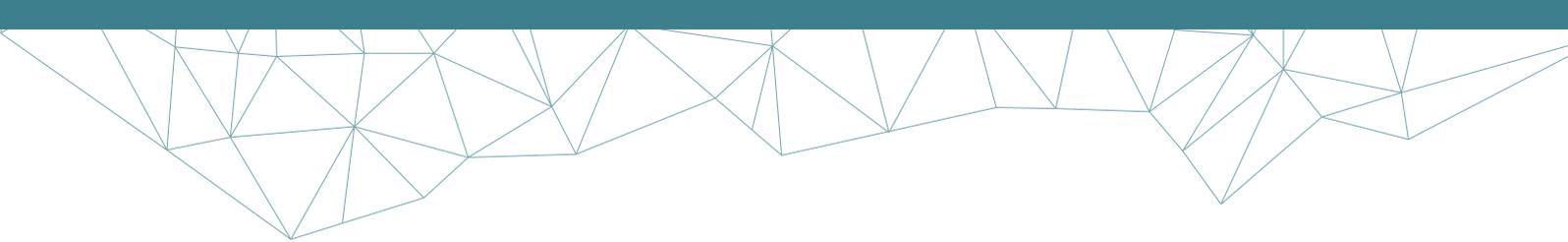


## Abbreviations

ANACONDA	Analysis of Causes of (National) Deaths for Action
ASFRs	age-specific fertility rates
BD4H	Bloomberg Philanthropies Data for Health Initiative
CDC	Centers for Disease Control and Prevention
COD	cause of death
CRVS	civil registration and vital statistics
DHS	Demographic and Health Survey
EMRO	WHO Regional Office for the Eastern Mediterranean
GBD	Global Burden of Disease study
HMN	Health Metrics Network
ICD-10	International Statistical Classification of Diseases and Related Health Problems, version 10
MICS	Multiple Indicator Cluster Surveys
SDG	Sustainable Development Goal
SOPs	standard operating procedures
TFR	total fertility rate
UN	United Nations
UNECA	United Nations Economic Commission for Africa
UNESCAP	United Nations Economic and Social Commission for Asia and the Pacific
WHO	World Health Organization

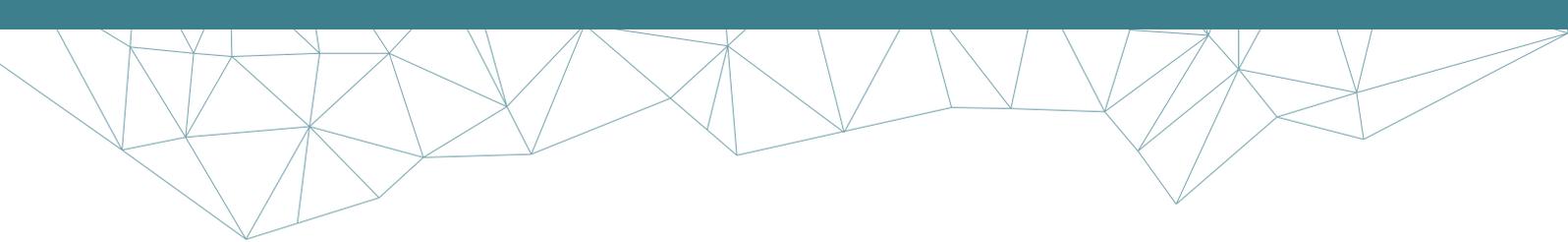
## Key terms

<b>Age-specific fertility rate:</b>	number of live births per 1000 women in a specific age group for a specific point in time, usually a calendar year
<b>Birth order:</b>	the numerical order of the live birth or foetal death being recorded in relation to all previous births of the mother
<b>Coverage:</b>	the fraction of the population with access to registration points
<b>Completeness:</b>	the percentage of actual births or deaths in a population that are registered
<b>Microdata:</b>	is the data contained in each individual record (such as name, date of birth, etc.)
<b>Standard operating procedures:</b>	detailed instructions compiled by an organisation to help workers carry out complex routine activities or tasks
<b>Tabulations:</b>	a way of arranging data in a systematic way for analysis, such as in a table
<b>Total fertility rate:</b>	the average number of children a woman would have assuming that current birth rates remain constant throughout her childbearing years



## Key points

- The registration of vital events and record compilation is not an end in itself – the quality and reliability of the data gathered must be evaluated if it is to become usable information and evidence for health policy and program action.
- Countries should aim to have the highest coverage possible for the registration of vital events, including all areas and all population groups, with sufficient registration points to service the entire population.
- Additionally, countries should aim to have complete registration of all births and deaths within the areas covered.
- Procedures to strengthen data quality should be implemented from the very beginning of the registration process. Civil registration staff should be trained and provided with Standard Operating Procedures.
- Introducing information technology and software for data entry can help prevent initial errors.
- Birth data should be assessed for the extent of its completeness, and, prior to their release, plausibility checks should be performed to ensure they are sound.
- Similarly estimations of completeness and plausibility checks for mortality data should be performed before their release. ANACONDA is a useful tool for mortality data to estimate the completeness of death registration as well as the quality of all-cause and cause-specific mortality indicators.
- For cause of death data, the percentage of deaths with ill-defined or unusable causes should be low for the data to be useful to guide health planning and policy.
- It is important that vital statistics are released in a timely fashion so that they can provide the greatest utility possible for policy makers.



# Improving vital statistics for informed policy: The importance of data quality

This paper describes the importance of data quality for vital statistics. It outlines the importance of coverage and completeness of birth and death registration, as well as simple procedures for checking data quality from the beginning of the registration process. Additional plausibility checks and software tools are highlighted for improved fertility, mortality, and cause of death statistics.

- 
- **The importance of data quality**
  - **Why coverage and completeness are key aspects of data quality**
  - **Strategies to improve the quality of data during notification and registration**
  - **Data quality checks for aggregated birth and death data**
  - **Timeliness**
  - **Summary**
- 

The quality of data collected must be evaluated before it is used in policy and planning.

## The importance of data quality

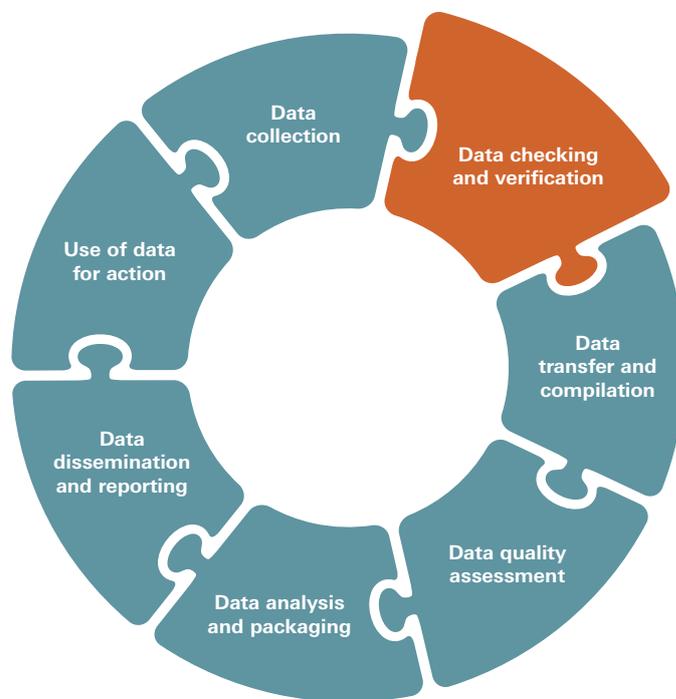
Civil registration and vital statistics (CRVS) system strengthening is gaining momentum worldwide, particularly as population-based data is needed to measure and evaluate progress towards achieving the 2030 Sustainable Development Goal (SDG) agenda<sup>1</sup>. Indicator 17.19.2, for example, measures the number of countries that, “have achieved 100 per cent birth registration and 80 per cent death registration.”<sup>2</sup> However, the improved notification and registration of vital events is not an end in itself – the quality and reliability of the data gathered must be evaluated if it is to become usable evidence for health policy and, most importantly, action.

The cyclical process of compiling and transforming raw data into knowledge for informed health and development decision-making is shown in Figure 1. This figure, originally developed by the Health Metrics Network (HMN), displays the key steps needed to prepare vital statistics for policy use, and also illustrates the different points where capacity building should occur to ensure that staff have the skills needed to perform their tasks. As shown in the figure, **data quality assessment is a critical first step before data can be analysed and released for wider use.**

<sup>1</sup> Brolan CE, Gouda H, AbouZahr C, et al. Beyond health: five global policy metaphors for civil registration and vital statistics. *The Lancet* 2017; 389; 1084-1085.

<sup>2</sup> <https://sustainabledevelopment.un.org/sdg17>

**Figure 1 The cycle of data collection, management, analysis, dissemination, and use**



Source: Health Metrics Network & World Health Organization (2008). *Framework and Standards for Country Health Information Systems* (2<sup>nd</sup> ed.). Geneva, World Health Organization

Several global and regional efforts to improve the quality of vital statistics are responsible for focusing attention on this topic, with the more prominent ones emanating from the HMN, World Health Organization (WHO), Health Information Systems Knowledge Hub (University of Queensland, Australia), Lancet, United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP), United Nations Economic Commission for Africa (UNECA), WHO Regional Office for the Eastern Mediterranean (EMRO), the World Bank, Centers for Disease Control and Prevention (CDC), Global Fund, and the Bloomberg Philanthropies Data for Health Initiative (BD4H).

Various data quality frameworks have been developed by such global and regional organisations and national statistical offices<sup>3</sup>. Common data quality concerns that are covered in most CRVS assessment frameworks, which will be discussed further below, include:

- if registration covers all the national territory and population groups (coverage)
- the extent to which all births and deaths that take place during the year are recorded (completeness)
- the presence of missing elements or characteristics, or only partially collected
- the percentage of deaths that do not have a useable cause of death (COD)
- the use of standard definitions and classifications
- timeliness and accessibility of the data.

<sup>3</sup> See, for example, the Data quality framework developed by the Australian Bureau of Statistics at <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+Data+Quality+Framework> (accessed 20 December 2017)

Coverage and completeness are often used interchangeably, however they measure different aspects of data quality and CRVS system performance.

## Coverage and completeness: key aspects of data quality

The quality of vital statistics is highly dependent on how well the data cover the entire national territory and population groups, and whether all vital events taking place are recorded and registered. The aim of any civil registration system should be to have universal **coverage**, ie enough functioning registration facilities available to the entire country and all population groups, and to have 100% **completeness**, ie register all births and deaths that occur.

Coverage and completeness are often used interchangeably in many countries. However, to know how well a national CRVS system is performing, it is useful to have information on both coverage and completeness since together they determine the overall reliability and utility of the vital statistics tabulations produced.

Furthermore, since the areas covered (usually urban) are likely to be different to those that are not covered (eg remote and/or rural), the population whose events are not covered (and not registered) is likely to experience different birth and death rates to the population whose events are registered. See Box 1 for how to measure coverage and completeness.

### Box 1: Measuring coverage and completeness

**Coverage can be measured by counting the population in the administrative areas that have access to local registration points as a fraction of the total population.**

$$\text{Coverage (\%)} = \frac{\text{Population in administrative areas served by the civil registry system}}{\text{Total population}} * 100$$

Conversely, lack of coverage can be measured by counting the population in those administrative areas that have no local registration facilities available to them as a fraction of the total population of the country.

**Completeness, on the other hand, measures the number of events registered in the population as a fraction of the total number of actual events (births or deaths) that occurred in that population.**

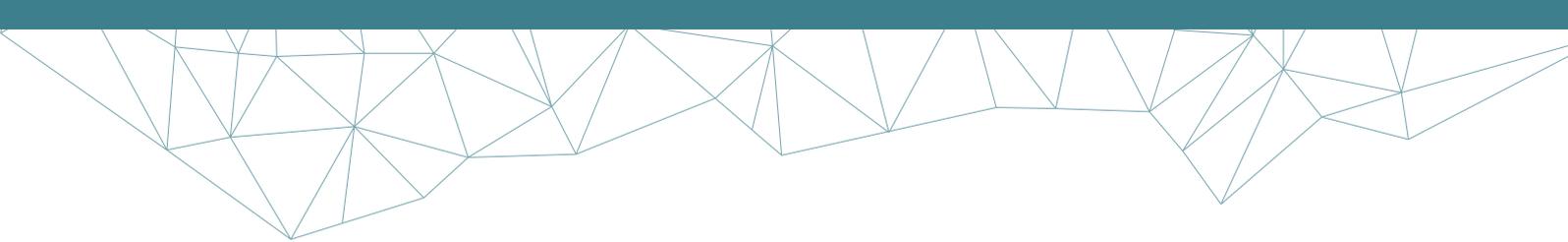
$$\text{Completeness (\%)} = \frac{\text{Number of registered events in the population}}{\text{Number of actual events in the population}} * 100$$

Countries should aim to have the highest coverage possible, and to have complete registration of all births and deaths within the population. Measuring completeness is discussed in more detail in the section on tabulation.

## Ensuring data quality during notification and registration

Even when CRVS system coverage and completeness are very high, the statistics produced from these systems are not always accurate and reliable. As shown in Figure 1, data quality checks should begin as soon as the data are collected – in the phase where civil registrars are registering the vital events and preparing the registration documents<sup>4</sup>. At this stage of the cycle, the focus is on **microdata** – that is, the accuracy and completeness of the information contained in each individual record.

<sup>4</sup> University of Melbourne. *CRVS systems need well-functioning civil-registry offices*. CRVS summaries. Melbourne, Australia: Civil Registration and Vital Statistics Improvement, University of Melbourne, Bloomberg Philanthropies Data for Health Initiative; 2017.



Procedures to strengthen data quality should be implemented from the beginning of the registration process.

There are many potential sources of errors in microdata that affect the quality of the vital statistics produced. Some of these can stem from poor form design, such as not having the date of birth of the mother (for births) or date of birth of the deceased as questions to complete. Other critical sources of error are:

- under-registration (not all births and deaths are registered)
- not reporting, or misreporting, the age at death or the age of the mother
- not recording the sex of the decedent or sex of a newborn baby
- not indicating the place of occurrence or usual residence
- not registering the event within the required time period following the event, and
- not indicating the cause of death, or using a non-standardised method to determine the cause of death (such as lay reporting), or poor certification practices.

Data quality checks at this phase are dependent on the civil registration staff carefully verifying the information received and entering it correctly into the registration papers or database. For this to happen in a consistent manner, it is necessary to have well thought out and rigorously implemented **standard operating procedures (SOPs)** for civil registration staff. Strategies to improve the quality of data at the time of collection (notification and registration) are listed below.

### **Strategy 1: Strengthen capacity in local registration and health offices to check the micro and compiled data before they are transferred further**

At the time of registration, simple checks should be implemented to verify the information entered on the record is correct and complete. SOPs should guide staff to check each data item for errors and omissions, and verify the accuracy of any supporting documentation (ie notification form, medical certificate of cause of death, or identity papers) and any other documentation requested for the registration to take place.

Besides checking that registration forms are complete, staff should also perform some **simple plausibility checks** on the individual records (Box 2). For example, staff may double-check:

- The age of a mother whose age is listed as younger than 15 or older than 49.
- The sex of a decedent whose COD does not seem plausible (ie men do not die from maternal causes, women do not die from prostate cancer, and adults generally do not die from causes associated with the neonatal period).



### Box 2: Plausibility checks on compiled data

Plausibility checks involve evaluating the data to determine if they are likely to be true (plausible) or if they are likely to be untrue (implausible). The final judgement about the plausibility of data can often only be arrived by using some comparator: that is, comparing recent data to previous years', or data from another source. The data are more plausible if they are similar to those from other years or sources.

For example, if the current year's vital statistics data show there were 120,000 deaths, and there were 118,000 deaths in the previous year, and the United Nations (or Global Burden of Disease or other relevant international source) estimated 122,000 deaths, then these results are similar enough to show the recent data are plausible.

With the increasing availability of computerised registration services, the possibility for basic electronic checks should be used when the data are compiled for transfer. However, even staff who are skilled at using computers to register events should receive **training in data cleaning and basic quality assessment** to check for potential errors in the data (Box 3)<sup>5</sup>. Such checks on the compiled data should be a standard part of their duties and should be performed routinely on all data to ensure that errors are corrected or fixed, and the data are correct and consistent before transferral and further use.

### Strategy 2: Use information technology to expand capacity and reduce errors

Customised software tools have been developed to assist data entry and compilation. These tools ensure all key data items are collected and are consistent. Sometimes these tools employ drop-down menus to assist with data entry, and they can also have built-in plausibility checks to help ensure the data being entered is as accurate as possible. For example, if the sex of the child is not entered, the computer may display an error message asking the user to confirm the sex of the child before continuing with the rest of the birth registration.

Additionally, employing information technology can improve the timeliness of data. The long delays often observed between the collection of the data and the production of periodic reports is often due to slow transmission of data to central offices for collation.

Introducing IT and software programs for data entry can prevent initial errors.

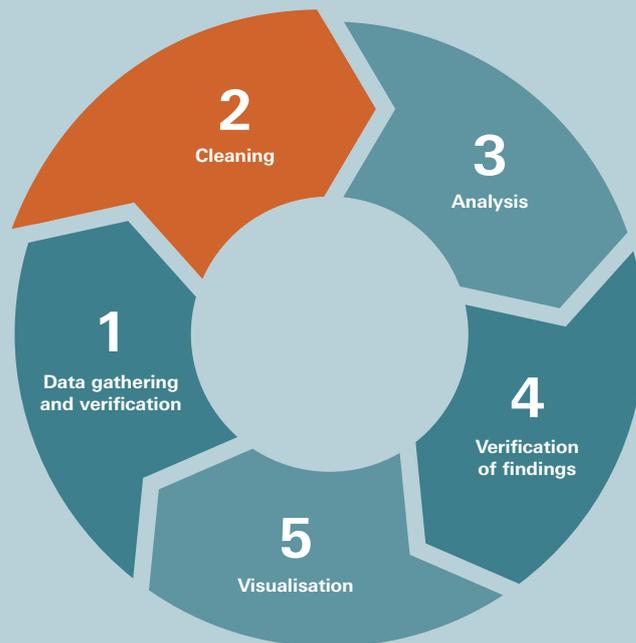
<sup>5</sup> See, for example, the Data Quality Review (DQR) Toolkit developed by the World Health Organization. Available at [http://www.who.int/healthinfo/tools\\_data\\_analysis/dqr\\_modules/en/](http://www.who.int/healthinfo/tools_data_analysis/dqr_modules/en/) (accessed 20 December 2017)

### Box 3: Data cleaning

Data cleaning involves a series of checks and actions that need to be applied to the raw data before it can be further analysed. These checks and actions should be specified in standard operating procedures and will vary between institutions and countries, but may include making sure:

1. there is only one record per line and one line per record
2. there is only one field per column
3. that each file has the variables or fields in the same order so they can be combined later
4. there are no duplicates in the data
5. date formats are consistent
6. as many fields are completed as possible.

Phases of an investigation with data



Source: adapted from Verification handbook for investigative reporting. Available at <http://verificationhandbook.com/book2/chapter5.php> [Accessed 20 December 2017]



## Data quality checks for aggregated birth and death data

Once birth and death events have been recorded, they need to be aggregated and produced as statistics to become useful for policy users<sup>6</sup>. However, once tabulated, data quality checks should be performed before the data is released. Different factors affecting the quality of tabulated birth and death data are discussed below.

## Data quality checks for tabulated birth data

Birth data should be assessed for the extent of its completeness.

**Coverage and completeness** are the two primary factors impacting the usefulness of birth data. Knowing the extent to which births are undercounted is important for informing all health programmes and social services. Similarly, when these techniques are used at the sub-national level, the results can inform which areas or population groups' registration drives should focus on, as the level of completeness is likely to be highest in urban areas and lowest in rural, remote, and marginalised areas with indigenous populations.

The completeness of birth registration is defined as follows:

$$\text{Completeness of birth registration (\%)} = \frac{\text{Number of registered births}}{\text{Actual number of births}} * 100$$

Knowing the percentage completeness of birth registration gives users an indication of the reliability of the data. Generally, a registration rate of 90 per cent or greater is considered complete enough for policy purposes.

### Box 4: Counting birth registrations

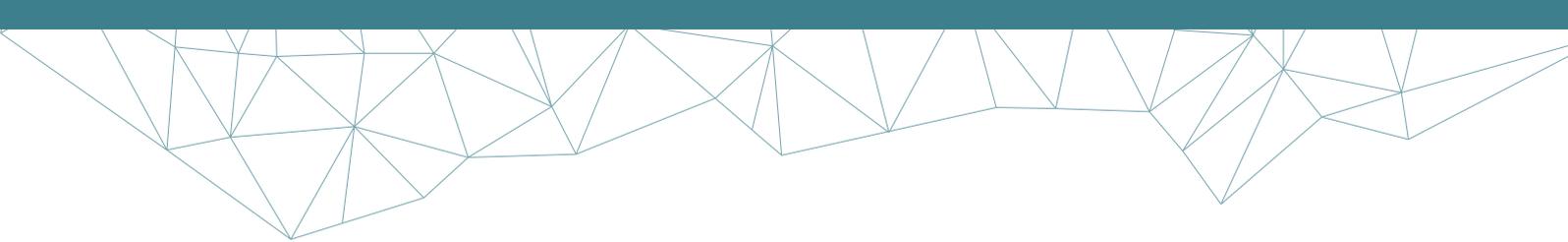
It is important to capture all births within 12 months of when they occur. Annual birth registration data are often a mix of children of different ages (both newborns and older children) who were registered that year. Sometimes the annual registration data may even include adult registrations. Some children may only be registered for the first time when they reach school age, due to the frequent requirement of a birth certificate for enrolment.

For reporting and analysis purposes, children whose birth did not occur in the year of registration should be separated from children who were born in the last year. The percentage of late registrations should be carefully monitored as these lower the utility of the data, and as the births of children who die before registration will be missed.

A high percentage of births with unspecified age of mother is an indicator of poor quality data.

Other errors or omissions in CRVS data such as the age of the mother and sex of the child impact the quality of fertility statistics. The mother's age is critical for calculating statistics related to fertility and teenage birth rates. Both the **total fertility rate (TFR)** and **age-specific fertility rates (ASFRs)** are key indicators in monitoring health. Adolescent birth rates are an important indicator to monitor the progress of SDG goal 3: 'Ensure healthy lives and promote well-being for all at all ages.' Recording the correct age of the mother at the time of registration is a critical component in calculating both total and age-specific fertility rates from vital statistics data. The percentage of births with unspecified age of mother should be very low; high percentages of records missing this information can be an indicator

<sup>6</sup> The United Nations has a list of recommended tabulations for civil registration data, available at <https://crvsgateway.info/learningcentre/deaths-cause-of-death-statistics/un-recommended-tabulations-for-births-and-deaths-from-civil-registration-data> (accessed 20 December 2017)



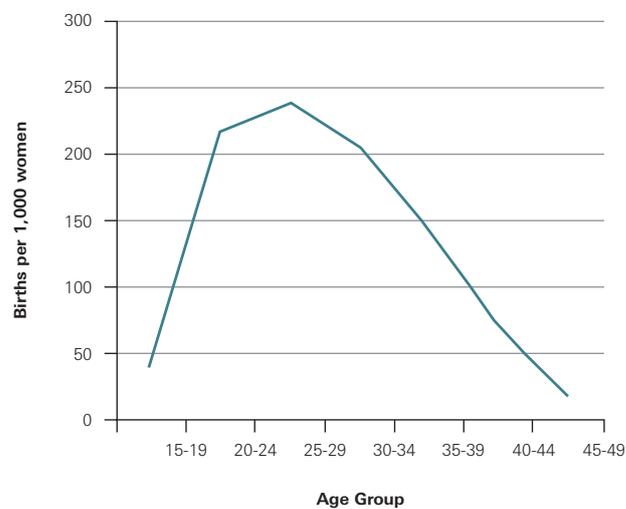
Plausibility checks are also about asking 'does this make sense'. If the calculated TFR is 10 per woman, this is the same as saying each woman would have 10 children: a strong indication that the data has errors in it.

of poor quality data. The sex of the child can also provide a measure to determine the quality of the data by monitoring the percent of records where this information is missing, and can also act as a signal of foetal sex selection and preference in countries where this takes place.

If needed, the total fertility rate (TFR) can also be used to check the level of completeness of birth registration. For instance, the TFR generated from the CRVS system can be compared to those derived from census and survey data. If the TFR derived from CRVS data is significantly lower than the TFR from other sources, this may indicate incomplete birth registration and signal poor data quality. Additionally, while the TFR tends to decline slowly over time, a large drop compared to previous years or other data sources should alert users that the data needs to be reviewed in more detail to assess data quality and completeness.

Age specific fertility rates (ASFRs) from registration data can be compared to those from other sources to further investigate likely under-registration in certain age groups or implausible patterns of fertility. A typical ASFR pattern in a country with high fertility is bell-shaped, and would look something like Figure 2 below. The peak in fertility is around age 20-24 or 25-29, and ages 20-29 have rates above 200 births per 1,000 women in countries with high fertility. If the ASFR pattern generated from CRVS data is widely different from the bell-shaped one shown in Figure 2 it may be an indication of missing births or of wrong age reporting of mothers.

**Figure 2 Example of plausible age-specific fertility rates for a high fertility country**



Other key pieces of information in the birth record such as birth order of the child, birth weight, and type of attendant at birth (and so on) can also be used to monitor data quality. The percentage of records where any of these characteristics are unspecified should be low; a large percentage of records with unspecified characteristics could signal poor data quality and should be further evaluated.



## Data quality checks for tabulated mortality data

As is the case with fertility, **coverage and completeness greatly impact the utility of mortality statistics**. In many countries, more than 50% of deaths occur outside of medical facilities, and many never get registered or identified by the health system<sup>7</sup>. Knowing how many deaths are missed and therefore the extent of incomplete registration is a primary concern for any civil registration systems as it seriously limits the utility of the data collected and statistics produced.

Unfortunately, completeness of death registration can be challenging to measure accurately.<sup>8</sup> Several indirect demographic and statistical techniques (for example the Generalised Growth Balance or Bennett-Horiuchi methods) can be utilised. However, these methods are best used by national statistical offices as they require the use of census data, and an understanding of the different techniques and the assumptions underlying their operation. More recently, an empirical method to estimate the completeness of death registration has been developed that is relatively accurate and can be readily used by national and subnational statistical offices<sup>9</sup>.

ANACONDA is a useful tool for assessing the quality of mortality data.

For countries to have a user-friendly tool to assess death registration completeness and the accuracy of mortality data, the Bloomberg Philanthropies Data for Health Initiative developed the Analysis of Causes of (National) Deaths for Action, or ANACONDA. The tool only requires the age and sex structure of the source population (denominator) and the International Statistical Classification of Diseases 10th Revision (ICD-10) codes by sex and five-year age groups (numerator). ANACONDA conducts all the tests and calculations needed for a comprehensive data quality review and automatically generates the associated figures and tables from which a report can be written, including estimated death registration completeness. It reviews both mortality as well as the cause of death patterns from the input data, and calculates standard mortality indicators<sup>10</sup>.

**The accuracy of age, sex, and COD are particularly important with regards to the quality of mortality statistics.** The percentage of records with unspecified age, sex, or COD should be very low. If a significant number of records are missing any of these characteristics, this is a sign of poor quality data and these omissions should be evaluated and corrected at the source.

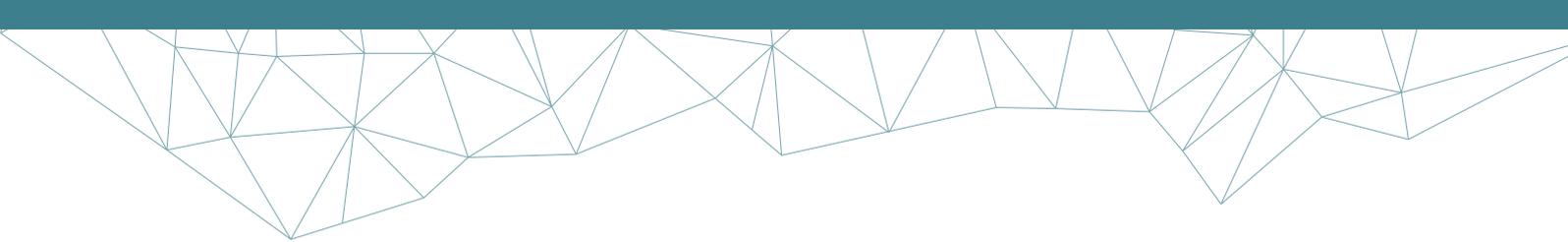
Age may sometimes be reported inaccurately in death records, especially when notification and registration forms are poorly designed. These errors are then reflected in age-related mortality patterns when the death records are aggregated. When this occurs, ‘age heaping’ is seen in certain age groups. It is well documented that family members or next of kin, when unsure about the exact age of the decedent, tend to report age digits ending at 0 or 5. For example, if a 68-year-old woman died, her family might report her as being “about 70” if they were not sure of her exact age or date of birth. These errors need to be detected as they will bias the data, and should be corrected applying demographic techniques for smoothing age heaping. See Figure 3 for an example of age heaping in the reported age at death from a country’s civil registration system.

<sup>7</sup> de Savigny D, Riley I, Chandramohan D, et al. Integrating community-based verbal autopsy into civil registration and vital statistics (CRVS): system-level considerations. *Global Health Action* 2017; 10: 1272882.

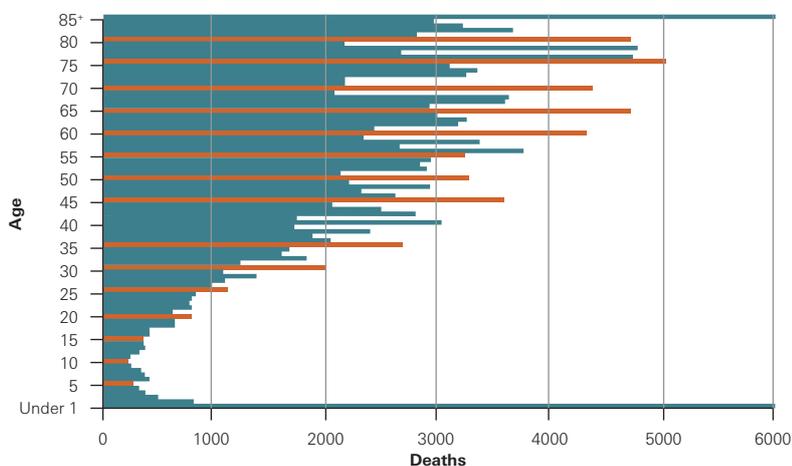
<sup>8</sup> University of Melbourne. *The importance of routinely measuring birth and death registration completeness*. CRVS summaries. Melbourne, Australia: University of Melbourne, Civil Registration and Vital Statistics Improvement, Bloomberg Philanthropies Data for Health Initiative; 2017.

<sup>9</sup> Adair T & Lopez AD. Estimating the completeness of death registration: an empirical method [submitted for publication].

<sup>10</sup> Mikkelsen L & Lopez AD. *Guidance for assessing and interpreting the quality of mortality data using ANACONDA*. CRVS resources and tools. Melbourne, Australia: University of Melbourne, Civil Registration and Vital Statistics Improvement, Bloomberg Philanthropies Data for Health Initiative; 2017.



**Figure 3 Example of age-heaping in civil registration data (red bars show deaths for each age that end in 0 or 5)**



Aggregating mortality data by age and sex may also reveal implausible sex distributions of deaths from what would be expected, which might signal under-registration of certain age groups, eg older women. The ANACONDA tool can help assess if such age under-reporting is present.

**Simple plausibility and consistency checks** should be carried out to assess how robust and credible aggregated data are before being released to the public. Mortality indicators derived from CRVS data such as the crude death rate, infant mortality rate, under five mortality rate, and age-specific mortality rates should be compared to other sources such as prior years' vital statistics data, data from the national census, or results from DHS, MICS, GBD or UN. Where possible, trend analysis should be performed and unexpected or sudden variations in rates compared with previous years or other sources should always be investigated, as major population indicators typically evolve slowly over time, unless there have been specific events which have caused sudden increases or reductions, or the population size is very small. Large changes in estimates compared to values from other sources may signal a problem with data quality.

Compiling the data in different ways and running a series of standard cross tabulations covering all the different variables is a useful way to reveal errors and inconsistencies. When possible, charts and maps can be used to display results and check for outlying values for a better overview of large datasets. Often the implausible values discovered are due to data compilation errors or misunderstood procedures and can be corrected accordingly.

COD is by far the most difficult characteristic to check at the individual record level. While some plausibility checks may be possible at the time of registration, it would not be possible for a lay person to validate the majority of CODs. The quality of COD statistics is best assessed at the aggregated level where cumulative errors or misuse of certain causes become evident. The ANACONDA tool can be used to help compare countries' data against international standards and benchmarks for this purpose.

**Ill-defined and poorly specified CODs** adversely affect data quality. Unfortunately, not all physicians have been trained in filling out death certificates correctly and completely, and many are not aware of its importance<sup>11</sup>. As a result, coders who receive the death certificate often assign ill-defined and poorly specified ICD-10 codes. These codes are not useful for

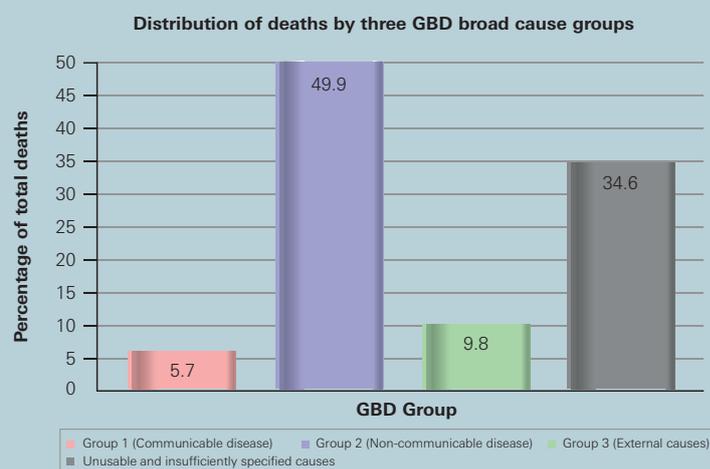
The percentage of deaths with ill-defined or unusable causes should be low for the data to be useful for policy.

<sup>11</sup> University of Melbourne. *Reducing barriers to the accurate medical certification of cause of death*. CRVS development series. Melbourne, Australia: University of Melbourne, Civil Registration and Vital Statistics Improvement, Bloomberg Philanthropies Data for Health Initiative; 2018.

public health purposes, and reduce the utility of a country's COD statistics. Ideally, less than 10% of deaths should be assigned these codes. The ANACONDA tool can help reveal the proportion of deaths assigned to such codes, as well as the most misused causes and what impact these have on death distributions, giving countries a starting point to improve COD data quality (Box 5).

#### Box 5: Example ANACONDA output

Step 6 in ANACONDA distributes all the causes of death in the input data into four broad categories: infectious diseases (Group I); noncommunicable diseases (Group II); external causes (ie accidents and injuries) (Group III); and unusable and insufficiently specified causes. The resulting graph can be used to assess how well the input data compares to the expected distribution of causes of death across the three groups; and also the extent of causes of death that cannot or should not be used as the underlying cause of death.



In addition to using ANACONDA, other basic plausibility checks include:

- producing a set of verification tables that consist of basic tabulations for the majority of variables in the database by province or territory of occurrence
- sending verification tables to each provincial/territorial registrar of regional statistical offices for their review to ensure that their registry obtained the same results
- checking for internal consistencies, for example, by running frequencies and looking for outliers on certain data elements
- comparing the most recent data year with past data years and other sources to detect any unusual or unexpected changes.

## Timeliness

Another common challenge in many countries is the timely release of vital statistics. Timeliness of data has several components including promptness of:

- registration of the vital event
- data transfer

It is important that vital statistics are released in a timely fashion if they are to be useful for policy and planning.

- processing, verification, generation, and dissemination of the vital statistics.

Timely registration of the vital event is largely dependent on the legally required timeframe – for example, if registration needs to happen within 24 hours or six months. Promptness of data transferral is mostly related to the time frames for when the local civil registrars are supposed to transmit the records to the central office and to the extent the system is computerised.

The final component of timeliness relates to how quickly the central office can process, verify, generate and disseminate the vital statistics. It is recommended that the basic counts of births and deaths are released as quickly as possible, and that a later dissemination is planned for those data characteristics which are more complex to clean, check, and tabulate.

Data limitations and delays should not keep data from being published, provided that the publishing agency is transparent about these delays and limitations. If reasonable data quality assessments have been undertaken but the data are incomplete, the data can still be useful if made available with appropriate explanations of its limitations – that is, that certain territories are not covered, or only major urban areas are covered, or data for some regions are estimated to be only 60% complete.

Similarly, if delays in publication occur because a number of external causes of death cannot be finalised within the usual period (perhaps because of lengthy legal procedures), it is preferable to publish the data in a timely way, and revise them once the legal outcome is known. If the revision is of the data is significant, an explanation needs to be given so as not to bias the annual data.

In the example below (Figure 4) from the Australian Bureau of Statistics, they have included a clause which reads: *Causes of death data for 2016 are preliminary and subject to a revision process*. This allows the Bureau to publish cause of death data in a timely way, while also being able to go back and update the data once more data is received and quality checks have taken place.

**Figure 4 Example vital statistics tabulation using preliminary cause of death data**

**2.4 Top 10 leading causes of death<sup>(a)</sup>, NSW, Qld, SA, WA and NT<sup>(b)</sup> - by Aboriginal and Torres Strait Islander status - 2016<sup>(c)(i)</sup>**

Cause of Death and ICD-10 code	Aboriginal and Torres Strait Islander		Non-Indigenous		Rate ratio <sup>(e)</sup>	Rate difference <sup>(f)</sup>
	no.	rate	no.	rate		
<b>All Causes</b>	<b>2919</b>	<b>921.9</b>	<b>108 925</b>	<b>567.0</b>	<b>1.6</b>	<b>354.9</b>
Ischaemic heart diseases (I20-I25)	367	113.0	13,127	67.8	1.7	45.2
Diabetes (E10-E14)	228	81.2	3,187	16.4	4.9	64.8
Chronic lower respiratory disease (J40-J47)	200	75.8	5,519	28.3	2.7	47.6
Malignant neoplasm of trachea, bronchus and lung (C33-C34)	184	62.6	5,836	29.6	2.1	33.1
Intentional self-harm [suicide] (X60-X84)(g)	162	23.6	1,906	11.4	2.1	12.3
Cerebrovascular diseases (I60-I69)	97	41.5	7,378	38.2	1.1	3.3
Cirrhosis and other diseases of the liver (K70-K76)	93	22.0	1,217	6.4	3.5	15.7
Land transport accidents (V01-V89)(h)	80	13.7	910	5.4	2.5	8.3
Accidental po...						
Dementia, inc...						
(F01, F03, G30)						

(c) Causes of death data for 2016 are preliminary and subject to a revisions process. See Explanatory Notes 55-58 and A More Timely Annual Collection: Changes to ABS Processes (Technical Note in this publication).

(a) Causes list...  
 Explanator...  
 (b) Data are re...  
 Capital...  
 id...  
 id... see Explanatory Notes 59-70.  
 (c) Causes of death data for 2016 are preliminary and subject to a revisions process. See Explanatory Notes 55-58 and A More Timely Annual Collection: Changes to ABS Processes (Technical Note in this publication).

Source: Australian Bureau of Statistics (2017). Australia's leading causes of death, 2016. Available at <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/3303.0-2016-Main%20Features-Australia's%20leading%20causes%20of%20death,%202016-3> [Accessed 20 December 2017]



## Summary

Vital statistics must be of high quality to have utility for health and development policy use. Poor quality data can be misleading and lead to bad decisions, especially when used for policy, guiding public health programs, or for monitoring the indicators needed for Sustainable Development Goal achievement. The quality of vital statistics is closely related to how well the data cover the entire national territory and whether all vital events taking place are recorded and registered. The aim of any civil registration system should therefore be to have universal coverage; to have functioning registration facilities available to the entire population that register every vital event. Furthermore, prompt registration and data transfer are key for increasing the timeliness, relevance, and utility of the data.

Some simple and effective capacity building strategies can be applied that can lead to significant improvement in the accuracy of data at both the micro and aggregated levels. Standard Operating Procedures need to be available, understood and routinely applied to ensure data are checked at the point of collection and compilation. The introduction of information technology is an effective way of reducing errors in both the micro and compiled records, and can lead to significant gains in improving timeliness of data transfer and dissemination.

All countries should integrate tools into their annual vital statistics production to help eliminate errors and reduce data inconsistencies. Completeness of birth registration should be calculated and monitored by geographic region and over time. Plausibility checks should be performed on fertility data to ensure it is of good quality before its release. Analysis programs to check plausibility, such as ANACONDA, can be used to routinely assess the completeness and quality of vital statistics, and provide an overview of what can be done to improve data quality.

Finally, it is important that vital statistics are released in a timely fashion to ensure they are useful and relevant for policy and planning. One practical solution countries can implement to balance the need for high-quality data while ensuring timeliness, is to release provisional data that can be updated once quality assurance checks are completed.

## Related resources and products

### University of Melbourne, BD4H Initiative, CRVS Knowledge Gateway: Library <https://crvsgateway.info/library>

*ANACONDA 10 steps: Quick reference guide.* CRVS summaries.

*Estimating the completeness of death registration.* CRVS summaries.

*Guidance for assessing and interpreting the quality of mortality data using ANACONDA.* CRVS Resources and tools.

*The importance of routinely measuring birth and death registration completeness.* CRVS summaries.

### University of Melbourne, BD4H Initiative, CRVS Knowledge Gateway: Learning Centre <https://crvsgateway.info/learningcentre>

Topic 5: Improving the quality and presentation of civil registration and vital statistics data.

Topic 6: CRVS tools –ANACONDA mortality data quality assessment tool.

### University of Melbourne, BD4H Initiative, CRVS Knowledge Gateway: Courses <https://crvsgateway.info/courses>

Enterprise architecture/business process mapping for countries.

Estimating the completeness of registration.

## Further reading

AbouZahr C, Mikkelsen L, Rampatige R, Lopez AD. *Mortality statistics: a tool to improve understanding and quality.* Working Paper No. 13. Queensland, Australia: University of Queensland, School of Population Health, Health Information Systems Knowledge Hub; 2010. Available at <https://crvsgateway.info/library> (accessed 22 February 2018).

Moultire T, Dorrington R, Hill A, Hill K, Timaeus I, and Zaba B (eds.). The relational Gompertz model, in '*Tools for demographic estimation*'. Paris, France: International Union for the Scientific Study of Population; 2013. Available at [http://demographicestimation.iussp.org/sites/demographicestimation.iussp.org/files/TDE\\_2013\\_2ndImpression\\_0.pdf](http://demographicestimation.iussp.org/sites/demographicestimation.iussp.org/files/TDE_2013_2ndImpression_0.pdf) (accessed 20 December 2017).

International Monetary Fund. *Data Quality Assessment Framework.* Washington DC, USA: International Monetary Fund; 2012. Available at <http://dsbb.imf.org/pages/dqrs/dqaf.aspx> (accessed 20 December 2017).

United Nations Statistics Division. Principles and Guidelines for Managing Statistical Confidentiality and Microdata Access. New York, USA: UNSD; 2007. Available at <https://unstats.un.org/unsd/statcom/doc07/BG-Microdata-E.pdf> (accessed 20 December 2017).

United Nations Statistics Division. *Principles and Recommendations for a Vital Statistics System*, rev. 3. New York, USA: UNSD; 2014. Available at <http://unstats.un.org/unsd/Demographic/standmeth/principles/M19Rev3en.pdf> (accessed 20 December 2017).

The program partners on this initiative include: The University of Melbourne, Australia; CDC Foundation, USA; Vital Strategies, USA; Johns Hopkins Bloomberg School of Public Health, USA; World Health Organization, Switzerland.

Civil Registration and Vital Statistics partners:



## For more information contact:

CRVS-info@unimelb.edu.au  
crvsgateway.info

CRICOS Provider Code: 00116K

Version: 1117-02

### Copyright

© Copyright University of Melbourne March 2018.

The University of Melbourne owns the copyright in this publication, and no part of it may be reproduced without their permission.

### Disclaimer

The University of Melbourne has used its best endeavours to ensure that the material contained in this publication was correct at the time of printing. The University gives no warranty and accepts no responsibility for the accuracy or completeness of information and the University reserves the right to make changes without notice at any time in its absolute discretion.

### Intellectual property

For further information refer to: [unimelb.edu.au/governance/statutes](http://unimelb.edu.au/governance/statutes)