Verbal autopsy coding: are multiple coders better than one?
Rohina Joshi, Alan D Lopez, Stephen MacMahon, Srinath Reddy, Rakhi Dandona, Lalit Dandona & Bruce Neal

Objective To assess the impact on the reported cause-of-death patterns of a verbal autopsy coding strategy based on a review of every death by multiple coders versus a single coder.

Methods Deaths in 45 villages (total population 180 162) in southern India were documented during 12 months in 2003–2004, and a standard verbal autopsy questionnaire was completed for each death. Two physician coders, each unaware of the other’s decisions, assigned an underlying cause of death in accordance with the causes listed in the chapter headings of the International classification of diseases and related health problems, 10th revision (ICD-10). For the three chapter headings that applied to more than 100 of the deaths, agreement for subsets of causes of death within the chapter was also analysed. In the event of discrepancies, a third coder was used to finalize a cause of death. Cohen’s kappa statistic (K) was used to measure levels of agreement between the two physician coders.

Findings In total, 1354 deaths were documented, and a verbal autopsy was completed for 1329 (98%) of them. At the chapter heading level of the ICD-10, physician coders assigned the same cause to 1255 deaths (94%) (K = 0.93; 95% confidence interval: 0.92–0.94). The patterns of death derived from the causes assigned by each physician were all very similar to the patterns obtained through the consensus process, with the rank order of the 10 leading causes of death being the same for all three coding methods.

Conclusion Duplicate coding of verbal autopsy results has little advantage over a single-coder system for mortality surveillance or for identifying population patterns of death. Resources could be better diverted to other parts of the mortality surveillance process, such as validation.

Introduction

Most high-income countries have well-established mortality reporting systems with complete recording of deaths and medical certification for more than 90% of them.¹ The information about the pattern of death in the population that is provided by these systems is important for planning health service provision.²³ However, most low-income countries lack systems to document mortality rates and causes of death. Most deaths occur at home,² many go unreported and information about the cause of death is typically absent or unreliable.²⁴ In India, about 7 million of the 9.5 million deaths that occur annually happen at home and fewer than half have a medically certified cause of death.⁵

Mortality surveillance systems based on verbal autopsy are a key source of data about causes of death in less developed countries such as India where limited resources are available for the vital registration process.⁶⁻⁷ The verbal autopsy method relies on information gathered from a standardized interview with a relative or caretaker; trained coders review the data and apply standardized diagnostic algorithms to arrive at a cause of death.⁸⁻¹⁰

The usual practice for cause of death assignment has been to either use a panel of expert physicians or to have two or more physician coders independently review the data and arrive at a diagnosis.¹¹⁻¹³ In the latter case, any discrepancies between the two diagnoses are resolved by consultation or through review of the verbal autopsy questionnaire by a third physician.¹⁴ There has been debate about the impact of different methods of cause of death assignment, but in the absence of empirical data no consensus has been reached.¹⁴

In this paper we compare verbal autopsy strategies using a single coder versus multiple coders and report on the impact of each method on the reported patterns of death in an Indian community.

Methods

This study was conducted as part of a research collaboration (the Andhra Pradesh Rural Health Initiative) between five Australian and Indian institutions (see the acknowledgements section). The data were collected between 1 October 2003 and 30 September 2004, and the main causes of death have been previously reported.¹⁵ Ethics approval for the project was received from the ethics committees of the CARE Hospital, Hyderabad, India, and the University of Sydney, Australia. Written informed consent was obtained from each respondent before the collection of any data, and the project was designed and conducted in line with the Declaration of Helsinki and its subsequent amendments. For participants who could not read or write, the participant information sheet and consent form were explained by a multipurpose primary healthcare worker (MPHW), and a thumb impression was taken instead of a signature. An MPHw is a female non-physician health worker who provides basic health care to the community served by a village health centre. Her role is to collect vital statistics, provide primary health care to the villagers and assist the visiting doctor during clinic hours in the village health centre. In this study, the MPHW carried out the
mortality surveillance work in addition to regular clinic activities.

Population
This project was conducted in 45 villages in East and West Godavari districts in the state of Andhra Pradesh in southern India. The population (N = 180,162) age and sex structure was defined by a population census conducted in 2002–2003. The age distribution of the population in the villages is characteristic of that found in places where fertility has recently decreased, with relatively low proportions of the population in the very young and very old age groups. A quarter of the population is below 15 years of age, and a tenth is older than 60 years. Most people are engaged in work related to agriculture or aquaculture, and the average household income is 2000.00 Indian rupees (Rs) (US$ 50.78) a month. The literacy level of the adult population 30 years of age and above is 54%.

Identification of deaths
Each MPHW had primary responsibility for identifying deaths in a village. Identification of deaths by the MPHW was facilitated by her daily contact with the villagers and a network of key informants, including the village headman, the panchayat (that is, the village governing body responsible for registration of deaths), priests, cremation staff and other community leaders. The completeness of identification of deaths was checked between 25 April 2005 and 30 May 2005 by the field supervisor or the MPHWs by visiting every house in every village and checking with residents that all deaths between 1 October 2003 and 30 September 2004 had been recorded.

Data collection
For each death recorded, the MPHW sought to visit the deceased person’s household within a month of the date of death. The family member or other caretaker best able to report on the events preceding the death was identified, and a systematic enquiry into the events leading up to the death was made using a semi-structured verbal autopsy tool in accordance with an established technique. The verbal autopsy tools used in this project were based on validated verbal autopsy tools used in studies in China, the United Republic of Tanzania and the Registrar General of India’s Sample Registration System, with modifications made to suit local terminology.

Separate questionnaires were used for deaths in each of three age groups (0–28 days, 29 days to < 15 years and ≥ 15 years) and all included a series of structured questions and an open narrative section. The open narrative section was completed with the aid of a defined symptom list with specific questions about treatments, medical procedures and associated documentation. The MPHWs were trained in data collection before the study commenced with refresher training provided every 6 months.

Cause of death assignment
Cause of death was assigned using validated materials and processes developed for the Registrar General of India’s Sample Registration System. In a system henceforth referred to as the consensus process, every verbal autopsy was assessed independently by two trained physician coders, each of whom assigned an underlying cause of death. Physician coders were unaware of the decisions about cause of death made by their colleagues. Immediate and contributory causes were assigned wherever possible. Causes of death were selected from the International classification of diseases and related health problems (ICD), 10th revision (ICD-10). Assignment of the causes of death by the physicians was facilitated by a series of algorithms developed for the Sample Registration System. In the event of disagreement between the two physicians as to the underlying causes of death, a third physician reviewed the evidence and decided on the underlying cause. Resolution of discrepant decisions was done throughout the study and at the earliest available opportunity.

Outcomes
The main outcome for this study was the proportion of deaths with underlying causes corresponding to those corresponding to chapter headings in the ICD-10. For the three chapter headings that included more than 100 deaths, we also conducted analyses for subsets of causes of death within that chapter. As a result, stroke, coronary heart disease, heart failure and other deaths from cardiovascular causes were analysed separately for the chapter on cardiovascular causes; suicide, falls, transport injuries and other external causes of injury-related deaths were analysed separately for the chapter on injury-related causes; and intestinal infections, tuberculosis, HIV infection and other infectious diseases were analysed separately for the chapter on infectious diseases.

We estimated the primary cost of the surveillance system, which included training meetings for MPHWs and physician coders, a salary component for the MPHWs to cover data collection (about 1 day per week), the salary cost for the project coordinator (employed at 0.5 of a full-time equivalent) and the field coordinator (employed at 0.5 of a full-time equivalent), administrative costs and a unit cost for each death reviewed by physician coders (Rs 250.00 or US$ 5.50).
Analysis

The level of agreement between the two physician coders for the underlying causes of death at the chapter heading level of the ICD-10 was calculated using Cohen’s kappa statistic (K). Values above 0.75 were considered to reflect good agreement; those between 0.40 and 0.75, fair to good agreement; and those below 0.40, poor agreement. To compare the population pattern of underlying cause of death as assigned by each physician coder with the pattern derived from the consensus process we prepared graphical presentations of the proportions of deaths assigned to each underlying cause and/or ranked tabulations of these data. The proportions of deaths attributed to each underlying cause were calculated by dividing the number of deaths attributed to a particular cause by the total number of deaths for which a verbal autopsy was performed, with the results expressed as a percentage. Unclassifiable deaths (ICD-10 “R” code) were included as a separate category. Analyses were carried out using SPSS, version 12 (SPSS Inc., Chicago, IL, United States of America).

Costs were estimated by multiplying the various components and adding to obtain subtotals for training, data collection, cause of death assignment and project management. A separate projection of the savings that might be achieved with a single coding strategy was made by subtracting the costs associated with one physician coder and one third of the project management costs. The project management costs included photocopying and couriering of the verbal autopsy reports and the salaries of one half-time field coordinator and one half-time senior project manager. Thus, having single coding would decrease the costs of photocopying and courier by half and the salary costs by a third.

Results

Over the 12-month study period, 1354 deaths were documented in the 45 villages. A verbal autopsy was completed for 1329 (98%) of these deaths, and an underlying cause of death was assigned for 1084 of them (82% of all verbal autopsies). The crude death rate was 7.5 per 1000 population (95% confidence interval, CI: 7.1–7.9), and the death rate for males exceeded that for females in all age groups except that of children 4 years of age or under. The top five causes of death were diseases of the circulatory system (32.4%), external causes (13.3%), infections (11.8%), neoplasms (7.3%) and respiratory diseases (5.3%).

Inter-coder agreement

Physician coders assigned the same underlying causes of death at the chapter-heading level of the ICD-10 for 1255 (94%) of the 1329 deaths for which a verbal autopsy was performed. These 1329 deaths represented 98% of the total; the remaining 2% could not be located. The kappa statistic for overall inter-coder agreement was 0.93 (95% CI: 0.92–0.94). The following were the kappa statistics for inter-coder agreement for the four main subdivisions of each chapter: cardiovascular diseases (K = 0.87; 95% CI: 0.83–0.91) (Fig. 1); external causes of injury (K = 0.94; 95% CI: 0.90–0.98) (Fig. 2); infectious diseases (K = 0.96; 95% CI: 0.90–0.98) (Fig. 3). For inter-coder agreement on the causes of death assigned in different age groups, the following K were obtained: 1.0 for the 11 deaths in children 0–28 days old; 0.65 for the 68 deaths in children 29 days to 14 years old; and 0.93 for the 1250 deaths in individuals 15 years old or above.

Single versus multiple coders

The patterns of the causes of death assigned by each of the physician coders were similar, and the pattern in the causes assigned by both was practically the same as that derived from the consensus process (Table 1). The rank order of the causes of death assigned in each case was identical for the 10 lead-
Table 1. Inter-physician agreement for the underlying cause of death* in a study of verbal autopsies in villages in Andhra Pradesh, India, 2003–2004

<table>
<thead>
<tr>
<th>Cause</th>
<th>Physician coder 1</th>
<th>Physician coder 2</th>
<th>Final consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perinatal (P00-P96)</td>
<td>14a</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Congenital (Q00-Q99)</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Blood (D50-D89)</td>
<td>13</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Circulatory system (I00-I99)</td>
<td>421</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Digestive system (K00-K93)</td>
<td>60</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Genitourinary system (N00-N99)</td>
<td>38</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>Respiratory system (J00-J99)</td>
<td>66</td>
<td>1</td>
<td>70</td>
</tr>
<tr>
<td>Skin and subcutaneous tissue (L00-L99)</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Endocrine disorders (E00-E99)</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Infectious diseases (A00-B99)</td>
<td>151</td>
<td>5</td>
<td>157</td>
</tr>
<tr>
<td>External causes of injury (S00-T98, V01-Y98)</td>
<td>179</td>
<td>3</td>
<td>182</td>
</tr>
<tr>
<td>Mental and behavioural (F00-F99)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Neoplasms (C00-D48)</td>
<td>93</td>
<td>3</td>
<td>98</td>
</tr>
<tr>
<td>Pregnancy (O00-O99)</td>
<td>9</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Not elsewhere classified (R00-R99)</td>
<td>202</td>
<td>228</td>
<td>245</td>
</tr>
</tbody>
</table>

a Number finally assigned to each cause after discrepancies in diagnosis between physician coder 1 and physician coder 2 were reviewed by a third physician.

b The diagonal series of numbers in bold typeface indicate agreement reached independently by the two physicians.

Costs incurred

Once the materials were developed, the main costs involved in establishing and running the surveillance system included training (of interviewers and physician coders), payment to health workers for data collection and to physicians for the coding of verbal autopsies, and project management. Although a formal economic analysis was not performed, we were able to estimate the costs of the surveillance system retrospectively (Table 2). Had there been one instead of two coders, the costs for coding would have been reduced by Rs 169 500 (US$ 3730), and there would have been an additional saving on project management fees of about Rs 103 662 (US$ 2281). Overall, we estimate that use of a single coding system would have reduced the total costs for the one-year surveillance system by about 31% from Rs 891 773 (US$ 19 625) to Rs 618 611 (US$ 13 614).

Discussion

Our findings suggest that verbal autopsy mortality surveillance systems that use trained physician coders gain little from duplicate coding of causes of death. Although there were some discrepancies in the causes of death assigned to individual cases, they had little impact on the overall pattern of reported mortality. This could be either because there were few errors in assigning deaths to the broad causes or because the duplicate coding process failed to identify any errors made. Either way, there seems to be little support for a system of duplicate coding, and guidelines about the design of verbal autopsy mortality surveillance systems should be reviewed accordingly.14,17,36

Our findings represent an important step forward because the impact of single versus multiple coding in mortal-
ity surveillance systems based on verbal autopsy has never been studied in detail, despite the fact that inter-observer and intra-observer variations in coding have been recognized for many years.18,19,27,28

The duplicate coding process would guard against random error and also systematic error. However, the high proportion of deaths for which the same cause was assigned by each of the two physician coders suggests that random errors in coding were rare. Most likely, this consistency in cause-of-death assignment reflects the marked difference between the causes of death listed in each of the ICD-10 chapter headings on which the primary analyses were based. The same reason probably also accounts for the strong agreement between the causes assigned by the coders for each of the major subcategories of deaths attributed to cardiovascular diseases, external causes of injury and infections.

An alternative explanation for the strong agreement between the causes assigned by the two coders is that the duplicate coding process may be a poor method for detecting systematic errors in the assignment of causes. For example, the system does not prevent errors that arise from biases in the data collection process: if the cause-of-death decision is made using information that was incorrectly recorded during the verbal autopsy interview, then duplicate coding will have little effect on the reliability of the final results. Similarly, if poor training of coders resulted in a bias towards a particular cause-of-death assignment, the skew would probably be repeated by subsequent coders because the training and supporting materials provided within a verbal autopsy mortality surveillance system are highly standardized. To better understand the effect of these biases on the final results of the mortality surveillance system, validation studies that compare gold standard diagnoses based on reliable medical records with diagnoses derived from the verbal autopsy method would be needed.

Verbal autopsy studies generally use lay interviewers for data collection and physician coders for interpretation and cause of death assignment.12,27,29,30 Coding of verbal autopsies by physicians has several potential drawbacks. First, physician coders may be biased by their prior knowledge of disease patterns in the community, hence two or more physicians drawn from the same population would be expected to have high agreement. Second, expert physician reviewers are an expensive resource, and using them for this type of work diverts them from clinical roles in settings where physicians are often scarce.

Large-scale mortality surveillance systems that use verbal autopsy, such as India’s Sample Registration System,3,31 the INDEPTH Network32 and the United Republic of Tanzania’s Adult Mortality and Morbidity Project,33 use dual coding for cause-of-death assignment.33,34 Our findings suggest that expenses in programmes such as these could be substantially reduced by switching to a system of single coding. Alternatively, the funds currently used for duplicate coding could be reassigned to conduct validation studies that compare the cause-of-death assignments from single-coding verbal autopsy systems with cause-of-death assignments derived from reliable medical records, diagnosis by autopsy, or physician-diagnosed deaths in the community.

Had a validation study been possible within this project, it would have provided important external confirmation of the pattern of death observed; however, the absence of such data has few implications for our conclusions about the value of a single-coding system versus a multiple-coding system. Changes in the assignment of cause of death resulting from a validation process would have had a broadly similar effect on the pattern reported by each coder and little impact on the level of agreement between them.

A final consideration in the interpretation of this study is that the same three physicians coded all the deaths. Verbal autopsy mortality surveillance systems with a greater number of physician coders may stand to gain from a duplicate cause of death assignment process.

In conclusion, the findings of our study indicate that a single trained physician coder may be as effective as two coders and raise important questions about the value of the system of multiple coders that is currently widely accepted. Since the mortality surveillance system studied here follows the same process as other mortality systems based on verbal autopsy methods, our findings are likely to be broadly applicable to other settings. We believe that workers on current and planned verbal autopsy mortality surveillance projects based on duplicate coding systems should review their need for repeat coding and consider whether investment in external validation studies might not be a better use of resources.

Acknowledgements

The Andhra Pradesh Rural Health Initiative is a collaboration between the Byraju Satyanarayana Raju Foundation (Hyderabad, India), the CARE Foundation (Hyderabad, India), the Centre for Chronic Disease Control (New Delhi, India), The George Institute for International Health (Sydney, Australia) and the School of Population Health, University of Queensland (Brisbane, Australia). We would like to thank all the Multipurpose Primary Healthcare Workers, the physician coders, the project staff and all the respondents who participated in the study.

Table 2. Breakdown of estimated annual cost of implementing a mortality surveillance system, Andhra Pradesh, India, 2003–2004

<table>
<thead>
<tr>
<th>Process</th>
<th>Cost per death</th>
<th>Total cost*</th>
<th>Total savings with single coding*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rs</td>
<td>US$</td>
<td>Rs</td>
</tr>
<tr>
<td>Training</td>
<td>50</td>
<td>1.09</td>
<td>67025</td>
</tr>
<tr>
<td>Data collection</td>
<td>106</td>
<td>2.34</td>
<td>144024</td>
</tr>
<tr>
<td>Cause of death assignment</td>
<td>250</td>
<td>5.51</td>
<td>339000</td>
</tr>
<tr>
<td>Project management</td>
<td>252</td>
<td>5.55</td>
<td>341724</td>
</tr>
<tr>
<td>Total cost</td>
<td>659</td>
<td>14.49</td>
<td>891773</td>
</tr>
</tbody>
</table>

US$, United States dollars; Rs, Indian rupees.
* Cost per death multiplied by total number of deaths (1354).
* With deaths coded only once, the cause-of-death assignment costs would be reduced by half, and project management costs would be reduced by one third.
* Conversion rate: 1 Rs = US$ 0.0220063 (average exchange rate for 2003–2004, Reserve Bank of India).25
Funding: Funding support for the India-based component of this project was provided by the Byrraju Foundation and the Wellcome Trust (grant number GR076471MF). The George Institute’s contribution to this project was made possible by an award from the George Foundation. Rohina Joshi is supported by an international post-graduate research scholarship and international post-graduate award from the University of Sydney, and Bruce Neal by a fellowship from the National Heart Foundation of Australia.

Competing interests: None declared.

Resumen
Codificación de las autopsias verbales: ¿varios codificadores mejor que uno?

Objetivo Evaluar el impacto en los perfiles de causas de defunción notificados de una estrategia de codificación de las autopsias verbales reposando en el examen de cada defunción por varios agentes codificadores para examinar cada caso de defunción.

Métodos A lo largo de 12 meses durante 2003-2004 se documentaron las defunciones registradas en 45 aldeas (población total: 180 162 habitantes) del Sur de la India, rellenándose para cada fallecimiento un cuestionario de autopsia verbal estándar. Dos codificadores médicos, sin conocer uno a uno las decisiones del otro, asignaron las causas de defunción basándose en las causas enumeradas en los títulos de los capítulos de la décima revisión de la Clasificación Internacional de Enfermedades y Problemas de Salud Conexas (CIE-10). Para los tres títulos de capítulos que se aplicaron a más de 100 defunciones, se analizó también el grado de coincidencia para subconjuntos de causas de defunción dentro del capítulo. En caso de discrepancia, un tercer codificador intervenía para dirimirlo. La medición del grado de acuerdo entre los dos codificadores médicos se realizó mediante el estadístico kappa de Cohen (K).

Resultados En total se documentaron 1354 defunciones, realizándose una autopsia verbal en 1329 (98%) de ellas. En lo que respecta a los títulos de los capítulos de la CIE-10, los codificadores médicos asignaron la misma causa a 1255 defunciones (94%) (K = 0,93; intervalo de confianza del 95%: 0,92-0,94). Los perfiles de mortalidad derivados de las causas asignadas por cada médico fueron muy similares a los perfiles obtenidos mediante el proceso de consenso, hasta el punto de que el orden de importancia de las 10 causas principales de defunción fue el mismo con los tres métodos de codificación.

Conclusión La duplicación de la codificación de los resultados de las autopsias verbales no ofrece grandes ventajas respecto al sistema de codificador único en el contexto de la vigilancia de la mortalidad o la identificación de perfiles demográficos de la mortalidad. Se podrían desviar recursos hacia otros aspectos del proceso de vigilancia de la mortalidad, como por ejemplo la validación.

ملخص
تقييم الصفة التشريحية اللفظية: هل المرمزون المتعدّدون أفضل من مرزم واحد؟

الهدف: تقييم أثر أداء أسباب الوفيات المنقولة في استراتيجيات الترميز للم되기ات التشريحيّة اللفظية، استنادًا إلى مراجعة كل وفاة من حيث تصنيفها من قبل مرزمين متعددين مقابل تصنيفها من مرزم واحد.

الطريقة: تم توثيق الوفيات في 45 قرية من قرى جنوب الهند التي يقطنها 126 180 من السكان خلال مدة 12 شهراً في عاميّ 2003-2004، مع استيفاء استبيان قياسي للصفة التشريحية اللفظية لكل وفاة. وقُدّم تحليل قياسي والبيانات للجيل، مما أدى إلى إعداد مجموعة من الأسباب الوفيات من حيث 두며ن. والنتائج جمعتً من مرزمين متعددين ومرزم واحد وواحد. 

النتائج: في أول عام، كانت النتائج حوالي 0,93 في معدل Agreement مع الأسباب الرئيسية 1255 في 94% من المجموعات المتعددة مقابل 92% من مرزم واحد. مرتين أي أن الحاجة إلى استخدام السلطات التشريحية اللفظية قد تكون ضرورية في الأسباب الوفيات بشكل كامل، وقد حدد طبيبان...
الطبيين الذين يعرضونقناعتهم بال명ى على طبيبيهم لتحديد النتائج النهائية للوفيات، وتعتمد الأدلة الإحصائية كابلدوسية للقياس لمتواتر
الوفيات بين طبيبيين المصابين.

الاستنتاج: ليس متاحة عمليات التزميم الناتجة عن الصعوبات التفصيلية أو ت POLITICO
الوفيات أو للتعرف على أنماط الوفيات لدى السكان.

1. References
