

Monitoring progress with national and sub-national health goals by integrating verbal autopsy and medically certified cause of death data

SUPPLEMENTARY FILE

ASSESSING THE CHARACTERISTICS AND QUALITY OF MCCOD AND VA DATA

The percentage of hospital and community deaths occurring in each age group needs to be assessed for the extent to which it follows the expected pattern of hospital deaths comprising a higher proportion of deaths at younger ages (e.g. neonatal deaths) than community deaths, which commonly comprise more older deaths.

The quality of each data source can be assessed using existing tools to identify weaknesses in the data that may bias CSMFs and which can inform remedial actions. For VA data derived using the SmartVA tool, this assessment can be undertaken using the Verbal Autopsy Interpretation, Performance and Evaluation Resource (VIPER) tool, which comprises a step-by-step approach that assesses different types of data quality.¹ For other VA outputs the assessment can be done by following the steps in the associated Guidelines document.² VIPER measures the completeness of VA reporting (i.e. the proportion of community deaths for which a VA was conducted) by using the empirical completeness method.³ The lower the completeness of VA reporting, then the less accurate that CSMFs will be because CSMFs of unreported deaths are likely to be different from reported deaths, especially if there are socio-economic or geographic differences in the two groups of deaths. A previous study found that at 70% completeness of death reporting, the accuracy of CSMFs is 77%; this declines to 57% where completeness is 50%.⁴ Where completeness of VA death reporting is below 50%, then the bias will be too great and the integration should not proceed. Another data quality issue identified by VIPER is the proportion of VAs for which there are 'undetermined' causes of death (where there is insufficient information in the VA interview responses to reliably estimate a cause of death). The higher the proportion of VAs with undetermined causes, the greater the bias in CSMFs; where more than approximately 20% of VAs have an undetermined cause, the data are likely to be too poor for the integration to proceed. Finally, VIPER also helps understand the plausibility of VA CSMFs and how they differ from expected levels and patterns given the demographic and epidemiological characteristics of the population.

The quality of MCCOD data can be assessed using the ANALYSIS of Causes of National Deaths for Action (ANACONDA) tool.⁵ This will identify the level of "unusable or insufficiently specified" causes (known as "garbage codes"), including the proportion that are of each level of severity, and the specific causes they comprise.⁶ For the purposes of the integration exercise, the percentage of MCCODs with a garbage code that is in either severity levels very high, high or medium should be assessed. One type of garbage code is R00-R99 (symptoms, signs and ill-defined conditions) in the ICD-10 classification; these are mapped to undetermined causes in SmartVA and not to any specific causes. Where a high proportion of MCCODs have a cause of death of poor quality, then the CSMFs will be biased. Where deaths with ICD codes R00-R99 comprise 15% or more of deaths with a MCCOD, then the integration should not proceed.

Whilst poor quality VA or MCCOD will compromise the quality of the data (and its use) when integrated, this assessment can also be used to inform training to improve data quality. For instance, a high proportion of undetermined causes in VA data may mean there are issues with the VA data

collection and the need for improved training and supervision of the data collection process. The ANACONDA assessment can be used to inform training to improve physicians' certification of causes of death and the ICD coding of causes of death.

References

1. Verbal Autopsy Interpretation, Performance and Evaluation Resource (VIPER) [program]. Melbourne: Bloomberg Philanthropies Data for Health Initiative, Civil Registration and Vital Statistics Improvement, the University of Melbourne, 2020.
2. D4H Technical Working Group. Guidelines for interpreting verbal autopsy data. CRVS Resources and Tools. Carlton, VIC: University of Melbourne, Civil Registration and Vital Statistics Improvement, Bloomberg Philanthropies Data for Health Initiative, 2020.
3. Adair T, Lopez AD. Estimating the completeness of death registration: An empirical method. *PLoS One* 2018;13(5):e0197047.
4. Mikkelsen L, Phillips DE, AbouZahr C, et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *Lancet* 2015;386(10001):1395-406.
5. Mikkelsen L, Moesgaard K, Hegnauer M, et al. ANACONDA: a new tool to improve mortality and cause of death data. *BMC Med* 2020;18(1):61. doi: 10.1186/s12916-020-01521-0 [published Online First: 2020/03/10]
6. Naghavi M, Richards N, Chowdhury H, et al. Improving the quality of cause of death data for public health policy: are all 'garbage' codes equally problematic? *BMC Med* 2020;18(1):55. doi: 10.1186/s12916-020-01525-w [published Online First: 2020/03/10]

Table S1: Mapping of SmartVA cause list to ICD-10 codes

Adult Cause	ICD-10 Code
Diarrhea/Dysentery	A00-A09
TB	A15-A19
AIDS	B20-B24
Malaria	B50-B54
Other Infectious Diseases	A10-A14, A20-B19, B25-B49, B55-B99, U07.1
Esophageal Cancer	C15
Stomach Cancer	C16
Colorectal Cancer	C18-C21
Lung Cancer	C34
Breast Cancer	C50
Cervical Cancer	C53
Prostate Cancer	C61
Leukemia/Lymphoma	C81-C85; C91-C96
Other Cancers	C00-C14, C17, C22-C33, C35-C49, C51-C52, C54-C60, C62-C80, C86-C90, C97-D48
Diabetes	E10-E14
Other Cardiovascular Diseases	I00-I19 I26-I59, I70-I99
Ischaemic Heart Diseases	I20-I25
Stroke	I60-I69
Pneumonia	J10-J22,J85
Chronic Respiratory diseases (COPD/Asthma)	J40-J46
Cirrhosis	K70-K76
Chronic kidney	N17-N19
Maternal	O00-O99
Undetermined	R00-R99
Road Traffic	V01-V89
Falls	W00-W19
Drowning	W65-W74
Fires	X00-X19
Bite of Venomous Animal	X20-X29
Poisonings (accidental)	X40-X49
Suicide (intentional self-harm)	X60-X84
Homicide (assault)	X85-Y09
Other Injuries	S00-T98, V90-V99, W20-W64, W75-W99, X30-X39, X50-X59, Y10-Y98
Other Non-communicable Diseases	All other ICD-10 codes

Bold: Measured in SDG Indicator 23.

Table S2: Calculation of probability of dying between exact ages 30 and 70 from any of cardiovascular disease, cancer, diabetes, chronic respiratory disease, by sex, 42 townships, Myanmar, 2019

Age group	VA CSMF				Number of VAs	MCCOD CSMF				Number of MCCODs
	Cardiovascular diseases	Cancers	Chronic respiratory	Diabetes		Cardiovascular diseases	Cancers	Chronic respiratory	Diabetes	
Males										
30-34	0.122	0.025	0.018	0.027	1,017	0.097	0.005	0.003	0.008	393
35-39	0.150	0.031	0.019	0.029	1,381	0.122	0.016	0.006	0.010	492
40-44	0.150	0.046	0.019	0.034	1,602	0.159	0.026	0.014	0.009	428
45-49	0.219	0.064	0.021	0.035	1,799	0.217	0.027	0.016	0.008	368
50-54	0.279	0.084	0.055	0.074	1,840	0.263	0.054	0.017	0.026	350
55-59	0.304	0.103	0.062	0.065	1,994	0.261	0.066	0.006	0.025	318
60-64	0.374	0.099	0.083	0.067	2,163	0.352	0.064	0.037	0.045	267
65-69	0.427	0.110	0.099	0.055	2,185	0.330	0.069	0.050	0.041	218
Females										
30-34	0.248	0.126	0.011	0.078	270	0.097	0.029	0.010	0.049	103
35-39	0.203	0.180	0.033	0.057	300	0.165	0.035	0.009	0.026	115
40-44	0.238	0.181	0.024	0.057	453	0.218	0.056	0.008	0.032	124
45-49	0.257	0.196	0.033	0.096	721	0.200	0.055	0.027	0.064	110
50-54	0.316	0.200	0.063	0.105	909	0.230	0.048	0.018	0.121	165
55-59	0.351	0.180	0.065	0.126	1,158	0.249	0.092	0.040	0.075	173
60-64	0.387	0.136	0.090	0.120	1,574	0.349	0.073	0.052	0.109	192
65-69	0.437	0.105	0.115	0.103	1,858	0.318	0.061	0.034	0.135	148