METHODOLOGIC ISSUES

Diagnosis based injury severity scaling: investigation of a method using Australian and New Zealand hospitalisations

S Stephenson, G Henley, J E Harrison, J D Langley

Objective: To assess the performance of the International Classification of Diseases (ICD) based injury severity score, ICISS, when applied to two versions of the 10th edition of ICD, ICD-10 and ICD-10-AM.

Design: ICISS was assessed on its ability to predict threat to life using logistic regression modelling. Models used ICISS and age as predictors and survival as the outcome.

Setting: Australia and New Zealand.

Patients or subjects: Hospitalisations with an ICD-10-AM principal diagnosis in the range S00–T89 from 1 July 1999 to 30 June 2001 (Australia) or 1 July 1999 to 31 December 2001 (New Zealand).

Interventions: None.

Main outcome measures: The models were assessed in terms of their discrimination, measured by the concordance score, and calibration, measured using calibration curves and the Hosmer-Lemeshow statistic.

Results: 523 633 Australian and 124 767 New Zealand hospitalisations were selected, including 7230 and 1565 deaths respectively. Discrimination was high in all the fitted models with concordance scores of 0.885 to 0.910. Calibration results were also promising with all calibration curves being close to linear, though ICISS appeared to underestimate mortality somewhat for cases with an ICISS score less than 0.6. Overall ICISS performed better when applied to the Australian than the New Zealand hospitalisations. Australian and New Zealand hospitalisations were very similar. ICISS was also only a little more successful when ICD-10-AM rather than mapped ICD-10 was used.

Conclusions: ICISS appears to be a reasonable way to estimate severity for databases using ICD-10 or ICD-10-AM. It is also likely to work well for other clinical variants of ICD-10.

Injury severity measurement is important for meaningful comparison of outcomes of trauma care, and assessment of burden of injury. Trends in hospitalisations are often used as a proxy for trends in population incidence. These trends can be misleading due to variations over time or between places in factors unrelated to incidence, such as admission policies and practices. Injury severity measurement is a promising basis for improving measurement of trends in the population burden of injury, as it provides a means of screening out the less serious injury cases which are most susceptible to extraneous factors.\(^1\)

Different injury severity criteria are potentially of interest depending on the application. These include threat to life (probability of death), cost, impairment, incapacity, impact on quality of life, and disability.\(^2,3\) This investigation has focused on probability of death, as it is the only criterion for which well established methods exist to estimate injury severity. Many countries, including Australia and New Zealand, maintain large administrative databases of injury hospitalisations. These databases include detailed information on the nature and circumstances of injury in the form of International Classification of Diseases (ICD) codes. During the 1990s both New Zealand and Australia used clinical modifications of the ninth revision of the ICD (ICD-9), firstly from North America (ICD-9-CM) and then Australia (ICD-9-CM-A).\(^4,5\)

Two methods have been shown to be capable of providing fairly reliable probability of death estimates based on ICD-9-CM diagnosis codes. One method involves translation of ICD-9-CM codes to Abbreviated Injury Scale (AIS) scores via the proprietary software package ICDMAP-90.\(^6\) AIS has long been regarded as the injury coding “industry standard” with respect to diagnosis based severity scores. AIS scores derived using ICDMAP-90 are referred to as ICD/AIS scores.

The other method, known as ICISS (ICD based injury severity score), involves estimating probability of death directly from ICD-9-CM injury diagnoses by examining a large set of cases for which survival status is known. The estimated probabilities obtained can (with certain caveats) be applied to cases in other similar data sets.\(^7\) There is more experience with ICD/AIS based measures than with ICISS, and early literature suggested that the former might be more effective.\(^8,9\) However, recent evidence suggests that the methods perform similarly.\(^11,12\) Neither seems to be as good as AIS scoring based on direct case note review. However, both offer an inexpensive way to derive quite good severity estimates for large databases if data with good quality ICD-9-CM coding is available.

In 1992 the World Health Organisation published the 10th revision of ICD (ICD-10).\(^13\) A clinical modification, ICD-10-AM, was developed in Australia.\(^14\) ICD-10-AM was initially introduced in four Australian jurisdictions in 1998, then in the rest of Australia and in New Zealand from 1 July 1999.

Transition to ICD-10 presents problems for both ICD/AIS and ICISS. Formal updating of ICDMAP for use in Australia or New Zealand would require development and validation of a map between ICD-10-AM and AIS. Alternatively, tables of equivalent categories could be used to map ICD-10-AM codes to ICD-9-CM-A, then to the relevant edition of the North American ICD-9-CM and then, via ICDMAP, to AIS. However, this would be complex, and

Abbreviations: AIS, Abbreviated Injury Scale; ICD, International Classification of Diseases; ICISS, ICD injury severity score; SRR, survival risk ratio.
predictive power would be lost at each of the translation steps.

ICISS appears to offer a more direct and certain route to severity scores based on ICD-10-AM. The algorithm to derive ICISS is not specific to a particular version of ICD. Only one previous study was identified that has examined the performance of ICD-10 based ICISS. This study indicated a lower performance for ICD-10 based ICISS when compared to ICD-9-CM based ICISS. However, the sample size used in this study was very small.\(^\text{15}\)

Data available in Australia and New Zealand provided a particularly good basis for testing the ICISS method in the context of ICD-10-AM coding. Mapping of the Australian and New Zealand data from ICD-10-AM to ICD-10 also allowed assessment of the applicability of ICISS to ICD-10. The aim of this investigation was to assess the performance of ICISS when applied to ICD-10-AM and ICD-10 using New Zealand and Australian data.

METHODS

Data were extracted from Australia’s National Hospital Morbidity Database, for 1 July 1999 to 30 June 2001, and New Zealand’s National Minimum Data Set, for 1 July 1999 to 31 December 2001. Hospitalisations with an ICD-10-AM principal diagnosis in the range S00-T89 were selected (that is, the injury and poisoning chapter excluding sequelae). Patients admitted and discharged on the same day were excluded unless they died. For Australian hospitalisations, in order to reduce multiple counting of repeat admissions for the same injury, discharges involving transfer to another hospital or a change in treatment regimen without leaving the hospital were excluded. For New Zealand hospitalisations, only diagnoses coded for the first hospitalisation were used, and repeat admissions for the same injury were excluded using an approach described previously.\(^\text{16}\) A map was created to translate ICD-10-AM codes back to ICD-10 codes primarily by reversing an ICD-10 to ICD-10-AM forward map.\(^\text{17}\)

The ICISS method involves calculating a survival risk ratio (SRR)—that is, the probability of survival, for each individual injury diagnosis code as the ratio of the number of patients with that injury code who have not died to the total number of patients diagnosed with that code. Thus, a given SRR represents the likelihood that a patient will survive a particular injury. Each patient’s ICISS score (survival probability) is then the product of the probabilities of surviving each of their injuries individually. A set of SRRs was calculated for each country with hospitalisations coded to each of ICD-10 and ICD-10-AM giving four sets of SRRs in total. Each set of SRRs was used to calculate an ICISS for each patient in the Australian and New Zealand data.

The performance of ICISS was assessed by its ability to predict severity, in terms of threat to life, using logistic regression modelling. Models used ICISS as the predictor variable and survival as the outcome. Models were fitted, both with and without age added as a second predictor variable. Age was coded as a categorical variable at five levels (0–14, 15–24, 25–64, 65–79, and 80+ years old). The models were assessed in terms of their level of discrimination, measured by their concordance, and calibration, measured using calibration curves and the Hosmer-Lemeshow statistic.\(^\text{18}–\text{20}\) Concordance measures the ability of the model to distinguish survivors from non-survivors on a scale from zero to one with one indicating perfect separation of the two groups. The concordance is equal to the area under the receiver operator characteristic curve.\(^\text{19}\) Calibration curves and the Hosmer-Lemeshow statistic indicate the accuracy of the model’s estimates of probability of death with a perfectly calibrated model giving a straight 45° line calibration curve and an Hosmer-Lemeshow statistic of zero. The higher the Hosmer-Lemeshow statistic the poorer the fit. Bootstrapping was used to correct the concordance values for bias caused by the use of a single data set for design and testing.\(^\text{18}–\text{21}\)

Hospitals were assigned to one of four categories based on their principal diagnosis. These categories were head injury (S00–S09), other mechanical trauma (S10–T14), complications (T80–T88), and other injuries (T15–T79, T89). Concordance scores were calculated separately for each of these categories.

The number of diagnoses per hospitalisation in the Australian and New Zealand data were compared. The SRR sets for Australia and New Zealand were compared.

Confidence intervals were constructed using the methods of Agresti and Coull for proportions.\(^\text{22}\) Calculation of SRR values and initial model fitting was undertaken in SPSS version 11.5.0 (1989–2002, SPSS Inc) for Australian data and SAS version 8.00 (1999, SAS Institute Inc) for New Zealand data. Concordance calculations were conducted using R version 1.6.2 (2003, The R Development Core Team).\(^\text{23}\)

RESULTS

Applying the selection criteria resulted in 523 633 hospitalisations including 7230 deaths (1.4%) from Australia and 124 767 hospitalisations including 1565 deaths (1.3%) from New Zealand. The Australian data had means of 1.59 and 2.56 and the New Zealand data 1.69 and 2.64 injury diagnoses per hospitalisation for survivors and deaths respectively. The bootstrap adjusted concordance values for the Australian and New Zealand hospitalisations are presented in table 1. The concordance values for both countries were high ranging from 0.885 to 0.906 in Australia and 0.902 to 0.910 in New Zealand. The concordance of the model was lower when the hospitalisation source differed from the SRR source, by 0.019 when New Zealand SRR values were used on Australian hospitalisations compared with 0.004–0.005 when Australian SRRs were used on New Zealand hospitalisations. For both Australian and New Zealand hospitalisations, the ICD-10-AM concordance scores were 0.002–0.004 higher than the ICD-10 values. When age was excluded from the regression models, concordance values for Australian hospitalisations were on average 0.049 lower, with differences ranging from 0.035–0.061 compared to New Zealand hospitalisations, where concordance values were on average 0.052 lower, with differences ranging from 0.046–0.058.

Concordance by diagnosis results are shown in table 2. In both countries, concordance was significantly higher for mechanical trauma injuries than other injuries. Head injuries had the highest concordance and complications the lowest. For Australian hospitalisations, the concordance dropped by 0.006–0.089 when the New Zealand SRRs were used with the largest drop being associated with “other” injuries. For New Zealand hospitalisations, the concordance dropped by

<table>
<thead>
<tr>
<th>Hospitalisations source</th>
<th>SRR source</th>
<th>Concordance</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Australia (ICD-10-AM)</td>
<td>0.906</td>
<td>0.903 to 0.909</td>
</tr>
<tr>
<td></td>
<td>Australia (ICD-10)</td>
<td>0.887</td>
<td>0.884 to 0.891</td>
</tr>
<tr>
<td></td>
<td>New Zealand (ICD-10-AM)</td>
<td>0.905</td>
<td>0.900 to 0.910</td>
</tr>
<tr>
<td></td>
<td>New Zealand (ICD-10)</td>
<td>0.902</td>
<td>0.896 to 0.908</td>
</tr>
</tbody>
</table>
0–0.014 when the Australian SRRs were used with the largest drop again being associated with “other” injuries.

The Hosmer-Lemeshow statistics for the Australian and New Zealand hospitalisations are presented in table 3. For both countries, a marked increase was observed in the Hosmer-Lemeshow statistic when comparing ICD-10 to ICD-10-AM derived scores. For Australian hospitalisations, when New Zealand SRRs were used in place of Australian SRRs, marked increases were observed in the Hosmer-Lemeshow statistic for both ICD-10-AM and ICD-10. In contrast for New Zealand hospitalisations, when Australian SRRs were used in place of New Zealand SRRs, slight decreases were observed in the Hosmer-Lemeshow statistics for both ICD-10-AM and ICD-10.

Calibration curves for the ICD-10-AM coded Australian hospitalisations are given in fig 1. Using the Australian SRRs produced a moderately better calibrated model than using the New Zealand SRRs. Calibration was generally better at lower mortality, and was particularly good below 30% observed mortality, in both models. Similar patterns were observed for the ICD-10 coded hospitalisations. For ICD-10-AM, the calibration curves for the models where age was excluded were poorer fits than for those with age included, in contrast to ICD-10 where similarly fitting models were observed.

The differences in SRRs between Australian and New Zealand SRR sets for the most common diagnoses were small with differences ranging from 0%–4%. Differences between the two sets were also generally small in terms of diagnoses with the most deaths, differences in SRRs ranging from 0%–14%. The Australian SRRs were generally lower than the New Zealand SRRs. Four diagnoses from “S36—injury to intra-abdominal organs” had significantly different SRRs, indicating there may be a difference in the use of this code between Australia and New Zealand.

**DISCUSSION**

In all of the fitted models discrimination was high with concordance values ranging from 0.885–0.910. These estimates are in line with concordance results reported for ICISS applied to ICD-9-CM data of 0.87–0.90 and somewhat higher.
than the only previous study of ICISS applied to ICD-10 where the reported concordance was 0.84.10–12

Concordance was higher for mechanical trauma injuries, particularly head injuries, than for complications and other injuries. Factors underlying the differences might include characteristics of the ICD-10-AM classification and differences between patients with each of these types of condition. Formal analysis of the differences was beyond the scope of this paper.

Calibration results were also promising with all the calibration curves being reasonably close to linear. The calibration curves indicated ICISS underestimated probability of death somewhat for those with an ICISS score of less than approximately 0.6 (that is, estimated mortality >40%). The calibration problems may be related to an assumption used in the ICISS method. By multiplying the SRRs together to arrive at an ICISS score the method assumes the survival probabilities are independent. This assumption will be violated by groups of diagnoses that are often coded together. A specific example is the ICD-10-AM coding of an open wound communicating with a fracture. The open wound code can only occur where an associated fracture code is present. This will lead to underestimated survival probabilities.

Overall, ICISS performed better when applied to the Australian than the New Zealand hospitalisations. This was expected given that the New Zealand data set was so much smaller than the Australian set. Nevertheless, ICISS worked surprisingly well for the New Zealand hospitalisations given the relatively small number of deaths.

Both the New Zealand and Australian data sets were limited by only including hospital deaths. For Australia, the in-hospital injury deaths among the cases used in this study comprise 44% of all external causes deaths registered in the same period. Available data suggest that the proportion was similar for New Zealand. Some deaths occurring outside hospital were included in the previous study of ICISS in New Zealand but more complete data on deaths among the cases studied were not available for the present study period for either country.11 As discussed elsewhere, adding deaths occurring outside hospital would enhance the value of the result of severity estimates.11

A limitation of the Australian data was the inability to differentiate between individuals who were admitted to hospital for the treatment of their injury in both the acute and rehabilitative phases resulting in a series of one or more readmissions for the same injury. All of these readmissions were included in the data set resulting in an overestimation of survivors relative to deaths. This limitation could be overcome by introducing a process of person based linkage that allows all readmissions for a particular injury to be linked to the original admission. A system to do this already exists in New Zealand and some parts of Australia.

ICD-10-AM was introduced in New Zealand and four Australian jurisdictions at the start of the study period, and had been in use in the rest of Australia for one year. While we are not aware of relevant published evaluations, coding quality might have been reduced during a period of familiarisation soon after introduction of the new classification. If this occurred, the most likely effect would be some reduction in concordance estimates.

Comparison of the SRRs suggested the Australian and New Zealand hospitalisations were very similar. This was reflected in the only slight drop in the performance of ICISS when SRRs from the other country replaced “local” SRRs. Previous studies have suggested ICISS may not be well calibrated when SRRs are obtained from a separate data set.10 This study suggests there will not be a substantial drop in calibration if the data set used to derive the SRRs is similar to the data set for which ICISS is to be calculated. It seems plausible that a given set of SRRs could have at least some capacity to be used as the basis for valid comparisons of injury severity between places and over time. This paper reports the first example of an international comparison. Further research is required to assess the stability of SRRs over time and between places.

As expected the extra detail in the coding of ICD-10-AM resulted in better ICISS performance than was seen for ICD-10. However, the difference was relatively small and in an absolute sense ICISS also appears to be a good estimate of probability of death for injury hospitalisations data coded to ICD-10. The main reason for the lack of a large difference appears to be that in many cases, the subdivided ICD-10-AM codes created from a single ICD-10 code had similar SRRs. The results reported here are likely to be a conservative estimate of the performance of ICISS when applied to ICD-10 as mapped ICD-10 codes were used. The assumptions that had to be made in order to produce the map are likely to have generated more non-specific codes than would have been the case had the data been coded directly in ICD-10.

Age was included in many of the fitted models reported here and as such these analyses do not assess the performance of ICISS alone. Age is acting as a proxy for resilience and, to some extent, comorbidity. A more direct approach may be to extend the ICISS methodology to incorporate non-injury diagnoses. The recently developed severity measure HARM (Harborview Assessment for Risk of Mortality) includes some comorbidities.24 One approach would be to calculate SRRs for each chapter of the non-injury diagnoses.

We have used the Hosmer-Lemeshow statistic to summarise and compare the relative calibration of the fitted curves. While it is a reasonable choice for this purpose, it has limitations and other approaches could be used.25

As noted by Meredith et al the ICISS score is unlikely to be linear in the log odds as assumed in constructing the logistic regression model. In fact, by definition ICISS should be linearly related to the outcome, probability of death, since ICISS is an estimate of probability of survival. This problem will only affect the calibration, not the discrimination results.11 26

table3

<table>
<thead>
<tr>
<th>SRR source</th>
<th>Australian hospitalisations</th>
<th>New Zealand hospitalisations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia (ICD-10-AM)</td>
<td>166.3</td>
<td>33.8</td>
</tr>
<tr>
<td>New Zealand (ICD-10-AM)</td>
<td>271.3</td>
<td>37.1</td>
</tr>
<tr>
<td>Australia (ICD-10)</td>
<td>246.5</td>
<td>64.1</td>
</tr>
<tr>
<td>New Zealand (ICD-10)</td>
<td>457.4</td>
<td>68.7</td>
</tr>
</tbody>
</table>

Key points

- ICISS is a reasonable method to estimate severity for databases using ICD-10 or ICD-10-AM and is likely to work well for other variants of ICD-10.
- ICISS is non-reliant on any mapping or manual coding processes, making it readily adaptable to any future modifications of the ICD coding system.
- ICISS exhibits a fair degree of stability and appears to be minimally affected if the data set used to derive SRRs is similar to the data set for which ICISS is calculated.
In conclusion, ICISS is a reasonable way to estimate severity for large databases using ICD-10 or ICD-10-AM. It is also likely to work well for other variants of ICD-10. ICISS could be used to construct indicators that would track changes in the incidence of serious injuries without undue influence from changes in extraneous factors.

ACKNOWLEDGEMENTS

The work reported in this paper was conducted as part of the program of the Australian Institute of Health and Welfare National Injury Surveillance Unit. The Injury Prevention Research Unit is jointly funded by the Health Research Council of New Zealand (HRC) and the Accident Compensation Corporation. The IPRU work reported in this paper was funded by the HRC. The views expressed in this paper are those of the authors and do not necessarily reflect those of the above organisations. The New Zealand Health Information Service supplied the New Zealand data used in this paper. The authors wish to thank Dr Colin Cryer for comments on an earlier draft of this paper.

Authors’ affiliations

S Stephenson, J D Langley, Injury Prevention Research Unit, University of Otago, New Zealand

G Henley, J E Harrison, AIHW National Injury Surveillance Unit, Flinders University, Australia

REFERENCES


