CRVS best-practice and advocacy
Summary: Redefining 'garbage codes' for public health policy

This CRVS summary is edited from 'Redefining 'garbage codes' for public health policy: Report on the expert group meeting, 27-28 February 2017', a CRVS technical outcome series paper available at crvsgateway.info/Library~23

The challenge of collecting accurate cause of death data

Issues and challenges with mortality statistics

All countries need accurate and up-to-date **mortality statistics** for a variety of purposes, including:

- Informing health and social policy debates
- Monitoring trends in diseases and injuries
- Evaluating policies designed to improve health outcomes
- Monitoring progress relative to national, regional and global development goals.

However, in many countries, the systems that produce mortality and cause of death (COD) data either **do not exist or are poorly developed**. As a result, the statistics they produce are often not reliable enough to be used for the purposes listed above. Common challenges across countries for collecting reliable COD data include:

- Incompleteness
- Tabulated cause lists and aggregated codes
- Variability in data format
- Hiding HIV/AIDS and other possible stigmatising diseases or injuries
- Mis-assignment of certain causes of death
- Various types of 'garbage codes'.

Garbage codes

System deficiencies can result in a high proportion of causes of deaths assigned to **garbage codes**,[1] which are codes that have no use in informing public health policy, as the related underlying cause of death (UCOD) is too vague or simply impossible. These include:

- Codes that do not identify underlying causes, eg 'heart failure'
- Impossible causes for specific age or sex groups, based on global medical and biological knowledge and epidemiological patterns.

Garbage codes bias the true pattern of mortality in a country, as it is unlikely they would be equally or proportionally distributed across the disease categories used in analysing COD data. Hence, the data will not represent the true health status of the population.

Global Burden of Disease (GBD) Study

In the 1990 GBD Study, Murray and Lopez were the first to attempt to identify and resolve the extent and pattern of garbage codes in mortality data.[2]

Subsequent iterations of the concept of garbage codes have been much more detailed, complex and encyclopaedic. For the 2010 GBD Study, Naghavi and colleagues developed a public health classification of garbage codes that would allow comparability across ICD revisions.

Other frameworks for classifying garbage codes also exist, such as the World Health Organization's shortlist,[3] and International Statistical Classification of Diseases, 10th Revision (ICD-10).[4]

1 Naghavi M, Makela S, Foreman K, et al. Algorithms for enhancing public health utility of national causes-of-death data. Population Health Metrics 2010; 8:9.

2 Murray CJL, Lopez AD (eds.). The Global Burden of Disease and Injury 1: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. London, UK: Harvard University Press on behalf of the World Health Organization and World Bank; 1996.

3 Mathers C, Stevens G, Ma Fat D, et al. *WHO methods and data sources for country-level causes of death 2000–2012.* Global Health Estimates Technical Paper WHO/HIS/HIS/GHE/2014.7. Geneva, Switzerland:WHO; 2014.

4 World Health Organization. Online *ICD-10 Version: 2016*. Available at http://apps.who.int/classifications/icd10/browse/2016/en#/XVIII

Strategies to reduce garbage codes and improve data quality

Identifying data problems

The best source of mortality and COD statistics for a population is a civil registration and vital statistics (CRVS) system, which registers and assigns a medically certified UCOD for all deaths. CRVS systems in many developing countries, however, achieve inadequate levels of coverage and completeness, resulting in millions of births and deaths going unrecorded each year.[5,6]

To improve CRVS systems, it is important to understand any existing problems with the data, particularly their completeness[7] and diagnostic accuracy.[8] A common concern with CRVS systems is their reliability in describing the actual mortality patterns in the population to which they refer.

Redefining garbage

About 30 per cent of COD data reported by countries around the world are 'garbage'.[9] The rationale for identifying garbage codes is that certifying physicians and coders should avoid any ICD code that is too vague to guide major national and global disease and injury control strategies or goals.

Medical certification requires resources to collect and code data, but determining the UCOD for community deaths – without an attending physician – is a challenge. However, many hospital deaths are also coded to intermediate causes, a form of garbage, due to reasons like poor training on certification or limited diagnostic equipment.

It must also be recognised that, at a broader level, the information required for some levels of disease prevention does not need the UCOD to be precise. For example, broad categories of diseases (eg communicable or noncommunicable) are sufficient to design and implement primary prevention strategies like health education or immunisation campaigns.[10]

ANACONDA

The Bloomberg Philanthropies Data for Health (D4H) Initiative is working to improve the quality of COD information from hospitals, and apply verbal autopsy for community deaths. As part of D4H's efforts to produce high-quality datasets and improve data analysis skills, an electronic tool for Analysis of Causes of National Deaths for Action – or ANACONDA – was developed by the universities of Melbourne and Basel. ANACONDA allows users to analyse the quality of their mortality data to better understand if the data are fit for their intended purpose.[11]

In addition to an overall analysis of the input mortality data, ANACONDA provides a detailed framework for assessing the plausibility and quality of COD data. ANACONDA has now been piloted in more than 50 countries since its 2015 launch.

Moving forward

An expert meeting

In February 2017, the University of Melbourne, as part of the D4H Initiative, convened a meeting with the aim of developing an alternative classification of garbage codes that better aligns with public health interventions and priorities. It was agreed that the classification should be able to identify, for each country dataset, which unusable codes with the most severe impact on disease and injury control strategies are being used frequently.

The meeting attendees wanted to address, for example, if it was sufficient for public health purposes to accept the use of 'unspecified pneumonia' as an UCOD. In this case, it would no longer be regarded as a garbage code (as in the GBD), as it provides sufficient information on the COD to guide future health interventions. Alternatively, it may be necessary to identify the main disease agents causing the pneumonia. In this case, it would remain classified as 'garbage'.

5 Setel PW, Macfarlance SB, Szreter S, et al. A scandal of invisibility: Making everyone count by counting everyone. *Lancet* 2007; 370(9598):1569–1577.

6 United Nations Children's Fund. *Every child's birth right: Inequities and trends in birth registration*. New York, USA: UNICEF; 2013.

7 University of Melbourne. *A new method for estimating the completeness of death registration*. CRVS summaries. Melbourne, Australia: Bloomberg Philanthropies Data for Health Initiative, and Civil Registration and Vital Statistics Improvement, University of Melbourne; 2018.

8 Worster A, Haines T. Advanced statistics: Understanding medical record review (MMR) studies. *Academic Emergency Medicine* 2004; 11(2).

9 Naghavi M, Makela S, Foreman K, et al. Algorithms for enhancing public health utility of national causes-of-death data. *Population Health Metrics* 2010; 8:9.

10 Last JM, Spasoff RA, Harris SS (eds). ***A dictionary of epidemiology***. Fourth edition. New York, USA: Oxford University Press; 2014.

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

Four levels of unusable codes

The meeting attendees agreed that any public health-oriented classification of garbage codes should be realistic about countries' diagnostic capacity at different levels. Such an approach must remember, however, that policymakers need precise mortality statistics to design effective policies and programs.

Attendees agreed that an additional classification of unusable codes should be added to ANACONDA. This is based on the concept of **'severity of the garbage'**, or the extent of bias the unusable codes would introduce to the overall COD distribution.

The new classification defines four levels of unusable or poorly specified codes that should be avoided. The classification depends on how serious their impact is for misguiding public policy. These four levels are:

1. **Level 1 (very high): codes with serious implications.** These are causes for which the true UCOD could in fact belong to more than one broad cause group. These are the most serious of the unusable codes, since they could potentially bias the true pattern of mortality in a population.
2. **Level 2 (high): codes with substantial implications.** These are causes for which the true COD is likely to belong to only one of the three broad groups. These unusable causes are less serious than Level 1 since they do not alter the understanding of the broad composition of causes of death in the population. They do, however, affect knowledge on leading causes.
3. **Level 3 (medium): codes with important implications.** These are causes for which the true underlying COD is likely to be one within the same ICD chapter. For instance, 'unspecified cancer' still provides enough information to know the COD was cancer. However, knowledge about the site of cancer is important for public health policy because different strategies are applied for different types (sites) of cancer.
4. **Level 4 (low): codes with limited implications**. These are diagnoses for which the true COD is likely to be confined to a single disease or injury category. The implications of unusable causes classified at this level will therefore generally be much less important for public policy. In this four-level classification, they are not included under the broad category of unusable causes, but defined as 'poorly specified'.

Using the new classification

The addition of this new classification of unusable codes to ANACONDA has allowed countries to see the comparative importance of these four levels of unusable codes. It is then up to the country to decide the level it is interested in investigating further.

For most countries, the important levels to work on are Levels 1–3, which contain most of the 'harmful garbage'. Level 4 is likely to contain those codes that demand considerable diagnostic sophistication and equipment to precisely determine COD, and might not be possible to resolve in all countries and circumstances. Furthermore, not all ill-defined codes can be eliminated. For some deaths, particularly in older age groups where comorbidities are common, a physician may be unable to determine the precise UCOD.

Because of different age structures, cultures and socioeconomic development, countries will show different patterns of unusable codes. As such, it is important to identify the most commonly used types of unusable codes for each level. Doing so will facilitate the development of targeted interventions aimed at improving medical certification and decreasing the use of unusable codes.

A hierarchical process to identify the actual ICD codes most commonly used within each level of unusable codes was developed. These are based on grouping similar unusable codes into 'packages' at each of the four levels:

- The packages at each level are ranked in order of importance so that users can immediately see within each level what practices are causing the most garbage codes.
- Within each of these packages, ANACONDA then offers the possibility to rank the top 10 ICD-10 codes that are causing the most unusable codes within a specific package. It is this detailed information that is likely to be most useful in guiding improvement strategies.

Using the information provided by ANACONDA (ie unusable causes ranked by severity level, package, and relevant ICD-10 code), countries are able to develop interventions aimed at improving data quality, such as additional physician training about the principles of medical certification and the UCOD.

Summary

Countries need accurate cause of death data to develop public health policy and practice, but the presence of 'garbage', or unusable, codes to classify causes of death can reduce data usefulness. ANACONDA is a data quality assessment tool that checks for common errors in mortality data, and provides a framework for identifying the type of garbage codes that reduce the utility of the data.

The University of Melbourne, as part of the D4H Initiative, has developed a new four-level classification of garbage codes that includes the most frequently used unusable and poorly specified codes, and this classification has been integrated into ANACONDA. The new classification gives users the option to identify the most frequently misused codes, and reflects public health disease patterns and intervention strategies in developing countries.

This will be helpful to guide elimination efforts, as these codes are the ones that have the largest effect on quality and produce the biggest bias in the data for public health purposes. Using this new classification, ANACONDA offers countries the opportunity to design focused strategies to improve the quality of cause of death data according to country needs and resources.

Bloomberg Philanthropies DATA FOR HEALTH INITIATIVE

Australian Government
Department of Foreign Affairs and Trade

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:

Vital Strategies

CDC Foundation
Together our impact is greater

THE UNIVERSITY OF MELBOURNE

For more information contact:

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

CRICOS Provider Code: 00116K

Version: 0718-01