

Can the ICD-based injury severity score (ICISS) be improved?

A report prepared for Official Statistics Research, Statistics New Zealand

June 2007

Gabrielle Davie
Colin Cryer
John Langley
Daniel Russell

Injury Prevention Research Unit
Department of Preventive and Social Medicine
University of Otago

Can the ICD-based injury severity score (ICISS) be improved?

Summary

Objective: To investigate whether the predictive ability of the International Classification of Disease (ICD)-based Injury Severity Score (ICISS) can be enhanced through the use of integrated hospitalisation and mortality data sources to calculate ICISS scores and/or whether taking account of comorbidity improves the predictive ability of the ICISS scores.

Design: Models using either the ICISS based solely on hospital discharge data or one of nine modified ICD-based injury severity scores as the predictor variable were assessed on their ability to predict survival using logistic regression modelling.

Setting: New Zealand

Patients or subjects: Hospitalisations with an ICD-10-AM principal diagnosis in the range S00-T89 with a date of injury within the years 2000 to 2003 and fatalities with any S00-T89 ICD-10-AM diagnosis and a date of death within the same period.

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. R^2 goodness-of-fit statistics were also compared.

Results: 186 835 cases including 9 968 deaths were included. The ICISS which included both mortality data and Charlson comorbid conditions at the ICD-10-AM level (ICISS9) had the best concordance. Although the calibration of ICISS9 was very similar, the lowest H-L statistic was obtained from ICISS10 which included both mortality data and Charlson comorbid conditions at the variable level. The calibration curves indicated that the scores that used hospital discharge data only to calculate SRRs underestimated mortality whereas the scores that used hospital discharge data and mortality data overestimated mortality. “Head injuries” had the highest concordance and “complications” the least.

Conclusions: Valid measurement of injury severity is important for both the meaningful monitoring of trends and to assist in classifying information to meet specific injury policy prevention and control needs. This study suggests that the predictive ability of ICISS would be improved if both mortality data and comorbidity were used in the calculation of the injury severity score.

Can the ICD-based injury severity score (ICISS) be improved?

INTRODUCTION

Valid measurement of injury severity is critical to producing valid indicators, as well as for the production of valid information from the analysis of injury data to inform policy and injury prevention practice. Research suggests that the indicators that most New Zealand (NZ) Government departments are currently using to monitor trends in non-fatal injury are potentially misleading because of their inability to remove service-provision and access effects from the underlying trend data.¹

One of the key criteria that injury indicators should satisfy is that they include only serious injuries.¹ That is, events associated with significantly increased risk of impairment, disability, functional limitation, or death, decreased quality of life or increased cost.

In 2003 the NZ Government signed off a New Zealand Injury Prevention Strategy (NZIPS) and indicated that it wished to have valid indicators for measuring performance in the priority areas. In a report commissioned by NZIPS, the Injury Prevention Research Unit (IPRU) developed a set of fatal and serious non-fatal injury indicators. To minimise threats to validity for the non-fatal injury indicators, a well established method known as ICISS (ICD based injury severity score) was used to estimate injury severity.² This method involves estimating probability of death directly from ICD injury diagnoses by examining a large set of cases for which survival status is known. Although the majority of research to date on the development of ICISS has been based on data collected in trauma centres, the same techniques have been successfully applied to national population-level statistics using injury hospitalisations databases.³ NZ maintains a large accurate administrative database of injury hospitalisations (National Minimum Data Set (NMDS)) that contains detailed information on the nature and circumstances of injury in the form of *International Classification of Diseases* (ICD) codes.⁴ The clinical modification of the 10th revision of ICD (ICD-10-AM) was introduced in NZ from 1 July 1999. ICISS has been shown to be capable of providing reliable probability of death estimates based on ICD-10-AM diagnosis codes.⁵

Determining which injuries are “serious” by the ICISS method involves calculating a Survival Risk Ratio (SRR) for each individual injury diagnosis code. An SRR is the proportion of cases with a certain injury diagnosis who do not die or in other words a given SRR represents the likelihood that a patient will survive a particular injury. Using the standard ICISS methodology, each patient’s ICISS (survival probability) is the product of the SRRs of all of that case’s diagnoses.² For cases that have an isolated injury, their ICISS is just the SRR of the particular injury diagnosis code they have. For those with multiple injuries, all listed diagnoses codes are considered equally with no consideration given to whether a particular injury diagnosis is the primary diagnosis or, for example, the 20th. This aspect of the ICISS methodology has been queried and alternative approaches suggested. In 2003, Kilgo et al. compared the standard multiple injury ICISS to a worst-injury ICISS that, rather than being the product of the SRRs, was just the smallest SRR amongst the diagnoses for an individual. Results of this research indicated that worst injury discriminated survival better, fitted better and explained more variance than the multiple injury ICISS.⁶

Another concern that has been raised is that rather than estimating the probability of death, previous research applying the ICISS methodology to injury hospitalisations has estimated the *probability of death in hospital given admission*.^{3 5} In some cases individuals die from their injuries after being discharged from hospital. If these individuals could be included in the set of cases the *probability of death given admission* would be estimated. In addition to these deaths are those where the individual is never admitted to hospital (e.g. they die at the scene of the injury). Thus the ideal set of cases would include all injury deaths irrespective of whether they were admitted to hospital or not as this gives an estimate of *probability of death*. As NZ has routinely collected mortality data with ICD-10-AM diagnosis codes since 1999⁷, the application of ICISS methodology to an integrated injury hospitalisation and mortality dataset is possible.

The first aim of this research was to determine whether the inclusion of injury deaths that occurred outside of hospital (i.e. those that were never admitted to hospital and those that occurred after discharge from hospital) improved the predictive ability of ICISS.

As the severity of an injury is conceivably partly dependent on any non-injury conditions the patient has, the second aim of this research was to determine whether taking account of comorbidity could improve the predictive ability of the ICISS. Although a number of approaches exist for assessing comorbidity in routinely-collected data, there is no obvious gold standard. A prominent measure in literature concerned with the effects of comorbidity in health outcomes and burden of care is the Charlson Comorbidity Index (CCI).⁸ Of the thirty clinically important comorbidities defined by Charlson et al., the proportional hazards model used indicated that nineteen were independent predictors of one-year mortality. The CCI was first adapted to administrative databases in 1993 with the resulting algorithm subsequently developed further to generate 17 yes/no Charlson comorbidity variables for each patient record.⁹

The Harborview Assessment for Risk of Mortality (HARM) score presents an alternative approach for including comorbidity.¹⁰ The HARM score includes ICD-10 diagnoses codes for anatomic injury, mechanism, intent, comorbidity as well as age. The 11 comorbid conditions considered for the HARM model were based on work by Morris et al in 1990.¹¹

The age of an individual has the potential to impact on the severity of their injuries. Including age in the ICISS model has the potential though to complicate the analysis as age is correlated with comorbidity. As one of the aims of this study was to look at the specific impact of comorbidity on the predictive ability of ICISS, the approach taken did not include age.

Investigating whether the measurement of injury severity can be improved by modifying the ICISS methodology is important as it presents the potential to produce more trustworthy injury statistics for the purposes of measuring the impact of policy and practice in reducing injury in NZ and overseas.

The purpose of this study was to determine whether the predictive ability of ICISS can be improved first, by including deaths that occurred outside of hospital and secondly, through the inclusion of either a Charlson or HARM comorbidity component.

Can the ICD-based injury severity score (ICISS) be improved?

METHODS

Data

The NMDS records all publicly funded inpatient treatment of injuries in NZ hospitals.⁴ During the period 1 January 2000 to 31 December 2003, there were 318,394 hospital discharges in the NMDS with a principal diagnosis in the range S00-T89 (Injury and Poisoning chapter excluding sequelae). Readmissions for the same injury were identified using an approach previously described elsewhere.¹²

In the Mortality Collection there were 111,278 fatalities registered between 1 January 2000 and 31 December 2003, of which 6,641 had at least one diagnosis in the range S00-T89. Linkage between the NMDS and Mortality Collection was deterministic using patients Master National Health Index (NHI).¹ For individuals that were not hospitalised but died, diagnoses from the Mortality Collection were used; NMDS diagnoses were used for all other cases.

The dataset on which the analysis was based was obtained from the union of the following:

- d1) Hospitalisations with an ICD-10-AM principal diagnosis in the range S00-T89 discharged dead at any admission within 90 days of their injury date where their date of injury was between 1 January 2000 and 31 August 2003 excluding those readmissions where the first admission did not have an ICD-10-AM principal diagnosis in the range S00-T89
- d2) First admission hospitalisations with an ICD-10-AM principal diagnosis in the range S00-T89 that had an injury date between 1 January 2000 and 31 August 2003 that either stayed at least one night in hospital or died within 90 days
- d3) Fatalities located in the Mortality Collection with date of death² between 1 January 2000 and 31 August 2003 who had an ICD-10-AM diagnosis in the range S00-T89 in any field.

¹ Records for individuals in the NMDS and the Mortality Collection may have both a Master NHI and an Event NHI. This is because one person may have more than one NHI. One of the main tasks of the NZHIS Data Quality team is to identify and link records such as these. In these situations both NHIs remain valid and attached to the person's details but the earlier NHI is referred to as the Master NHI and is the preferred one for future use. Instances where 2 people are using the exact same NHI are extraordinarily rare and are resolved by the creation of a new Master NHI.

² Date of injury is not recorded in the NZHIS Mortality Collection.

Can the ICD-based injury severity score (ICISS) be improved?

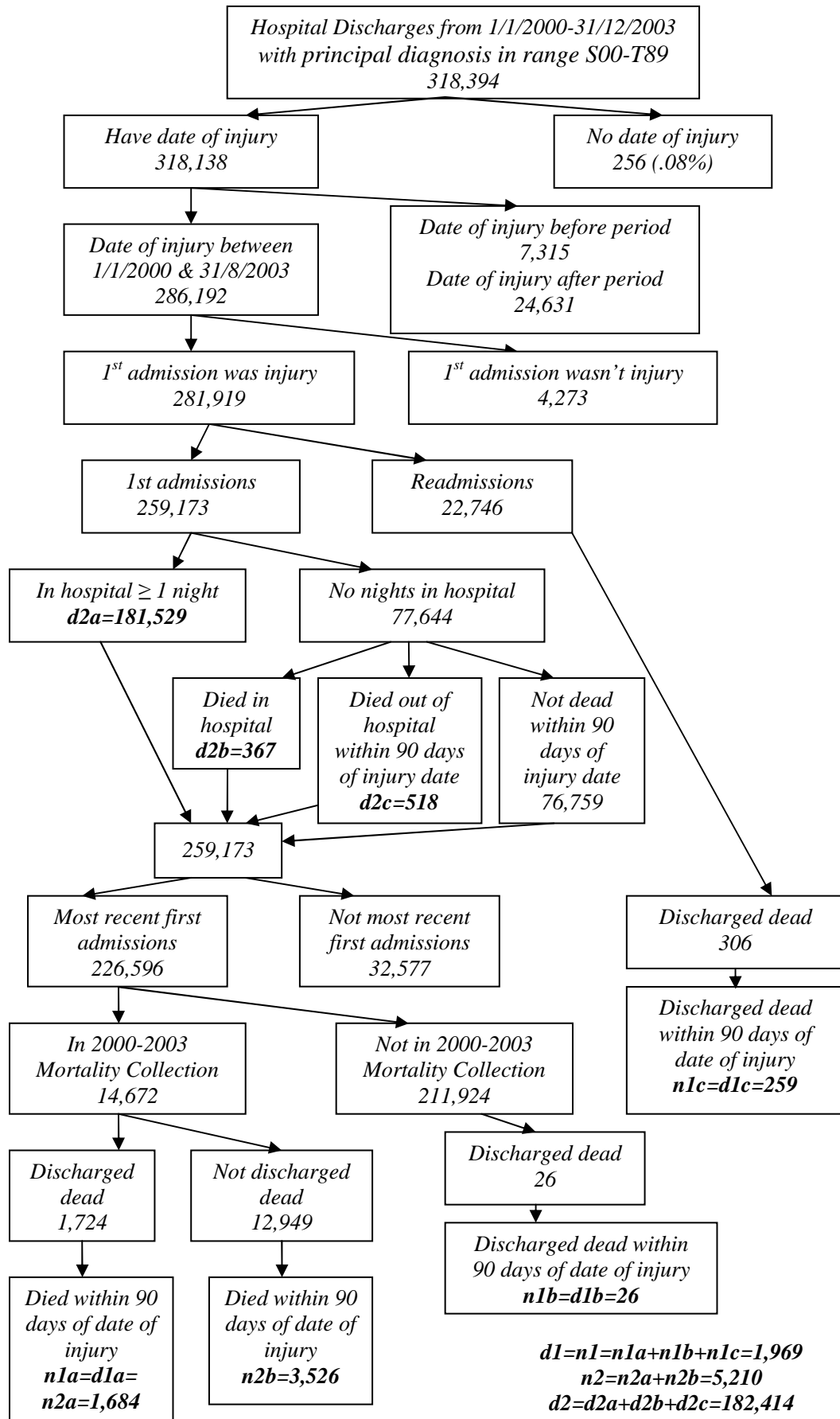


Figure 1. Selection of cases from National Minimum Data Set

There were 1,969 cases that satisfied criteria d1 and 182,414 and 6,008 that satisfied d2 and d3 respectively (Figures 1 and 2). The distribution and intersection of these is presented in Figure 3. Overall there were 186,835 injury cases identified in this way. These were from 167,479 people.

Outcome

The outcome measured for each case was survival/death.

As an international convention could not be located (except for Motor Vehicle Traffic Crashes) on the maximum time lapse between date of injury and date of death to use when classifying injury deaths, an enquiry was sent to international colleagues via the International Collaborative Effort (ICE) discussion list. This did not elicit a clear response so an approach based on empirical evidence was developed. This evidence was that among injury hospitalisations that were discharged dead, 95% of them died in hospital within 90 days of admission for injury. When a 90 day cut-off was proposed to international colleagues it was well accepted; it was thus adopted for this study.

Cases were coded as ‘dead’ if they satisfied any of the following criteria:

- n1) Hospitalisations with an ICD-10-AM principal diagnosis in the range S00-T89 discharged dead at any admission within 90 days of their injury date where their date of injury was between 1 January 2000 and 31 August 2003 excluding those readmissions where the first admission did not have an ICD-10-AM principal diagnosis in the range S00-T89
- n2) Most recent first admission hospitalisations with an ICD-10-AM principal diagnosis in the range S00-T89 and an injury date between 1 January 2000 and 31 August 2003 that could be located in the subset of the 2000-2003 Mortality Collection where the date of death was within 90 days of the injury date
- n3) Fatalities located in the Mortality Collection with date of death³ between 1 January 2000 and 31 August 2003 who had an ICD-10-AM diagnosis in the range S00-T89 in any field.

³ Date of injury is not recorded in the NZHIS Mortality Collection.

There were 1,969 cases that satisfied criteria n1 and 5,210 and 6,008 that satisfied n2 and n3 respectively (Figures 1 and 2). The distribution and intersection of these is presented in Figure 3. Overall 9,968 (5.3%) were classified as dead.

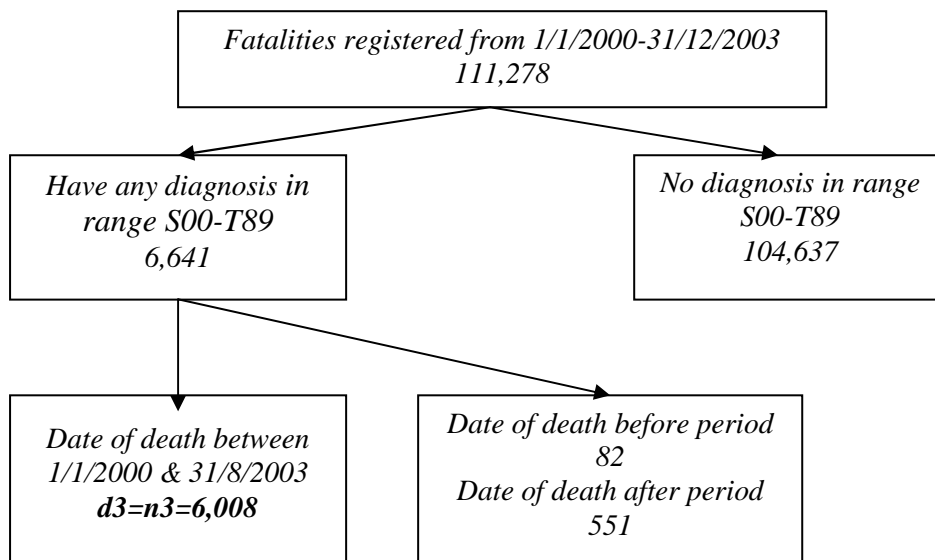


Figure 2. Selection of cases from the Mortality Collection

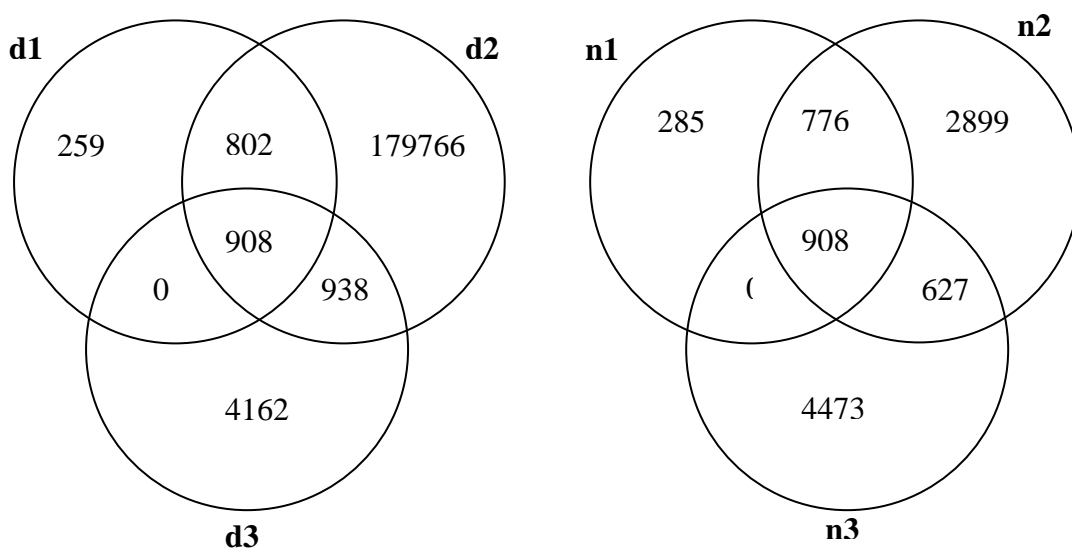


Figure 3. Distribution of denominators (d) and numerators (n) that satisfy the selection criteria

Comorbidity

Two approaches to including comorbidity were considered: the comorbidity component of the HARM score and the comorbidity variables considered by the CCI.

The HARM score incorporated the comorbid conditions listed in Table 1.¹⁰ These conditions were defined using ICD-9-CM diagnoses codes so forward and back-mapping was used to determine the appropriate ICD-10-AM codes.

Table 1. Comorbid conditions as defined by Charlson and HARM

Charlson	HARM
1 Myocardial Infarction	1 Ischaemic Heart Disease
2 Congestive Heart Failure	2 Hypertension
3 Peripheral Vascular Disease	3 Psychoses
4 Cerebrovascular Disease	4 Chronic Obstructive Pulmonary Disease
5 Dementia	5 Diabetes mellitus
6 Chronic Pulmonary Disease	6 Cirrhosis
7 Connective tissue disease	7 Congenital Coagulopathy
8 Peptic Ulcer Disease	8 Epilepsy
9 Mild Liver Disease	9 Obesity
10 Diabetes without complications	10 Alcohol or drug dependence
11 Diabetes with complications	11 Neurological degenerative disease
12 Paraplegia and Hemiplegia	
13 Renal Disease	
14 Cancer	
15 Moderate or severe liver disease	
16 Metastatic carcinoma	
17 AIDS/HIV	

The CCI, which reflects the cumulative increase in likelihood of one-year mortality due to the severity of the effect of comorbidities, was calculated using a Stata module ‘charlson’ that maps ICD-9-CM Charlson comorbidity diagnoses codes to ICD-10.¹³ Minor modifications were required to adapt the Stata module for ICD-10-AM.¹⁴ The 17 Charlson comorbid conditions are listed in Table 1. Although the data contained multiple patient records for some individuals, comorbidity information was calculated independently for each case.

Comorbidity SRRs were calculated both at the ICD-10-AM code level (i.e. one SRR calculated for each ICD-10-AM code within a given comorbidity) and at the variable level (i.e. one SRR calculated for a given comorbidity).

Table 2. Distinctions between the 10 ICD-based Injury Severity Scores

ICISS	Mortality Collection	Comorbidity	
		Approach	Level
ICISS1 (traditional)*	No	-	-
ICISS2	No	HARM	ICD-10-AM code
ICISS3	No	HARM	Variable
ICISS4	No	Charlson	ICD-10-AM code
ICISS5	No	Charlson	Variable
ICISS6	Yes	-	-
ICISS7	Yes	HARM	ICD-10-AM code
ICISS8	Yes	HARM	Variable
ICISS9	Yes	Charlson	ICD-10-AM code
ICISS10	Yes	Charlson	Variable

*The traditional ICISS is calculated from all diagnoses available in the hospital discharge data only. Survival is obtained from discharge status.

Calculation of ICISS

Ten ICISS scores were calculated for each case. The distinctions between these are listed in Table 2.

SRRs for the ICISS scores that did not use data from the Mortality Collection were calculated using a subset of the data that included n1, d1 and d2 only (Table 3). All cases with a particular diagnosis code (e.g. S01.1) listed anywhere on the hospital discharge record or Mortality Collection, where applicable, were included in the calculation of the SRR for that particular diagnosis code. All ICISS scores were calculated using the full dataset (Table 3).

Table 3. Data used to calculate the 10 SRRs and their corresponding ICD-based Injury Severity Scores

Measure	Numerator	Denominator	Deaths	Total	% dead
SRR1 - SRR5	n1	d1, d2	1,969	182,673	1.08
SRR6 - SRR10	n1, n2, n3	d1, d2, d3	9,968	186,835	5.34
ICISS1 - ICISS10	n1, n2, n3	d1, d2, d3	9,968	186,835	5.34

For the 13,349 linked cases, a comparison of the average number of diagnoses per person in the NMDS (mean = 6.2) to that from the Mortality Collection (mean=1.9) indicated that the number of diagnoses recorded was different in these two administrative databases. Due to concern about the bias this may introduce, Kilgo's 'worst injury' methodology was applied.⁶ The logic being that for cases where there were multiple injuries but variable recording of the number of diagnoses, it is not unreasonable to assume that as a minimum the worst injury would be recorded. Thus the ICISS1 was calculated as follows:

$$\text{ICISS} = \text{smallest}(\text{injurySRR}_1, \text{injurySRR}_2, \dots, \text{injurySRR}_n)$$

For the ICISS scores that included comorbidity, different sets of SRRs were calculated at the HARM and Charlson variable-level and at the ICD-10-AM code level. Consistency between the two parts of the method was obtained by having the comorbid SRRs contribute to the ICISS scores as follows:

$$\text{ICISS} = \text{smallest}(\text{injurySRR}_1, \text{injurySRR}_2, \dots, \text{injurySRR}_n) * \\ \text{smallest}(\text{comorbiditySRR}_1, \text{comorbiditySRR}_2, \dots, \text{comorbiditySRR}_m)$$

Statistical Analysis

The full dataset (186,835 cases including 9,968 deaths) was used to validate the 10 injury severity scores.

Logistic regression models using ICISS as the predictor variable and survival as the outcome variable were used to assess the performance of the severity scores.

According to Harrell et al.¹⁵, there are two grounds on which the different severity measures can be compared: discrimination and calibration.

Discrimination is the ability of the model to distinguish survivors from non-survivors. This was measured by concordance 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 Operating Characteristic (ROC) curve.¹⁵ Bootstrapping was used to correct the concordance values for bias caused by the use of a single data set for design and testing.^{16 17} Calibration curves (plots of observed against estimated mortality with cases grouped by estimated mortality) and the Hosmer-Lemeshow (H-L) 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 a H-L statistic of zero.¹⁸ The higher the H-L statistic the poorer the fit.¹⁹ The statistical significance of the H-L statistic has not been assessed, as it is inappropriate to do so with such a large sample.²⁰

From the logistic regression models, R^2 values were derived. The R^2 is a descriptive goodness-of-fit measure between 0 and 1 that describes the proportion of variance explained by the model; higher values are better.

Stata version 9.2 was used for all statistical analysis.²¹

RESULTS

The bootstrap adjusted concordance values, H-L statistics and R^2 statistics for all 10 ICISS models are listed in Table 4.

Table 4. Outcome Statistics from the Logistic Regression Models

ICISS	Concordance	95% CI*	H-L	R^2
ICISS1	0.777	(0.772 , 0.783)	2757	0.123
ICISS2	0.800	(0.795 , 0.806)	1352	0.163
ICISS3	0.794	(0.790 , 0.798)	1361	0.150
ICISS4	0.818	(0.813 , 0.823)	1673	0.184
ICISS5	0.816	(0.811 , 0.821)	1710	0.175
ICISS6	0.851	(0.848 , 0.855)	2222	0.227
ICISS7	0.874	(0.870 , 0.877)	1233	0.282
ICISS8	0.866	(0.863 , 0.870)	1256	0.262
ICISS9	0.891	(0.888 , 0.894)	926	0.328
ICISS10	0.885	(0.882 , 0.888)	910	0.301

*Bootstrap adjusted 95% confidence intervals (CI)

ICISS9 which included both mortality data and Charlson comorbid conditions at the ICD-10-AM level had the best concordance. Scores calculated using the hospitalisation and mortality data (ICISS6-ICISS10) all had better concordance than those calculated using only the hospitalisation data (ICISS1-ICISS5). The score that was calculated in the traditional way using hospitalisations only (ICISS1) had the worst concordance. The scores calculated using comorbidity data (ICISS2-ICISS5; ICISS7-ICISS10) had higher concordance than their corresponding score that didn't include comorbidity (ICISS1; ICISS6). The inclusion of comorbidity using the Charlson conditions (ICISS4-ICISS5, ICISS9-ICISS10) produced scores with better concordance than those that used HARM approach to including comorbidity (ICISS2-ICISS3, ICISS7-ICISS8). The scores calculated using comorbidity SRRs calculated at the ICD-10-AM level (ICISS2, ICISS4, ICISS7, ICISS9) had higher concordance than their respective scores calculated using SRRs at the comorbidity variable level (ICISS3, ICISS5, ICISS8, ICISS10).

Calibration was assessed by the H-L statistics and the calibration curves presented in Figure 4 (ICISS1-ICISS5) and Figure 5 (ICISS6-ICISS10). Previous work by Stephenson produced calibration curves that presented all observations.⁵ We believe presenting all observations is potentially misleading and irrelevant for the broad purpose of this research – identifying a threshold for defining serious injuries. Currently in work commissioned by NZIPS, injuries with an ICISS ≤ 0.941 (about 15% of all non-fatal hospitalisations) are classified as serious.²² Since the vast majority of cases have ICISS scores close to 1 (low estimated mortality), only cases with estimated mortality of 30% or less are presented in Figures 4 and 5. This corresponds to presenting 90-99% of the data depending on which of the 10 ICISS measures is being referred to.

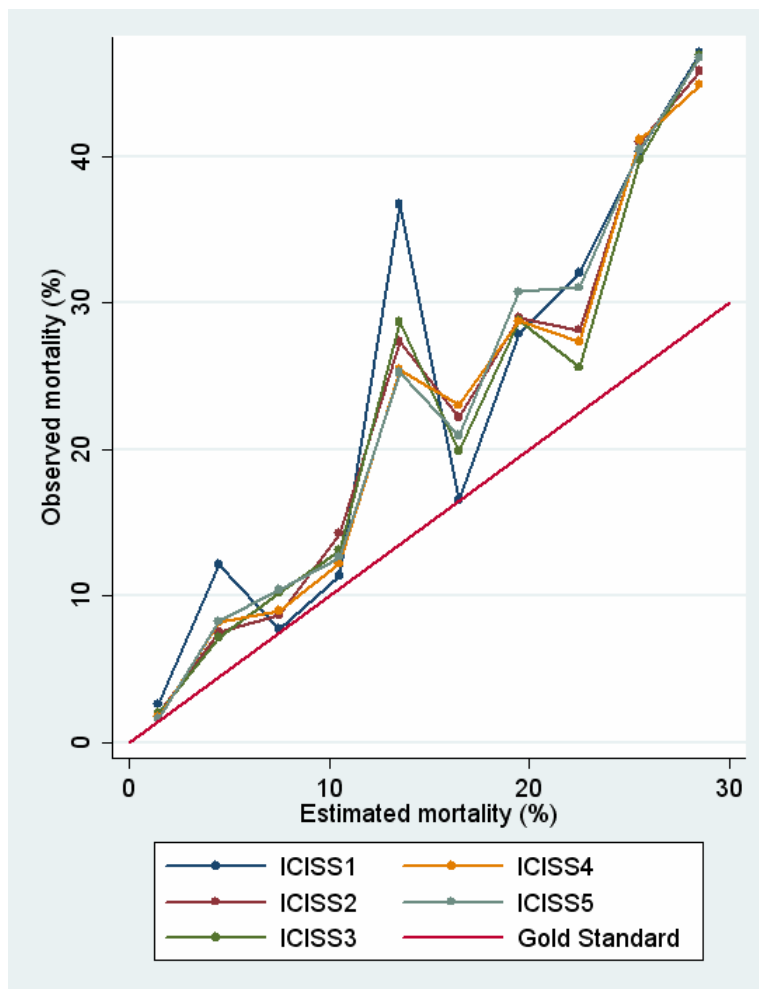


Figure 4. Calibration curves for ICD-based Injury Severity Scores calculated from Survival Risk Ratios based on hospital discharge data only

The differences in performance between the scores are difficult to assess through the calibration curves. It is clearly evident though through the curves that the scores that used hospital discharge data only to calculate SRRs (ICISS1-ICISS5) underestimated mortality whereas the scores that used hospital discharge data and mortality data (ICISS6-ICISS10) overestimated mortality. Calibration was generally better at lower levels of estimated mortality.

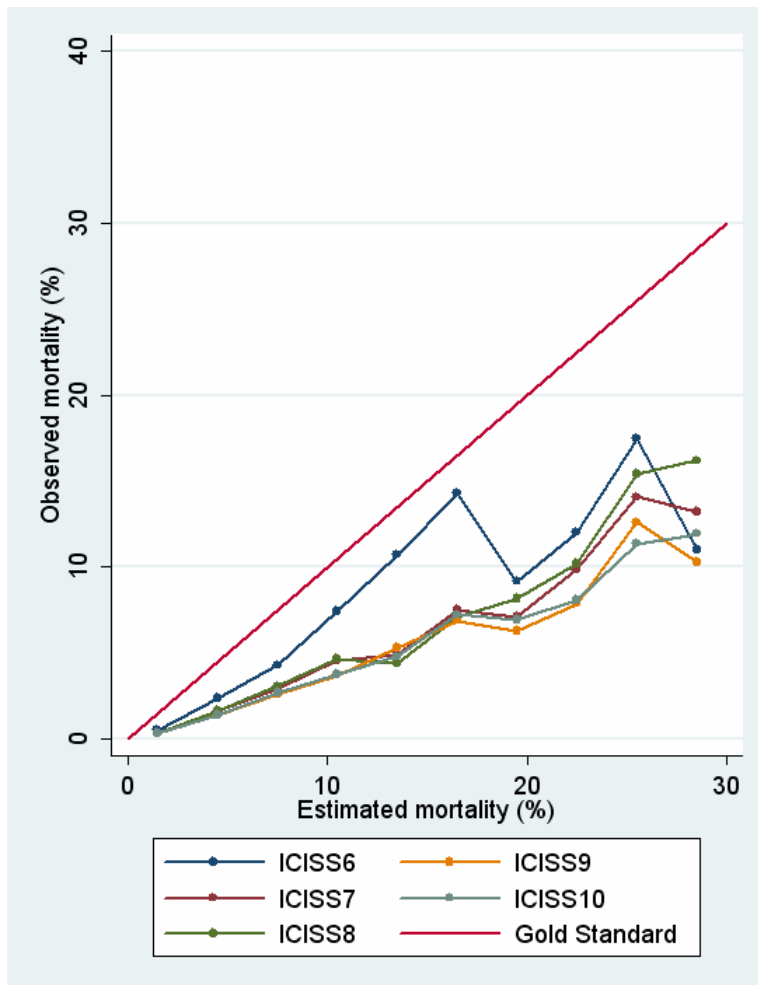


Figure 5. Calibration curves for ICD-based Injury Severity Scores calculated from Survival Risk Ratios based on hospital discharge and mortality data

ICISS10 which included both mortality data and Charlson comorbid conditions at the variable level had the lowest H-L statistic although the calibration of ICISS9 was very similar. Scores calculated using the hospitalisation and mortality data and taking account of comorbidity (ICISS7-ICISS10) had lower H-L statistics than the other

scores. The score that was calculated in the traditional way using hospitalisations only (ICISS1) had the worst calibration. The scores calculated using comorbidity data (ICISS2-ICISS5; ICISS7-ICISS10) had lower H-L statistics than their corresponding score that didn't include comorbidity (ICISS1; ICISS6). For estimated mortality less than 20%, the calibration curve for ICISS6 was noticeably closer to the Gold Standard than for other scores using hospitalisation and mortality data (ICISS7-ICISS10). The inclusion of comorbidity using the Charlson conditions produced scores with better calibration than those that used HARM approach to including comorbidity for ICISS7-ICISS10. For ICISS2-ICISS5, this was reversed. For all except ICISS9-ICISS10, the scores calculated using comorbidity SRRs calculated at the ICD-10-AM level (ICISS2, ICISS4, ICISS7) had lower H-L statistics than their respective scores calculated using SRRs at the comorbidity variable level (ICISS3, ICISS5, ICISS8).

The same order of performance observed for concordance was seen for the R^2 statistics. ICISS9 which included both mortality data and Charlson comorbid conditions at the ICD-10-AM level explained 33% of the variation in the model. This was close to three-times higher than that observed for the score with the lowest R^2 statistic, ICISS1.

The results by diagnosis group are listed in Table 5. "Head injuries" represented 13% of the cases with "other mechanical trauma" representing 61%. 13% were "complications" and 12% "other injuries".

"Head injuries" had the highest concordance and "complications" the least. The range of concordance scores for "other mechanical trauma" and "other injuries" was similar. Within each diagnostic group, ICISS9 had the best concordance. Most of the trends in concordance by diagnostic groups followed those reported in Table 4.

Within the diagnostic groups, calibration, measured by the H-L statistic, did not follow the same patterns as observed in Table 4. For "head injuries", ICISS7 had best calibration whereas for "other mechanical trauma", ICISS3 had the lowest H-L statistic. For "complications", the best calibrated score was ICISS9 and for "other injuries" the score with the best calibration was ICISS8.

Table 5. Outcome Statistics from the Logistic Regression Models by Diagnosis group

Diagnosis group	ICISS	Concordance	95% CI*	H-L**	R ²
Head Injuries (S00-S09)	ICISS1	0.834	(0.824 , 0.843)	217	0.309
	ICISS2	0.852	(0.842 , 0.862)	257	0.316
	ICISS3	0.850	(0.838 , 0.862)	238	0.313
	ICISS4	0.852	(0.842 , 0.863)	273	0.318
	ICISS5	0.853	(0.844 , 0.862)	280	0.318
	ICISS6	0.886	(0.878 , 0.894)	139	0.410
	ICISS7	0.907	(0.900 , 0.914)	103	0.428
	ICISS8	0.904	(0.897 , 0.911)	115	0.421
	ICISS9	0.909	(0.902 , 0.915)	126	0.439
	ICISS10	0.907	(0.900 , 0.914)	121	0.433
Other mechanical trauma (S10-S14)	ICISS1	0.784	(0.776 , 0.791)	1170	0.098
	ICISS2	0.818	(0.811 , 0.824)	732	0.169
	ICISS3	0.812	(0.803 , 0.820)	598	0.159
	ICISS4	0.828	(0.820 , 0.836)	836	0.193
	ICISS5	0.825	(0.818 , 0.832)	647	0.191
	ICISS6	0.848	(0.842 , 0.854)	1711	0.159
	ICISS7	0.880	(0.876 , 0.885)	701	0.249
	ICISS8	0.873	(0.869 , 0.877)	796	0.221
	ICISS9	0.893	(0.888 , 0.898)	678	0.284
	ICISS10	0.887	(0.884 , 0.891)	726	0.261
Complications (T80-T88)	ICISS1	0.650	(0.632 , 0.668)	283	0.023
	ICISS2	0.678	(0.665 , 0.690)	200	0.067
	ICISS3	0.671	(0.659 , 0.683)	398	0.049
	ICISS4	0.753	(0.742 , 0.764)	179	0.116
	ICISS5	0.751	(0.739 , 0.763)	271	0.091
	ICISS6	0.719	(0.706 , 0.733)	108	0.065
	ICISS7	0.736	(0.723 , 0.749)	74	0.104
	ICISS8	0.725	(0.712 , 0.739)	112	0.082
	ICISS9	0.820	(0.809 , 0.830)	46	0.229
	ICISS10	0.790	(0.777 , 0.804)	70	0.143
Other injuries (T15-T79, T89)	ICISS1	0.802	(0.790 , 0.814)	906	0.101
	ICISS2	0.783	(0.774 , 0.793)	1038	0.090
	ICISS3	0.753	(0.742 , 0.764)	736	0.084
	ICISS4	0.801	(0.790 , 0.812)	2174	0.094
	ICISS5	0.794	(0.783 , 0.805)	1025	0.096
	ICISS6	0.888	(0.879 , 0.896)	226	0.347
	ICISS7	0.888	(0.879 , 0.896)	142	0.350
	ICISS8	0.878	(0.870 , 0.887)	116	0.339
	ICISS9	0.897	(0.889 , 0.905)	127	0.367
	ICISS10	0.895	(0.887 , 0.903)	121	0.362

*Bootstrap adjusted 95% confidence intervals (CI)

** H-L statistics should be compared within, and not between, diagnostic groups

Although within diagnostic groups, the order of performance for the R² statistics did not always follow that of concordance, the best performing score in terms of concordance had the highest R² statistic and the score with the lowest concordance

had the lowest R^2 statistic. For “head injuries”, ICISS9 explained 44% of the variation in the model. In comparison, ICISS9 explained 28%, 23% and 37% for “other mechanical trauma”, “complications” and “other injuries” respectively.

DISCUSSION

The first aim of this study was to investigate whether modifying the ICISS method to include out-of-hospital deaths improves the validity of ICISS. Comparing the performance of ICISS1 and ICISS6 clearly indicates that the predictive ability of ICISS can be enhanced through the use of integrated hospitalisation and mortality data sources to calculate ICISS scores.

Assessing whether the predictive ability of ICISS could be improved through the inclusion of comorbidity was the second aim of this study. Comparison of the performance of the ICISS scores that include comorbidity with those that don't (ICISS2-ICISS5 cf ICISS1 and ICISS7-ICISS10 cf ICISS6) indicates a marked improvement in the performance of ICISS when non-injury diagnoses are taken into account.

Determining the appropriate dataset that would enable both the aims of this research to be met and potential sources of error to be limited was not an easy task and involved a number of unforeseen considerations. It became apparent that the recording and coding of the deaths that occur outside hospital was very different to that which occurs for deaths in hospital. Of concern was that for individuals who were discharged dead from hospital, the average number of diagnoses per person on the NMDS was considerably higher than that from the Mortality Collection. Some additional key problems with the diagnosis information from the Mortality Collection were the lack of diagnosis ordering and the observation that the average number of diagnoses per person depended on whether a post-mortem was carried out. Because of this, for cases where both data from the Mortality Collection and the NMDS was available, only the latter was used. For injury deaths that were not hospitalised prior to the death, the diagnosis information had to be obtained from the Mortality Collection. The use of diagnosis information from different sources depending on whether hospitalisation had occurred prior to the death has the potential to introduce error but no alternative method was apparent.

A linkage of injury deaths that occurred in hospital to injury deaths in the Mortality Collection highlighted some additional concerns. Using a definition of a 'hospitalised

injury death' as one in which a patient with a principal ICD-10-AM diagnosis in the range S00-T89 was discharged dead from a first admission, only 48% of the hospitalised injury deaths had an external cause code as the 'underlying cause' of death in the Mortality Collection. This raises questions as typically injury fatalities are defined as those with an external cause code listed as the underlying cause of death. Similarly only 50% of the hospitalised injury deaths had an S00-T89 diagnosis listed on the electronic record in the Mortality Collection. For this work injury fatalities had to have at least one S00-T89 diagnosis so that SRRs for these diagnoses could be obtained. These findings are not unexpected as it has been previously reported that a range of factors may lead to the exclusion of important or predominant causes of the process leading to the death.²³ Thus for cases with both NMDS and Mortality data, the NMDS diagnoses were used. For cases identified solely from the Mortality Collection, it was assumed that the diagnoses in were recorded correctly and fully. This may not always be the case but in the absence of a proper review of case notes it is impossible to know. Further work that explores the reliability of the diagnoses in the Mortality Collection would be valuable.

With the 10 ICISS models, discrimination was between 0.78 and 0.89. These estimates are in line with one of the two previous studies that have assessed the concordance of ICISS applied to ICD-10 in which the reported concordance was 0.84.²⁴ In the other study by Stephenson et al., the reported concordance was 0.91 for NZ hospitalisations using SRRs based on ICD-10-AM.⁵ In contrast to this current work, the study of Stephenson and colleagues used all diagnoses rather than just the 'worst injury' to calculate ICISS and included age as a categorical variable in the models. In this study, as in that by Stephenson and colleagues, "Complications" had the lowest concordance of the diagnostic groups. It is unclear why concordance should be lower for this diagnostic group.

To validate the injury severity scores a dataset that included both hospital discharge data and data from the Mortality Collection was used. As ICISS1-ICISS5 were calculated using SRRs from hospital discharge data only, it was expected that these models would be poorly calibrated due to this data only including 20% of the deaths attributed to injury. The high H-L statistics for ICISS1 – ICISS5 and the underestimation of mortality in the calibration curves are thus not surprising.

It appears that the injury diagnostic codes used in the Mortality Collection tend to be less specific than those in the NMDS. For example, in hospitalised injury death cases with both NMDS and Mortality Collection diagnoses, S09.9 which represents unspecified head injury was used more than twice as often as the first diagnosis in the Mortality Collection than in the NMDS. This phenomenon may explain the overestimation of mortality by models ICISS6 – ICISS10 that include deaths from the Mortality Collection.

In this study the Hosmer-Lemeshow statistic was used to summarise calibration and to compare the relative performance of the models. The authors accept the Hosmer-Lemeshow statistic has limitations. Hosmer and Lemeshow, in describing the statistic, suggested that it should be used where at least one of the predictor variables is continuous so that almost all observations have a unique set of predictors.¹⁹ The consequences of not having a continuous predictor are unclear. Although other approaches could have been used, it was thought this statistic was a reasonable choice for this purpose.²⁵

Rather than using the standard ICISS methodology and obtaining the product of all SRRs for every injury diagnosis listed for a case, this work used the only the SRR relating to the worst injury. This restricted the variation in ICISS scores and limited the number of covariate patterns in the ten models. For ICISS1, the 186,835 cases produced only 458 unique injury severity scores. The inclusion of comorbidity increased the number of covariate patterns to between 3508 and 9231. This lack of variation may have limited the predictive ability of the ICISS scores.

Using the Mortality data enabled the inclusion of deaths that occurred in individuals who were never admitted to hospital and deaths that occurred after discharge from hospital. The number of deaths detected in this way far outweighed the number located using the hospital discharge data alone. Given 29% of the deaths in this study were included solely because they satisfied our criteria that any death that occurred within 90 days of an injury be attributed to the injury, it would be interesting to determine whether the conclusions reached by this study would vary if a shorter time lapse between date of injury and death was used.

In this study only the smallest SRR from the comorbid conditions was included in the calculation of the injury severity score. This approach differs greatly from that of HARM who included, in addition to other terms, separate terms in a logistic regression model for each of the 11 comorbidity conditions used by West and colleagues. The coefficients of this model which also included mechanism, intent and age were used to obtain the HARM score. Additional research into the predictive ability of ICISS using an approach similar to that of West and colleagues may indicate further gains.

Rather than selecting the worst SRR from a set of 11 SRRs calculated for the HARM comorbid variables or from a set of 17 SRRs for the Charlson comorbid variables, allowing a separate SRR for each of the ICD-10-AM codes considerably increased the set of SRRs from which the worst was chosen. In this dataset the 11 HARM comorbid variables produced 331 SRRs corresponding to the ICD-10-AM codes. Over 500 SRRs at the ICD-10-AM level were obtained from the 17 Charlson comorbid variables. Not surprisingly, the larger the set of SRRs from which the worst was chosen, the better the ICISS performed.

The age of individuals was not included in this analysis. Further work should look at the importance of including age in ICISS, whether the inclusion of comorbidity is usefully better than the simpler item, age and whether there is justification for including both age and comorbidity into ICISS. One possible way of including age in ICISS would be to calculate age SRRs either for individual ages or for age groups and multiply these by the worst comorbid SRR and the worst diagnostic SRR.

Valid measurement of injury severity is important for both the meaningful monitoring of trends and to assist in classifying information to meet specific injury policy prevention and control needs. This study suggests that the predictive ability of the standard ICISS method of determining injury severity would be improved if both mortality data and comorbidity were considered. Whether other researchers have the data available to use these additional sources and whether they consider that the gains will outweigh the extra effort depends on their individual situation.

ACKNOWLEDGEMENTS

This research was supported by a grant from Official Statistics Research, Statistics New Zealand (SNZ). The authors wish to thank Paul Brown, Information Manager, SNZ for sponsoring this research and the anonymous reviewer for his helpful comments on the draft.

Can the ICD-based injury severity score (ICISS) be improved?

REFERENCES

1. Cryer C, Langley J, Jarvis S, MacKenzie S, Stephenson S, Heywood P. Injury outcome indicators: the development of a validation tool. *Injury Prevention* 2005;11:53-57.
2. Osler T, Rutledge R, Deis J, Bedrick E. ICISS: An International Classification of Disease-9 Based Injury Severity Score. *Journal of Trauma Injury, Infection and Critical Care* 1996;41(3):380-388.
3. Stephenson SCR, Langley JD, Civil ID. Comparing measures of injury severity for use with large databases. *Journal of Trauma Injury, Infection and Critical Care* 2002;53(2):326-332.
4. New Zealand Health Information Service. National Minimum Data Set Dictionary. Version 6.6: Wellington: New Zealand Ministry of Health, 2006.
5. Stephenson S, Henley G, Harrison JE, Langley JD. Diagnosis based injury severity scaling: investigation of a method using Australian and New Zealand hospitalisations. *Injury Prevention* 2004;10:379-383.
6. Kilgo PD, Turner MO, Meredith W. The Worst Injury Predicts Mortality Outcome the Best: Rethinking the Role of Multiple Injuries in Trauma Outcome Scoring. *The Journal of Trauma Injury, Infection, and Critical Care* 2003;55(4):599-606.
7. New Zealand Health Information Service. Mortality Collection Data Dictionary. Version 1.2. Wellington: Wellington: New Zealand Ministry of Health, 2004.
8. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *Journal of Chronic Diseases* 1987;40(5):373-383.
9. Romano R, Roost L, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *Journal of Clinical Epidemiology* 1993;46:1075-79.
10. West TA, Rivara FP, Cummings P, Jurkovich GJ, Maier RV. Harborview Assessment for Risk on Mortality: An Improved Measure of Injury Severity on the Basis of ICD-9-CM. *The Journal of Trauma Injury, Infection, and Critical Care* 2000;49(3):530-541.
11. Morris JA, MacKenzie EJ, Edelstein SL. The effect of preexisting conditions on mortality in trauma patients. *Journal of the American Medical Association* 1990;263:1942-6.
12. Langley JD, Stephenson SCR, Cryer C, Borman B. Traps for the unwary in estimating person based injury incidence using hospital discharge data. *Injury Prevention* 2002;8:332-7.
13. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Medical Care* 2005;43(11):1130-1139.
14. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson Comorbidity Index predicted in-hospital mortality. *Journal of Clinical Epidemiology* 2004;57:1288-1294.
15. Harrell F, Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy and measuring and reducing errors. *Statistics in Medicine* 1996;15:361-87.
16. Steyerberg EW, Harrell FEJ, Borsboom GJ, Eijkemans MJC, Vergouwe Y, Habberma DF. Internal validation of predictive models: efficiency of some

- procedures for logistic regression analysis. *Journal of Clinical Epidemiology* 2002;54:774-81.
17. Harrell FEJ, Lee KL, Mark DB. Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 1996;15:361-87.
 18. Rowan K, Kerr J, McPherson K, Short A, Vessey MP. Intensive Care Society's APACHE II study in Britain and Ireland – II: outcome comparisons of intensive care units after adjustment for case-mix by the American APACHE II method. *British Medical Journal* 1993;307:977-981.
 19. Hosmer D, Lemeshow S. *Applied Logistic Regression*. New York: John Wiley and Sons Inc, 1989.
 20. Sacco W, MacKenzie E, Champion H, Davis E, Buckman R. Comparison of alternative methods for assessing injury severity based on anatomic descriptors. *Journal of Trauma Injury, Infection and Critical Care* 1999;47:922-925.
 21. Stata Statistical Software: Release 9 [program]. College Station, TX: StataCorp LP, 2005.
 22. Cryer C, Davie G, Langley JD. A chartbook of the New Zealand Injury Prevention Strategy serious injury outcome indicators 1994-2004. Dunedin: Injury Prevention Research Unit and Accident Compensation Corporation, 2006:1-106.
 23. Goldacre M, Roberts S, D Y. Mortality after admission to hospital with fractured neck of femur: database study. *British Medical Journal* 2002;325(7369):868-9.
 24. Kim Y, Jung Y, Kim C, Y K, Y S. Validation of the international classification of diseases 10th edition-based injury severity scores (ICISS). *Journal of Trauma Injury, Infection and Critical Care* 2000;48:280-285.
 25. Hosmer DW, Hosmer T, Le Cressie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine* 1997;16:965-980.