Automatic assignment of underlying cause of death based on verbal autopsy instrument

Lucia Pereira Barroso, Carmen Saldiva, Paulo Saldiva
University of São Paulo

Abstract

Many of the deaths in low or middle income countries occur without medical attention, especially in rural areas. Verbal Autopsy (VA) are being used to provide information on cause of death. The verbal autopsy questionnaire has a reduced version, which is being validated in Brazil for 22 causes of death in adults. This questionnaire has been applied previously in other countries in a visit to the family of the deceased few months after the death. In Brazil, the application is being made at the Death Verification Service of São Paulo (SVOC), shortly after the death of the individual. The attribution of the underlying cause of death can be made based on the answers given by the family, using an automatic method based on the Tariff Score, applied in SmartVA software, and using weights obtained in a study previously developed in other countries. The Tariff Method was developed by the Population Health Metrics Research Consortium (PHMRC). The SmartVA software was developed by the Institute for Health Metrics and Evaluation. The objectives of this study are to present the automatic method of assigning the underlying cause of death based on a verbal autopsy instrument and to show preliminary results.

Keywords

Verbal autopsy; Tariff Method; SmartVA; cause of death

Introduction

Many of the deaths in low- or middle-income countries occur without medical attention, especially in rural areas. Verbal Autopsy (VA) are being used to provide information on cause of death. The verbal autopsy questionnaire has a reduced version, which is being validated in Brazil for 22 causes of death in adults. This questionnaire has been applied previously in other countries in a visit to the family of the deceased few months after the death. In Brazil, the application is being made at the Death Verification Service of São Paulo (SVOC), shortly after the death of the individual.

The attribution of the underlying cause of death can be made based on the answers given by the family, using an automatic method based on the Tariff Score, applied in SmartVA software, and using weights obtained in a study previously developed in other countries. The Tariff Method was

developed by the Population Health Metrics Research Consortium (PHMRC), (see Murray et al., 2011). The SmartVA software was developed by the Institute for Health Metrics and Evaluation (see Serina et al., 2015).

The objectives of this study are to present the automatic method of assigning the underlying cause of death based on a verbal autopsy instrument and to show preliminary results.

Tariff Score

The Tariff-score was calculated based on information collected from 12,501 individuals in the University of Washington's Population Health Metrics Research Consortium (PHMRC) in 6 cities in 4 countries - India (Andhra Pradesh and Uttar Pradesh), the Philippines (Bohol) , Mexico (Mexico) and Tanzania (Dar es Salaam and Pemba Island).

The questionnaire contains open and closed questions about Closed questions:

- ✓ Symptoms of terminal illness;
- ✓ Diagnosis of chronic diseases - records of previous health services;
- ✓ Risk behavior (alcohol, tobacco);
- ✓ Details of any interaction with health services.

Open questions: open narrative by the relatives through which one can identify words or groups of words associated with the cause of death. The underlying cause of death was known based on medical records and the authopsy made in the SVOC (gold standard). The input worksheet contains values 0 and 1, being 1 when the observed symptom occurred for a specific individual. With this, it is possible to calculate the endorsement rate, given by:

x_{ij} = fraction of VAs for which there is a positive response to deaths from cause i for item j

The Tariff is the endorsement rate standardized by the median and interquartile range of all causes of death, fixed a symptom. For cause *i*, symptom *j* is given by

$$Tariff_{ij} = \frac{x_{ij} - median(x_{ij})}{IIQ(x_{ij})}$$

where median(x_{ij}) is the median fraction with a positive response for item j across all causes, and IIQ(x_{ij}) is the interquartile range of positive response rates averaged across causes. Note that Tariff can assume positive or negative values. As a final step, Tariffs are rounded to the nearest 0.5 to avoid overfitting and to improve predictive validity.

For the calculation of the Tariff-Score of case *k*, cause *i*, the 40 symptoms with the highest absolute value of the Tariff are considered

$$TS_{ki} = \sum_{r=1}^{40} Tariff_{ir} I_{kr}$$

where I_{kr} is the response for death *k* on item *j*, taking on a value of 1 when the response is positive and 0 when the response is negative.

Assignment of underlying cause of death

The assignment of the underlying cause of death to a new individual (k) is made on the basis of the Tariff-ordered score for each cause. The simplest idea would be to assign the cause for which the Tariff-Score was the largest. However, some causes have a naturally higher score and then classification is based on ranks. The Tariff Score of individual k is calculated for each of the possible causes.

The original sample is sorted from lowest to highest value. The Tariff-score of the new individual is included in this order for each possible cause. The new individual is assigned the cause that presents the highest rank.

For this analysis we consider the chapter classification of ICD10 - International Statistical Classification of Diseases (World Health Organization - WHO). Table 1 shows the counts of cases classified according to the gold standard and by the Tariff method.

Table 1: Classification: Gold Standard (row) *versus* Tariff Method (column)

Cause	BC	AIDS	Stroke	DB	IHD	CRD	OD	Total
BC	10	0	0	0	0	0	4	14
AIDS	1	7	0	1	0	0	1	10
Stroke	3	1	44	5	7	3	13	76
DB	0	1	18	30	17	1	11	78
IHD	1	0	98	44	150	20	101	414
CRD	0	0	16	5	10	20	20	71
OD	10	3	90	63	125	12	296	599
Total	25	12	266	148	309	56	446	1262

BC = Breast Cancer
DB = Diabetes
IHD = Ischemic Heart Disease
CRD = Chronic Respiratory Diseases
OD = Other Diseases

Conclusion

- ✓ The most common underlying cause is Ischemic Heart Disease, totaling 414 cases out of 1262 observed, that represents, 32.8%.
- ✓ Considering the 414 cases of Ischemic Heart Disease, 150 were classified correctly by the Tariff method, which means 36.20% accuracy.
- ✓ Considering 633 cases (excluding Other Diseases), 261 cases were correctly classified, 39.4%.

- ✓ The percentages of correct classification are: Breast Cancer (71.4%), AIDS (70.0%), Stroke (57.9%), Diabetes (38.5%), Ischemic Heart Disease (36.2%), Chronic Respiratory Diseases (28.2%).

References

1. Serina P, Riley I, Stewart A, James SL, et al. Improving performance of the Tariff Method for assigning causes of death to verbal autopsies. BMC Medicine. 2015 Dec 8; 13:291.
2. Murray CJL, Lopez AD, Black R, Ahuja, et al. Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets. Population Health Metrics. 2011 Aug 4; 9:27.

Acknowledgment

This study is supported by grant #2013/21728-2, São Paulo Research Foundation (FAPESP) and grant 815781/2014, Ministry of Health Brazil.

1) University of São Paulo
2) Sírio-Libanês Hospital
3) Ministry of Health Brazil
4) Secretary of Health, São Paulo
5) Federal University of Minas Gerais
6) Federal University of Rio Grande do Norte