

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

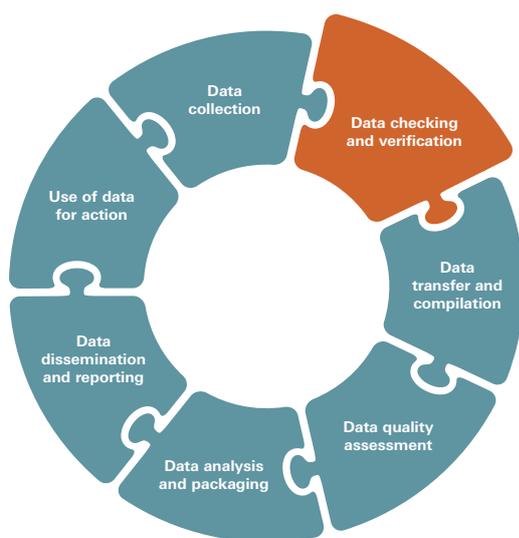
This *CRVS summary* is edited from 'Improving vital statistics for informed policy: The importance of data quality', a *CRVS development series* paper available at <https://crvsgateway.info/library>

Why does data quality matter?

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) agenda¹. However, the improved notification and registration of vital events is not an end in itself – the quality and reliability of the data must be evaluated if it is to become usable evidence for health policy and, most importantly, action.

Figure 1 shows 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.**

Figure 1 The cycle of data collection, management, analysis, dissemination, and use²



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

² Adapted from: Health Metrics Network & World Health Organization (2008). *Framework and Standards for Country Health Information Systems* (2nd ed.). Geneva, World Health Organization.

What affects the quality of vital statistics?

A number of different data quality frameworks have been developed by global and regional organisations, and national statistical offices. Common data quality concerns 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 partially collected or missing elements or characteristics (ie sex, age, residence etc.)
- 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.

Coverage and completeness

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 occur³.

How can quality be assured during data collection?

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. At this stage of the cycle, the focus is on **microdata** – that is, the accuracy and completeness of the information

³ University of Melbourne (2017). The importance of routinely measuring birth and death registration completeness. *CRVS summaries*. Bloomberg Philanthropies Data for Health Initiative: Melbourne.

contained in each individual record.

Data quality checks at this phase depend 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 **standard operating procedures** (SOPs) for civil registration staff.

Besides checking that registration forms are complete, staff should also perform some **simple plausibility checks** on the individual records (Box 1). 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).

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 2)⁴. 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 transfer and further use.

Box 1: Plausibility checks

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 estimated 122,000 deaths, then these results are similar enough to show the recent data are plausible.

⁴ 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/

Box 2: 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 SOPs 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.

What are some quality checks for aggregated fertility and mortality statistics?

Fertility

Coverage and completeness are the two primary factors impacting the usefulness of fertility statistics. Knowing the percentage completeness of birth registration gives users an indication of the reliability of the data. Examining completeness at the sub-national level, can also show any population groups or areas that are being undercounted.

Checking for errors or omissions in the data. 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. Other key pieces of information in the birth record such as birth order of the child, birth weight, and type of attendant at birth 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.

If needed, the total fertility rate (TFR) can 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. Age specific fertility rates (ASFRs) from registration data can also be compared to those from other sources to further investigate

likely under-registration in certain age groups or implausible patterns of fertility.

Mortality

As is the case with births, **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 system⁵. 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.

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.

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 should be compared to other sources. Where possible, trend analysis should be performed and unexpected or sudden variations in rates should be investigated, as major population indicators typically evolve slowly over time. Large changes in estimates compared to values from other sources may signal a problem with data quality.

Ill-defined and poorly specified CODs adversely affect data quality. These causes are not useful for public health purposes, and reduce the utility of a country's COD statistics. However, 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.

Use of electronic tools. Many of the data quality checks described above can be performed by the Analysis of Causes of (National) Deaths for Action (ANACONDA) tool. 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. It reviews both mortality as well

as the cause of death patterns from the input data, and calculates standard mortality indicators⁶.

How does timeliness affect data quality?

Timeliness of data has several components including promptness of:

- registration of the vital event
- data transfer
- 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.

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

⁵ de Savigny D, Riley I, Chandramohan D et al (2017). Integrating community-based verbal autopsy into civil registration and vital statistics (CRVS): system-level considerations. *Global Health Action* 10:1272882.

⁶ Mikkelsen L & Lopez AD (2017). Guidance for assessing and interpreting the quality of mortality data using ANACONDA. CRVS Resources and tools. University of Melbourne, Bloomberg Philanthropies Data for Health Initiative: Melbourne.

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, or guiding public health programs.

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 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.

For more information contact:

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

CRICOS Provider Code: 00116K

Version: 0218-01