Systematic review of predictive performance of injury severity scoring tools

Hideo Tohira
Ian Jacobs
David Mountain
Nicholas Gibson
Allen Yeo

Follow this and additional works at: https://ro.ecu.edu.au/ecuworks2012

Part of the Nursing Commons

10.1186/1757-7241-20-63
Tohira, H., Jacobs, I., Mountain, D., Gibson, N. P., & Yeo, A. (2012). Systematic review of predictive performance of injury severity scoring tools. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 20(art. no. 63), 1-12. Available here
This Journal Article is posted at Research Online. https://ro.ecu.edu.au/ecuworks2012/214
Systematic review of predictive performance of injury severity scoring tools

Hideo Tohira1*, Ian Jacobs1, David Mountain1, Nick Gibson1 and Allen Yeo2

Abstract
Many injury severity scoring tools have been developed over the past few decades. These tools include the Injury Severity Score (ISS), New ISS (NISS), Trauma and Injury Severity Score (TRISS) and International Classification of Diseases (ICD)-based Injury Severity Score (ICISS). Although many studies have endeavored to determine the ability of these tools to predict the mortality of injured patients, their results have been inconsistent. We conducted a systematic review to summarize the predictive performances of these tools and explore the heterogeneity among studies. We defined a relevant article as any research article that reported the area under the Receiver Operating Characteristic curve as a measure of predictive performance. We conducted an online search using MEDLINE and Embase. We evaluated the quality of each relevant article using a quality assessment questionnaire consisting of 10 questions. The total number of positive answers was reported as the quality score of the study. Meta-analysis was not performed due to the heterogeneity among studies. We identified 64 relevant articles with 157 AUROCs of the tools. The median number of positive answers to the questionnaire was 5, ranging from 2 to 8. Less than half of the relevant studies reported the version of the Abbreviated Injury Scale (AIS) and/or ICD (37.5%). The heterogeneity among the studies could be observed in a broad distribution of crude mortality rates of study data, ranging from 1% to 38%. The NISS was mostly reported to perform better than the ISS when predicting the mortality of blunt trauma patients. The relative performance of the ICISS against the AIS-based tools was inconclusive because of the scarcity of studies. The performance of the ICISS appeared to be unstable because the performance could be altered by the type of formula and survival risk ratios used. In conclusion, high-quality studies were limited. The NISS might perform better in the mortality prediction of blunt injuries than the ISS. Additional studies are required to standardize the derivation of the ICISS and determine the relative performance of the ICISS against the AIS-based tools.

Keywords: ISS, NISS, TRISS, ICISS, AUROC, Outcome prediction

Background
Many scoring systems to assess injury severity have been developed over the past few decades. The need to improve the quality of trauma care has led researchers to develop more accurate tools that allow physicians to predict the outcomes of injured patients. The Abbreviated Injury Scale (AIS) was the first comprehensive injury severity scoring system to describe injuries and to measure injury severity [1]. Because the AIS cannot measure the overall injury severity of a patient with multiple injuries, tools that can measure the overall severity of multiple injuries were developed using the AIS. These tools include the Injury Severity Score (ISS) [2], the New Injury Severity Score (NISS) [3] and the Trauma and Injury Severity Score (TRISS) [4].

AIS codes are not always available for all injured patients because of the limited resources available for maintaining a trauma registry. However, the International Classification of Diseases (ICD) codes are routinely collected in administrative databases, including morbidity or mortality databases. Osler et al. introduced the ICD-based Injury Severity Score (ICISS) to overcome the unavailability of AIS-based tools [5]. It was reported that the ICISS performed as well as the AIS-based tools in predicting trauma patient outcomes [6-17].

* Correspondence: soba@tororo.net
1School of Primary, Aboriginal and Rural Health Care, The University of Western Australia, M516 The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia
Full list of author information is available at the end of the article

© 2012 Tohira et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Many researchers have studied the predictive performances of injury severity scoring tools. Their results, however, were inconsistent. This disparity may be due to the differences in study populations and the differences between the formulas used to calculate severity. For example, the ICISS employs at least two types of formulas for its calculation: a product of multiple survival risk ratios (SRRs), referred to as the traditional ICISS, and the use of a single worst injury (SWI) risk ratio. In 2003, Kilgo et al. first reported that the SWI performed better than the traditional ICISS [18]. The superiority of the SWI remains uncertain because of conflicting results from another researcher [19].

The predictive performances of the TRISS and ICISS have rarely been compared. The TRISS utilizes a logistic regression model that incorporates the ISS as a predictor; therefore, the TRISS intuitively outperforms the ISS. In contrast to the TRISS, the ICISS is based on a multiplicative model that uses SRRs. Because the TRISS and ICISS are based on different mathematical models, the superiority of one tool over the other remains inconclusive.

Although several traditional narrative reviews have been conducted, these reviews typically focused on how each tool was established and how each score was derived [20-24]. A few reviews have addressed the predictive performances of injury severity scoring tools [21,23]. However, the methodologies used for selecting studies were unclear, and the interpretation of the results was subjective. It is more appropriate to review studies in a systematic manner to best integrate all of the evidence. Currently, there is no systematic review that evaluates the predictive performances of injury severity scoring tools.

**Aims**

In this systematic review, we aimed to summarize the ability of the injury severity scoring tools that are currently in use to predict the mortality of injured patients. We also aimed to explore the potential sources of the heterogeneity among studies to better understand comparative studies of injury severity scoring tools.

**Methods**

**Injury severity scoring tools**

To investigate predictive performances, we chose the following injury severity scoring tools: the ISS [2], the NISS [3], the TRISS [4] and the ICISS [25]. These tools are hereafter referred to as the “target tools.” These target tools were selected because they were frequently found in injury research articles. We included the TRISS in the target tools to determine the superiority of the TRISS over the ICISS or vice versa.

We subdivided the TRISS and the ICISS further when reporting their predictive performance. We classified the TRISS into two types: the TRISS that used coefficients derived from the Major Trauma Outcome Study (MTOS TRISS) and one that used coefficients derived from non-MTOS populations (non-MTOS TRISS).

We categorized the ICISS into four subgroups based on the type of formula and SRR. There are two types of formula, as described previously. There also are two types of SRR: traditional SRR and independent SRR. An SRR is calculated by dividing the number of survivors from another researcher [19].

The predictive performances of the TRISS and ICISS have rarely been compared. The TRISS utilizes a logistic regression model that incorporates the ISS as a predictor; therefore, the TRISS intuitively outperforms the ISS. In contrast to the TRISS, the ICISS is based on a multiplicative model that uses SRRs. Because the TRISS and ICISS are based on different mathematical models, the superiority of one tool over the other remains inconclusive.

We conducted an online database search in June 2009 using MEDLINE and Embase with predetermined search words (Additional file 1). We set no language restrictions and sought a translation if required. We checked the references of relevant articles. Conference abstracts, letters and unpublished studies were not included.
Finding relevant articles
We carefully examined the titles and abstracts of all of the articles retrieved from the online databases. We read the entire article if the relevance was unclear. After this first screening, we carefully read the complete articles to identify additional relevant articles that fulfilled the predetermined criteria described above.

Information extraction
We extracted information relating to methodologies, study population, injury severity scoring tools and performance scores from relevant articles [see Additional file 2, which includes the information that we sought].

Evaluating articles
We evaluated the quality of each relevant article using a quality assessment questionnaire. Because there is no widely used quality assessment tool for this type of systematic review, we developed a questionnaire to meet the needs of our study by referring to a systematic review of diagnostic tools and outcome prediction models [27-30]. Our questionnaire contains ten questions, which include two questions that were only applicable to the TRISS or the ICISS (Additional file 3). The total number of positive answers to the eight questions that could be applied to all of the target tools was reported as the overall quality score of the study. We did not sum the weights of each question because there is no consensus on how to do so.

Statistical analysis
We did not conduct a meta-analysis in this review because we could not control the heterogeneity among studies by employing random effect models and performing subgroup analyses.

A protocol did not exist for this systematic review, and this review was not prospectively registered.

Results
Relevant articles
In total, we retrieved 5,608 potential articles from the online database search. After carefully reading all of the titles and abstracts and checking references using relevant journals, we finally identified 61 relevant articles written in English [3,7-17,19,31-78] and three non-English articles [79-81]. The total number of relevant articles was 64, with 157 AUROCs of the target tools (Figure 1, Table 1).

In these articles, the ISS was the most frequently studied target tool (58%), followed by the TRISS (53%), the ICISS (31%) and the NISS (25%). The MTOS TRISS were more frequently reported than the non-MTOS TRISS (see Additional file 2 for the details). Regarding the formulas used in ICISS calculation, 32 out of 39 AUROCs were derived from the traditional ICISS, whereas 7 AUROCs were derived from the SWI. There were 33 and 6 AUROCs of the ICISS using traditional and independent SRRs, respectively (Table 1).

Of the 64 relevant studies, 26 studies were conducted in the U.S., and 26 studies used data from a single hospital. Only three studies included data from hospitals in multiple countries (see Additional file 2).

Quality assessment
The results of the quality assessment are shown in Table 2. The distribution of the number of positive answers was positively skewed, and the median was 5 out of 8, ranging from 2 (4 studies) to 8 (2 studies) (Figure 2).

Most studies described the selection criteria for the study subjects and the demographics of the subjects. In contrast, less than half of the studies reported the following items: the version of AIS and/or ICD used (37.5%); the quality assurance measure for collecting and measuring scores (37.5%); and the precision of the AUROCs (48.4%).

Regarding the two questions that were only relevant to the TRISS and ICISS, the majority of studies reported the origin of the coefficients of the TRISS or SRRs of the ICISS (41 out of 52 studies). The TRISS and ICISS that used newly derived coefficients or SRRs were internally or externally validated in 25 out of 28 studies.

Discussion
We identified 64 relevant articles with 157 AUROCs. The ISS was most frequently reported (48 AUROCs), followed by the TRISS (45 AUROCs), ICISS (40 AUROCs) and NISS (24 AUROCs). We could not pool the AUROCs because of the heterogeneity among the studies.

Study quality
There was a scarcity of high-quality studies that investigated the performance of the target tools. Specifically, the version of the injury code system and any quality assurance measure were poorly described.

Most studies described their selection criteria and reported the demographic data of the study population; however, key information that can influence the predictive performance was underreported. An AUROC can be affected by two types of factors: factors that influence the measurement of injury severity scores and those that affect the outcome [82]. The former include the version of the injury codes, type of formula and derivation of coefficients and/or SRRs; the latter includes the distribution of age, mechanism of injury and inclusion/exclusion of special cohorts (e.g., elderly patients with an isolated hip fracture, dead on hospital arrival). One of these
factors (the version of the injury code system) was found to be underreported by the questionnaire. These factors may need to be fully described as much as possible to improve the quality of studies on injury severity scoring tools.

**Source of heterogeneity**

The sources of the heterogeneity among the relevant studies could be found in the different characteristics of their study populations. For instance, we found a wide range of crude mortality rates of the study populations, ranging from 1.1% [15] to 38% [81]. This wide distribution of the rates might be due to the difference in the type of database used between the ICISS and AIS-based tools. Studies that investigated the ICISS mostly used administrative databases, whereas studies that analyzed AIS-based tools generally used a trauma registry. Because the majority of studies of the ICISS used such a database without considering the mechanism of injury, severity of injury or, sometimes, age groups, the crude mortality rates of these studies were lower than those of the studies of AIS-based tools. Among 19 studies of the ICISS, only two studies reported that their crude mortality rates were more than 10%, whereas the rates of all of the other studies were less than 10%. In contrast, among 45 articles that did not study the ICISS, 22 studies reported more than 10% as the crude mortality rate. These high mortality rates of studies of the AIS-based tools may be explained by the fact that these studies used trauma registries that generally have inclusion and/or exclusion criteria that prevent many minor injuries from being registered.

**ISS vs. NISS**

We identified 16 studies that reported 24 pairs of AUROCs of the ISS and NISS [3,7,9,10,14,15,37,39,51,53,58,66,72,74,78,79]. Among the 24 pairs of AUROCs, eight pairs demonstrated that the ISS had a greater AUROC than the NISS [9,15,39,58,74,78], whereas the other 16 pairs showed greater AUROCs for the NISS than for the ISS. There were seven pairs of AUROCs that were derived using only blunt trauma patients [7,37,39,51,58,74,78]. Among these seven pairs, only one pair had a higher AUROC for the ISS than the NISS [39]. There were four pairs of AUROCs that were derived using penetrating trauma patients [51,58,74,78]. Only one study had a higher AUROC for the NISS than the ISS [51]. Although further studies are required, the NISS might be better at predicting the outcomes of blunt trauma patients than the ISS, and vice versa for penetrating trauma patients. Because the mechanism of injury might affect the predictive performance of the ISS and the NISS, researchers should clearly describe the mechanism of injury of the study population and analyze blunt and penetrating trauma patients separately when investigating the predictive performance.

---

**Figure 1 Flow diagram of study selection process.** In total, we retrieved 5,608 potential articles from the online database search. We finally identified 64 relevant articles.
| Author      | Year | Country    | Hospital | Tool  | AUROC  | N          | Mortality | Mechanism | Code          | Age |
|------------|------|------------|----------|-------|--------|------------|-----------|-----------|---------------|-----|
| Aydin      | 2008 | Turkey     | 1        | ISS   | 0.907  | 550        | 21.6%     | ND        | AIS90         | >16 |
|            |      |            |          | NISS  | 0.914  |            |           |           |               |     |
|            |      |            |          | TRISS (M) | 0.934 |            |           |           |               |     |
| Barbieri   | 2004 | Italy      | 1        | TRISS (M) | 0.946 | 93         | 28.0%     | ND        | ND            | ND  |
| Becalick   | 2001 | UK         | 1        | TRISS (N) | 0.941 | 677        | 16.2%     | B + P     | ND            | ND  |
| Bergeron   | 2004 | Canada     | 1        | TRISS (M) | 0.873 | 5,672      | 6.9%      | B         | ND            | ≥15 |
|            |      |            |          | TRISS (N) | 0.878 |            |           |           |               |     |
| Bijima     | 2004 | Netherlands | 1        | ISS   | 0.84   | 668        | 18.4%     | ND        | ND            | ND  |
| Bonaventura| 2001 | Czechoslovakia | 1     | ISS   | 0.57   | 1,113      | 18.0%     | ND        | ND            | ND  |
|            |      |            |          | NISS  | 0.852  |            |           |           |               |     |
| Bouamra    | 2006 | UK         | 106      | TRISS (N) | 0.937 | 20,895     | 4.4%      | B         | ND            | adult |
| Bouillon   | 1997 | Germany    | 32       | ISS   | 0.961  | 612        | 30.9%     | ND        | AIS90         | all |
|            |      |            |          | TRISS (M) | 0.974 |            |           |           |               |     |
| Brenneman  | 1998 | Canada     | 1        | ISS   | 0.799  | 2,328      | 13.0%     | B         | ND            | all |
|            |      |            |          | NISS  | 0.852  |            |           |           |               |     |
| Burd       | 2008 | US         | Multiple | ICISS (T) | 0.726 | 276,366$^1$ | 2.8%      | B + P     | ICD9CM        | all |
|            |      |            |          | SWI (T) | 0.743 |            |           |           |               |     |
|            |      |            |          | ICISS (I) | 0.793 |            |           |           |               |     |
|            |      |            |          | ICISS (T) | 0.855 | 312,592$^2$ | 5.1%      |           |               |     |
|            |      |            |          | SWI (T) | 0.866 |            |           |           |               |     |
|            |      |            |          | ICISS (I) | 0.867 |            |           |           |               |     |
| Chytra     | 1999 | Czechoslovakia | 1       | ISS   | 0.89   | 165$^3$   | 17.0%     | ND        | ND            | ≥18 |
|            |      |            |          | TRISS (M) | 0.85  |            |           |           |               |     |
|            |      |            |          | ISS   | 0.76   | 109$^4$   | 38.0%     |           |               |     |
|            |      |            |          | TRISS (M) | 0.83  |            |           |           |               |     |
| Davie      | 2008 | New Zealand | ND       | ICISS (T) | 0.777 | 186,835    | 5.3%      | ND        | ICD10AM       | ND  |
|            |      |            |          | ICISS (T) | 0.851 |            |           |           |               |     |
| Dillion    | 2005 | UK         | ND       | ISS   | 0.832  | 53,286     | 4.1%      | B         | ND            | ≥16 |
|            |      |            |          | NISS  | 0.827  |            |           |           |               |     |
|            |      |            |          | TRISS (N) | 0.939 | 12,606     | 4.4%      |           |               |     |
| DiRusso    | 2000 | US         | 25       | ISS   | 0.766  | 2,768      | 8.4%      | all       | ND            | all |
|            |      |            |          | TRISS (M) | 0.895$^5$ |            |           |           |               |     |
|            |      |            |          | TRISS (M) | 0.918$^6$ | 2,673      | 8.3%      |           |               |     |
|            |      |            |          | TRISS (N) | 0.92   | 2,768      | 8.4%      |           |               |     |
| Effekhar   | 2005 | Iran       | 6        | ISS   | 0.944  | 7,226      | 3.8%      | all       | AIS90         | all |
|            |      |            |          | TRISS (M) | 0.969 |            |           |           |               |     |
| Fischler   | 2007 | Switzerland | 1        | ISS   | 0.75   | 208        | 13.0%     | ND        | ND            | adult |
| Frankema   | 2002 | Netherland | 1        | TRISS (M) | 0.975 | 1,024      | 6.9%      | B + P     | AIS90         | ≥15 |
| Frankema   | 2005 | Netherland | 1        | TRISS (M) | 0.94   | 1,102      | 11.0%     | B + P     | AIS90         | ≥15 |
| Gabbe      | 2005 | Australia  | Multiple | TRISS (M) | 0.87  | 1,387      | 4.4%      | B         | NR            | all |
| Glance     | 2009 | US         | 359      | ISS   | 0.868  | 66,214     | 4.2%      | B + P     | ND            | ≥1  |
|            | 2009 | US         | 68       | SWI (T) | 0.862 | 749,374    | 5.0%      | B + P     | ICD9CM        | ≥1  |
| Study          | Year | Country          | Methodology | Sample Size | Sensitivity | Specificity | Null Hypothesis | Additional Information |
|---------------|------|------------------|-------------|-------------|-------------|-------------|------------------|-------------------------|
| Glance [46]   | 2005 | US               | ISS         | 2,412       | 0.85        |             | ND               | all                     |
| Guzzo [48]    | 2005 | US               | TRISS (M)   | 20,883      | 0.85        |             | B               | ND                     |
| Hannan [50]   | 1999 | US               | TRISS (M)   | 192         | 0.85        | 15.1%       | all             | ND                     |
| Hannan [7]    | 2005 | US               | ISS         | 39,534      | 0.776       |             | B               | ND                     |
| Hannan [49]   | 2007 | US               | SWI (I)     | 117,630     | 0.754       |             | 2.9%            | all                     |
| Harwood [51]  | 2006 | 4 countries (Germany, Netherlands, Switzerland, Austria) | ISS         | 10,062      | 0.78        |             | B               | ND                     |
| Hunter [52]   | 2000 | UK               | TRISS (M)   | 7,831       | 0.941       |             | ND              | ND                     |
| Jamulitrat [53] | 2001 | Thailand         | ISS         | 2,044       | 0.966       |             | all             | AIS90                   |
| Kilgo [54]    | 2006 | US               | TRISS (M)   | 310,958     | 0.939       |             | B + P            | ND                     |
| Kim [8]       | 1999 | Korea            | ISS         | 367         | 0.892       |             | B               | ND                     |
| Kroezen [55]  | 2007 | Netherlands      | TRISS (M)   | 349         | 0.806       |             | 14.0%           | B + P                   |
| Kuhls [56]    | 2002 | US               | TRISS (M)   | 179         | 0.93        | 3.5%        | B + P            | ND                     |
| Lane [57]     | 1996 | Canada           | TRISS (N)   | 1,793       | 0.908       |             | 7.9%            | ND                     |
| Lavoie [58]   | 2004 | Canada           | ISS         | 23,306      | 0.818       |             | 6.3%            | B                      |
| Macleod [59]  | 2003 | Uganda           | ISS         | 24,263      | 0.819       |             | 6.6%            | B + P                   |
| Meredith [9]  | 2002 | US               | ISS         | 170,853     | 0.89        |             | 5.4%            | B + P                   |
| Meredith [60] | 2003 | US               | ICISS (T)   | 170,853     | 0.89        |             | 5.4%            | B + P                   |
| Study          | Year | Country | Dataset | Method | Reported AUROC | Sample Size | Mortality | Missing Data | AUROC | Mortality | Missing Data |
|---------------|------|---------|---------|--------|----------------|-------------|-----------|--------------|-------|-----------|--------------|
| Meredith [61] | 2003 | US      | ND      | ICISS (I) | 0.892          | 192,347     | 5.1%      | B + P        | ICD9CM | all       |              |
|               |      |         |         | ICISS (T) | 0.875          |             |           |              |        |           |              |
| Millham [62]  | 2004 | US      | ND      | TRISS (M) | 0.837          | 31,000      | 4.6%      | B            | ND     | all       |              |
|               |      |         |         | TRISS (N) | 0.936          |             |           |              |        |           |              |
|               |      |         |         | TRISS (M) | 0.982          | 5,200       | 9.9%      | P            |        |           |              |
|               |      |         |         | TRISS (N) | 0.981          |             |           |              |        |           |              |
| Moore [10]    | 2008 | Canada  | ND      | ISS      | 0.822          | 25,111      | 7.3%      | B + P        | AIS90   | ≥16       |              |
|               |      |         |         | NISS     | 0.831          |             |           |              |        |           |              |
|               |      |         |         | ICISS (T) | 0.852          |             |           |              |        |           |              |
| Moore [63]    | 2009 | Canada, US | Multiple | TRISS (N) | 0.928          | 178,377     | 6.2%      | B            | ND     | >16       |              |
| Osler [11]    | 1996 | US      | 1       | ISS      | 0.866          | 2,337       | NR        | B            | ICD9   | all       |              |
|               |      |         |         | ICISS (T) | 0.918          |             |           |              |        |           |              |
|               |      |         |         | ISS      | 0.906          | 805         | NR        | P            |        |           |              |
|               |      |         |         | ICISS (T) | 0.93          |             |           |              |        |           |              |
|               |      |         |         | ISS      | 0.87          | 3,142       | 9.0%      | B + P        |        |           |              |
|               |      |         |         | ICISS (T) | 0.921         |             |           |              |        |           |              |
| Osler [3]     | 1997 | US      | 2       | ISS      | 0.869          | 3,136       | 9.0%      | B + P        | AIS90   | all       |              |
|               |      |         |         | NISS     | 0.896          |             |           |              |        |           |              |
|               |      |         |         | ISS      | 0.896          | 3,449       | 7.0%      |              |        |           |              |
|               |      |         |         | NISS     | 0.907          |             |           |              |        |           |              |
| Osler [12]    | 1997 | US      | 1       | ISS      | 0.843          | 1,812       | 2.5%      | all          | ND     | for AIS | all           |
|               |      |         |         | ICISS (T)*13 | 0.884         |             |           |              |        |           | ICD9         |
|               |      |         |         | ICISS (T)*14 | 0.872         |             |           |              |        |           |              |
| Osler [64]    | 2008 | US      | 206     | ISS      | 0.871          | 140,000     | 4.1%      | all          | ND     | ≥1        |              |
| Rabbani [65]  | 2007 | Iran    | 3       | TRISS (N) | 0.93          | 2,514       | 6.0%      | all          | ND     | all       |              |
| Raum [66]     | 2009 | 4 countries (Germany, Netherlands, Switzerland, Austria) | 97 | ISS | 0.722 | 1,292 | 18.9% | all | ND | ≥16 | |
|               |      |         |         | NISS     | 0.764          |             |           |              |        |           |              |
|               |      |         |         | TRISS (M) | 0.851          |             |           |              |        |           |              |
| Reiter [67]   | 2004 | Australia | Multiple | TRISS (M) | 0.84          | 5,538       | 12.3%     | all          | ND     | ≥18       |              |
| Rhee [68]     | 1990 | US      | 6       | ISS      | 0.7967         | 691         | 15.8%     | all          | ND     | ≥11       |              |
| Rutledge [69] | 1997 | US      | ND      | ISS      | 0.939          | 44,032      | 6.5%      | all          | ND     | for AIS | all           |
|               |      |         |         | ICISS (T)*15 | 0.939         |             |           |              |        |           | ICD9CM       |
|               |      |         |         | ICISS (T)*16 | 0.929         |             |           |              |        |           |              |
|               |      |         |         | ICISS (T)*17 | 0.858         |             |           |              |        |           |              |
| Rutledge [70] | 1998 | US      | Multiple | ICISS (T) | 0.957          | 9,438       | 5.1%      | all          | ICD9CM  | all       |              |
| Rutledge [13] | 1998 | US      | 13      | ISS      | 0.667          | 7,276       | 3.8%      | all          | ND     | for AIS | all           |
|               |      |         |         | ICISS (T) | 0.916          |             |           |              |        |           | ICD9CM       |
|               |      |         |         | TRISS (M) | 0.877          |             |           |              |        |           |              |
| Sacco [14]    | 1999 | US      | 26      | ISS      | 0.86           | 30,287      | 7.1%      | all          | ND     | for AIS | all           |
|               |      |         |         | NISS     | 0.86           |             |           |              |        |           | ICD9CM       |
|               |      |         |         | ICISS (T)*15 | 0.87          |             |           |              |        |           |              |
|               |      |         |         | ICISS (T)*18 | 0.88          |             |           |              |        |           |              |
ICISS vs. AIS-based tools

We could not clearly determine the relative performance of the ICISS against the AIS-based tools because of the scarcity of comparative studies. We identified 11 studies that reported AUROCs of the ICISS and ISS and/or NISS [8-15,17,49,69]. Most of these studies reported greater AUROCs for the ICISS than for the ISS/NISS, with one exception [17]. In contrast, the ICISS was rarely compared with the TRISS. We could find three studies that reported AUROCs of both the ICISS and the TRISS [8,13,16]. Among these studies, two studies showed that the TRISS performed better than the ICISS [8,13], and one study demonstrated the opposite [16]. Based on these results, the ICISS is better at predicting outcomes than the ISS/NISS, but the superiority of the TRISS over the ICISS was inconclusive.

Instability of the ICISS

The ICISS appeared to be unstable in terms of its predictive performance for two reasons: the multiplicity of

| Table 1 List of included studies and reported AUROCs (Continued) |
|---------------------------------------------------------------|
| Sammour [71] 2009 New Zealand 1 ISS 0.8547 1,197 3.7% all ND ≥15 |
| Stephensen [15] 2002 New Zealand ND ISS 0.847 340,000 1.1% all AIS90 all |
| Suarez-Alvarez [73] 1995 Spain 1 TRISS (M) 0.85 404 19.6% B + P ND ND |
| Tamin [72] 2008 Lebanon 1 ISS 0.881 891 3.6% all ND all |
| Tay [74] 2004 US 1 ISS 0.922 NR NR B ND all |
| Ulvik [75] 2007 Norway 1 ISS 0.61 325 16.9% ND AIS98 >18 |
| Vassar [76] 1999 US 6 TRISS (M) 0.82 2,414 12.3% B + P AIS90 ≥16 |
| West [16] 2000 US 1 ICISS (T) 0.94 9,923 NR B + P ICD9CM, all |
| Wong [77] 1996 Canada 1 TRISS (M) 0.89 470 13.0% all ND all |
| Wong [17] 2008 Hong Kong 1 ISS 0.8677 1,166 13.8% B ND for AIS, all |
| Zhao [78] 2008 China 1 ISS 0.922 1,532 NR B ND ≥16 |

AIS, Abbreviated Injury Scale; ISS, Injury Severity Score; NISS, New ISS; ICISS, International Classifications of Diseases-based (ICD-based) ISS; TRISS, Trauma and ISS: TRISS(M), TRISS with coefficients based on the Major Trauma Outcome Study (MTOS); TRISS(N), TRISS with coefficients based on non-MTOS population; ICISS–(T), product of traditional Survival Risk Ratios (SRRs); ICISS(I), product of independent SRRs; SWI, Single Worst Injury; SWI (T), SWI of traditional SRRs; SWI(I), SWI of independent SRRs.

*1, Nationwide Inpatient Sample; *2, National Trauma Data Bank (NTDB); *3, polytrauma without brain injury; *4, polytrauma with brain injury; *5, data in 1995; *6, data in 1996.
*7, SRRs derived from ICD10; *8, SRRs derived from ICD9CM; *9, SRRs derived from North Carolina Hospital Discharge Database; *10, SRRs derived from NTDB; *11, Albuquerque data; *12, Portland data; *13, SRRs derived from Hospital Information System; *14, SRRs derived from trauma registry; *15, SRRs derived from North Carolina Trauma Registry; *16, SRRs derived from Agency for Health Care Policy Healthcare Cost Utilization Project; *17, SRRs derived from San Diego Trauma Registry; *18, SRRs derived from Pennsylvania Trauma Outcome Study.

UK, United Kingdom; US, United States of America; NZ, New Zealand.

1, Nationwide Inpatient Sample; 2, National Trauma Data Bank (NTDB); 3, polytrauma without brain injury; 4, polytrauma with brain injury; 5, data in 1995; 6, data in 1996.
7, SRRs derived from ICD10; 8, SRRs derived from ICD9CM; 9, SRRs derived from North Carolina Hospital Discharge Database; 10, SRRs derived from NTDB; 11, Albuquerque data; 12, Portland data; 13, SRRs derived from Hospital Information System; 14, SRRs derived from trauma registry; 15, SRRs derived from North Carolina Trauma Registry; 16, SRRs derived from Agency for Health Care Policy Healthcare Cost Utilization Project; 17, SRRs derived from San Diego Trauma Registry; 18, SRRs derived from Pennsylvania Trauma Outcome Study.

B, blunt injury; P, penetrating injury; B + P, blunt and penetrating injuries; ND, not described; NR, not reported; NA, not applicable.
its formula and SRR and the dependency on the data source for the SRRs. We divided the AUROCs of the ICISS into four subgroups based on the formula and type of SRR. There was only one study that reported AUROCs of all four types of the ICISS [49]. According to this study, the SWI with traditional SRRs performed best (AUROC = 0.764), followed by the SWI with independent SRRs (0.754), the traditional ICISS with traditional SRRs (0.745) and the traditional ICISS with independent SRRs (0.744). Glance et al. supported the superiority of the SWI over the traditional ICISS [47], but Burd et al. reported a greater AUROC for the traditional ICISS than for the SWI [19]. Regarding the type of SRRs, there were three studies that compared the traditional SRRs with independent SRRs [17,49,61]. The results were inconclusive; one reported that independent SRRs were better than traditional SRRs, but the other two reported the opposite.

The predictive performance of the ICISS was also dependent on the data sources from which the SRRs were derived. Rutledge et al. reported AUROCs of traditional ICISSs using different sets of SRRs derived from four different databases [69]. One of the four AUROCs was greater than that of the ISS, but the other three AUROCs were the same as or less than that of the ISS. Kim et al. demonstrated another type of difference in the source data of SRRs. These authors showed that the traditional ICISS based on the ICD9CM performed better than the ISS but that the traditional ICISS using the ICD10 performed worse than the ISS. As a whole, the type of data used for SRR derivation appeared to be a crucial factor in determining the predictive performance of the ICISS.

**Generalizability**

It is difficult to draw broad generalizations from this study because 41% of the studies evaluated were conducted in the U.S., and 41% of the studies contained data from a single hospital (see Additional file 2 for the details). In short, the results derived from narrowly recruited study populations cannot be readily applied to other populations. One can increase the generalizability of results with data from multiple hospitals and/or

**Table 2 Results of quality assessment**

| Internal validity | Yes | No |
|-------------------|-----|----|
| Q1 Were selection criteria clearly described? | 61  | 3  |
| Q2 Were any quality assurance measures for managing and/or collecting data described? | 24  | 40 |
| Q3 Were missing data adequately managed? | 38  | 28 |
| Q4 Was the length of follow-up described? | 35  | 29 |
| Q5 Was the version of the reference code systems used described? | 24  | 40 |
| Q6 Was the derivation of coefficients of TRISS or weights of ICISS described? | 41  | 11 |
| Q7 Were the new coefficients or weights validated? | 25  | 3  |
| External validity | Yes | No |
| Q8 Was the description of the study population reported? | 62  | 2  |
| Q9 Was the study conducted using multi-institutional population? | 28  | 36 |
| Q10 Was the precision of AUROC, such as standard error, reported? | 31  | 33 |

NA, not applicable; AUROC, area under the Receiver Operating Characteristic curve; TRISS, Trauma and Injury Severity Score; ICISS, International Classification of Diseases-based.
multiple countries. Trauma registries in which multiple countries take part have recently been developed [83,84]. The use of such registries might constitute an alternative way to increase the generalizability of study results.

Potential biases
We searched relevant articles using two major online databases, MEDLINE and Embase. We set no language restrictions and checked the references of the relevant articles. These processes enabled us to identify as many relevant articles as possible and to reduce dissemination bias. We might have been able to reduce the bias further if we used other databases (e.g., CINAHL), although the effect of adding another database might have been minimal.

Limitations
We only focused on four injury severity scoring systems. We acknowledge that there are other tools, including A Severity Characterization of Trauma (ASCOT) [85], the Anatomic Profile Score (APS) and the modified Anatomic Profile (mAP) [86]. However, because these tools were not widely used when this study was conducted, we excluded these tools from this review.

Future research directions
Future studies might need to focus more on statistical models that incorporate an injury severity scoring tool with a risk adjustment. Such models could potentially yield a higher predictive performance than the tools in this review. Moore et al. reported on the Trauma Risk Adjustment Model (TRAM), which was superior to the TRISS with regard to both discrimination and calibration [63]. Such high performance predictive models play a key role in hospital performance rankings (e.g., the Trauma Quality Improvement Program) [87]. Furthermore, although systematic reviews studying predictive models for brain trauma injury have been conducted [28], a review that focuses on predictive models for general trauma populations, including the TRAM, has not yet been performed. Reviewing the statistical models used to predict the outcomes of injured patients would provide researchers with clues for important predictors and appropriate statistical techniques.

Conclusions
We could not pool reported evidence because of the heterogeneity among the relevant studies. The NISS appeared to be better at predicting the mortality of blunt trauma patients than the ISS. We could not determine the relative performance of the ICISS against the TRISS. The ICISS appeared to be less stable in its predictive performance than the AIS-based tools because of the many variations in its computational method. Additional studies are required to standardize the derivation of the ICISS and determine the relative performance of the ICISS against the AIS-based tools.

Additional files

Additional file 1: MEDLINE and Embase search strategy using OvidSP version 3.
Additional file 2: Study characteristics of the relevant studies. Details of study characteristics.
Additional file 3: Quality assessment questionnaire of injury severity scoring tools.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
HT and U designed the study and directed its implementation. DM, NG and AY helped to design the statistical strategy of the study. UJ helped to extract data and assess the quality of the studies. HT analyzed the data and wrote the paper. All authors read and approved the final manuscript.

Acknowledgements
We acknowledge that there are other tools, including A Severity Characterization of Trauma (ASCOT) [85], the Anatomic Profile Score (APS) and the modified Anatomic Profile (mAP) [86]. However, because these tools were not widely used when this study was conducted, we excluded these tools from this review.

Received: 26 March 2012 Accepted: 28 August 2012
Published: 10 September 2012

References
1. AMA: Rating the severity of tissue damage. I. The abbreviated scale. JAMA 1971, 215:277–280.
2. Baker SP, O’Neill B, Haddon W Jr, Long WB: The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma 1974, 14:187–196.
3. Osler T, Baker SP, Long W: A modification of the injury severity score that both improves accuracy and simplifies scoring. J Trauma 1997, 43:922–925.
4. Champion HR, Sacco WJ, Hunt TK: Trauma severity scoring to predict mortality. World J Surg 1983, 7:4–11.
5. Osler T: ICISS: An International Classification of Disease-9 Based Injury Severity Score. J Trauma 1996, 41:380.
6. Al-Salamah MA, McDowell I, Stiell IG, Wells GA, Perry J, Al-Sultan M, Nesbitt L: Initial emergency department trauma scores from the OPALS study: the case for the motor score in blunt trauma. Acad Emerg Med 2004, 11:834–842.
7. Hannan EL, Waller CH, Farrell LS, Cayten CG: A comparison among the abilities of various injury severity measures to predict mortality with and without accompanying physiologic information. J Trauma 2005, 58:244–251.
8. Kim Y, Jung KY, Kim CY, Kim YI, Shin Y: Validation of the International Classification of Diseases 10th Edition-based Injury Severity Score (ICISS). J Trauma 2004, 56:280–285.
9. Meredith JW, Evans G, Kilgo PD, Mackenzie E, Osler T, McGwin G, Cohn S, Esposto T, Gennarelli T, Hawkins M, et al: A comparison of the abilities of nine scoring algorithms in predicting mortality. J Trauma 2002, 53:621–628.
10. Moore L, Lavoie A, Le Sage N, Bergeron E, Emond M, Abdous B: Consensus or data-derived anatomic injury severity scoring?. J Trauma 2008, 64:420–426.
11. Osler T, Rutledge R, Des J, Bedrick E: ICISS: an international classification of disease-9 based injury severity score. J Trauma 1996, 41:380–386.
12. Osler TM, Cohen M, Rogers FB, Camp L, Rutledge R, Shackford SR: Trauma registry injury coding is superfluous: a comparison of outcome prediction based on trauma registry International Classification of
13. Rutledge R, Osler T, Emeny S, Kromhout-Schro S. The end of the Injury Severity Score (ISS) and the Trauma and Injury Severity Score (TRISS): ICISS, an International Classification of Diseases, ninth revision-based prediction tool, outperforms both ISS and TRISS as predictors of trauma patient survival, hospital charges, and hospital length of stay. J Trauma 1998; 44:41–49.

14. Sacco WJ, MacKenzie EJ, Champian HR, Davis EG, Buckman RF. Comparison of alternative methods for assessing injury severity based on anatomic descriptors. J Trauma 1999; 47:441–466.

15. Stephenson SC, Langley JD, Civil ID. Comparing measures of injury severity for use with large databases. J Trauma 2002; 53:326–332.

16. West TA, Rivara FP, Cummings P, Jurkovich GJ, Maier RV. Harborview assessment for risk of mortality: an improved measure of injury severity on the basis of ICD-9-CM. J Trauma 2000; 49:530–540.

17. Wong SSN, Leung GKK. Injury Severity Score (ISS) vs. K-Derived Injury Severity Score (ICISS) in a patient population treated in a designated Hong Kong trauma centre. Megg J Med 2008; 119:13–18.

18. Klugo PD, Osler TM, Meredith W, Klugo PD, Osler TM, Meredith W. The worst injury predicts mortality outcome the best: rethinking the role of multiple injuries in trauma outcome scoring. J Trauma 2003; 55:599–606.

19. Burd RS, Ouyang M, Madigan D. Bayesian logistic injury severity score: a model for predicting mortality using international classification of disease-9 codes. Acad Emerg Med 2008; 15:466–475.

20. Chawda MN, Hildebrand F, Pape HC, Giannoudis PV. Predicting outcome after multiple trauma: which scoring system?. Injury 2004; 35:347–358.

21. Lefering R. Trauma score systems for quality assessment. European Journal of Trauma 2002; 28:52–63.

22. Osler T. Injury severity scoring: perspectives in development and future directions. Ann J Surg 1993; 165:435–515.

23. Osler T, Nelson LS, Bedrick EJ. Comparison of artificial intelligence techniques in trauma: a simplified technique based upon ICD-9 coding. J Trauma 1999; 47:441–466.

24. Sacco WJ, Jameson JW, Copes WS, Lawnick MM, Keast SL, Champion HR. Progress toward a new injury severity characterization: severity profiles. Comput Biol Med 1988; 18:419–429.

25. Rutledge R, Falhoy S, Baker C, Oller D. Injury severity grading in trauma patients: a simplified technique based upon ICD-9 coding. J Trauma 1993; 35:497–506. discussion 506–497.

26. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143:29–36.

27. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J, Whiting P, Rutjes AWS. Intraobserver agreement: a comparison of current instruments for survival chance estimation: an evaluation comparing the predictive performance of the ISS, NISS, and AP scores in a Dutch local trauma registration. J Trauma 2005; 58:596–604.

28. Gabbe BJ, Cameron PA, Wolfe R, Simpson P, Smith KL, McNeill JJ. Predictors of mortality, length of stay and discharge destination in blunt trauma. ANZ J Surg 2005; 75:650–656.

29. Guccione LG, Osler TM, Mukamel DB, Meredith W, Dick AW. Expert consensus vs empirical estimation of injury severity effect on quality measurement in trauma. Arch Surg 2009, 144:326–332.

30. Guance LG, Osler TM, Mukamel DB, Meredith W, Wagner J, Dick AW. TMMP-ICD9: a trauma mortality prediction model based on ICD-9-CM codes. Ann Surg 2009, 249:1032–1039.

31. Harwood PJ, Giannoudis PV, Probst C, Van Griensven M, Krettek C, Pape HC, The Polytrauma Study Group of the German Trauma Center S, Harwood PJ, Giannoudis PV, Probst C, et al. Which AIS based scoring system is the best predictor of outcome in orthopaedic blunt trauma patients?. J Trauma 2006, 60:334–340.

32. Hunter A, Kennedy L, Henry J, Ferguson I. Application of neural networks and sensitivity analysis to improved prediction of trauma survival. Comput Methods Prog Biomed 2000; 62:11–19.

33. Januartinis S, Sangkred P, Thongyappong S, Na Narong M. A comparison of mortality predictive abilities between NISS and ISS in trauma patients. J Med Assoc Thai 2001; 84:1416–1421.

34. Klugo PD, Meredith JW, Osler TM, Klugo PD, Meredith JW, Osler TM. Incorporating recent advances to make the TRISS approach universally available. J Trauma 2006, 60:1002–1008. discussion 1008–1009.

35. Koozzen B, Bijurma TS, Liem MLS, Meeuwis J, Leenh LPH. Base deficit-based predictive modeling of outcome in trauma patients admitted to intensive care units in Dutch trauma centers. J Trauma 2007; 63:908–913.

36. Kuhl D, Milton, D, McComb J, Paulo M, Kullus D, Milton D, McLoughlin DJ, Nihle D. Base deficit-based predictive modeling of outcome in trauma patients admitted to intensive care units in Dutch trauma centers. J Trauma 2007; 63:908–913.

37. Lane PL, Doss G, Mikrogianakis A, Charyk ST, Stefanits T. An evaluation of Ontario trauma outcomes and the development of regional norms for Trauma and Injury Severity Score (TRISS) analysis. J Trauma 1996; 41:731–734.

38. Lavoie A, Moore L, LeSage N, Liberman M, Sampalis JS. The New Injury Severity Score: a more accurate predictor of in-hospital mortality than the Injury Severity Score. J Trauma 2004, 56:1312–1320.
Comparison of the Injury Severity Score and ICD-9 diagnosis codes to trauma patients in a setting different from “Major Trauma Outcome Study.” Arch Intern Med 1990, 150:383–386.

Raum MR, Nijsten MW, Vogelzang M, Schuring F, Lefering R, Bouillon B, Robijn D, Neugebauer EA, Ten Duis HJ: Polytrauma Study Group of the German Trauma Study et al.: Emergency trauma score: an instrument for early estimation of trauma severity. Crit Care Med 2009, 37:1972–1977.

Reiter A, Mauritz W, Jordan B, Lang T, Pold A, Pellinka L, Metnitz PG: Improving risk adjustment in critically ill trauma patients: the TRISS-SAPS Score. J Trauma 2004, 57:373–380.

Rutledge R, Hoyt DB, Eastman AB, Sise MJ, Velky T, Canty T, Wachtel T, Osler TM: Comparison of the Injury Severity Score and ICD-9 diagnosis codes as predictors of outcome in injury: analysis of 44,032 patients. J Trauma 1997, 43:477–487.

Rutledge R, Osler T: The ICD-9-based illness severity score: a new model that outperforms both DRG and APR-DRG as predictors of survival and resource utilization. J Trauma 1998, 45:791–799.

Sammour T, Kahokehr A, Caldwell S, Hill AG: Venous glucose and arterial lactate as biochemical predictors of mortality in clinically severely injured trauma patients–A comparison with ISS and TRISS. Injury 2009, 40:104–108.

Tamin H, Al Hazouzi AZ, Mahfoud Z, Atoul M, El-Chemaly S: The injury severity score or the new injury severity score for predicting mortality, intensive care unit admission and length of hospital stay: experience from a university hospital in a developing country. Injury 2008, 39:115–120.

Suarez-Alvarez JR, Mical J, Del Rio FJ, Ortega P: Epidemiologic aspects and results of applying the TRISS methodology in a Spanish trauma intensive care unit (TICU). Intensive Care Med 1995, 21:729–736.

Tay S-Y, Sloan EP, Zun L, Zaret P: Comparison of the New Injury Severity Score and the Injury Severity Score. J Trauma 2004, 56:162–164.

Ulvik A, Wentzel-Larsen T, Flaatten H: Trauma patients in the intensive care unit: short- and long-term survival and predictors of 30-day mortality. Acta Anaesthesiol Scand 2007, 51:171–177.

Vassar MJ, Lewis FR Jr, Chambers JA, Mullins RL, O’Brien PE, Weigel JA, Hoang MT, Holcroft JW: Prediction of outcome in intensive care unit trauma patients: a multicenter study of Acute Physiology and Chronic Health Evaluation (APACHE), Trauma and Injury Severity Score (TRISS), and a 24-hour intensive care unit (ICU) point system. J Trauma 1999, 47:324–329.

Wong DT, Barrow PM, Gomez M, McGuire GP: A comparison of the Acute Physiology and Chronic Health Evaluation (APACHE) II score and the Trauma-Injury Severity Score (TRISS) for outcome assessment in intensive care unit trauma patients. Crit Care Med 1996, 24:1642–1648.

Zhao X-G, Ma Y-F, Zhang M, Gan J-X, Xu S-W, Jiang G-Y: Comparison of the new injury severity score and the injury severity score in multiple trauma patients. Chin J Traumatol 2008, 11:368–371.