


Are multiple coders needed for cause of death assignment: results from telephonic verbal autopsy interviews conducted in 2021 in South Africa

DIGITAL HEALTH
Volume 10: 1–12
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241282395
journals.sagepub.com/home/dhj



Ian Neethling^{1,2} , Diane Morof^{3,4}, Tracy Glass¹, Natasha Kallis¹, Chalapati Rao⁵, Debbie Bradshaw^{1,6} and Pam Groenewald¹

Abstract

Introduction: Verbal autopsy (VA) methods have emerged to estimate causes of death in populations lacking robust civil registration and vital statistics (CRVS) systems. Despite World Health Organization endorsement of routine VA use, cost and efficiency concerns persist. Telephonic verbal autopsies (teleVAs) can reduce cost. Physician coding offers a valuable approach, but the expertise required makes it resource-intensive, often involving multiple coders for consensus.

Objective: To assess inter-coder agreement for cause of death (CoD) in South African teleVAs using Kappa statistics, evaluating if agreement surpasses a 0.8 cut-off (very high) potentially allowing single coders.

Methods: A cross-sectional study employed telephonic VA interviews on non-facility deaths in Cape Town (December 2020–September 2021). Trained fieldworkers administered a standard VA questionnaire. Each case's VA responses were reviewed independently by two physicians, medically certifying the CoD. A panel was used to solve disagreements. Cohen's kappa-statistic (*k*-statistic) tested agreement levels.

Results: Decedents were aged between 18 and 98 years. In total, 228 teleVAs (16.6% response rate) were conducted. Physician coding agreement was good overall (*k*-statistic: 0.63). Diabetes mellitus (47%) and other non-communicable disease (42%) had initial agreement between physician coders in less than 50% of cases in comparison to consensus totals. COVID-19 (89%) and acute cardiac disease (83%) showed initial agreement in more than 80% of cases compared to consensus totals. A chi-square test revealed a significant difference in the number of causes listed on death notification forms for cases with and without agreement in Part 1 ($\chi^2 = 14.71$, $p < 0.01$), but not in Part 2 ($\chi^2 = 4.97$, $p = 0.17$).

Conclusion: CoD agreement might not be high enough to infer that single coders can be used instead of multiple coders. Challenges with co-morbidities and specific CoDs with multiple sequelae highlight the need for further research and refinement of VA methodologies for reliable CoD determination in routine practice.

Keywords

Telephonic verbal autopsy, cause of death, multiple physician coders, agreement, kappa analysis, co-morbidities, South Africa

Submission date: 28 March 2024; Acceptance date: 23 August 2024

¹Burden of Disease Research Unit, South African Medical Research Council, Cape Town, South Africa

²Institute for Lifecourse Development, University of Greenwich, London, UK

³U.S Centers for Disease Control and Prevention, Division of Global HIV and TB, Durban, South Africa

⁴U.S. Public Health Service Commissioned Corps, Rockville, MD, USA

⁵National Centre for Epidemiology and Population Health, Australian National University, Canberra, Australia

⁶School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa

Corresponding author:

Ian Neethling, Burden of Disease Research Unit, South African Medical Research Council, P.O. Box 19070, Tygerberg 7505, Cape Town 8000, South Africa.
Email: ian.neethling@mrc.ac.za



Introduction

Accurate and dependable mortality data are crucial for informing healthcare policies and priorities.¹ In South Africa, the civil registration and vital statistics (CRVS) system is the primary source of cause of death (CoD) data. While registration rates have improved significantly, from 50% to over 90% for adults since 1994, there are still challenges with the quality of CoD data.² HIV/AIDS deaths are frequently misclassified, injuries are often not categorized by the manner of death, and a high proportion of causes are poorly defined based on poor medical certification.^{2,3} The SARS CoV-2 pandemic highlighted further issues in the CRVS system, including a shortage of timely data and challenges associated with gathering medical information for deaths occurring outside of healthcare facilities, which account for roughly 45% of all deaths.⁴

Verbal autopsy (VA) is an indirect method used to determine the CoD in the absence of formal medical certification where informants of decedents are interviewed about individuals who died outside of a health facility. This information is either reviewed by physicians or included in computer algorithms to determine the likely CoD. VA can facilitate the completeness of death registration and CoD information.⁵ It is used in Demographic and Surveillance sites in multiple countries globally, including three sites in South Africa, i.e. Agincourt (located in rural Mpumalanga Province), the African Health Research Institute (rural KwaZulu-Natal Province) and Dimamo (Limpopo Province).⁶⁻⁸

The World Health Organization (WHO) has advocated for the adoption of the routine use of VA into CRVS in countries with incomplete and unreliable mortality data.⁹ However, current VA methods and data processing can be costly and require additional human resources to ensure CoD are ascertained. Efficiencies may be needed to ensure integration of VA into CVRS can be accomplished and sustained. Automated CoD coding using probabilistic and algorithmic models can aid the routine use of VAs by processing large volumes of data with 100% internal consistency and validity.¹⁰ Existing literature offers no conclusive evidence regarding the relative effectiveness of physician coding compared to alternative methods. For instance, Jha et al.¹¹ comparing various automated coding methods (Naïve Bayes Classifier, King-Lu, InSilicoVA, InSilicoVA-NT, InterVA-4, and SmartVA) with physician assignment of CoD for VAs showed an average concordance of 62%, 56%, and 59% for adult, child and neonatal deaths, respectively. Another systematic review¹² evaluated the performance of several automated coding methods against physician coding using hospital-based deaths as the reference standard. This review found that both methods achieved an overall chance-corrected concordance of around 50% or lower. Finally, Chandramohan, Fottrell¹³ explores the trade-offs between physician coding (accurate but biased/slow) and automated coding (fast/scalable but

potentially less accurate) in VA methods' evolution, ultimately finding it difficult to definitively recommend one over the other. Despite the inconclusive evidence, physician coding remains the standard practice for CoD assignment in VAs. For reliability purposes, physician coding is usually done by multiple coders or a panel of physician experts who independently review the data from a VA and assign a CoD.¹⁴ Compensating physicians to review and assign a CoD from a VA can be a costly exercise.

However, Joshi et al.¹⁵ showed a very high agreement on all CoD coding between 2 physician coders, indicating that duplicate coding has little advantage over a single-coder system which has huge cost saving implications. That study was conducted in 45 small villages across Eastern and Western India with a combined population of 180,000. Overall agreement in the study was assessed at the International Statistical Classification of Disease and Related Health Problems 10th Revision (ICD-10) chapter level, which might have inflated the level of agreement between physician coders compared to an assessment at specific underlying cause level.

Telephonic verbal autopsies (teleVAs) can also reduce the data collection cost of conducting routine verbal autopsies. In 2021, we conducted a teleVA study to assess the feasibility of implementing teleVA in Western Cape,¹⁶ understand its acceptability to interviewers and families of the deceased, and determine the quality of the CoD information collected using teleVA (vs. face-to-face VA) and physician coding. Interviewers and respondents found the teleVA approach to be acceptable and feasible. TeleVAs experienced technical difficulties in 15.7% of cases, primarily due to poor connectivity (11.4%).¹⁶ Successful interviews were achieved with 62% of respondents,¹⁶ and initial analysis suggests that the quality of teleVA responses to questions were comparable to face-to-face interviews that were conducted in the National Cause-of-Death Validation Study (NCODV).¹⁷ The quality of teleVA responses were assessed by comparing proportions of non-informative responses ("don't know" and "missing" as oppose to "yes" and "no") between teleVA and face-to-face VAs for each root question using a chi-squared test and revealed only one question with a statistically significant difference.

In addition to providing a cost saving alternative, teleVAs also offers several potential advantages over traditional face-to-face VAs in the context of routine CRVS, particularly in resource-limited settings. TeleVAs overcome geographical limitations and reach geographically dispersed populations, particularly in remote areas with limited healthcare infrastructure.¹⁸ Interviews can be easily scheduled, offering greater flexibility in both interview time and location.¹⁹ Finally, teleVAs offer a safer alternative for fieldworkers in situations where in-person visits might pose safety risks.

There are also potential issues that could hinder the wider implementation of teleVAs. TeleVAs may face

difficulties in reaching all potential participants due to the reliance on telephone communication. This can potentially lead to lower response rates, which could affect the representativeness and generalizability of the collected data. Response rates from telephone surveys typically range between 5% and 30%.²⁰ Furthermore, obtaining accurate contact information for next-of-kin (NoK) can be an issue, hindering the initial recruitment process.

This analysis focused on physician coding and investigated the necessity of utilizing multiple physician coders for CoD assignment in teleVAs conducted in South Africa. We analyzed inter-coder agreement using Kappa statistics, aiming to determine if the agreement level surpasses a pre-defined cut-off (0.8 for very high agreement) that would suggest a single coder may be sufficient for accurate CoD assignment.

Methods

Study design

This was a cross-sectional study that collected CoD information through teleVAs on a convenience sample of study participants who died outside of a health facility.

Setting and inclusion criteria

TeleVAs were conducted with NoK of decedents that resided in Cape Town and passed away outside of a health facility between December 2020 and September 2021. NoK were interviewed if they were aged 18 years or older and if their deceased family members were aged 18 or older and died from natural causes outside a health-care facility (excluding accidents and suicides). The study team sought to interview the best respondent (caregiver during last illness or closest NoK); however, when no relative was available for interview, any person with knowledge of the decedent's health condition in the period leading to death was interviewed. Therefore, NoK in this study refers to any person that was interviewed.

Data source

Three different data sources were utilized to access the name and contact details of NoK. Data sources included the prospective collection of details by (a) participating funeral parlors which identified out-of-facility deaths, contacted NoK, and secured interview consent for project staff; (b) emergency medical services (EMS) between August and September 2021 which obtained contact details of the individual who contacted EMS at the time of the death and (c) retrospective identification of decedents from a list of person-level clinical data between December 2020 and July 2021 consolidated across government services by the Western Cape Provincial Health Datacentre (WCPHDC)

and NoK contact details were sought from other data center resources. A convenience sample of 1571 decedents with complete information on name and contact details for the NoK were selected from a participant list of 4342 NoK. Calls were made to 1347 NoK in total due to time-constraints of the fieldwork (please see Groenewald et al., 2022 for a more detailed description of the recruitment process).¹⁶

Ethics

Ethics approval for the study was obtained from the SAMRC Ethics Committee and the Switzerland Tropical Health Institute. NoK identified through funeral parlors and EMS were briefed about the study, before telephonic contact by interviewers, and informed that they may be contacted telephonically for participation in the study, whereas those identified through the WCPHDC were directly contacted by interviewers. All NoK who agreed to participate in the study were provided with a written study information sheet about the study via WhatsApp. The study process was then explained verbally at the start of the interview, after which all study participants provided verbal consent. The ethics committees approved the use of verbal consent for participant enrollment. This approval was granted due to the urgent public health crisis posed by the COVID-19 beta wave in South Africa and its resultant lockdown measures, which included restriction on movement, social distancing etc. Verbal consent was audio recorded and saved on an online data collection platform (Kobotools) hosted by the WCPHDC. TeleVA interviews were then conducted by trained fieldworkers during a 7-week period between August and October of 2021. Interviewers underwent 5 days of refresher VA training, which included the newly added COVID-19 questions, as well as training on confidentiality, informed consent, human subject research, and research integrity. They also received training on conducting telephonic VAs. This activity was reviewed by CDC, deemed not research, and was conducted consistent with applicable federal law and CDC policy.

Verbal autopsy interviews:

To capture the impact of the COVID-19 pandemic, we used a modified version of the validated World Health Organization (WHO) 2016 VA questionnaire (see Questionnaire in Supplementary Appendix).^{21,22} This modified version, developed by the WHO itself, included additional questions specifically designed to assess experiences related to COVID-19. The questionnaire was further supplemented with a narrative from the NoK on the circumstance of death. The questionnaire was translated by a registered company into nine official languages (isiXhosa, isiZulu, isiNdebele, Afrikaans, SePedi, SeSotho,

SeTswana, TshiVenda, XiTsonga) for the KoBoToolbox platform. Independent reviewers ensured translation accuracy, resolving discrepancies through discussion. Interviews were conducted in the participant's preferred language: English ($n = 146$), Afrikaans ($n = 44$), and isiXhosa ($n = 38$), out of a total of 228 participants. The VAs commenced with an open narrative, during which respondents described the events leading to the relative's death in their own words, before answering structured questions from the VA questionnaire. Interviewers recorded handwritten notes in English in provided notebooks during this narrative section. The data were captured on KoboToolBox, an online platform, via tablets.²³

Quality control

Quality control checks on the captured answers to the VA questions were done throughout the data collection period by the fieldwork manager and two study investigators to review completeness, accuracy, and consistency. For example, the eligibility of each case, with respect to age, place of death and residence of decedent was checked, verifying whether questions pertaining to maternal deaths were answered in accordance with the sex of the deceased, and the captured relationship of decedent to respondent were checked against age and sex. Narratives were checked for legibility and used as a check of the answers provided to the VA standard questions. Anomalies or any case that did not meet the inclusion criteria was referred to the fieldworkers who rechecked the information against their notes and with NoK if necessary.

Cause of death coding

Nine physicians were recruited to independently review the VA responses and narrative and medically certify the CoD according to ICD-10 guidelines.²⁴ Physician coders participated in a prior intensive training exercise on medical certification of CoD using the WHO 2016 Verbal Autopsy tool, with additional COVID-19 questions. This training ensured competency in medical certification of CoD according to ICD-10 guidelines and interpreting VA data.²⁵ All medical certificate of cause of death forms (MCCD) were reviewed by physician pairs, working independently of each other. The VA responses and images of the VA narratives were accessible through a web portal to the investigators on the study and saved to a MS Teams folder for access by physician coders. The physicians completed a VA review form developed on KoboToolBox using the Kobocollect app on mobile devices.²³ The review form included the internationally recommended MCCD²⁶ with the causal sequence in Part 1 and contributing causes in Part 2 (see medical certificate of CoD in Supplementary Appendix). Two physicians independently certified the CoD for each case. In cases where the underlying COD

were different, consensus was initially sought by discussion between the two reviewers, or through a panel of physician coders where consensus could not be reached. Using IRIS software,²⁷ the underlying CoD was coded to ICD-10 automatically. Where IRIS was unable to determine an underlying CoD, manual coding was done by the principal investigator, a qualified physician.

Data cleaning

The data were downloaded from Kobotools into Microsoft Excel[®], and imported into STATA 16 (Stata Corporation, College Station, TX) for further analysis. Data were cleaned by recoding the assigned ICD four-digit codes to three digits and checking the validity of ICD codes and causes inconsistent with age and sex. Cases where it was not possible to determine a CoD with the available information were removed for the agreement analyses.

The ICD-10 codes were aggregated to the national burden of disease (NBD) list with the addition of COVID-19 and further aggregated to a shortlist containing 26 causes.³

Analysis

Cohen's kappa-statistic (k -statistic) was used to test for levels of agreement of cause-of-death coding between the physician coders. A 95% confidence interval (CI) was used to estimate the lower and upper limits around the k -statistic. The Landis and Koch²⁸ rating scale was used to indicate the strength of agreement, i.e. very good for k of 0.80 and above, good for k between 0.60 and 0.79, moderate for k between 0.40 and 0.59 and poor for k of 0.39 and below. A very good agreement was considered an indication that employing multiple coders provided minimal advantages over a single-coder system.

The distribution of the number of assigned causes in Part 1 and Part 2 were analyzed to assess whether multiple assigned causes influenced physician agreement. A Chi-squared test was performed to examine potential differences in agreement based on the number of conditions specified in Part 1 and Part 2, with significance set at $p < 0.05$.

Results

From the 1371 NoK contacted, a total of 228 teleVA interviews were conducted (16.6% response rate). On average, the VA interviews were conducted 75 days after death. The majority of 228 decedents were female (54.0%) and aged 50 years or older (76.8%), as shown in Table 1. Age of the decedents ranged between 19 and 98 years. Respondents were mostly a child of the decedent (37.3%), other family (22.4%), the spouse (18.9%), or a sibling (10.1%). Only 7.5% of respondents were not family of the decedent.

The five specific leading causes of death identified through the consensus process were COVID-19 (27.2%), followed by HIV/AIDS (9.7%), stroke (7.0%), diabetes mellitus (6.6%), and acute cardiac disease (5.3%), as shown in Table 2. Other non-communicable diseases (NCDs; 8.3%), other cardiac diseases (7.9%), and other cancers (7.5%) contributed significantly to the mortality profile but are not categorized as single underlying causes of death. NCDs were prominent, with stroke in particular contributing 7% of deaths, while different types of cancer contributed 12.4%. Injuries (suicide, other accidents, homicide and transport accidents) contributed towards 4.8% of all deaths.

Table 3 shows the agreement pattern for specific causes of death between independent coders. The same underlying CoD was assigned between different pairs of physician coders for 153 (67.1%) out of a total of 228 reviews. Agreement between coders varied across cause-of-death when comparing the number of initial agreements

between reviewers for each CoD to the total number identified through the consensus process. Diabetes mellitus (47%) and other non-communicable disease (42%) were the causes-of-death with more than 10 cases that had initial agreement in less than 50% of cases in comparison to consensus totals. COVID-19 (89%), acute cardiac disease (83%) and other cancers (82%) were the causes with initial agreement in more than 80% of cases compared to consensus totals.

Table 2. Distribution of underlying cause of death using NBD categories ($N=228$).

Cause list	<i>N</i> (%)
COVID-19	62 (27.2)
HIV/AIDS	22 (9.7)
Other NCDs	19 (8.3)
Other cardiac disease	18 (7.9)
Other cancers	17 (7.5)
Stroke	16 (7.0)
Diabetes mellitus	15 (6.6)
Acute cardiac disease	12 (5.3)
Pulmonary TB	7 (3.1)
Digestive cancer	7 (3.1)
Obstructive pulmonary disease	7 (3.1)
Undetermined	6 (2.6)
Respiratory cancer	4 (1.8)
Suicide	4 (1.8)
Other accidents	3 (1.3)
Homicide	3 (1.3)
Other infectious	2 (0.9)
Pneumonia	1 (0.4)
Diarrhea	1 (0.4)
Renal failure	1 (0.4)
Transport accidents	1 (0.4)

Table 1. Background characteristics of decedents ($N=228$).

Characteristics	Number (%)
Sex	
Male	104 (45.6)
Female	123 (54.0)
Transgender	1 (0.4)
Age category (years)	
18-29	7 (3.1)
30-49	46 (20.2)
50-69	103 (45.2)
70+	72 (31.6)
Respondent-to-decedent relationship	
Spouse	43 (18.9)
Child	85 (37.3)
Parent	9 (4.0)
Sibling	23 (10.1)
Other family	51 (22.4)
Not family	17 (7.5)

Table 3. Underlying cause of death coding agreement between coders.

	HIV/ Pneumonia	AIDS	Diarrhea	TB	Pulmonary infectious	Other COVID-19	Maternal cancer	Digestive cancer	Respiratory cancer	Other cancers	Diabetes mellitus	Acute cardiac disease	Stroke	Other cardiac disease	Renal failure	Other NCDs	Transport accidents	Homicide	Suicide	Other external cause	Undetermined	Total
Pneumonia	0																					1
HIV/AIDS	15				1	1						2										20
Diarrhea	0					1																1
Pulmonary TB	1		4		1	1		1				1								1		9
Other infectious					1	1			1													3
COVID-19	1				50			1				1										55
Maternal			1			0				1												2
Digestive cancer								6														6
Respiratory cancer									3													3
Other cancers							2		1	14			1									18
Diabetes mellitus			1		4					7		2										16
Acute cardiac disease					3				1	10		2										17
Stroke					1					1		9	1									12
Other cardiac disease									1	1	1	10	1									15
COPD					1						1	1	3									8
Renal failure												1		1								2
Other NCDs					3					1	1	3			8							19
Transport accidents																	1					1

(continued)

Table 3. Continued.

	HIV/ Pneumonia	AIDS	Diarrhea	TB	Pulmonary infectious	Other infectious	COVID-19	Maternal	Digestive cancer	Respiratory cancer	Other cancers	Diabetes mellitus	Acute cardiac disease	Other cardiac disease	Stroke	Renal failure	Other NCDs	Transport accidents	Other accidents	Homicide	Suicide	Other external cause	Undetermined	Total
Other accidents																		1				1		2
Homicide																				1		2		4
Suicide																						4		4
Other external cause																						0		0
Undetermined																								10
Total	3	15	1	6	1	1	65	0	9	4	16	12	16	15	22	5	1	14	1	1	1	4	4	228
Consensus	1	22	1	7	2	2	62	0	7	4	17	15	12	16	18	7	1	19	1	3	4	0	6	6

Strengthening mortality surveillance (SMS) for improved monitoring of HIV/AIDS, TB, and COVID-19 in South Africa, September 2021.
 HIV/AIDS = human immunodeficiency virus autoimmune deficiency syndrome, TB = tuberculosis, COVID-19 = coronavirus disease 2019, NCDs = non-communicable disease, COPD = chronic obstructive pulmonary disorder.

The kappa statistic for the overall initial agreement between physician coders was 63% (CI: 59–67%).

The majority of cases (63%) had only 1 condition reported in the causal sequence in Part 1 with 37% reporting 2 or more conditions, (Figure 1, see also Supplementary Appendix in Table A1). A high proportion of cases (75%) had contributing conditions indicated in Part 2 of the death form with almost half of the cases (47%), having 2 or more conditions.

Figure 1(a) and (b) shows the number of agreements and disagreements by the number of causes specified in part 1 and part 2. In cases with disagreements in the assigned underlying CoD, 49% had 1 condition reported in the causal sequence in part 1, while 51% had two or more conditions reported in part 1. The disagreement cases had a higher proportion with 2 or more conditions specified in part 1 compared to the agreement cases (51% vs. 31%). The contributing conditions reported in part 2 with 2 or more conditions were slightly higher in those cases with disagreements (83% vs 75%).

The results of the chi-square test suggest that there is a statistically significant difference in the number of causes specified on the death notification form between cases with agreement and disagreements in part 1 ($\chi^2 = 14.71$, $p = 0.002$), while in part 2 the difference is not statistically significant ($\chi^2 = 4.97$, $p = 0.17$).

Discussion

There was good overall agreement (k -statistic: 0.63) on underlying CoD assignment between physician coders and the consensus process. The observed level of agreement ($\kappa = 0.63$) falls short of our threshold of 0.80 for recommending single coders in routine VAs. This highlights the potential challenges associated with relying solely on single coders, particularly considering the need for reliable CoD assignment across diverse categories. However, other aspects also warrant discussion in considering multiple vs single coders in routine CRVS systems.

The relatively small sample size of this study may limit the generalizability of the findings. The overall agreement is also not as strong as that found by the Joshi et al.¹⁵ study (k -statistic: 0.93). It's important to consider that the Joshi, Lopez¹⁵ study analyzed agreement at the ICD-10 chapter level, which would have facilitated higher agreement (k -statistic: 0.93), especially for cardiovascular disease, external causes of injury and infectious disease chapters. To assess the impact of agreement at the ICD-10-chapter level, we conducted similar analysis (result not presented) and found a slightly higher agreement (k -statistic: 0.67; see Supplementary Appendix in Table A2) compared to the agreement of more specific NBD level underlying causes but still lower than found in Joshi et al.¹⁵ This reiterates that the level of aggregation of CoD assignment should be taken into account in the design of future teleVA

systems. Using more granular coding allows for a deeper understanding of specific causes of death, which is valuable for research, targeted interventions, and improving health-care outcomes. ICD 3 digit character codes enable aggregation to the local NBD list of codes for national use; and other lists for international comparisons such as the WHO Global Health Estimates,²⁹ ICD Mortality Tabulation List 1,³⁰ and the Institute for Health Metrics and Evaluation Global Burden of Disease List.³¹

In addition to the granularity of CoD assignment, the routine introduction of teleVAs into the CRVS is also likely to be resource dependent. A single coder system, as opposed to multiple coders would save on costs. However, a consensus approach may be needed for cases with co-morbidities as there were disagreements in the consensus process for 48% (38/85) of cases with 2 or more causes reported in part 1. There was a significant difference between cases with agreement and disagreement on assigned underlying CoD based on the number of conditions specified in part 1 ($p = 0.002$) which may suggest that the cases with disagreements had more complicated causal sequences. In addition, CoDs with multiple associated conditions or sequelae, may pose challenges for accurate CoD assignment. For instance, diabetes mellitus, which often presents with various complications, required the consensus process for agreement in a significant number of cases. Our analysis revealed that only 7 out of 15 cases identified as diabetes mellitus through the consensus process had initial agreement between the physician coders.

In contrast, CoDs like COVID-19 exhibited high initial agreement. Diseases with clear diagnostic criteria, like COVID-19 with its established testing protocols, might lead to more consistent coding compared to conditions like diabetes mellitus which can have a wider spectrum of presentations. A nuanced approach may be needed in the development of CoD assignment algorithms for cases with co-morbidities and conditions with multiple sequelae. The identification and assessment of co-morbidities and multi-sequelae conditions may benefit from a multiple coder consensus approach, alongside the incorporation of additional clinical information. Refining physician training to enhance coding accuracy and consistency would also be important. Therefore, in a routine system, a hybrid approach that combines single coding with a consensus process could be beneficial. Over time, experiences from the hybrid approach could guide the development of a standardized process for physician assignment of CoD from VA, resulting in optimal levels of accuracy and consistency in coding.

In a hybrid approach optimal staffing for physician coding may be an important consideration to ensure adequate allocation of resources. This study used nine physician coders to assign a CoD. However, it remains uncertain how many coders were used in the study conducted by Joshi et al.¹³ as it is indicated that two coders were assigned

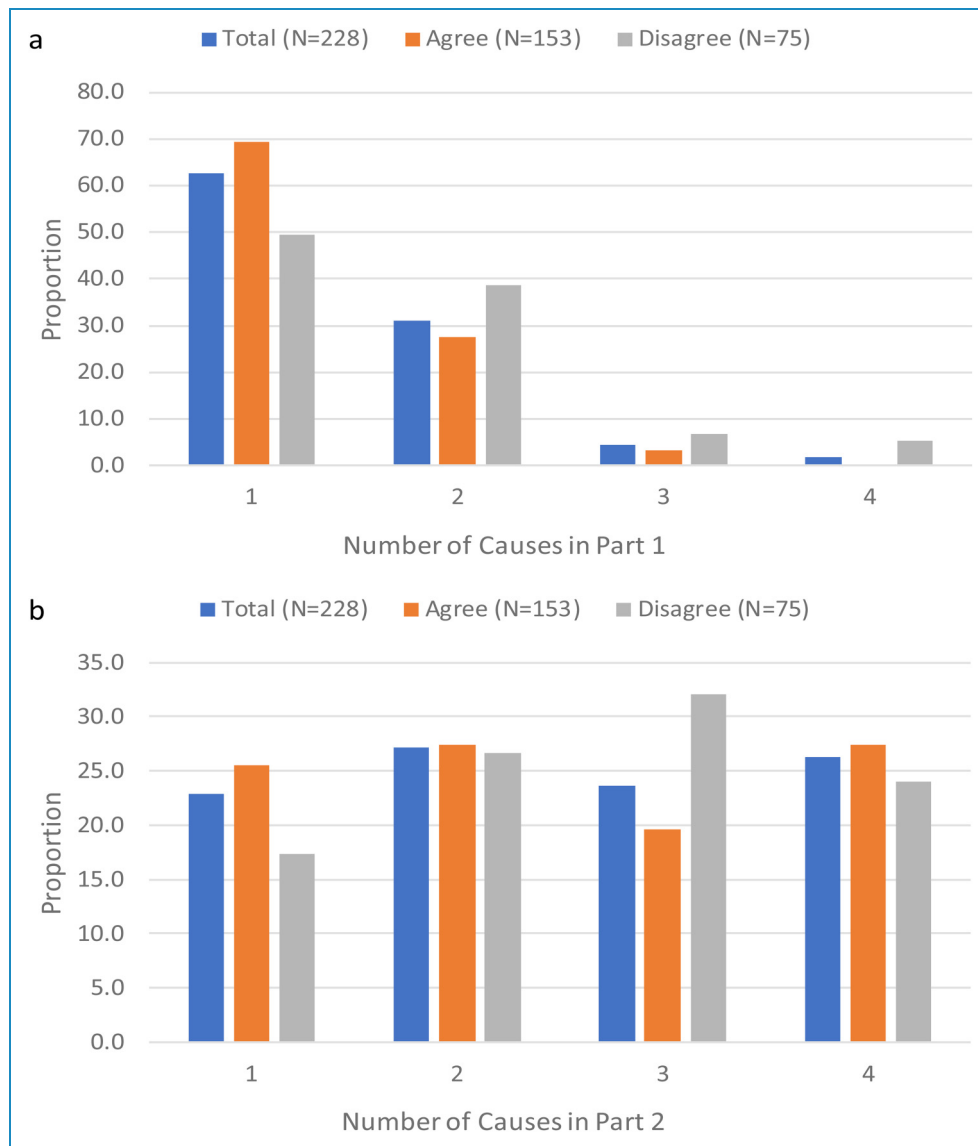


Figure 1. Agreements and disagreements in consensus process by number of causes assigned to (a) Part 1 and (b) Part 2. Strengthening mortality surveillance (SMS) for improved monitoring of HIV/AIDS, TB and COVID-19 in South Africa, September 2021.

to each case. The chance of a random error in CoD assignment occurring might increase as the number of physicians increase. This could be an important consideration for a routine system where physicians of various demographics and backgrounds would be needed and might suggest that duplicate coding would be necessary to guard against random error. The trade-off between accuracy, resource requirements (including physician workload and potential need for duplicate coding), and cost-effectiveness associated with various automated and machine learning assignment approaches needs careful consideration, particularly in resource-limited settings like South Africa.

In the context of economic constraints in South Africa, using VAs for CoD assignment of only home deaths might be preferable. However, in South Africa, there

were an estimated 493,602 deaths in 2020 after adjusting for completeness of deaths on the National Population register³² and an estimated 45%⁴ of these deaths occurred at home, suggesting that the implementation of VAs with physician coding in a routine system would require a large number of physician coders to review and assign a CoD. Considering the resource constraints, automated coding may offer a fast, cost-effective, and scalable approach to process such large volumes of data. While physician coding remains a valuable tool, automated coding algorithms warrant further exploration and evaluation, particularly to assess their suitability for integration into resource-limited routine CoD assignment systems. Advancements in AI hold the potential to further refine and improve the accuracy of these algorithms.³³ In addition

to addressing deaths that occur outside of a healthcare facility, teleVAs could also be used to address unreliable vital registration mortality data.

In South Africa, vital registration CoD data are blighted by various quality issues, including misclassified HIV/AIDS deaths, a high proportion of ill-defined causes, and a lack of manner of injury deaths.³⁴ The routine use of VAs could mitigate these issues, as demonstrated by the NCODV study that conducted in-person VAs on a nationally representative sample of respondents, which showed a much higher proportion of HIV/AIDS deaths compared to StatsSA data (28.9% vs. 4.9%).¹⁷ Similarly, Brazil has used VAs since 2005 to determine the underlying CoD in cases where death data sources are insufficient for investigating ill-defined causes.^{35,36} This method has contributed to a decrease of 68% in age-standardized ill-defined causes of death between 2000 and 2015. Automated and machine learning algorithms, when incorporated into VA analysis, have the potential to further improve data accuracy by identifying patterns and inconsistencies that may not be apparent to human coders.

Additionally, to establish a routine system, additional engagement is needed to ensure NoK are identified in advance, and aware of the VA process. Addressing this challenge could involve collaboration with the Department of Home Affairs to source contact details and enhance awareness of the VA process.

The relatively small sample size in this study is a limitation as it constrained the number of cases per CoD and limited the variability in different CoDs, such as specific cancer sites and injuries. This limitation may impact the generalizability and robustness of the study results, particularly regarding the variability within individual CoD categories. This study employed a convenience non-probability sampling approach. Given the time constraints for fieldwork and the time required to obtain contact information for home deaths identified by the WCPHDC, a convenience sampling approach was chosen to expedite data collection. Since readily available cases with contact information were included as they became available, the sample may not be fully representative of the target population (recent deaths outside healthcare facilities in the Western Cape). We used nine physicians to review the VAs which might have increased the chances of random error. The study is not nationally representative as it was conducted in only one metropolitan area. The VAs was conducted during the beta wave of COVID-19 (August to October/2021) when most people were confined to their homes which might have made telephonic interviews easier to conduct. In addition, the high prevalence of COVID-19 deaths (approximately 27% of cases) may limit the generalizability of the findings to situations with a more diverse cause-of-death profile. Future research exploring inter-coder agreement in a broader range of causes of death is needed.

The time elapsed between the deaths and telephonic VAs may have introduced recall bias. Interviews were conducted

between August and October 2021 for deaths occurring between December 2020 and September 2021, with a lag ranging from 16 to 290 days (average: 75 days). This extended timeframe could potentially affect the accuracy of NoK recollections regarding the decedents health and circumstances surrounding death.

Although this paper did not explicitly assess the feasibility of telephonic VAs, the response rate highlights a potential limitation. Refusals accounted for 10.2% alongside challenges reaching participants due to unanswered calls (45.1%) and invalid numbers (27.7%). Future studies designed to assess the feasibility of telephonic VAs in similar contexts might explore strategies to improve contact rates and address potential participant concerns. In addition, further studies exploring how teleVA data quality impacts the performance of automated coding algorithms would be valuable, particularly in resource-limited settings.

Conclusion

This study offers valuable insights into the agreement by physician coders in the assignment of CoDs for teleVAs in a South African context. While overall agreement on CoD determination was encouraging, improvements can be achieved by addressing challenges in cases with comorbidities and multiple sequelae. Research and refinement of VA methodologies are needed to ensure reliable CoD determination in routine practice. To make a more informed decision about the routine use of teleVAs in South Africa, a larger study, representative of the national population, is needed. Such research would provide a better idea of the applicability and benefits of teleVAs in improving the accuracy of CoD determination in the country's vital registration system.

Acknowledgments: The authors acknowledge the project team who assisted with logistics and fieldwork monitoring of the pilot study and the fieldwork team who collected data.

Contributorship: Study conception and design were done by IN, DM, TG, NK, CR, DB and PG. Data collection was performed by IN, TG, NK, DB and PG. Analysis and interpretation of results were done by IN, DM, TG, NK, CR, DB and PG. Draft manuscript preparation was done by IN, DM and PG. All authors have reviewed and accepted the final version of the paper and given their permission for publication.

Data accessibility statement: The data is available from the corresponding author upon request and subject to permission from the funders. The anonymized data collected during this study was subject to consent from participants.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (SAMRC Human Research Ethics Committee, EC005-2/2021) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: President's Emergency Plan for AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention (CDC) under the terms of cooperative agreement GH002193 (grant number CDC-RFA-GH19-1911-03CONT21/INU2GGH002193-03-01).

Guarantor: As the corresponding author, IN takes full responsibility for the article's accuracy and the reference list's appropriateness.

ORCID iD: Ian Neethling  <https://orcid.org/0000-0003-2778-4995>

Supplemental material: Supplemental material for this article is available online.

References

- Lopez AD and Setel PW. Better health intelligence: a new era for civil registration and vital statistics? *BMC Med* 2015; 13: 73.
- Joubert J, Rao C, Bradshaw D, et al. Evaluating the quality of national mortality statistics from civil registration in South Africa, 1997–2007. *PLoS One* 2013; 8: e64592.
- Wyk VP-V, Laubscher R, Msemburi W, et al. *Second South African National burden of disease study: data cleaning, validation and SA NBD list*. Cape Town: Burden of Disease Research Unit, South African Medical Research Council, Report no. 978-1-920618-17-9, 2014.
- Statistics South Africa. *Mortality and causes of death in South Africa: findings from death notification 2017*. Pretoria: StatsSA, 2020.
- World Health Organization. *Verbal autopsy standards: the 2014 WHO verbal autopsy instrument*. Geneva: WHO, 2015.
- Soleman N, Chandramohan D and Shibuya K. Verbal autopsy: current practices and challenges. *Bull WHO* 2006; 84: 239–245.
- D'Ambrosio L, Price J, Cowan E, et al. Refining circumstances of mortality categories (COMCAT): a verbal autopsy model connecting circumstances of deaths with outcomes for public health decision-making. *Glob Health Action* 2021; 14: 2000091.
- Alberts M, Dikotope SA, Choma SR, et al. Health & demographic surveillance system profile: the dikgale health and demographic surveillance system. *Int J Epidemiol* 2015; 44: 1565–1571.
- Shibuya K and Gilmour S. Civil registration as a means to promote human security. *Lancet* 2015; 386: 1311–1418.
- Byass P, Fottrell E, Dao LH, et al. Refining a probabilistic model for interpreting verbal autopsy data. *Scand J Public Health* 2006; 34: 26–31.
- Jha P, Kumar D, Dikshit R, et al. Automated versus physician assignment of cause of death for verbal autopsies: randomized trial of 9374 deaths in 117 villages in India. *BMC Med* 2019; 17: 16.
- Leitao J, Desai N, Aleksandrowicz L, et al. Comparison of physician-certified verbal autopsy with computer-coded verbal autopsy for cause of death assignment in hospitalized patients in low- and middle-income countries: systematic review. *BMC Med* 2014; 12: 20140204.
- Chandramohan D, Fottrell E, Leitao J, et al. Estimating causes of death where there is no medical certification: evolution and state of the art of verbal autopsy. *Glob Health Action* 2021; 14: 1982486.
- Garenne M and Fauveau V. Potential and limits of verbal autopsies. *Bull WHO* 2006; 84: 164.
- Joshi R, Lopez AD, MacMahon S, et al. Verbal autopsy coding: are multiple coders better than one? *Bull WHO* 2009; 87: 51–57.
- Groenewald P, Neethling I, Bradshaw D, et al. *Strengthening mortality surveillance (SMS) for improved monitoring of HIV/AIDS, TB and COVID-19 in South Africa: TeleVA pilot workshop report 22 November 2021*. Cape Town: South African Medical Research Council, 2022.
- Bradshaw D, Joubert JD, Maqungo M, et al. *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.
- Sant Fruchtmann C, Kabadi G, Mwanza J, et al. *Telephone verbal autopsy (TeleVA) for cause of death estimation in CRVS systems where there is no doctor - a proof of concept*, <https://www.researchgate.net/publication/372891607> (2019, accessed 2 February 2023).
- Nasaruddin NH, Ganapathy SS, Awaluddin SM, et al. Conducting verbal autopsy by telephone interview during the pandemic to support mortality surveillance: a feasibility study in Malaysia. *West Pac Surveill Resp J* 2022; 13: 1–7.
- Dutwin D and Lavrakas P. Trends in telephone outcomes, 2008–2015. *Surv Pract* 2016; 9: 1–9. DOI: 10.29115/SP-2016-0017.
- Nichols EK, Byass P, Chandramohan D, et al. The WHO 2016 verbal autopsy instrument: an international standard suitable for automated analysis by InterVA, InSilicoVA, and tariff 2.0. *PLoS Med* 2018; 15: e1002486.
- World Health Organization. Verbal autopsy sample questionnaires v.1.4., <http://www.who.int/healthinfo/statistics/verbalautopsystandards/en/>. (Accessed 22 May 2021) (2016, accessed 22 May 2021).
- Lakshminarasimhappa M. Web-based and smart mobile app for data collection: kobo toolbox/kobo collect. *J Indian Libr Assoc* 2022; 57: 72–79.
- World Health Organization. *International classification of diseases and related health problems. 10th revision, volume 2. Instruction manual*. Geneva: World Health Organization, 2016.
- SAMRC. National cause-of-death validation, <https://www.samrc.ac.za/research/intramural-research-units/BOD-national-cause-death> (accessed 02 March 2024).
- World Health Organization. *Medical certification of cause of death. Instructions for physicians on use of international form of medical certificate of cause of death*. Geneva: World Health Organization, 1979.

27. Johansson L, Pavillon G, Pelikan L, et al. *Iris automated coding system for causes of death: User's reference manual. Version 4.1.3*. 2012. Cologne, Germany: German Institute of Medical Documentation and Information (DIMDI): IRIS Institute.
 28. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.
 29. World Health Organization. *WHO methods and data sources for global burden of disease estimates 2000–2019*. Geneva: World Health Organization, 2020. https://cdn.who.int/media/docs/default-source/gho-documents/global-health-estimates/ghe2019_daly-methods.pdf?sfvrsn=31b25009_7 (Accessed 30 October 2023).
 30. World Health Organization. *Special tabulation lists for mortality and morbidity; mortality tabulation list 1. International statistical classification of diseases and health related problems - tenth revision volume 1*. Geneva: World Health Organization, 1993.
 31. Institute of Health Metrics and Evaluation. Global health data exchange. Global Burden of disease study 2019 cause list mapped to ICD codes, <https://ghdx.healthdata.org/record/ihme-data/gbd-2019-cause-icd-code-mappings> (accessed 29 October 2023) (2019, accessed 29 October 2023).
 32. Dorrington RE, Bradshaw D, Laubscher R, et al. *Rapid mortality surveillance report 2019 & 2020*. Cape Town: South African medical Research council, 2021. ISBN: 978-1-928340-58-4.
 33. Jeblee S, Gomes M, Jha P, et al. Automatically determining cause of death from verbal autopsy narratives. *BMC Med Inform Decis Mak* 2019; 19: 127.
 34. Bradshaw D, Pillay-Van Wyk V, Laubscher R, et al. *Cause of death statistics for South Africa: challenges and possibilities for improvement* 2012. Cape Town: South African Medical Research Council.
 35. França E, Ishitani L, Teixeira R, et al. Changes in the quality of cause-of-death statistics in Brazil: garbage codes among registered deaths in 1996–2016. *Popul Health Metr* 2020; 18: 20.
 36. Teixeira R, Naghavi M, Guimarães M, et al. Quality of cause-of-death data in Brazil: garbage codes among registered deaths in 2000 and 2015. *Bras J Epidemiol* 2019; 22: e19002.
-