Introduction

The poor state of health information systems and, particularly, mortality statistics in many countries is widely documented in the literature and in country reports to the World Health Organization (WHO). However, health systems worldwide depend on reliable information about causes of mortality to be able to respond effectively to changing epidemiological circumstances. Such responses depend critically on accurate data to guide decision-making. Within a health information system, accurate and timely data on the cause of death are fundamental for programme and policy development and for measuring change in the magnitude and distribution of ill-health and disease in populations. Assessments of vital registration systems in low- and middle-income countries consistently reveal substantial weaknesses in the generation of cause-of-death statistics. In a recent editorial, we drew attention to the fact that even hospital statistics on cause of death cannot be assumed to be correct – a fact that is not widely appreciated by governments and other users of these data.

The gold standard for cause-of-death reporting is to have the cause certified by a medical practitioner using the rules and procedures of the International classification of diseases and related health problems (ICD), which is currently available in its tenth revision (ICD-10). Although most countries with statistical systems for cause of death now use the ICD classification for coding, not all countries have introduced the international standard certificate for reporting cause of death. Furthermore, physicians often do not receive adequate training in standard ICD death certification practices. It is, therefore, not surprising that comparative assessments commonly find that the quality of medical certification of the cause of death is poor. Cause-of-death statistics of poor quality have limited policy utility and may even seriously mislead policy debates.

In most developing countries, more than half of all deaths occur outside hospitals. Since out-of-hospital deaths are rarely medically certified, most of the physician-certified deaths come from hospitals. Can we automatically assume that the cause assigned to a death in hospital is accurate? Unfortunately, even in countries where hospital data are the only source of cause-of-death information, data quality is rarely evaluated. Research in different countries has repeatedly identified substantial misclassification of the cause of death of people who die in hospitals – with attendant implications for the use of cause-of-death data in informing policy.

To carry out a validation study of cause-of-death data collected in hospitals, we need a gold standard against which the hospital cause-of-death reports can be compared. While autopsy findings provide the ideal gold standard for cause-of-death evaluations, this approach is prohibitively expensive, rarely applied and likely to be based on a biased sample of deaths assigned to coroners. It would not be practical to carry out autopsies for all of the deaths occurring in a country – or even for all of the hospital deaths in a country. Instead, researchers have reviewed the medical records of people who have died in hospitals as a reference standard for validating the accuracy of the causes of deaths recorded by the hospitals. Although all hospitals have medical records for their patients, such records are rarely used for carrying out routine assessments of the extent and nature of any diagnostic misclassifications among
hospital deaths. In part, this reflects a lack of awareness of the existence of such misclassification or – because there is no standard method and framework for carrying out routine evaluations of the quality of cause-of-death data – a lack of knowledge of how such misclassification might be identified. Here, we perform a systematic review of studies that used medical records to assess the quality of hospital cause-of-death data to ascertain the pattern and extent of diagnostic misclassification of the cause of death. We propose a standard method for the future use of medical record reviews for assessing the accuracy of hospital-based cause-of-death data.

**Systematic review**

**Search strategy**

We searched for published articles on studies that used medical record reviews to validate cause-of-death data for hospital deaths in English, Spanish or Portuguese, between 1983 and 2013. Search terms and databases are shown in (Fig. 1). All 199 studies identified from the initial investigation were screened for specific content and only the 29 studies that met our inclusion criteria were subjected to further scrutiny (Fig. 1, Appendix A available from: http://www.uq.edu.au/hishub/docs/Appendix%20A_final.pdf). Essentially, to be included in our systematic review, a study had to be primary research, published in a peer-reviewed journal and involve the validation of cause-of-death data originating from one or more hospitals against reference cause-of-death data obtained by review of the corresponding medical records.

**Results**

The 29 selected studies were categorized according to timing, geographical location, scope and basic study features, age at death and range of causes of death included (Table 1). Wherever possible, we also assessed methodological issues such as the characteristics of the reviewers, the quality of the medical records used and the diagnostic facilities available in the study hospitals. Since the methods, scopes and objectives of the selected studies were diverse, no review protocol was systematically applied and no comparative analysis of our findings was possible.

![Flowchart for the selection of studies on medical record reviews to validate cause-of-death data for hospital deaths](image)

**Table 1.** Wherever possible, we also assessed methodological issues such as the characteristics of the reviewers, the quality of the medical records used and the diagnostic facilities available in the study hospitals. Since the methods, scopes and objectives of the selected studies were diverse, no review protocol was systematically applied and no comparative analysis of our findings was possible.

Only nine of the 29 selected studies were published after 2005. The number of deaths included in each study ranged from 23 in a study in Mexico to 3316 in a nationally representative study of cause-of-death accuracy in Thai hospitals. More than half of the selected studies were carried out in the Americas; Mexico contributed three studies and Brazil six. There was only one study from the Pacific – in Tonga – and one from Africa – in South Africa. Eighteen studies considered all causes of death while the remainder investigated selected causes only. The large majority (n = 24) covered all age groups. However, two studies included only adult deaths, one study focused on deaths in the elderly and another two studies focused on infant deaths (Table 1).

The main purpose of most of the medical record reviews was to identify the degree of misclassification of cause-of-death data...
of death at the individual level – by comparing the hospital or vital registration diagnosis with a reference diagnosis based on a review of the dead patient’s medical records. A misclassification matrix is the primary outcome of several studies.\(^3,16–18,20,22,27\) An example of a misclassification matrix is the primary outcome of several medical records. A misclassification diagnosis with a reference diagnosis can be applied to routine cause-of-death data – from vital registration systems – which had been adjusted for underreporting – to estimate the probable true pattern of causes of death in the country in that year.\(^44\) For some causes of death, such as human immunodeficiency virus infection/acquired immunodeficiency syndrome and ischaemic heart disease, the corrected numbers of deaths in this study were 3- to 4-fold higher than the numbers recorded in the vital registration system – with huge implications for Thailand’s health policies. Fifteen of our selected studies\(^18\) provided information to correct specific mortality fractions based on vital registration data – where these fractions were known or suspected to be incorrect.

The utility of any medical record review depends on the ability of the reviewers to identify the underlying cause of death correctly from the medical records.\(^18\) The reviewers in most of our selected studies (n = 18) were physicians who had – reportedly – been trained in death certification practices (n = 15). One study simply reported that professionals had been used as reviewers, while nine studies provided no information about reviewer qualifications. A Brazilian study\(^34\) that used both a physician and a researcher was classified as physician-review in this analysis. The number of reviewers per study depended on sample size and ranged from one in each of four studies to 84 in the Thai study.\(^17\)

The accuracy of the reference cause-of-death diagnosis depends on the quality of information available in the medical records. Ten of the reviewed studies included such an assessment – in which the records were categorized either simply, as adequate or inadequate, or, with more qualification, as excellent, good, average, weak or poor. One study excluded deaths when the corresponding medical records were judged to be incomplete.\(^23\) The remainder provided no information about the quality of the medical records used.

In nine of our selected studies, standard diagnostic criteria were predetermined – i.e. the clinical evidence required to classify a death as being due to a particular cause was specified in advance of the review.\(^14,16,18,21,22,23,24,41,43\) In case of diagnostic uncertainty, nine studies used a physician panel to discuss diagnostic problems. Three studies ordered verbal autopsies when the cause of death was in doubt. One study referred diagnostic problems to a second physician, and two others referred to external experts. Fourteen studies provided no information about the resolution of diagnostic uncertainty.

The diagnostic capacity of hospitals will clearly influence the accuracy of their cause-of-death diagnoses, both in vital registration systems and in medical record reviews. Only the Chinese study\(^38\) specifically mentioned that urban hospitals were selected to ensure that the study hospitals had adequate diagnostic facilities. None of the other studies included in our analysis referred to hospital diagnostic capacity.

The correct assignment of the underlying cause of death on the death
Table 2. Misclassification of causes of deaths, all ages, both sexes combined, Colombo, Sri Lanka, 2012

<table>
<thead>
<tr>
<th>Vital registration diagnosis</th>
<th>Diagnosis based on medical records review (no. of deaths)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Certain infectious and parasitic diseases</td>
</tr>
<tr>
<td></td>
<td>All cancers</td>
</tr>
<tr>
<td></td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td></td>
<td>Other diseases of the nervous system</td>
</tr>
<tr>
<td></td>
<td>Hypertensive diseases</td>
</tr>
<tr>
<td></td>
<td>Ischaemic heart diseases</td>
</tr>
<tr>
<td></td>
<td>Cerebrovascular diseases</td>
</tr>
<tr>
<td></td>
<td>Other heart diseases</td>
</tr>
<tr>
<td></td>
<td>Pneumonia</td>
</tr>
<tr>
<td></td>
<td>Chronic lower respiratory diseases</td>
</tr>
<tr>
<td></td>
<td>Other diseases of the respiratory system</td>
</tr>
<tr>
<td></td>
<td>Diseases of the liver</td>
</tr>
<tr>
<td></td>
<td>Diseases of the skin</td>
</tr>
<tr>
<td></td>
<td>External causes</td>
</tr>
<tr>
<td></td>
<td>All other causes</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

| Certain infectious and parasitic diseases | 9 3 1 0 0 0 2 0 1 0 1 2 4 2 |
| All cancers                               | 1 34 1 0 0 2 3 0 0 4 0 2 0 0 |
| Diabetes mellitus                         | 3 3 34 1 4 22 9 1 3 2 0 3 1 1 |
| Other diseases of the nervous system       | 2 0 0 3 0 2 0 0 1 0 1 2 0 0 |
| Hypertensive diseases                      | 4 0 2 0 12 9 10 0 1 3 0 1 0 0 |
| Ischaemic heart diseases                   | 2 0 9 0 2 54 5 3 1 5 0 4 0 1 |
| Cerebrovascular diseases                   | 0 1 1 1 2 1 17 1 0 1 1 1 0 0 |
| Other heart diseases                       | 1 4 3 0 2 17 4 21 1 4 1 3 1 0 |
| Pneumonia                                  | 0 1 1 2 0 1 1 0 9 2 0 0 0 0 |
| Chronic lower respiratory diseases         | 1 1 1 0 0 7 1 1 1 1 1 3 0 0 |
| Other diseases of the respiratory system   | 0 1 1 1 0 1 1 0 1 0 3 2 0 1 |
| Diseases of the liver                      | 4 2 2 1 2 2 0 3 0 1 0 39 1 3 |
| Diseases of the skin                       | 0 1 0 0 0 1 0 0 0 0 0 1 0 1 |
| External causes                            | 3 3 3 0 2 3 0 1 0 1 0 2 0 1 |
| All other causes                           | 2 2 2 2 1 5 3 2 1 1 0 3 0 0 |
| Total                                      | 32 56 62 11 27 127 54 35 19 38 6 64 3 6 |

Data source: Based on data collected by Rampatige et al.22
certificate requires not only that the sequence and relationship of pathological conditions on the certificate are correctly certified but also that the underlying cause of death is correctly coded in terms of the ICD. Despite the critical role of coding, only three studies assessed coding accuracy.17,26,37

All studies reported some degree of diagnostic misclassification. Due to the use of different classification lists, the misclassification reported in these studies are not directly comparable. For the studies that have used ICD-10’s three-character codes for comparison of cause of death, the range of misclassification varied from 25.0% to 62.1%.16 The concordance at the level of ICD-10’s three-character codes was 41.4% in Sri Lanka21 while concordance at the level of the Basic Tabulation List in the ninth revision of the ICD was found to be 77% in Sweden.13 In South Africa, concordance at the level of ICD-10’s mortality tabulation level was 55.3%.14 The pattern and extent of misclassification found in the studies were generally reported either as matrices or as percentages. Other metrics that were used included simple concordance-based measures of sensitivity and specificity, and measures that attempt to correct for the probability of getting a concordant diagnosis merely by chance, such as Kappa statistics and chance-corrected concordance.46 More usefully, the analysis of misclassification matrices allows specific insight into the patterns of cause-of-death misallocation and is thus a critical input in activities designed to strengthen health information systems as well as in the correct interpretation of national mortality statistics. Common causes of death that were frequently misclassified included ischaemic heart disease, stroke, diabetes mellitus and external causes of injury. In a study based on a medical record review in the Islamic Republic of Iran, ill-defined causes of death were mainly reclassified to ischaemic heart disease (33.5%) and cerebrovascular accidents (17.1%) while a study in Kuwait reported that original death certificates underestimated cerebrovascular disease by 69%, diabetes mellitus by 60% and ischaemic heart disease by 33%.17 Surprisingly, in our selected studies, ill-defined causes of death were frequently reported even in hospital settings. The policy impact of allocating these non-specific codes to more definitive causes are reported in two studies;16-18 key findings from a meta-analysis of misclassification patterns can be found in the article by Rampatige et al.47

### Framework for reviewing medical records

Given the importance and cost of accurate cause-of-death statistics for health policy and priority setting – and the effort and cost of undertaking medical record reviews – evaluation studies using this method could benefit enormously from some simple, evidence-based lessons. Based on this systematic review and experience with our own field research, we provide a synthesis of some of the key issues that should guide national applications of medical record reviews.

Our systematic review confirms that a variety of approaches have been used to assess the accuracy of cause-of-death data. Rather than trying to evaluate whether there is one approach that is better than the others, we have tried to draw lessons from each study about the specific steps that added value to the study and could be repeated elsewhere. Building on these findings, we propose a framework that addresses the critical choices that need to be carefully considered when using medical record reviews for assessing cause-of-death data quality (Box 2). Further detail on each of the topics included in our framework has been published elsewhere – with specific guidance on the selection of hospitals and cases, the development and use of diagnostic criteria, interpreting misclassification matrices, evaluating the quality of the source documentation, and the choice and training of the reviewers used.47

### Discussion

Medical record reviews can and have been carried out to serve many different purposes. Perhaps the most common application is the independent assessment of the reliability of hospital cause-of-death data – particularly when expert review or other use of mortality data has revealed these to be deficient. Another common purpose is to assess whether deaths from specific causes – e.g. various cancers or traffic accidents – are being reliably recorded in hospital settings. If such an assessment involves a reasonably representative national sample of hospital deaths, correction factors can be calculated from the misclassification patterns.
Box 2. Proposed framework and key elements for conducting medical record reviews

Select hospitals to be reviewed
- Determine scope of investigation
- Obtain agreement for hospital cooperation
- Conduct census of diagnostic facilities available in included hospitals.

Select diagnostic categories and develop diagnostic criteria
- Set up a small expert group of physicians to develop standard diagnostic criteria
- Establish a list of the diseases that are the most important for the review
- Develop and pre-test diagnostic criteria on a sample of medical records.

Select sample death certificates
- Determine sample size
- Determine the sampling method and identify the number of death certificates to be included in the study
- Draw the sample of death certificates from the vital registration database or hospital mortality register
- Retrieve corresponding medical records from the hospitals
- Validate the quality of coding – according to the tenth revision of the International Classification of Diseases and Related Health Problems (ICD-10) – for the sample.

Select physicians to rediagnose causes of death
- Provide training in cause of death certification.

Trace the relevant medical records
- Decide on criteria to assess the quality of the records
- Decide on rules to determine which records can be used and which are too incomplete
- Reassess the sample size because of losses arising from poor or untraceable records
- Prepare a summary of the quality and availability of medical records and the corresponding maintenance practices.

Review medical records
- Design form for extracting information from medical records. An example of such medical data extraction forms is given in the study by Rampatige et al.\(^1\)
- Diagnose cause of death using pre-defined standard diagnostic criteria
- Code the new cause according to ICD-10
- Check that coding is correct.

Compare the cause of death originally reported from vital registration with that finally assigned after medical record review, and analyse findings
- Determine the extent of misclassification
- Draw up a misclassification matrix for all ages and both sexes – and by age and sex, if numbers allow – to identify patterns of misclassification
- Reassign the ill-defined and misclassified causes based on the misclassification matrix
- Compare the new distribution of causes of death with the original, to identify major discrepancies.

Write final report and initiate policy dialogue to strengthen health information system
- Describe the study design and methodology, including sampling strategy
- Discuss findings and implications
- Propose improvement steps as needed – e.g. improved certification, coding and keeping of medical records.

Matrices and applied to national cause-of-death data to estimate the likely true cause-specific mortality pattern in the population.

Despite an extensive literature review, we were only able to identify a small number of peer-reviewed studies that had used medical records to investigate the quality of cause-of-death data \((n = 29)\). Given the critical role of cause-of-death data in informing debates around national health policies, this is both surprising and disappointing. Our main observation – of extensive misclassification of causes of death in hospitals wherever studies have been carried out – indicates that inaccurate cause-of-death data are pervasive. However, it is important to consider potential biases that might have affected the limited number of studies that we were able to analyse. First, we were only able to review articles that had been published in a language in which at least one member of our review team was competent – i.e. English, Portuguese or Spanish. It is possible that articles published in other languages might have led to different conclusions about the pattern and extent of cause-of-death misclassification in hospitals. More comprehensive studies covering other major language groups would be welcome but there is no reason to suspect that these would not reinforce our main findings – which seemed fairly consistent irrespective of the language of the article involved. Thus, our conclusions and recommendations for policy action should still be valid. It is also possible that the literature we screened was affected by selective biases – such as publication bias – but we suspect that this would only operate to suppress the publication of more extreme findings about misclassification. Finally, the studies we selected had diverse aims and the set of causes of death investigated in each study was presumably biased towards the priorities of the interest groups involved. We found only one study that involved the comprehensive investigation of cause-of-death misclassification across all major causes of death. It is therefore quite likely that the extent of cause-specific misclassification across all countries is different to that indicated by the studies we selected. Again, however, that does not alter our fundamental conclusion that hospital statistics on causes of death – whatever the universe of causes investigated – are likely to be grossly inaccurate in many countries and ought to be the focus of an immediate policy response.

All of the studies we selected for review assessed the quality of the hospital-reported cause-of-death data by comparing them with cause-of-death data derived by reviewing medical records. However, the lack of a standard framework or method for conducting and reporting the findings of medical record reviews has resulted in substantial variation in the approaches used and difficulties in interpreting and comparing results. Building on these findings and our own empirical experience, we propose a series of steps to guide future studies on the accuracy of cause-of-death data that involve the review of
The synthesis of lessons learnt, summarized in Box 2, outlines a clear and tested method and identifies the most important steps in the process – e.g. the design of the study, development of diagnostic criteria, selection of the sample, review of medical records, and the analysis and interpretation of findings. However, the framework that we have proposed – like all frameworks – has limitations. In particular, it would not be feasible to apply our framework in settings where medical records are poorly kept or where the diagnostic capacity of hospitals is generally poor. An understanding – by hospital and health administrators – of the importance of studies based on our framework for policy-makers is crucial, as is a commitment, of those responsible for the development of national health information systems, to implement any resultant recommendations.

While the proposed framework needs to be tested in its entirety in further empirical studies, we consider it sufficiently robust to be applied in its current state in different settings. We recommend its periodic application in all countries, to identify deficiencies in national mortality statistics. Using the misclassification matrix that the study method yields, health authorities should be able to identify the diseases that are commonly misclassified as causes of death and also ascertain whether the causes of death that are being recorded in hospitals are of sufficient quality to be fit for purpose. This knowledge can and should be used to guide strategies for the strengthening of health information systems – e.g. increasing awareness among medical associations about the value of doctors correctly certifying deaths, improving the keeping of medical records in hospitals, and ensuring that cause-of-death coders are appropriately trained.

Systems for the determination and recording of the causes of hospital deaths represent a purposeful and costly investment that countries make. It is imperative that such systems perform to the standard required to support good public policy.

Funding: Funding for this research was provided by the former Australian Agency for International Development under a grant provided to the University of Queensland, Brisbane, Australia, to establish a health information systems knowledge hub.

Competing interests: None declared.
Objetivo

Examinar sistemáticamente la fiabilidad de los datos de mortalidad según el género en los hospitales europeos para que se puedan adaptar a la comparabilidad de los resultados.

Métodos

Se realizaron búsquedas de artículos en inglés, español y portugués en Google Scholar, PubMed y la Biblioteca Virtual de la Salud sobre estudios que informaron acerca de la mortalidad según el género en hospitales. Se incluyeron los resultados de 28 estudios seleccionados, que se publicaron entre 1983 y 2013.

Resultados

De los 28 estudios seleccionados, 13 informaron sobre la mortalidad según el género en hospitales de España, Portugal, Francia, Italia, Bélgica, Alemania y Reino Unido. Se observó una variación considerable en la mortalidad según el género entre los diferentes hospitales y países, lo que sugiere una falta de consistencia en la declaración de las causas de muerte. Para mejorar la comparabilidad de los datos, se recomienda la realización de estudios de seguimiento regular y la implementación de normas de registro uniformes en los hospitales.
utilidad de las causas de muerte notificadas, los gobiernos nacionales deberán llevar a cabo revisiones periódicas de los históricos médicos con objeto de validar la calidad de los datos hospitalarios sobre las causas de muertes por medio de un método estándar para lograr resultados comparables a nivel internacional.

References


Rasika Rampatigae et al.


