THE UNIVERSITY OF MELBOURNE

Bloomberg Philanthropies

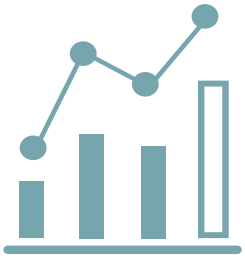
DATA FOR HEALTH INITIATIVE

CRVS TECHNICAL OUTCOME SERIES

Redefining 'garbage codes' for public health policy:

Report on the expert group meeting, 27–28 February 2017

May 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

Made possible through funding from Bloomberg Philanthropies
www.bloomberg.org

Acknowledgments

This paper is based on a technical meeting held at the Melbourne School of Population and Global Health, University of Melbourne, on 27–28 February 2018. The authors acknowledge all those who attended the meeting as part of the Bloomberg Philanthropies Data for Health Initiative: Saman Gamage, Deirdre Mclaughlin, Tim Moore, Rasika Rampatige and Ian Riley, University of Melbourne; Mohsen Naghavi, Institute of Health Metrics and Evaluation; Elisabeth Franca, University of Minas Gerais, Brazil; Pamela Groenewald, South Africa Medical Research Council; Ardeshir Khosravi, Ministry of Health, Iran; James Eynstone-Hinkins and Lauren Moran, Australian Bureau of Statistics; Jomilynn Rebenal, Ministry of Health, the Philippines; and Michael Hegnauer, Swiss Tropical and Public Health Institute, University of Basel.

Authors

Lene Mikkelsen, Nicola Richards, and Alan Lopez, University of Melbourne.

Suggested citation

Mikkelsen L, Richards N, Lopez AD. *Redefining 'garbage codes' for public health policy: Report on the expert group meeting, 27¬–28 February 2017.* CRVS technical outcome series. Melbourne, Australia: Bloomberg Philanthropies Data for Health Initiative, and Civil Registration and Vital Statistics Improvement, University of Melbourne; 2018.

Contents

Abbreviations

AIDS	acquired immunodeficiency syndrome
ANACONDA	Analysis of Causes of (National) Deaths for Action
BD4H Initiative	Bloomberg Philanthropies Data for Health Initiative
COD	cause of death
CRVS	civil registration and vital statistics
GBD Study	Global Burden of Disease Study
HIV	human immunodeficiency virus
ICD	International Statistical Classification of Diseases and Related Health Problems
ICD-10	International Statistical Classification of Diseases and Related Health Problems, 10th Revision
UCOD	underlying cause of death

Key terms

ANACONDA:	An electronic tool that assesses the accuracy and completeness of mortality and COD data. It checks for potential errors and inconsistencies in the data and provides users with an understanding of basic epidemiological and demographic concepts to interpret their data.
Cause of death:	Refers to 'all those diseases, morbid conditions or injuries which either resulted in or contributed to death and the circumstance of the accident or violence which produced any such injuries' (Twentieth World Health Assembly, 1967).
Underlying cause of death:	'The disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury' (World Health Organization, 1994).
Unusable code:	(also referred to as 'garbage code'): Any code that cannot or should not be an underlying cause of death, such as septicaemia, senility or headache; a cause that belongs in some other part of the morbid sequence of events leading to death; or a cause of death that is insufficiently specified.

Redefining 'garbage codes' for public health policy

Executive summary

Since the introduction of The Global Burden of Disease (GBD) Study a quarter of a century ago, there has been increasing interest in the use, and public policy implications, of a defined set of vague, incorrect diagnoses in national cause of death statistics, collectively known as 'garbage codes'. While the specification of the likely universe of garbage codes as used in the GBD may be technically correct, and appropriate for the cause of death estimation goals of the GBD, it is likely to prove too exacting for informing and guiding medical certification improvements in countries, one of the key goals of the Bloomberg Philanthropies Data for Health (D4H) Initiative.

To better support efforts in countries to improve medical certification practices, an action-oriented categorisation of garbage codes was developed, which includes the most frequently used garbage codes, and which also reflects the public health reality of disease patterns and disease control intervention strategies in developing countries. This new classification of such 'unusable' codes has been integrated into ANACONDA, a data quality assessment tool that checks for common errors in mortality data and provides a framework for identifying the kind and type of unusable codes that reduce the utility of mortality statistics.

The approach of this classification is on 'harm minimisation', by focussing on the unusable codes that cause most harm for guiding policy. In practical terms the new classification, with an option for users to identify the most frequently misused codes, will be extremely helpful to guide elimination efforts, as these codes are the ones having the greatest impact on quality and produce the biggest bias in the data for public health purposes. Using this new classification, ANACONDA offers countries the possibility to design focused strategies for improving the quality of cause of death data, according to their needs and resources.

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.** According to the 2010 Global Burden of Disease (GBD) Study, sensitivity tests showed that countries with very low levels of completeness (below 70 per cent) had biased, unrepresentative COD data.[1]
- **Tabulated cause lists and aggregated codes.** At the global level, comparing levels and patterns in mortality among countries is difficult, because countries use different versions of the International Statistical Classification of Diseases (ICD). This problem is compounded when countries use tabulated cause lists and aggregated codes.
- **Variability in data format.** It is common for countries to use different:
  - age-group aggregations
  - early, late or postneonatal age groups
  - age termination points.
- **Hiding HIV/AIDS and other possible stigmatising diseases or injuries.** These causes are usually allocated to other, correct, diseases such as meningitis, opportunistic diseases, skin cancer, tuberculosis, or inflammatory bowel disease; or given a vague or non-specific cause.

1 Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. ***Lancet 2012***; 380:2095–2128.

- **Mis-assignment of certain causes of death.** For example, classifying deaths due to overdoses as 'unintentional poisoning' among adults.
- **Various types of 'garbage codes'**, such as
  - **those that do not identify underlying causes**, such as 'heart failure'
  - **impossible causes for specific age or sex groups**, based on global medical and biological knowledge, and epidemiological patterns.

Garbage codes

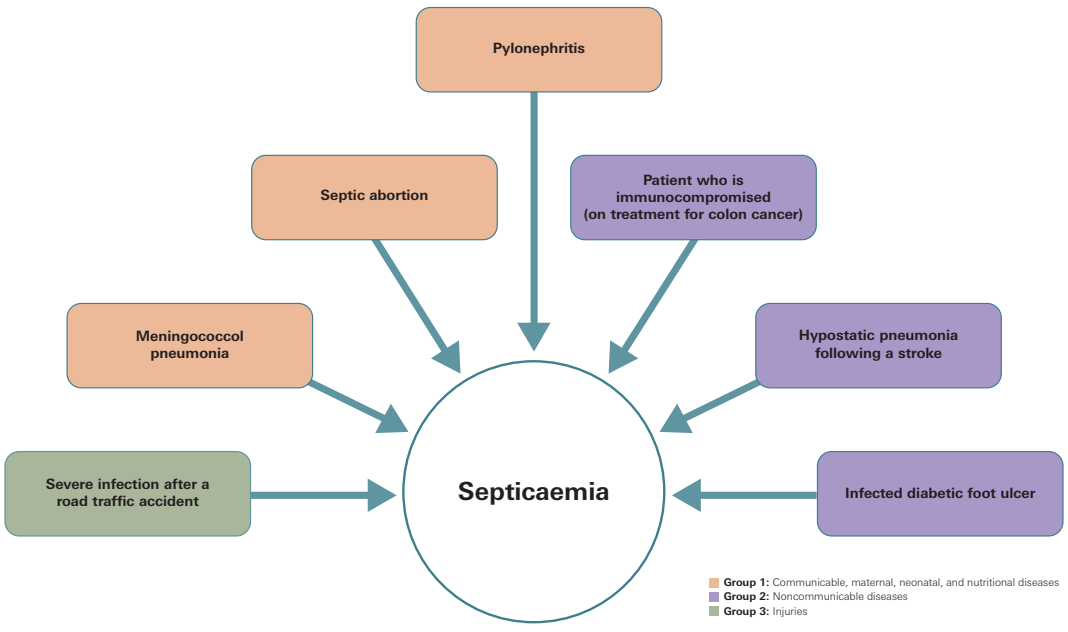
System deficiencies usually result in a high proportion of causes of deaths assigned to garbage codes.[2] Essentially, a garbage code is one that has no use in informing public health policy, as the related underlying cause of death (UCOD) is too vague, or simply impossible.

Nosologists[3] have identified causes in the ICD-10 that either cannot or should not be used as an UCOD (eg 'old age'), or that contain no information about the probable UCOD (eg 'ill-defined causes'). The ICD-10 also contains a number of codes that – although broadly useful for public health purposes (such as 'ill-defined sites of cancer') – are not specific enough to guide public policy dialogue.

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. A study of vital statistics data from the Islamic Republic of Iran,[4] for example, found that the true COD pattern of the population was considerably different from the pattern of mortality reported by the civil registration and vital statistics (CRVS) system. Many deaths originally coded to 'senility' or 'unknown' were found to be injury deaths, thus greatly underestimating the importance of external causes of death for the country. Similarly in Thailand, many deaths coded to 'septicaemia' were found to be due to cerebrovascular disease, HIV/AIDS and pneumonia.[5]

The bias introduced into the COD distribution will be more serious if the type of garbage code given is one where even the broad category of the underlying cause cannot be correctly determined. For example, when 'septicaemia' is written on a medical certificate of death, the death may be the result of a communicable disease, noncommunicable disease, or accident and injury (**Figure 1**).

Figure 1: Possible underlying causes of death leading to septicaemia



2 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.

3 Nosology is a branch of medicine that deals with the classification of diseases.

4 Khosravi A, Rao C, Naghavi M, et al. Impact of misclassification on measures of cardiovascular disease mortality in the Islamic Republic of Iran: a cross-sectional study. *Bulletin of the World Health Organization* 2008; 86(9):688–96.

5 Pattaraarchachai J, Rao C, Polprasert W, et al. Cause-specific mortality patterns among hospital deaths in Thailand: validating routine death certification. *Population Health Metrics* 2010; 8(1):12. 10.1186/1478-7954-8-12.

Global Burden of Disease 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.[6]

Subsequent iterations of the concept of garbage codes have been much more detailed, complex and encyclopedic. For the 2010 GBD Study, Naghavi and colleagues developed a public health classification of garbage codes that would allow comparability across ICD revisions (**Table1**). The classification contained four types:

- **Type 1 –** Causes that cannot or should not be considered as UCODs. Included are
  - essentially all the codes used to describe health services
  - 'essential primary hypertension' (which should be considered a risk factor)
  - 'atherosclerosis'
  - causes described as long-term sequelae (eg paraplegia, or complications of pregnancy and childbirth)
- **Type 2 –** Intermediate CODs, such as
  - heart failure
  - septicaemia
  - peritonitis
  - osteomyelitis
  - pulmonary embolism
- **Type 3 –** Immediate CODs, such as cardiac arrest and respiratory failure
- **Type 4 –** Insufficiently specified causes within ICD chapters. Although these may not be important for assessing aggregate causes for a category, the lack of the site of the cancer or the factor that caused the injury, for example, make them useless for public health preventions.

Subsequent iterations of the specification of garbage codes are based on the 2010 classification and categorisation.

Table1: Garbage codes by type and ICD-10 code[7]

GC Type	ICD-10 Codes
Type 1	A31.1, A59, A60.0, A71-A74, A63.0, B00.0, B08.1, B08.8, B30, B35-B36, F32-F33.9, F40-F42.9, F45-F48.9, F51-F53.9, F60-F98.9, G43-G45.9, G47-G52.9, G54-G54.9, G56-G58.9, H00-H04.9, H05.2-H69.9, H71-H80.9, H83-H93, J30, J33, J34.2, J35, K00-K11.9, K14, L04-L08.9, L20-L25.9, L28-L87.9, L90-L92, L94, L98.0-L98.3, L98.5-L98.9, M03, M07, M09-M12, M14-M25, M35.3, M40, M43.6-M43.9, M45.9, M47-M60, M63-M71, M73-M79, M95-M99, N39.3, N40, N46, N60, N84-N93, N97, Q10-Q18, Q36, Q38.1, Q54, Q65-Q74, Q82-Q84, R00-R99, B94.8, B949.9, G80-G83, Y86, Y87.2, Y89, I10, I15, I70
Type 2	A40-A41, A48.0, A48.3, E85.3-E85.9, E86-E87, G91.1, G91.3-G91.8, G92, G93.1-G93.6, I26, I27.1, 144-I45, I49-I50, I74, I81, J69, J80-J81, J86, J90, J93, J93.8-J93.9, J94, J98.1-J98.3, K65-K66, K71-K72 (except K71.7), K75, K76.0-K76.4, K92.0-K92.2, M86, N14, N17
Type 3	D65, 145-146, J96
Type 4	C80, C26, C39, C57.9, C64.9, C76, D00-D13, D16-D18,D20-D24, D28-D48, A49.9, B83.9, B99, E88.9 I51, I99, X59,Y10-Y34

Note: These garbage codes are based on the public health analysis cause list of 56 causes.

6 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.

7 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.

Other frameworks for classifying garbage codes

World Health Organization

The World Health Organization (WHO) uses the following shortlist of garbage codes, but does not explain the criteria used to derive them:

- Symptoms, signs and ill-defined conditions (ICD-10 codes R00-R99)
- Injuries undetermined whether intentional or unintentional (Y10-Y34 and Y87.2)
- Ill-defined cancers (C76, C80 and C97)
- Ill-defined cardiovascular diseases (I47.2, I49.0, I46, I50, I51.4, I51.5, I51.6, I51.9 and I70.9).

They also use an expanded list that includes:

- A40–A41 Streptococcal and other septicaemia
- D65 Disseminated intravascular coagulation
- E86 Volume depletion
- I10 Essential (primary) hypertension
- I269 Pulmonary embolism without mention of acute cor pulmonale
- I99 Other and unspecified disorders of circulatory system
- J81 Pulmonary oedema
- J96 Respiratory failure, nec [not elsewhere classified]
- K72 Hepatic failure, nec
- N17 Acute renal failure
- N18 Chronic renal failure
- P285 Respiratory failure of newborn.

WHO calculates the proportion of garbage codes received from all countries that supply COD data to the organisation. This is used, together with estimated completeness of the data, to classify countries according to the quality of their mortality data.[8]

International Statistical Classification of Diseases, 10th Revision

The ICD-10 provides excellent guidance on what the UCOD is and why it is important to use the concept. The ICD manuals also provide a detailed description of the rules and procedures for selecting the UCOD, and the immediate and intermediate causes. Limited guidance is given on garbage codes, apart from 'Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified', which the ICD says should not be used as an UCOD.[9]

The ICD-10 also suggests a list of alternative codes that can be used in cases where the medical certificate ends with certain codes, but where other conditions stated on the certificate would be more appropriate.

Strategies to reduce garbage codes and improve data quality

Identifying data problems

A functioning civil registration and vital statistics (CRVS) system that registers and assigns a medically certified UCOD for all deaths is the optimal source of mortality and COD statistics for a population. However, the literature shows that CRVS systems in many developing countries are struggling to achieve adequate levels of coverage and completeness, resulting in millions of births and deaths going unrecorded each year.[10,11]

As a first step to improving these systems, it is critical to gain a detailed understanding of problems with the data, particularly regarding completeness[12] and diagnostic accuracy.[13] A common concern with any mortality statistics produced from CRVS systems is how reliable they are in describing the actual mortality patterns in the population to which they refer. For example, even if a dataset includes all deaths in hospitals, it is important to remember that they are different from deaths that occur in the community. Hence, the data cannot be considered to represent the national mortality situation.[14]

8 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: Department of Health Statistics and Information Systems, World Health Organization; 2014.

9 WHO. Online ICD-10 Version: 2016. Available at http://apps.who.int/classifications/icd10/browse/2016/en#/XVIII.

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

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

12 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.

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

14 Murray CJL. Towards good practice for health statistics: lessons from the Millennium Development Goal health indicators. *Lancet* 2007; 369:862–873.

Redefining garbage

Globally, approximately 30 per cent of COD data reported by countries are 'garbage'. The use of garbage codes is likely to decline as CRVS systems are strengthened. However, countries with ageing populations, such as Japan, have experienced increases in garbage coding over time. This is due to multi-morbidities in older people and the resulting difficulties in defining the single underlying cause that led to death.

The rationale behind identifying garbage codes is that certifying physicians and coders should avoid any ICD code that is unlikely to be specific enough to guide major national and global disease and injury control strategies or goals. However, recent experience with countries in defining and categorising garbage codes has led to concern that the concept of garbage codes is overly 'academic'. Countries are potentially dissuaded from improving data quality simply because the size of the task seems daunting. Furthermore, and particularly in low to middle-income countries, the resources and capacity available in health facilities decreases as one moves down the hierarchy of government services, and from urban to rural areas. This can limit the available data.

Medical certification depends on good clinical practice and sufficient resources to collect and code the data. At the most basic level, determining the precise UCOD for a death that occurred in the community without an attending physician is a significant challenge. However, many hospital deaths are also coded to intermediate causes (such as sepsis or heart failure), a form of garbage, due to a number of reasons, including:

- Poor training on certification for medical students, interns and practising physicians
- Limited diagnostic equipment.[15] For example, 'unspecified stroke' is a garbage code according to the GBD. But a computed tomography scan or magnetic resonance image is needed to determine the difference between a haemorrhagic and an ischaemic stroke. These technologies are simply not available in all hospitals in many countries.

It is also important to recognise that, at a broader level, the information required for some levels of disease prevention does not need the UCOD to be precise. Tertiary prevention, for example, requires a precise understanding of disease aetiology to ensure correct late-stage management (eg knowing if it was an ischaemic or haemorrhagic stroke).

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

At a primary prevention level, broad categories of disease (eg communicable or noncommunicable) are sufficient to design and implement, for example, health education or immunisation campaigns.[16]

ANACONDA

The Bloomberg Philanthropies Data for Health (D4H) Initiative is working to improve the quality of mortality statistics from hospitals and apply verbal autopsy to better understand probable COD for community deaths. This will produce high-quality datasets and improve data analysis skills for policy and program analysis.

As part of the D4H Initiative, the Melbourne School of Population and Global Health, University of Melbourne, in collaboration with the Swiss Tropical and Public Health Institute, University of Basel, developed ANACONDA (Analysis of Causes of National Deaths for Action). ANACONDA is an easy-to-use electronic tool for checking the quality of mortality data, designed to help users analyse the quality of datasets to better understand if the data are fit for their intended purpose.[17] It contains a series of steps that apply the GBD definition of garbage codes, and uses these to assess the quality of mortality data produced from routine CRVS systems, including at the sub-national level.

ANACONDA is a second-generation data quality assessment tool, which built on and expanded the 10 data quality assessment principles first published by the Health Information Systems Knowledge Hub at the University of Queensland.[18] The WHO turned these principles into an Excel tool named ANACoD (analysing mortality levels and causes of death).[19]

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. This analysis is important because public health interventions target specific diseases or injuries, as do health policy and even some health services. Thus, they need accurate and timely data on mortality due to those diseases. For example, the public health strategies to control lung cancer are very different from those developed to control cervical cancer.

ANACONDA uses the term 'unusable codes' instead of the more pejorative, but widely accepted, term 'garbage codes'.

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

17 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.

18 AbouZahr C, Mikkelsen L, Rampatige R, et al. Mortality statistics: A tool to improve understanding and quality. Working Paper 13. Brisbane, Australia: HIS Knowledge Hub, University of Queensland; 2010.

19 Available at www.who.int/healthinfo/anacod/en/

The classification of unusable codes in ANACONDA is based on the methods used for the 2015 GBD Study, which are an updated classification from the 2010 GBD. The only difference is that Type 1 codes (as developed by Naghavi et al, see **Table 1**) are broken down into two groups, with the 'R' codes being separately identified. Consequently, ANACONDA allows users to distinguish five categories of unusable codes, namely:

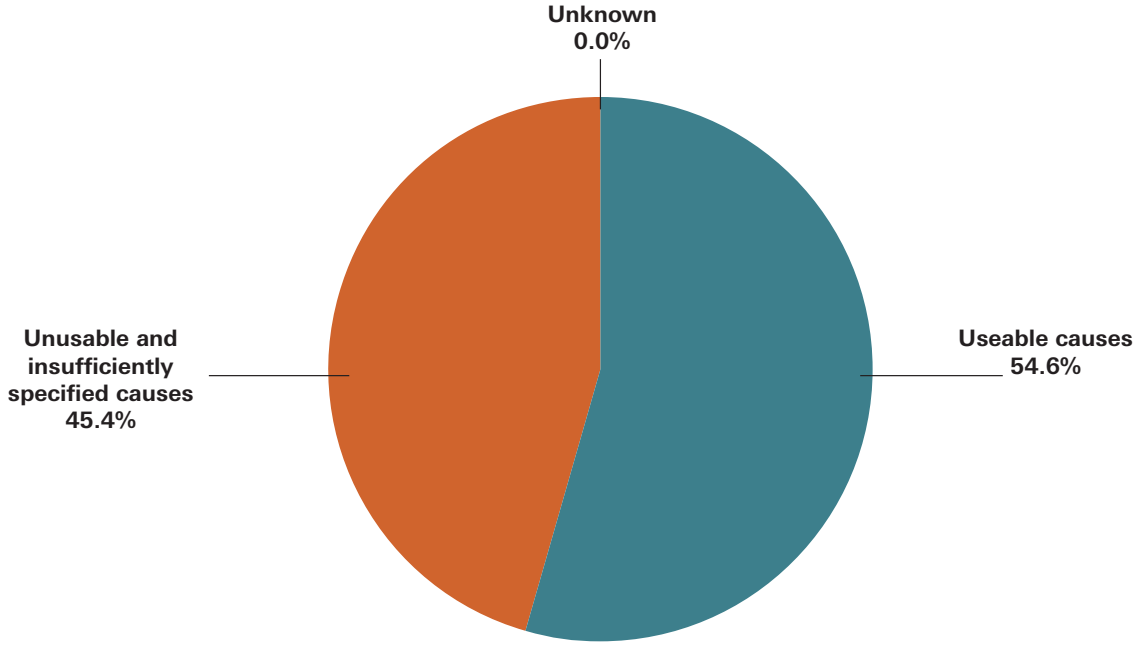
- **Category 1 –** Symptoms, signs and ill-defined conditions
- **Category 2 –** Impossible as underlying causes of death
- **Category 3 –** Intermediate causes of death
- **Category 4 –** Immediate causes of death
- **Category 5 –** Insufficiently specified causes within ICD chapters.

Collectively, ANACONDA refers to these as 'total unusable and insufficiently specified causes'. Categories 1–4 are considered to be completely unusable codes. Category 5 has some value in understanding what the person died from and for aggregating to larger COD groups – for example, cancer or heart diseases. These categories are weighted differently when calculating summary indices of overall data quality in ANACONDA.[13]

Categories 3 and 4 are clearly defined codes of clinical use, but they are of little value for public health prevention because they do not provide information on the injury or morbid condition that initiated the sequence leading to death. In ANACONDA, they are classified as totally unusable diagnoses rather than merely being insufficiently specified.

To illustrate how garbage codes can limit the utility of data to policy and seriously mislead users about the real health problems of the population, all the unusable and insufficiently specified ANACONDA codes found in a sample dataset have been extracted and shown in **Figure 2.** The figure shows that any policy based on the dataset used to generate this graph will have made use of only 55 per cent of the available data. With such a high fraction of unusable codes, the whole dataset is of questionable value and may skew policy conclusions that are based on it.

Figure 2: Example ANACONDA output – distribution of deaths by usability



ANACONDA was first piloted in 2015 as part of the BD4H Initiative and, as of 2018, it has been applied in more than 50 countries in national and inter-regional workshops. Some countries, such as Brazil, China, Philippines, Peru and Colombia, have used it for checking their sub-national COD data. Many countries have also gone on to offer training in ANACONDA at lower administrative divisions and levels.

Moving forward

An expert meeting

Given the points raised in the previous sections and experiences from applying ANACONDA in-country, 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 is better aligned with public health interventions and priorities.** The classification should also 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'.

Four levels of unusable codes

The meeting attendees agreed that any public health–orientated classification of garbage codes should be realistic about countries' diagnostic capacity at different levels. However, such an approach must remember that policy-makers need precise mortality statistics to design and implement effective public health policies and programs.

After much discussion, it was agreed that an additional classification of unusable codes should be added to ANACONDA. This classification is based on the concept of 'severity of the garbage' or the extent of bias the particular unusable codes would introduce to the overall COD distribution. For example, unusable codes such as heart failure and septicaemia, where the true underlying cause could be in any of the three major disease groups, were considered the most severe or harmful. Less harmful unusable codes would be those causes that are likely to have an impact on just one of the three broad groups, or which can be confidently reassigned within an ICD chapter. Finally, the least harmful are those that have a usable UCOD, but could have been better specified.

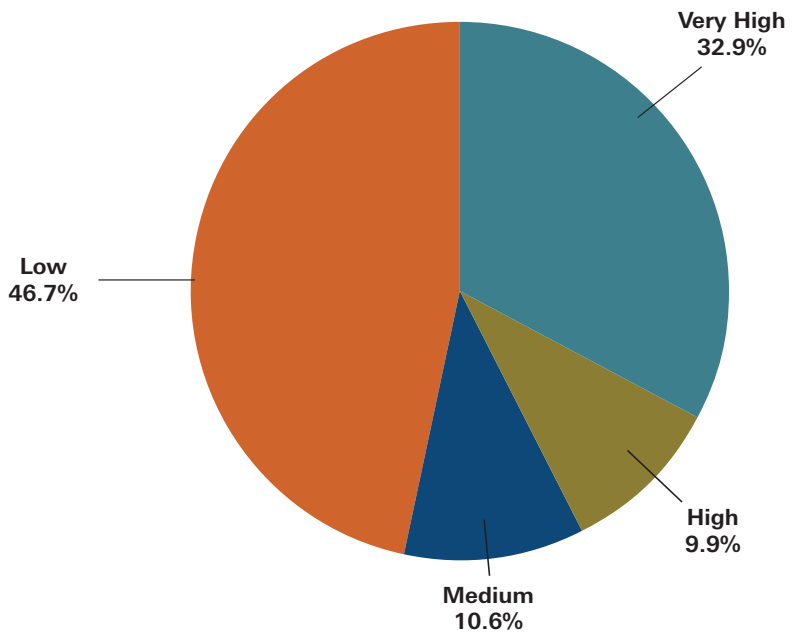
This new classification defines four levels of ICD codes that should be avoided. The classification depends on how serious their impact is for misguiding public policy. These four levels are:

- **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 (ie users of the mortality statistics cannot establish whether the true cause was a communicable disease, a noncommunicable disease or as a result of an injury) (see **Figure 1**). These are the most serious of the unusable codes, since they could potentially bias the true pattern of CODs in the population.
- **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 CODs in the population. They do, however, affect knowledge on leading CODs.
- **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 UCOD 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 (eg breast versus lung cancer).
- **Level 4 (low) – codes with limited implications.** These are diagnoses for which the true UCOD is likely to be confined to a single disease or injury category (eg unspecified stroke would still be assigned as a stroke death, and not to some other disease category). The implications of unusable causes classified at this level will therefore generally be much less important for public policy. Hence, in this four-level classification, they are not included under the broad category of unusable causes.

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 (**Figure 3**). 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' that is critical to reduce. 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, it is important to keep in mind that not all ill-defined codes can be eliminated. For some deaths, particularly in the older age groups where comorbidities are common, a physician may not be able to determine the precise UCOD.

Figure 3: Example ANACONDA output – distribution of unusable causes by severity



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. This is a significant improvement to the practical utility of ANACONDA. It will allow 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 unusable 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.

As demonstrated in **Figure 4**, within Level 1 (codes with serious implications) the 'shock and cardiac arrest' package ranked second. Further investigation reveals that more than 11,000 deaths in the dataset were coded to 'cardiac arrest' – an immediate COD, with no public policy use. With this information, 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.

Figure 4: Example ANACONDA output – unusable causes by severity level, package and relevant ICD-10 code

Leading packages - severity Very High						
Rank	**Package name**		**Rank**	**ICD code**	**Name of category**	**Total causes**
1	All, ill-defined		1	I46.-	Cardiac arrest	11,171
2	Shock & Cardiac arrest		2	I46.-	Shock, not elsewhere classified	503
3	Left HF		3	I95.-	Hypotension	47
4	Sepsis		4	R55.-	Syncope and collapse	3
5	Senility					
6	Renal failure					
7	Pneumonitis					
8	Other and unspecified disorders of fluid, electrolyte and acid-base balance					
9	Anaemia others					
10	Respiratory failure acute					
11	Dehydration					
12	Hydrocephalus					
13	Cerebral palsy					
14	Convulsions					
15	Pulmonary embolism					
16	Hepatic failure					
17	Peritonitis					
18	Cachexia					
19	Pneumothorax					
20	Psychotic disorder					
21	CNS abscess					
22	Accidental poisoning with: nonopioid analgesics, antipyretics and antirheumatics, other drugs acting on the automatic nervous system, other and unspecified					

Summary

Accurate cause of death data are important to allow countries to develop public health policy and practice. But the use of 'garbage', or unusable, codes to classify causes of death can threaten the usefulness of the data. A garbage code is one that has no use in informing public health policy, as the related underlying cause of death (UCOD) is too vague, or simply impossible. Garbage codes are any code that:

- Cannot or should not represent an UCOD, such as septicaemia, senility or headache.
- Represents a symptom or condition that belongs in some other part of the sequence of events leading to death.
- Insufficiently specifies a cause of death.

Garbage codes can significantly distort the true pattern of mortality in a country, which, in turn, can affect policy and practice based on those data.

Several frameworks have been developed for classifying garbage codes, to help countries to reduce their use, most notably that of the Global Burden of Disease Study. These approaches may be technically correct, but can be too cumbersome or exacting to inform and guide medical certification improvements in countries.

ANACONDA, a data quality assessment tool developed by the universities of Melbourne and Basel, checks for common errors in mortality data. It also provides a framework for identifying the type of garbage codes that reduce the utility of the data.

The Bloomberg Philanthropies Data for Health (D4H) Initiative has recently revisited ANACONDA, with the aim of developing an alternative classification of garbage codes that is better aligned with public health interventions and priorities. The BD4H Initiative has developed a new, four-level classification of garbage codes that includes the most frequently used garbage codes, and reflects the public health reality of disease patterns and disease control intervention strategies in developing countries. This new classification of such 'unusable' codes has been integrated into ANACONDA.

The new classification uses a harm minimisation approach. It focuses on the most frequently used unusable codes that cause the most harm. For example, unusable codes that potentially affect the distribution of deaths into their three broad groups (communicable diseases, noncommunicable diseases, and external causes and injuries) are the most serious (classified as Levels 1–3). This is because they have the greatest implications for policy.

The least harmful are likely to be codes that are confined to a specific disease (classified as Level 4).

To help countries eliminate garbage codes, ANACONDA now allows the identification of the leading ICD-10 codes that are misused within the most important 'packages' of garbage codes listed in Levels 1–3. ICD-10 codes listed in Level 4, although relevant for some countries, are given lower priority in national COD data improvement strategies.

In practical terms, the new classification gives users the option to identify the most frequently misused codes. 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 their needs and resources.

Related resources and products

University of Melbourne, D4H Initiative, CRVS Knowledge Gateway: Library
https://crvsgateway.info/library

Action guide on reducing barriers to medical certification. CRVS action guide.

ANACONDA 10 steps: Quick reference guide. CRVS summaries.

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

Improving vital statistics for informing policy: The importance of data quality. CRVS development series.

Reducing barriers to the accurate medical certification of cause of death. CRVS development series.

University of Melbourne, D4H Initiative, CRVS Knowledge Gateway: Learning Centre
https://crvsgateway.info/learningcentre

Topic 4: Cause of death in CRVS.

Topic 6: CRVS tools – ANACONDA mortality data quality assessment tool; ICD training tools; Medical certificate of cause of death assessment tool.

University of Melbourne, D4H Initiative, CRVS Knowledge Gateway: Courses
https://crvsgateway.info/courses

Analysis of Causes of (National) Deaths for Action (ANACONDA).

ICD-10 coding.

Medical certification of cause of death.

Further reading

Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380:2095–2128.

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.

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.

World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, 10th revision,* vol. 2, 10th edn. Geneva, Switzerland: WHO; 2016.

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

The University of Melbourne recognises the Swiss Tropical and Public Health Institute for their partnership and contribution

Swiss TPH

For more information contact:

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

CRICOS Provider Code: 00116K

Version: 0418-01