SOUTH AFRICAN

NATIONAL CAUSE-OF-DEATH VALIDATION PROJECT

VERSION 2 | REPORT ON THE QUALITY OF PHYSICIAN REVIEWED VERBAL AUTOPSIES CAUSE-OF-DEATH DATA USING ANACONDA

SAMRC Burden of Disease Research Unit | November 2022

CDC Centers for Disease Control and Prevention
saMRC advancinglife
GeoSpace measure it. manage it.
Bloomberg Philanthropies DATA FOR HEALTH INITIATIVE

SUGGESTED CITATION

Awotiwon OF, Nojilana B, Maqungo M, Groenewald P, Nannan N, Morof D, Rao C, Bradshaw D on behalf of the National Cause-of-Death Validation Project team. *Report on the quality of Physician Reviewed Verbal Autopsies cause-of-death data using ANACONDA for the South African National Cause-of-Death Validation Project.* Cape Town: South African Medical Research Council, 2022.

COLLABORATORING ORGANISATIONS

South African Medical Research Council
Statistics South Africa
National Department of Health
Department of Home Affairs
Centers for Disease Control and Prevention
South African Funeral Practitioners Association
National Funeral Directors Association

PRINCIPAL INVESTIGATORS

Prof Debbie Bradshaw, Dr Pam Groenewald, Dr Jané Joubert

CO-INVESTIGATORS

Prof Chalapati Rao, Ms Cherie Cawood, Dr Estevão Afonso, Mr Francois Bezuidenhout, Dr Jessica Price, Prof Johan Dempers, Dr Kassahun Ayalew, Prof Kathleen Kahn, Prof Lorna Martin, Dr Megan Prinsloo, Ms Mireille Cheyip, Dr Nadine Nannan, Mr Nesbert Zinyakatira, Prof Sam Clark, Dr Tshilidzi Muthivhi, Mr Thabo Molebatsi

TECHNICAL ADVISORS

Dr Beatrice Nojilana, Prof Carl Lombard, Dr Diane Morof, Dr Erin Nichols, Dr Frank Odhiambo, Dr Ian Neethling, Dr Lyn Hanmer, Ms Mmamokete Mogoswane, Ms Mosidi Nhlapo, Prof Sam Notzon, Ms Sizzy Ngobeni, Dr Oluwatoyin Awotiwon, Dr Victoria Pillay-van Wyk

PROJECT COORDINATORS

Dr Monique Maqungo and Ms Noluntu Funani

IMPLEMENTATION PARTNERS

Phase 1: Epicentre Health Research
Phase 2: GeoSpace International

FUNDING

This study has been supported in part by the President's Emergency Plan for AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention (CDC) under the terms of Cooperative Agreement SAMRC-CDC CoAG 1U2GGH01150. It has also been supported by the Bloomberg Philanthropies Data for Health Initiative through the CDC Foundation Civil Registration and Vital Statistics Project, and the South African Medical Research Council.

DISCLAIMER

The content of the report is that of the authors and does not necessarily represent the official position of the funding agencies.

 ISBN: 978-1-928340-51-5

A copy of this resource is available at: www.samrc.ac.za/intramural-research-units/BOD-resource-materials

Contents

List of Tables

List of Figures

Glossary

Analysis of Causes of (National) Death for Action tool (ANACONDA)

ANACONDA is an electronic tool that assesses the accuracy and completeness of mortality and cause of death data by checking for potential errors and inconsistencies.

Broad cause groups

To assist with understanding the data, the cause of death has been categorised into three groups: Group 1 (Communicable, maternal, neonatal, and nutritional diseases), Group 2 (Non-communicable diseases, including mental health conditions), and Group 3 (External causes and injuries (e.g., accidents, homicide, suicide, war deaths and natural disasters).

Garbage codes

Garbage codes (also referred to as 'unusable codes') are any ICD codes that cannot or should not be considered an underlying cause of death, such as septicaemia, senility, or headache. They may also be the codes for a cause that belongs in some other part of the morbid sequence of events leading to death such as the immediate or intermediate cause; or a cause of death that is insufficiently specified.

Global Burden of Disease Study (GBD)

A comprehensive worldwide observational epidemiological study— undertaken by the Institute for Health Metrics and Evaluation (IHME) and collaborators— that describes mortality and morbidity from major diseases, injuries and risk factors to health at global, national and regional levels. In this report, the GBD study refers to the study conducted by IHME.

International Classification of Diseases, 10th Revision (ICD-10)

The International Classification of Diseases and Related Health Problems is a classification and coding system developed by the World Health Organization (WHO) and defines the universe of diseases, disorders, injuries and other related health conditions, listed in a comprehensive, hierarchical fashion. The 10th revision, updated in 2016, is currently used as the international standard for reporting diseases and health conditions and can be found online. The next revision of ICD has been completed and it is anticipated that over the next few years, ICD-11 will be adopted.

IRIS

IRIS is an automated system for coding multiple causes of death and for the selection of the underlying cause of death based on the ICD-10 coding rules. It can be used in batch or interactively.

Medical doctor/ Physician

The terms Medical doctor and Physician are used interchangeably in this report.

Packages of garbage codes

This is an aggregation of the causes of death considered unusable. These are packaged to facilitate communication with clinicians and to assist with targeted effort in reducing their use. For example, Unspecified renal failure (N19) and Acute renal failure (N17) are combined as a package of Renal failure.

Underlying cause of death

The World Health Organization has defined the underlying cause of death as "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".

Unusable code

See garbage codes

Verbal Autopsy (VA)

A method of determining an individual's cause/s of death using a trained interviewer to administer a questionnaire to collect information about the signs, symptoms, treatment, and demographic characteristics of a recently-deceased person during a face-to-face interview with another individual – ideally a close caregiver or family-member – with knowledge about the deceased during his/her terminal illness/event.

1. Introduction

The **Analysis of Causes of (National) Deaths for Action, or ANACONDA**,[1] software tool used in this analysis was developed by the Melbourne School of Population and Global Health at the University of Melbourne, with the support of the Swiss Tropical and Public Health Institute at the University of Basel. ANACONDA is specifically designed to be applied to large datasets such as those from civil registries, or any other sources that routinely collect and generate cause of death data coded to the International Standard Classification of Diseases and Related Health Problems, 10th Revision (ICD-10), classified by age and sex.

We have used ANACONDA to assess the quality of the cause of death data from the South African National Cause-of-Death (SA NCOD) Validation Project based on the physician-reviewed verbal autopsy (VA) interviews. Several steps of ANACONDA related to assessing the completeness of death registration have been omitted as they were not relevant for these data.

The quality of input (cause of death) data is assessed by considering:

- the distribution of death by broad causes (Groups 1, 2 and 3) and garbage codes,
- the proportion of usable codes and garbage codes in the data assigned to ICD-10 Chapters,
- the type and severity of garbage codes,
- the packages of garbage codes, and
- biologically implausible codes.

In addition to the assessment of the coded data using ANACONDA as a starting point, the doctors' evaluation of the quality and sufficiency of the VA information has been analysed. This report is based on the analysis of the ICD-10 coded underlying cause of death, together with the doctors' assessment of the 5,387 VAs.

ANACONDA also provides a cause of death profile based on a redistribution of the garbage codes.

2. Data Source

Seventy-five medical doctors were appointed following training on correct certification of cause of death according to ICD guidelines, interpretation of VA narratives and interviews obtained using the WHO 2016 VA tool, and successful completion of assignments (practice case scenarios) and competency tests.[2] There were a total of 5,415 VA records and narratives, including 28 duplicate records. Each VA record, provided in batches of about 40 records, was independently reviewed by 2 doctors. Furthermore, a team of 5 medical doctors (quality assurance reviewers) assessed the reviewed cases; where different underlying cause(s) of death were reported on the causal sequence in Part 1 of the medical certificate of cause of death by the two reviewers, these doctors were asked by the independent reviewers to reach a consensus. When they could not come to consensus, the quality assurance reviewer either resolved the disagreement or brought the case to all the quality assessment reviewers and a Co-principal investigator (PI) for a panel review. Following de-duplication of the dataset, a total of 5,387 VA records were reviewed. In 3,115 (57.5%) of cases, there was agreement on the causal sequence of death between the independent reviews.

[1] 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, Civil Registration and Vital Statistics Improvement, University of Melbourne; 2017.

[2] Bradshaw D, Joubert JD, Maqungo M, Nannan N, Funani N, Laubscher R, Cheyip M, Zinyakatira N, Awotiwon O, Nojilana B, Bezuidenhout F, Martin L, Dempers J, Kahn K, Price J, Lombard C, Morof D, Nichols E, Rao C, Groenewald P. *South African National Cause-of-Death Validation Project: Methodology and Description of a National Sample of Verbal Autopsies*. Cape Town: South African Medical Research Council, 2020.

Two doctors reached a consensus review without input from the quality assurance (QA) team in 1,744 (42.8%) of cases; 557 (10.3%) cases had to be referred to the panel for review and ascertainment of the causal sequence.

The data were cleaned to remove incorrectly assigned scores. For example, one of the responses to the question "what was the quality of information" was given a score of 33 instead of 3. Furthermore, where a score of 999 (referring to "I don't know") was assigned, it was changed to a blank or missing.

Iris automated software[3] was used in the ICD-10 coding of both the multiple- and underlying cause(s) of death from the medical certificates of cause of death produced from the reviews of VA interviews and narratives by the doctor. Rejects were manually coded by two researchers and a Co-PI who had training in ICD-10 coding. The initial batch processing in Iris of the full dataset produced about 60% rejects. These rejects were mainly due to differences in spelling (e.g., TB, PTB, P TB), additional words (e.g., alcohol, alcohol use, alcohol ingestion etc.; cancer, carcinoma, ca, Ca, etc.), addition of brackets when recording duration of illness before death etc. In order to resolve these issues, we cleaned the data further by removing brackets, correcting typos etc., and updating the dictionary developed for the Western Cape local mortality surveillance system to include additional medical terms and their ICD-10 codes. Several re-runs of the dataset using the updated dictionary and manual coding, when required, resulted in coding of the full dataset.

The study protocol was reviewed and approved by the SAMRC Ethics Committee (EC004-2/2017). The project was also reviewed in accordance with CDC human research protection procedures and was determined to be research, but CDC investigators did not interact with human subjects or have access to identifiable data or specimens for research purposes.

3. Quality of information from VAs

Table 1 shows an assessment by the doctors on how sufficient the VA information was to certify the cause of death. About 65% of the VA records reviewed were reported to be of good quality (scored 3 - 5), and about 30% of the records were assessed to have provided insufficient information for the certification of cause of death.

Table 1: Description of how sufficient the information was from the VA narratives and questionnaires, SA NCOD Validation Project 2017/2018.

How sufficient was the Information	Number	%
1 (very poor)	576	10.7
2 (poor)	1,008	18.7
3 (good)	2,034	37.8
4 (very good)	1,156	21.5
5 (excellent)	324	6.0
Missing	289	5.4
Total	**5,387**	**100.0**

Table 2 shows an assessment of the quality of information from the VA narratives and questionnaire responses reviewed by the doctors. This was scored on a scale of 1 (very poor) – 5 (excellent) based

[3] Iris Institute. *Iris - Automated coding system for causes of death*. Heidelberg, Germany: German Institute of Medical Documentation and Information (DIMDI), 2018.

on the consistency between information provided in the narratives and the questionnaire responses, contradictions etc.

Of the 5,387 VA records reviewed, 4,392 (81.5%) were reported to be of good to excellent quality and about 13% of the records were assessed to have provided poor-quality information.

Table 2: Quality of information from VA narratives and questionnaires assessed by the doctors, SA NCOD Validation Project 2017/2018.

What was the quality of information	Number	%
1 (very poor)	105	2.0
2 (poor)	609	11.3
3 (good)	2,453	45.5
4 (very good)	1,699	31.5
5 (excellent)	240	4.5
Missing	240	5.2
Total	**5,387**	**100.0**

Out of the 5,098 cases with complete information on both criteria, 63% had exactly the same score for quality and sufficiency. The kappa statistic was 0.475 (95% CI: 0.459 - 0.491), indicating a moderate level of agreement between the two dimensions. We then assessed the association of these scores with the age of the deceased.

Figure 1 shows the age distribution of the sufficiency scores assigned to the VA information reviewed by the doctors. This shows that the doctors considered the sufficiency of the information provided was poorer for the infants, 45-49 years, and the 70 years and older.

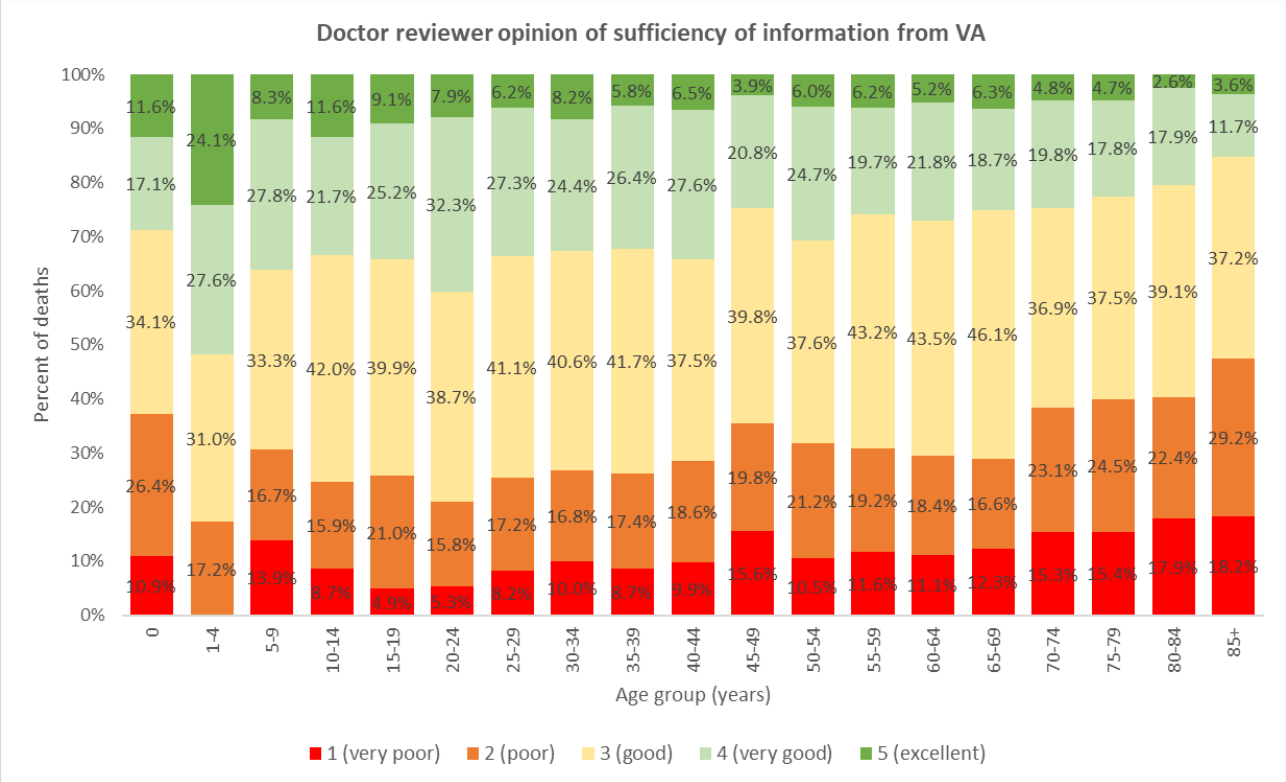

Figure 1: Sufficiency of information from verbal autopsy by age distribution, SA NCOD Validation Project 2017/2018.

In contrast, there does not seem to be any difference in the quality of the information by age of the deceased (**Figure 2**).

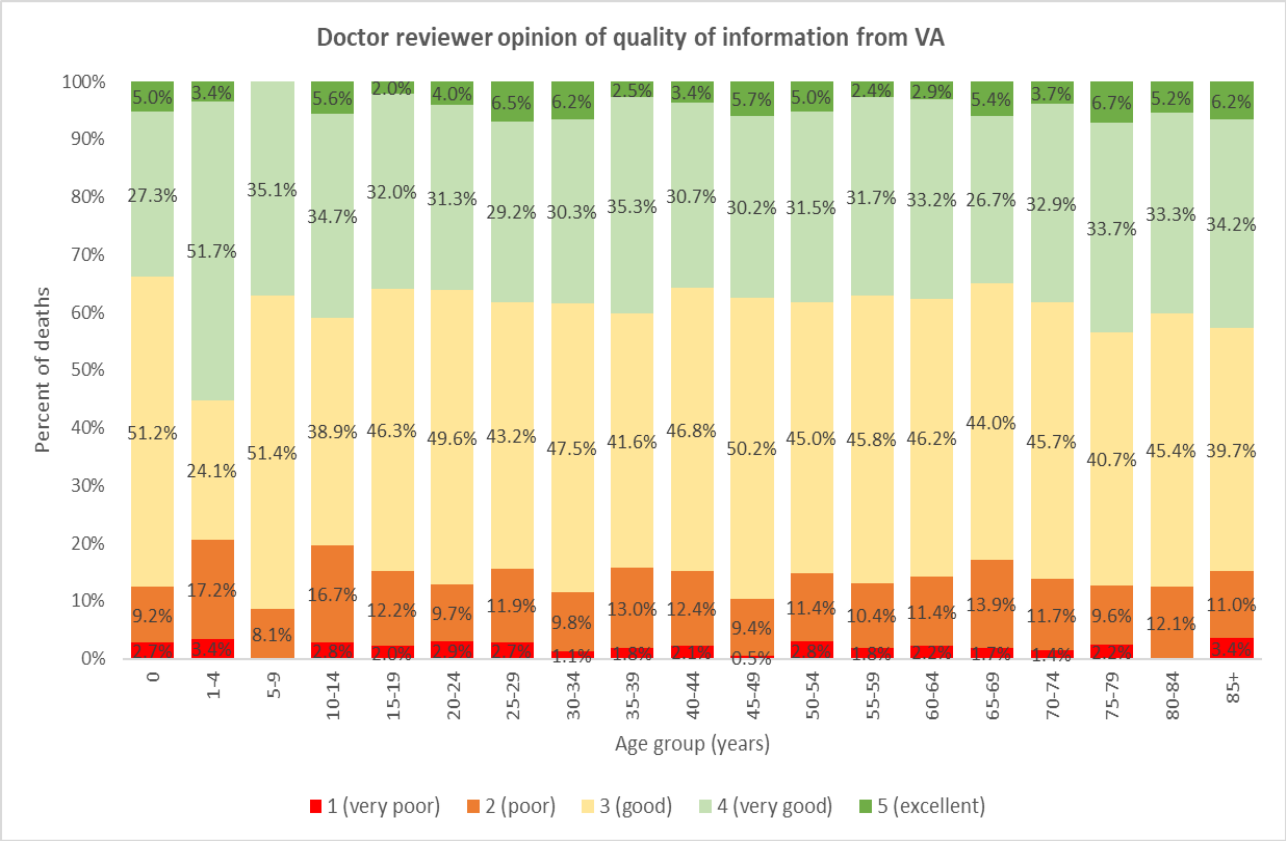

Figure 2: Quality of information from VA by age distribution, SA NCOD Validation Project 2017/2018.

4. Quality assessment of coded data

The data coded using IRIS were used as input in ANACONDA to evaluate the quality.

4.1 Garbage codes

From a data quality perspective, the proportion of deaths that have been coded as unusable or insufficiently specified is critically important. Also referred to as 'garbage codes', these codes contain no or very little useful information about the probable cause of death, and hence are of dubious value for guiding public health policy. When the proportion of garbage causes is large (above 10–15%) it will bias the distribution of deaths shown in the three broad cause groups, because the data will not represent the true health status of the population. [1] ANACONDA has identified codes that are insufficiently specified but would not have a substantial impact on the value of the data for public health policy and planning (e.g., unspecified stroke).

Overall, it appears that the doctors were successful in applying medical judgement and clinical experience in ascertaining causes of death from the VAs. Two-thirds (67.1%) of the deaths had useable causes; however, the proportion of garbage codes in this study was 21.6% (**Figure 3**). Furthermore, 11.3% of the deaths were classified as having insufficiently specified causes with limited impact. Until validated against other sources of clinical information, it is unknown whether the true underlying cause

of death has been identified.

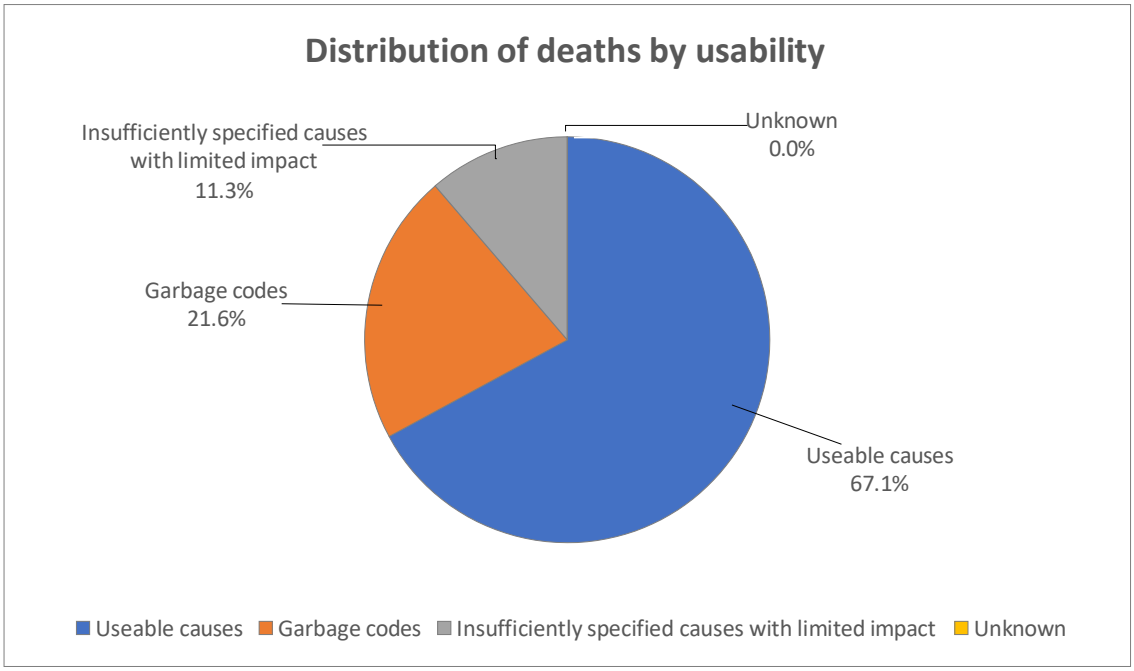

Figure 3: Proportion of deaths by 3 broad Groups, including unusable and insufficiently specified codes, SA NCOD Validation Project 2017/2018.

Figure 4 shows the breakdown of the useable causes of death according to the GBD broad cause groups. Most of the deaths (32.8%) were from Group 1 (Communicable diseases, maternal, neonatal, and nutritional diseases) causes.

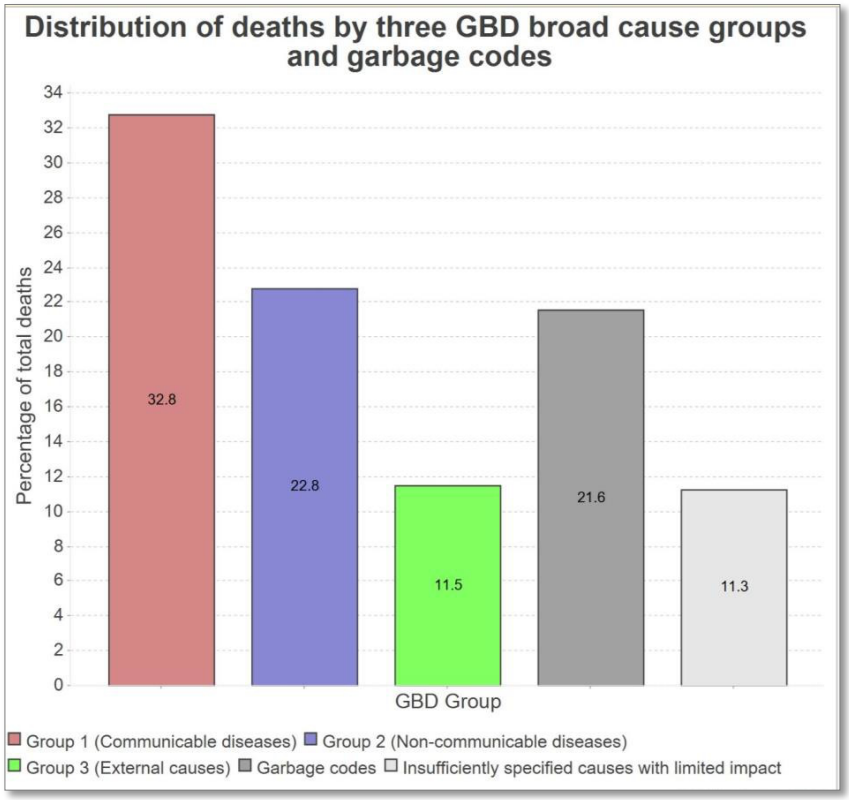

Figure 4: Distribution of deaths by utility of the cause of death, SA NCOD Validation Project 2017/2018.

4.2 Garbage codes by ICD-10 Chapter

The ICD-10 classifies mortality codes into 22 broad chapters. Displaying the proportion of deaths and fraction of garbage codes belonging to the leading 5 chapters of the ICD-10 (accounting for 82.8% of the records) is a useful first step to know where these codes are coming from and where the major areas of concern are (**Table 3**).

Chapter I (Certain infectious and parasitic diseases) contained most of the deaths (30.8%) but had the fewest garbage codes (1.1%); Chapter XVIII (Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified) had the most garbage codes in the data (48.2%) with 15.9% of the deaths were assigned. Garbage codes were also common in Chapter IX (Diseases of the circulatory system) with 19.7% and Chapter II (Neoplasms) with 8.2%.

Table 3: ICD Chapters with the highest amount of garbage codes, SA NCOD Validation Project 2017/2018.

Rank	Chapter	ICD-10 code range	Deaths (%)	Garbage codes (%)
1	Chapter I: Certain infectious and parasitic diseases	A00-B99	30.8	1.1
2	Chapter XVIII: Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	R00-R99	15.9	48.2
3	Chapter IX: Diseases of the circulatory system	I00-I99	15.0	19.7
4	Chapter XX: External causes of morbidity and mortality	V01-Y98	12.6	3.2
5	Chapter II: Neoplasms	C00-D48	8.5	8.2

4.3 Garbage codes by category

To further understand the origin and types of garbage codes in the cause of death data, ANACONDA classifies all unusable and insufficiently specified causes into five different categories based on ICD concepts, as follows:

- ***Category 1:*** Codes relating to **symptoms, signs and ill-defined conditions** (mostly drawn from R00–R99 in ICD-10).
- ***Category 2****:* Codes that have an **impossible underlying cause of death**.
- ***Category 3:*** Codes relating to **intermediate causes of death**.
- ***Category 4:*** Codes relating to **immediate causes of death**, such as heart or respiratory failure.
- ***Category 5:*** **Insufficiently specified codes** within ICD chapters within a larger disease category. These include causes like 'ill-defined site of cancer' and 'ill-defined injuries'. Use of these codes is unhelpful in guiding prevention efforts. Such efforts are usually cause-specific (e.g., lung cancer, early diagnosis of breast and prostate cancer).

As shown in **Table 4**, of the 5,387 deaths, 1,770 (32.9%) were assigned to garbage codes, with Category 1: Symptoms, signs and ill-defined conditions accounting for almost half of the assigned garbage codes (48.2%) and Category 2: Impossible underlying causes of death accounting for the least at only 0.1% (2 deaths). No garbage codes were classified to Category 4: Immediate causes of death. The fact that there were no garbage codes assigned in Category 4 shows that our doctors were well trained; however, higher percentages in Category 1 and 5 may be due to the quality of information provided in the VA.

Table 4: Garbage codes by category, SA NCOD Validation Project 2017/2018.

Category of garbage code	Number of deaths with a garbage code	Deaths assigned to garbage code (%)	Relative share of total garbage codes (%)
1: Symptoms, signs and ill-defined conditions	854	15.9	48.2
2: Impossible underlying causes of death	2	0.0	0.1
3: Intermediate causes of death	109	2.0	6.2
4: Immediate causes of death	0	0.0	0.0
5: Insufficiently specified causes within ICD-10 Chapters	805	14.9	45.5
TOTAL	1,770	32.9	**100%**

4.4 Garbage codes by severity

Unusable or insufficiently specified codes can be classified according to their potential impact for guiding or misguiding public policy to prevent premature deaths. In this classification, four levels of ICD-10 codes that should be avoided are defined, depending on how serious their impact is for misinforming public policy.[1] These four levels are:

- **Level 1 –** Codes with ***serious implications*** likely to have a ***very high impact*** for health policy. These are codes relating to such vague causes, that the true underlying cause of death could belong to more than one broad cause group.
- **Level 2 –** Codes with ***substantial implications*** likely to have a ***high impact***. These are codes relating to vague causes, where the true cause of death is likely to belong to only one of the three broad groups.
- **Level 3 –** Codes with ***important implications*** likely to have a ***medium impact***. These are codes for which the true underlying cause of death is known to be within the same ICD chapter. For instance, a death assigned to 'ill-defined site of cancer' indicates that the true cause of death was cancer but does not specify the site.
- **Level 4 –** Codes with ***limited implications*** likely to have ***low impact***. These are codes for which the true cause of death is likely to be confined to a single disease or injury category. For example, 'unspecified stroke' would still be assigned as a stroke death, and not to any other disease category. The implications for public policy of unusable causes classified at this level will generally be minor.

Table 5 provides the break-down of unusable codes by severity for the input data. More than half of the garbage codes (54.4%) are categorised as having very high severity (**Level 1)** with serious implications likely to have a ***very high impact*** for health policy. The assumption is that most of these causes come from the Symptoms, signs and ill-defined conditions. These codes have very vague causes, and the true underlying cause of death could belong to more than one broad cause group. This can misguide policy having serious public health implications.

Table 5: Garbage codes by severity, SA NCOD Validation Project 2017/2018.

Severity of garbage codes	Number of deaths with a garbage code	Deaths assigned to garbage code (%)	Relative share of total garbage codes (%)
Very high (Level 1)	962	17.9	54.4
High (Level 2)	26	0.5	1.5
Medium (Level 3)	175	3.2	9.9
Low (Level 4)	607	11.3	34.3
TOTAL	1,770	32.9	**100%**

4.5 Age-sex distribution of garbage codes

Examining the age and sex distribution of the four levels of garbage codes helps to assess whether certain garbage codes are more frequent in different population groups. As people age, they are more likely to suffer from diseases and conditions concurrently (co-morbidities). This makes it difficult for physicians to determine the true sequence of events leading to death and identify a single underlying cause of death. As a result, it is expected that garbage codes increase with age. In **Figure 5**, children aged under 5 years have a much lower proportion of garbage codes. However, while the proportions for the 5-64 years age group were similar for males and females, the proportion for females 65+ years was higher than the proportion for males 65+ years old. The fact that women usually survive to higher ages than men may be an explanation for the taller bar for women usually seen in the 65+ year age group.

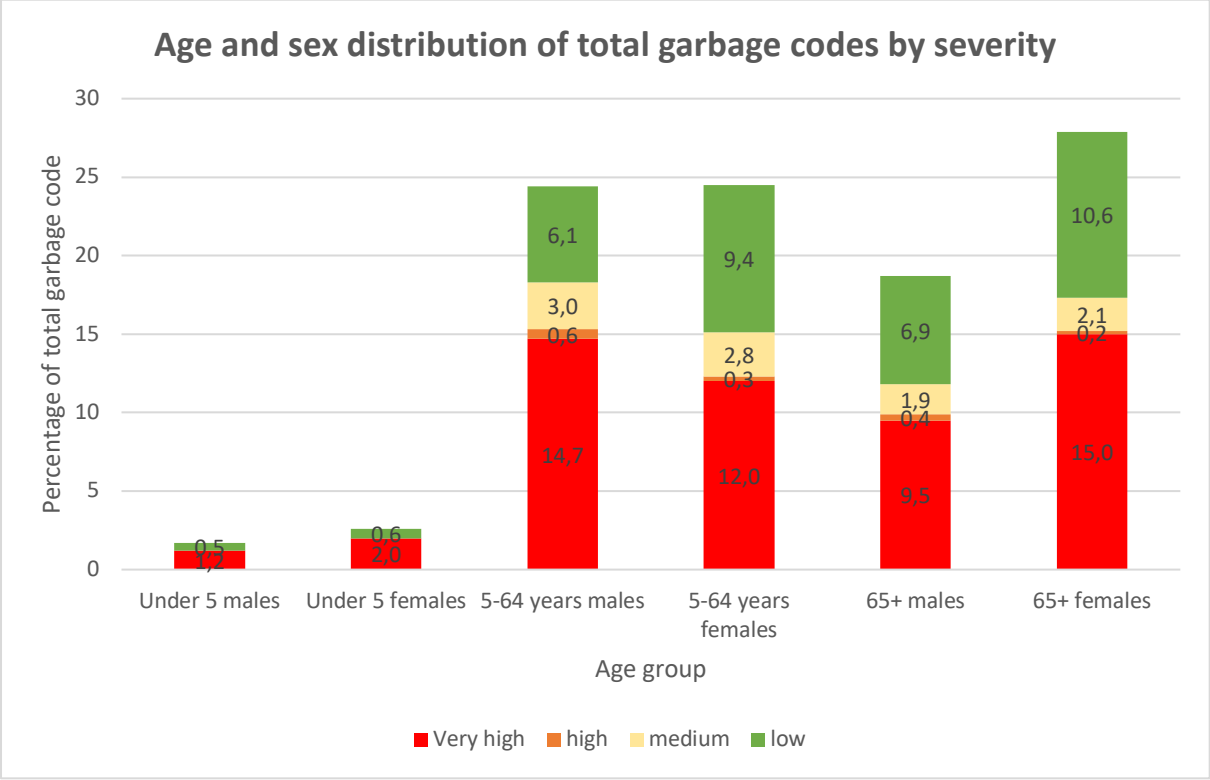

Figure 5: Age and sex distribution of total garbage codes by severity, SA NCOD Validation Project 2017/2018.

Females had a higher level of *garbage codes* with very high severity particularly in the 65+ year age group (15%). However, males in the 5-64-year age-group had a higher percentage of garbage codes with very high severity compared to the females (14.7% vs 12%, respectively).

If more than 15-20% of all garbage codes occur before the age of 65 or 70 years, this is indicative of serious certification and/or coding problems with the input data. However, this corresponds to the findings provided by the doctors based on insufficient VA information to assign a cause of death, as shown in Figure 1. Comparing the share of total deaths with the share of garbage codes in each age-sex group provides some insight into whether doctors have more difficulties in certifying the cause of death of older people because of comorbidities. Alternatively, it could be that many older people are dying at home without medical attendance, leading to vague and ill-defined diagnosis of the underlying cause of death.

4.6 Redistribution of garbage codes

As discussed previously, unusable or insufficiently specified causes of death and their corresponding codes have a significant impact on the true mortality composition of the population. **Figure 6** shows two pie charts; the first is based on the input data and shows the distribution of deaths across the three broad groups including unusable or insufficiently specified codes. The second displaces the data without garbage codes as they are redistributed to their most likely group using a redistribution algorithm developed by the IHME for use in the GBD Studies.[4]

The size of each of the lightly coloured proportions (21.7% and 12.9%) in the second pie chart indicates the proportion of deaths that were underestimated for each of the three broad groups as a result of garbage codes in the input data. After redistributing the garbage codes, the proportion of deaths in Group 1 (Communicable diseases) increased from 32.8% to 45.7%. A higher increase from 22.8% to 44.5% was seen in Group 2 (Non-communicable diseases) after redistribution.

The presence of garbage codes in mortality data can lead to serious misjudgment about the extent of the epidemiological transition in the country resulting in poor allocation of limited health resources, often underestimating the extent of mortality being caused by non-communicable diseases (Group 2) as shown in **Figure 6**.

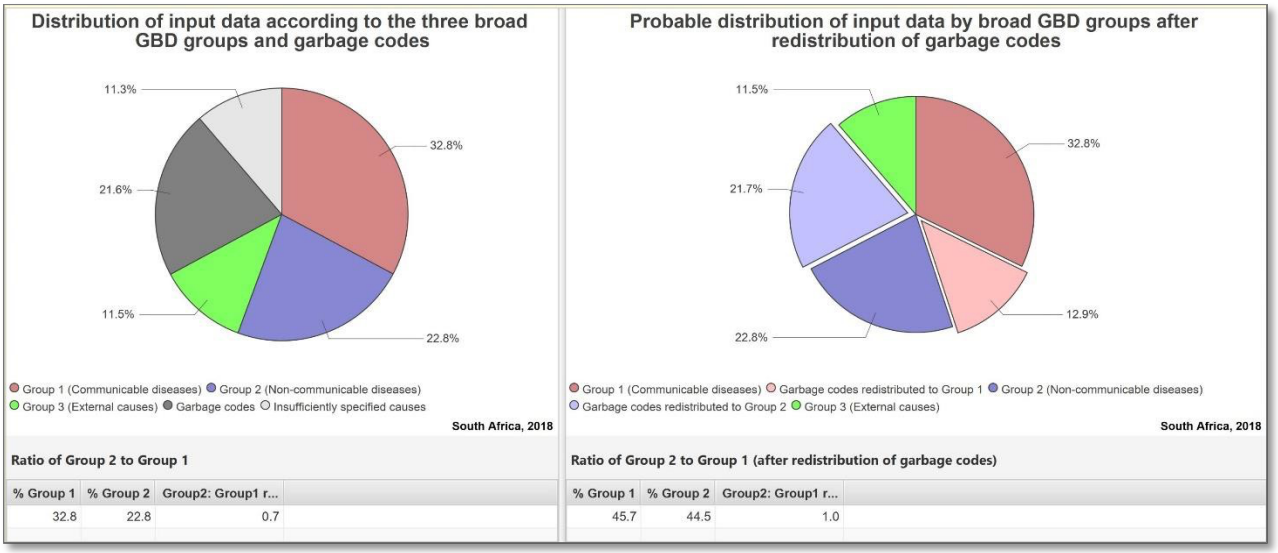

Figure 6: Distribution of causes of death data according to three broad groups, before redistribution of garbage codes, SA NCOD Validation Project 2017/2018.

4.7 Leading packages of garbage codes within Level 1 - very high severity

The identification of the most important "packages" of *garbage codes* within each severity level is crucial to facilitate communication and target efforts in order to reduce their use. Each package is a collection of poor diagnostic practices that are being applied during medical certification and coding, resulting in specific, identifiable misdiagnoses that are resulting in unusable and insufficiently specified

[4] Lozano R, 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* 2013: 380(9859);2095–128.

codes. The number of deaths assigned to each specific ICD garbage code in the package are also reported.

Table 6 provides a summary of the ICD-10 codes and category names (causes of death) for the top five packages within Level 1 (very high severity), which can be the focus of any improvement efforts. Ill-defined causes of death (R99) accounted for 89% (854 deaths) of causes in the very high severity group, followed by heart failure with 35 deaths (4%). Priority can be given to reducing the ill- defined and unspecified causes of mortality.

Table 6: Top five packages of garbage codes, Level 1, SA NCOD Validation Project 2017/2018.

Rank	Package name	Number of deaths	Causes in level (%)	Category name and ICD-10 code	Number of deaths
1	Impossible cause for death	854	88.77	Other ill-defined and unspecified causes of mortality (R99)	854
2	Left heart failure	35	3.64	Heart failure (I50)	35
3	Renal Failure	19	1.98	Unspecified renal failure (N19) Acute renal failure (N17)	17 2
4	Hepatic Failure	17	1.77	Hepatic failure, not elsewhere classified (K72)	17
5	Poisoning Accidental type 2	8	0.83	Accidental poisoning by and exposure to other and unspecified chemicals and noxious substances (X49) Accidental poisoning by and exposure to other and unspecified drugs, medicaments and biological substances (X44)	7 1
TOTAL (top five packages)		**955**	**99.3**	**TOTAL (top five categories)**	**955**
TOTAL (Level 1)		**962**	**100.0**		

Table 7 provides a detailed breakdown of the frequency of specific ICD-10 codes and category names (causes of death) for all the packages of garbage codes within Level 1 (very high severity) to Level 4 (low severity level) described above in **Table 4, 5 and 6**. Category 5 (Insufficiently specified causes within ICD-10 Chapters) accounted for 100% (782 deaths) of causes in Level 3 (medium severity) and Level 4 (low severity) and about 89% (23 deaths) of causes in Level 2 (high severity). Category 3 (Intermediate causes of death) accounted for the remainder of causes 11% (108 deaths) in Level 1 (very high severity) and 3.8% (1 death) of causes in Level 2 (high severity). Appropriate strategies can be put in place to address the use of insufficiently specified causes in death certification in order to improve public health utility of mortality data.

Table 7: Breakdown of garbage codes by severity levels and packages, SA NCOD Validation Project 2017/2018.

Leading packages - Very high severity level						
Rank	**Package name**	**Total causes**	**% of causes in level**	***ICD codes***	***Name of category***	***Total causes***
1	impossible cause for death	854	88.77	R99.-	Other ill-defined and unspecified causes of mortality	*854*
2	Left Heart failure	35	3.64	I50.-	Heart failure	*35*
3	Renal Failure	19	1.98	N19.-	Unspecified renal failure	*17*
				N17.-	Acute renal failure	*2*
4	Hepatic Failure	17	1.77	K72.-	Hepatic failure, not elsewhere classified	*17*
5	Poisoning Accidental type 2	8	0.83	X49.-	Accidental poisoning by and exposure to other and unspecified chemicals and noxious substances	*7*
				X44.-	Accidental poisoning by and exposure to other and unspecified drugs, medicaments and biological substances	*1*
6	Sepsis	7	0.73	A41.-	Other septicaemia	*7*
7	Anaemia Others	5	0.52	D64.-	Other anaemias	*5*
8	Poisoning Accidental type 1	4	0.42	X42.-	Accidental poisoning by and exposure to narcotics and psychodysleptics [hallucinogens], not elsewhere classified	*4*
9	Poisoning UDI type 5	3	0.31	Y14.-	Poisoning by and exposure to other and unspecified drugs, medicaments and biological substances, undetermined intent	*3*
10	Pulmonary Embolism	3	0.31	I26.-	Pulmonary embolism	*3*
11	Schizophrenia	2	0.21	F29.-	Unspecified nonorganic psychosis	*2*
12	Toxic liver disease	2	0.21	K71.-	Toxic liver disease	*2*
13	Anaemia Unspecified	1	0.1	D50.-	Iron deficiency anaemia	*1*
14	Peritonitis	1	0.1	K65.-	Peritonitis	*1*
15	Pneumonitis	1	0.1	J69.-	Pneumonitis due to solids and liquids	*1*
		962	**100**			***962***
Leading packages - High severity level						
Rank	**Package name**	**Total causes**	**% of causes in level**	***ICD codes***	***Name of category***	***Total causes***

1	Hypertension	13	50	I10.-	Essential (primary) hypertension	*13*
2	Unspecified external factor X59	4	15.38	X59.-	Exposure to unspecified factor	*4*
3	Cor pulmonale	3	11.54	I27.-	Other pulmonary heart diseases	*3*
4	External Causes UDI, type unspecified	3	11.54	Y34.-	Unspecified event, undetermined intent	*3*
5	Fire UDI	1	3.85	Y26.-	Exposure to smoke, fire and flames, undetermined intent	*1*
6	Gun UDI-type 2	1	3.85	Y24.-	Other and unspecified firearm discharge, undetermined intent	*1*
7	Sharp Objects UDI	1	3.85	Y28.-	Contact with sharp object, undetermined intent	*1*
		26	**100**			***26***
Leading packages - Medium severity level						
Rank	**Package name**	**Total causes**	**% of causes in level**	***ICD codes***	***Name of category***	***Total causes***
1	Unspecified Site Cancers	92	52.57	C80.-	Malignant neoplasm without specification of site	*77*
				C76.-	Malignant neoplasm of other and ill-defined sites	*15*
2	Unspecified Oropharynx Cancer	30	17.14	C14.-	Malignant neoplasm of other and ill-defined sites in the lip, oral cavity and pharynx	*30*
3	Unspecified Heart Diseases	18	10.29	I51.-	Complications and ill-defined descriptions of heart disease	*18*
4	Unspecified GI Cancer	13	7.43	C26.-	Malignant neoplasm of other and ill-defined digestive organs	*13*
5	Unspecified Uterus Cancer	10	5.71	C55.-	Malignant neoplasm of uterus, part unspecified	*10*
6	Unspecified Urinary Diseases	3	1.71	N40.-	Hyperplasia of prostate	*3*
7	Portal Phlebitis	2	1.14	K75.-	Other inflammatory liver diseases	*2*
8	Unspecified Chronic respiratory diseases	2	1.14	J98.-	Other respiratory disorders	*2*
9	Urinary Obstruction Diseases	2	1.14	N13.-	Obstructive and reflux uropathy	*2*
10	Unspecified Haemorrhagic Fever	1	0.57	A99.-	Unspecified viral haemorrhagic fever	*1*
11	Unspecified Maternal disorders	1	0.57	O95.-	Obstetric death of unspecified cause	*1*
12	Unspecified brain diseases	1	0.57	G93.-	Other disorders of brain	*1*
		175	**100**			***175***
Leading packages - Low severity level						
Rank	**Package name**	**Total causes**	**% of causes in level**	***ICD codes***	***Name of category***	***Total causes***

1	Unspecified Stroke	271	44.65	I64.-	Stroke, not specified as haemorrhage or infarction	*268*
				I67.-	Other cerebrovascular diseases	*3*
2	diabetes Unspecified type	209	34.43	E14.-	Unspecified diabetes mellitus	*209*
3	Unspecified lower respiratory infectious	75	12.36	J18.-	Pneumonia, organism unspecified	*55*
				J22.-	Unspecified acute lower respiratory infection	*15*
				P23.-	Congenital pneumonia	*5*
4	Unspecified Assault	30	4.94	Y09.-	Assault by unspecified means	*30*
5	Unspecified Malaria	11	1.81	B54.-	Unspecified malaria	*11*
6	Unspecified Cardiomyopathy	6	0.99	I42.-	Cardiomyopathy	*6*
7	Unspecified Meningitis	4	0.66	G00.-	Bacterial meningitis, not elsewhere classified	*4*
8	Unspecified Road injuries-1	1	0.16	V89.-	Motor- or nonmotor-vehicle accident, type of vehicle unspecified	*1*
		607	**100**			***607***
Total		**1,770**				***1,770***

4.8 Non-standard codes and biologically implausible causes of death

The dataset did not contain non-standard codes. All codes in the dataset were valid ICD codes. **Table 8** provides a summary of the 2 biologically implausible causes of death in the dataset. There were 3,616 total observations in the dataset (5,387 records). Eighteen (18) of these were Neonatal encephalopathy due to birth asphyxia and trauma, of which 2 (11.1%) were flagged as unlikely (implausible) observations due to age of decedent being above 10 years. Other neonatal deaths with 11 observations also had one (9.1%) flagged as implausible due to age above 10 years. These were possibly due to pre-existing conditions from childhood that significantly contributed to death later in life and the certifier was convinced that it followed a logical causal sequence in Part 1.

Table 8: The 2 biologically implausible causes of death, SA NCOD Validation Project 2017/2018.

Rank	Cause of death	Observations	Unlikely observations	Unlikely observations (%)	Reason for implausibility
1	Neonatal encephalopathy due to birth asphyxia and trauma	18	2	11.1	Above 10 years
2	Other neonatal disorders	11	1	9.1	Above 10 years
Total		**3,616**	**3**	**0.1**	

4.9 Distribution of deaths by broad group and age

Figure 7 shows the age distribution of deaths by broad cause group (as discussed above) merged into one figure, and the potential bias introduced by the proportion of deaths with an unusable or insufficiently specified code in each age group. An increase in the proportion of garbage codes was observed in the older age group with those aged 85+ years accounting for most (39.5%) of the garbage codes with high severity level.

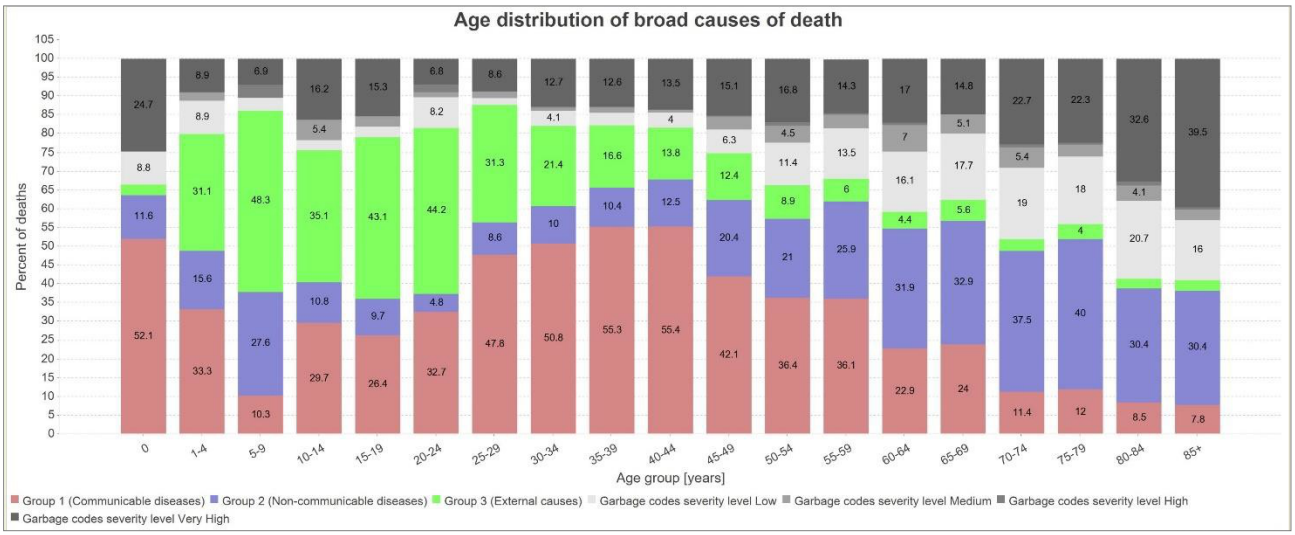

Figure 7: Distribution of deaths by age group and broad Group, including unusable and insufficiently specified causes, SA NCOD Validation Project 2017/2018.

5. Conclusion

- More than 80% of the VA information was assessed by the physician reviewers to be of good quality (scoring 3-5). About 60% of the VA information provided were assessed as being sufficient to certify cause of death by the doctors. Information became less sufficient the older the deceased. However, there was no relationship between the quality (consistency of information) and the age of the deceased.
- The percentage of useable codes in the dataset was high (67%), which corresponds well with the doctors' assessment about the quality and sufficiency of information from the VA interviews.
- Ill-defined and unspecified causes of death (garbage codes) were also present in the dataset (33%), most of which were of very high severity impact (55%), especially in the older age groups (65+ years). This may be due to comorbidities associated with the elderly, making it difficult for doctors/certifiers to correctly identify the cause of death based on the information obtained through the VA interview.
- The pattern of the garbage is consistent with the source of information being a VA interview. Half of the garbage codes were high severity, due to insufficient information provided for the doctors to assign cause of death certification. Most of the remaining garbage codes were classified as low severity, consistent with a VA interview having limited clinical information.
- The low number of intermediate levels of garbage codes might reflect high quality certification done by the project doctors. Provided there was sufficient information from the VA interview, the doctors identified an acceptable underlying cause of death.
- Three cases were identified by ANACONDA as biologically implausible. However, on review, these cases were not changed as they appeared to reflect accurately the underlying cause of death. They all related to long term consequences of conditions arising in the neonatal period. The lack of biologically implausible cases reflects the high-quality medical certification and coding achieved in this project.
- Although there are limitations with using VA, it still provides a high proportion of useable codes.

6. Recommendations

- Training the field workers has provided good quality information for doctors to certify, and future studies should incorporate adequate training.
- Training of medical doctors on how to certify deaths using information from VAs also contributed to good quality data. Hence, such training should be incorporated in future studies. Furthermore, it is important to train doctors in correct certification of death as this will contribute to good quality mortality statistics.
- Quality assurance (QA) process played an important role in getting good quality data. Hence, QA processes can be considered for incorporation into routine certification processes.

7. Next steps

- Conduct further analysis of the agreement between the independent reviewers to understand which conditions are more challenging for the medical doctors to certify using VA information.
- Validate the cause of death identified through VA against medical/forensic record information.
- Consider or investigate how to implement VA in the routine Civil Registration and Vital Statistics (CRVS) system.

Contact

Tel: +27 21 938 0911
Fax: +27 21 938 0200
Email: info@mrc.ac.za