



Published in final edited form as:

J Trauma Nurs. 2017 ; 24(1): 4–14. doi:10.1097/JTN.0000000000000255.

Validating the Use of ICD-9 Code Mapping to Generate Injury Severity Scores

Dr. Ross J. Fleischman, MD, MCR, Dr. N. Clay Mann, PhD, MS, Dr. Mengtao Dai, PhD, Dr. James F. Holmes, MD, MPH, Dr. N. Ewen Wang, MD, Dr. Jason Haukoos, MD, MSc, Dr. Renee Y. Hsia, MD, MSc, Dr. Thomas Rea, MD, MPH, and Dr. Craig D. Newgard, MD, MPH
Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, California (Dr Fleischman); Dept. of Pediatrics, University of Utah School of Medicine, Salt Lake City (Drs Mann and Dai); Department of Emergency Medicine, University of California Davis School of Medicine, Sacramento (Dr Holmes); Stanford University Medical School, Stanford, California (Dr Wang); Department of Emergency Medicine, Denver Health Medical Center, University of Colorado School of Medicine, Denver (Dr Haukoos); Colorado School of Public Health, Aurora (Dr Haukoos); Department of Emergency Medicine, University of California San Francisco (Dr Hsia); Department of Medicine, University of Washington, Seattle (Dr Rea); and Department of Emergency Medicine, Oregon Health and Science University, Portland (Dr Newgard)

Abstract

The Injury Severity Score (ISS) is a measure of injury severity widely used for research and quality assurance in trauma. Calculation of ISS requires chart abstraction, so it is often unavailable for patients cared for in nontrauma centers. Whether ISS can be accurately calculated from International Classification of Diseases, Ninth Revision (ICD-9) codes remains unclear. Our objective was to compare ISS derived from ICD-9 codes with those coded by trauma registrars. This was a retrospective study of patients entered into 9 U.S. trauma registries from January 2006 through December 2008. Two computer programs, ICDPIC and ICDMAP, were used to derive ISS from the ICD-9 codes in the registries. We compared derived ISS with ISS hand-coded by trained coders. There were 24,804 cases with a mortality rate of 3.9%. The median ISS derived by both ICDPIC (ISS-ICDPIC) and ICDMAP (ISS-ICDMAP) was 8 (interquartile range [IQR] = 4–13). The median ISS in the registry (ISS-registry) was 9 (IQR = 4–14). The median difference between either of the derived scores and ISS-registry was zero. However, the mean ISS derived by ICD-9 code mapping was lower than the hand-coded ISS in the registries (1.7 lower for ICDPIC, 95% CI [1.7, 1.8], Bland–Altman limits of agreement = –10.5 to 13.9; 1.8 lower for ICDMAP, 95% CI [1.7, 1.9], limits of agreement = –9.6 to 13.3). ICD-9-derived ISS slightly underestimated ISS compared with hand-coded scores. The 2 methods showed moderate to substantial agreement.

Correspondence: Ross J. Fleischman, MD, MCR, Department of Emergency Medicine, Harbor-UCLA Medical Center, 1000 W. Carson St, Box-21, Torrance, CA 90509 (rossf@emedharbor.edu).

The authors declare no conflicts of interest.

Author Contributions: (1) Study concept and design: Fleischman, Mann, Newgard. (2) Acquisition of the data: Mann, Holmes, Wang, Haukoos, Hsia, Rea, Newgard. (3) Analysis and interpretation of the data: Fleischman, Mann, Dai, Newgard. (4) Drafting of the manuscript: Fleischman. (5) Critical revision of the manuscript for important intellectual content: Mann, Holmes, Wang, Haukoos, Hsia, Rea, Newgard. (6) Statistical expertise: Fleischman, Mann, Dai, Holmes, Wang, Haukoos, Hsia, Newgard. (7) Obtained funding: Fleischman, Newgard. (8) Administrative, technical, or material support: Dai. (9) Study supervision: Fleischman, Newgard.

Although hand-coded scores should be used when possible, ICD-9-derived scores may be useful in quality assurance and research when hand-coded scores are unavailable.

Keywords

ICD-9; Injury severity; Injury Severity Score; Methodology-research concepts; Trauma

The Injury Severity Score (ISS) is a measure of anatomical injury severity associated with mortality, surgical intervention, and need for intensive care (Baker, O'Neill, Haddon, & Long, 1974; Chawda, Hildebrand, Pape, & Giannoudis, 2004). The ISS is typically abstracted manually from the injuries described in the patient's chart and therefore generally only available for injured patients cared for in trauma centers. Chart abstraction is labor-intensive, time-consuming, and expensive. More automated methods to derive ISS from International Classification of Diseases, Ninth Revision (ICD-9), codes have been developed and incorporated into software programs, although the accuracy of such measures remains unclear. The validation of ICD-9 code mapping programs to generate ISS would improve the efficiency in collecting this measure, enhance trauma research for patients cared for in nontrauma hospitals, and catalyze population-based injury research.

ICD-9 diagnosis codes are widely available in administrative databases, including those not specifically designed for trauma. Since the 1980s, a number of studies have described computerized techniques for deriving ISS from ICD-9 codes (Kingma, TenVergert, Werkman, Ten Duis, & Klasen, 1994). Mackenzie et al. developed and validated a table to map ICD-9 codes to Abbreviated Injury Scale (AIS) scores and ISS (MacKenzie, Steinwachs, & Shankar, 1989; MacKenzie, Steinwachs, Shankar, & Turney, 1986). Barnard, Loftis, Martin, and Stitzel (2013) compared mappings of ICD-9 codes to AIS scores derived from the Crash Injury Research Engineering Network and the National Trauma Databanks with those generated by ICDMAP. These techniques were published and utilized in peer-reviewed research but were not widely implemented.

Two additional techniques were developed and widely released as publicly available products. The first of these was a stand-alone application developed by investigators at the Johns Hopkins School of Hygiene and Public Health with federal grant support and subsequently marketed by the Tri-Analytics Corporation (Bel-Air, MD) as ICDMAP-90 (Help for ICDPIC, 2014). ICDMAP has not been distributed since 2001. A more recently released program is the ICDPIC procedure for STATA (Statacorp, College Station, TX; Clark, Osler, & Hahn, 2015). ICDPIC 3.0 was developed by investigators at Maine Medical Center and the University of Vermont and is freely available. These techniques do not replace the need for manual abstraction of ISS in trauma centers, as this is required by the American College of Surgeons for reporting to the National Trauma Data Bank; however, validating both programs remains important because they continue to be used for research and could be used for risk adjustment for quality improvement comparison purposes in data sets that span trauma and nontrauma centers. Although the United States is currently transitioning from ICD-9 to International Classification of Diseases, Tenth Revision (ICD-10), this work remains important for validating past research done using these

techniques as well as to support future studies of data collected up to the point at which ICD-9 is no longer used.

Both of these programs have undergone limited external validation studies. Meredith et al. (2002) found that ISS derived from ICD-9 codes present in the National Trauma Data Bank using ICDMAP were inferior to hand-coded ISS at predicting mortality (area under the receiver operating characteristic [ROC] curve = 0.84 vs 0.88). The most recent and comparable study with ours was a retrospective multicenter study comparing ISS derived from ICD-9 codes by ICDPIC in a European trauma registry (Di Bartolomeo, Tillati, Valent, Zanier, & Barbone, 2010). Our study adds to the current body of knowledge because of important differences between European and American trauma data sets including inclusion criteria, injury severity, and incentives for coding accuracy.

PURPOSE

Our objective in this study was to compare ISS derived from ICD-9 codes by ICDPIC and ICDMAP with the ISS manually coded by trauma registrars. We used a large, multicenter sample of U.S. trauma patients served by emergency medical services (EMS) and compared both of the commonly utilized conversion programs.

HYPOTHESIS

We hypothesized that ISS derived by ICDPIC and ICD-MAP would be similar to manually coded injury scores.

METHODS

Study Design

This was a multiregion retrospective cohort study of injured patients transported by EMS and captured in hospital trauma registries from January 1, 2006, through December 31, 2008. Two computer programs, ICDPIC and ICDMAP-90, were used to derive ISS from the ICD-9 codes present in the trauma registries. We then compared derived ISS to hand-coded ISS. Institutional review boards at all sites reviewed this protocol and waived the requirement for informed consent.

Study Setting and Population

The study included injured adults and children transported by 48 EMS agencies in seven regions across the Western United States and entered into nine trauma registries. They included Portland, OR/Vancouver, WA (four counties); King County, WA; Sacramento, CA (two counties); San Francisco, CA; Santa Clara, CA (two counties); Salt Lake City, UT (three counties); and Denver County, CO. Receiving hospitals included 15 Level I, eight Level II, and eight Level III–V trauma centers. These sites are part of the Western Emergency Services Translational Research Network. All sites have established trauma systems. The inclusion criteria between the registries varied.

Study Protocol

Patients with no ICD-9 codes were excluded from analysis. The lack of ICD-9 codes was assumed to be a coding failure, as most patients should at least have diagnosis codes for minor injuries. Patients with no ISS recorded in the trauma registry were also excluded. ICDPIC could not generate ISS scores from burn codes 940–949.5, so these diagnoses were not included in the calculation of ISS. Patients with only burn diagnoses were excluded. ICDMAP requires a patient age below 100 years, so patients missing age or those older than 99 years were also excluded.

Measurements

The ISS is an anatomical scoring system based on the severity of injury in six body regions according to the AIS score ranging from 0 (no injury) to 6 (not survivable) (Association for the Advancement of Automotive Medicine, 2008; Baker et al., 1974). The ISS is the sum of the squares of the AIS scores of the three most severely injured body regions. A patient with an AIS score of 6 (not survivable) in any region is assigned the maximal ISS of 75. An ISS of more than 15 is considered “serious injury” and identifies patients most likely to benefit from care in major trauma centers (Boyd, Tolson, & Copes, 1987; MacKenzie et al., 2006; Mullins et al., 1994; Sampalis et al., 1999).

We used probabilistic linkage to match trauma registry records with state hospital emergency department (ED) and discharge data (LinkSolv 8.2; Strategic Matching, Inc., Morrisonville, NY). Construction of this database, including the analytical methods used to compile the data, has been previously described (Newgard et al., 2012). Only the ICD-9 diagnosis codes from the trauma registries were used to derive ISS. Other variables included age and sex, ED disposition, inpatient mortality, length of hospital stay, and ICD-9-coded surgical procedures (including interventional radiology).

Data Analysis

We used STATA Version IC 13 to clean all ICD-9 codes. We then used ICDPIC to generate ISS-ICDPIC and ICDMAP to derive ISS-ICDMAP. ICDPIC assigned each ICD-9 code to the AIS body region and severity with which it is most frequently associated in the National Trauma Data Bank (Help for ICDPIC, 2014). For the 37 ICD-9 codes that were not clearly associated with AIS scores, the authors of ICDPIC assigned a body region and the AIS score by consensus. The proprietary algorithms used by ICDMAP were not explicitly stated.

The primary analysis used Bland–Altman plots to assess the agreement between methods (Bland & Altman, 1986). The Bland–Altman plots show the average ISS obtained by two methods for each subject versus the difference between those two methods and are useful for displaying how the difference between two methods varies across the range of ISS. The Bland–Altman limits of agreement show the range of 2 standard deviations from the mean difference, which was chosen a priori.

Secondary analyses included weighted kappas and the median difference in ISS calculated by different methods. We also collapsed scores into previously studied ranges: 0–8, 9–15, 16–24, 25–40, 41–49, 50–75 (Copes et al., 1988). The distribution of scores in each range

was depicted as a frequency graph, with calculation of the percentages of scores falling into the same range or adjacent ranges. The percentage of cases classified as “serious injuries” (ISS >15) was examined. We also compared ISS measures using the following age strata: 0–18, 19–64, and 65–99 years. Children were further stratified by collapsing U.S. Centers for Disease Control and Prevention age categories into 0–4, 5–14, and 15–18 years (U.S. Centers for Disease Control and Prevention, 2016).

RESULTS

There were 25,926 cases during the 3-year period, of which 24,804 were analyzed after excluding cases for the reasons given in Figure 1. Table 1 characterizes the demographics, injury severity, and outcomes of the study patients. Figure 2 shows the scatterplot of ISS-ICDPIC versus ISS-registry for each patient. Table 2 and Figure 3 provide the overall median injury severity scores calculated by each of the three methods. The median ISS calculated by ICDPIC and ICDMAP (8) was 1 point lower than hand-coded ISS in the trauma registries (9). ICDPIC and ICDMAP classified 19% of patients as seriously injured on the basis of an ISS of more than 15 versus 24.4% using hand-coded ISS from the trauma registries. Agreement between ISS-ICDPIC and ISS-registry for the categorical outcome of an ISS of more than 15 was 91.0% with a weighted $\kappa = 0.75$. Agreement between ISS-ICDMAP and ISS-registry for an ISS of more than 15 was 92.3% and a weighted $\kappa = 0.79$. Agreement between ISS-ICDPIC and ISS-ICDMAP was 96.2%, with a weighted $\kappa = 0.88$. The sensitivity, specificity, positive predictive value, and negative predictive value for determining an ISS of more than 15 by ICDPIC versus ISS-registry were as follows: 71.8% (70.6–72.9), 98.1% (97.9%–98.3%), 92.5% (91.7%–93.2%), and 91.5%, respectively. For ISS-ICDMAP versus ISS-registry, these values were 75.2% (74.0%–76.2%), 98.7% (98.6%–98.9%), 95.1% (94.4%–95.7%), and 92.5% (92.1%–92.8%), respectively.

Table 3 shows the difference between the three methods of calculating ISS. The median difference between any two methods of calculating ISS for each patient was zero in all age groups. Figures 4–6 show Bland–Altman plots for the differences between ISS-ICDPIC and ISS-registry, ISS-ICDMAP and ISS-registry, and ISS-ICDPIC and ISS-ICDMAP.

Table 4 shows the agreement between methods of calculating ISS as the percentage with the same ISS from two methods, the percentage with ISS that fell within the same range, and the percentage that fell within the same or adjacent ranges. Any two methods showed reasonable agreement by the criteria of the same ISS (59.2%–69.9% agreement) or ISS range (76.4%–88.9%) and excellent agreement by having the two ISS values within the same or adjacent ranges (95.6%–98.9%).

Among the 111 patients with discrepancies between ISS-ICDPIC and ISS-registry of more than 30, 55 (50%) had an ISS-registry of 75. The AIS score of 6, signifying an unsurvivable injury, was assigned to 126 patients by coders versus 100 patients by ICDPIC. A sensitivity analysis excluding patients with an ISS-registry of 75 improved the agreement between methods, narrowing the 95% CI for the Bland–Altman plot between ISS-ICDPIC and ISS-registry to –12.9 to 9.7 versus –13.9 to 10.5 for the entire data set. Similar findings were seen for the 104 patients assigned an ISS-ICDMAP of 75. The same sensitivity analysis

excluding patients with an ISS-registry of 75 narrowed the Bland–Altman limits of agreement for ISS-ICD9 versus ISS-registry by a similar amount.

Code 850.11 “concussion with loss of consciousness 30 min or less,” assigned an AIS score of 1 by ICD9, was present in 17.1% of cases with a discrepancy of more than 30 between ISS-ICD9 and ISS-registry, whereas it was present in only 2.4% of cases in the overall cohort. This code was overrepresented in children (3.7% of cases vs. 2.0% of adults). Code 800.90, “open fracture of vault of skull with intracranial injury of other and unspecified nature, unspecified state of consciousness,” assigned an AIS score of 3, was assigned to 9% of patients with discrepancies of more than 30. Code 805.4, “closed fracture of lumbar vertebra without mention of spinal cord injury,” assigned an AIS score of 2, was present in 7% of high discrepancy cases.

DISCUSSION

This study demonstrated reasonable agreement between ISS derived from ICD-9 codes and hand-coded ISS, although both ICD9 and ICD9 underestimated ISS compared with hand-coded scores. The Bland–Altman limits of agreement were fairly wide, demonstrating variability of scores among a portion of patients. Other measures suggested better accuracy of the ICD-9-derived ISS. Weighted kappas indicated substantial intermethod agreement.

The importance of ISS is arguably greatest around the cutoff of ISS of more than 15, which is commonly used to define severe injury (Boyd et al., 1987). Agreement between methods for classifying severe versus nonsevere injury was excellent, although the ICD-9-derived scores underestimated it.

The Bland–Altman analysis revealed a number of areas for further analysis. The Bland–Altman charts for both ISS-ICD9 versus ISS-registry and ISS-ICD9 versus ISS-registry show a bias toward lower ICD-9-derived ISS than hand-coded scores. This negative bias was more apparent with higher ISS values. Analysis of outliers with the largest discrepancies between methods showed that the majority had been assigned an “unsurvivable” AIS score of 6. By definition, an AIS score of 6 results in an ISS of 75. This reflects the inherent inability of the ISS mapping tools to reflect the range of severities that can be associated with a particular injury, which manual coders can do. This problem will be most pronounced with the most severe or fatal injuries. Very few ICD-9 diagnoses are uniformly fatal and would receive an AIS score of 6 from one of the mapping programs; however, manual coders have more detailed knowledge of the patient’s injuries and subsequent demise, enabling them to apply the “unsurvivable” AIS score of 6 when appropriate, resulting in the maximal ICD code of 75.

Although the Bland–Altman plot for ISS-ICD9 versus ISS-ICD9 looks to have the same pattern of outliers, there were actually far fewer of these cases, with only 20 having a discrepancy of greater than 30 in either direction. As in the other plots, these high discrepancy cases can largely be attributed to the assignment of an ISS of 75 (usually due to an AIS score of 6), as this occurred in 17 of the 20 cases.

The appearance of diagonal lines within the plot is an artifact of the definition of ISS as the sum of the squares of the three highest AIS body region scores. This limits the possible ISS values to those which are sums of perfect squares. This is seen in the histogram of scores in Figure 3, where there is a great deal of variation from one bin to the next. The increasing distance between points near the top right and bottom right borders of the graph is due to the few observations there. The diagonal lines forming the top right and bottom right borders are cases where the patient received a score of 75 for an unsurvivable injury by only one of the scoring systems.

Many of the cases with the largest differences between two methods were due to an insufficient number of ICD-9 codes being entered. For all severely injured patients (ISS-registry >15), the median number of ICD-9 codes was six. For the cohort with large discrepancies between ISS-registry and ISS-ICDPIC (difference >30), the median number of ICD-9 codes was three. Given that having a discrepancy between methods greater than 30 requires having an ISS of at least 30, this high discrepancy group should have been expected to have more diagnoses than the less severely injured group defined by ISS of less than 16. This suggests that some of the underestimation of ISS in patients with high discrepancies between systems was due to coding an insufficient number of diagnoses.

The code 850.11 “concussion with loss of consciousness of 30 min or less” was significantly overrepresented in patients with large discrepancies between either of the ICD-9-derived scores and ISS-registry. This was one of the 37 codes that was assigned an AIS score by consensus of the authors of ICDPIC rather than empirically. This suggests a weakness in ICDPIC in assigning a head AIS score of 1 (minor injury) to this diagnosis. Given our evolving understanding of the long-term consequences of concussion, a score of 1 is likely too low, necessitated by the fact that this code could be applied to a very wide range of presentations, from a momentary loss of consciousness to one with up to 29 min of unresponsiveness. Another possibility is that the availability of this non-specific code resulted in the failure to use more specific codes that would have given a higher derived AIS score (e.g., coding as a concussion when the patient actually had a subdural hematoma). Although the algorithms used by ICDMAP were proprietary so that we cannot know exactly how code 850.11 was treated, the same correlation between that code and high discrepancies between ISS-ICDMAP and ISS-registry was observed. None of the other codes assigned a consensus AIS score by the authors of ICDPIC were overrepresented in the cases with high discrepancies.

Two codes that were not assigned an AIS score by consensus were also frequent in the cases with high discrepancies. These were code 800.90, “open fracture of vault of skull with intracranial injury of other and unspecified nature, unspecified state of consciousness,” and code 805.4, “closed fracture of lumbar vertebra without mention of spinal cord injury,” assigned an AIS score of 2. Similar to the concussion code, this nonspecific code could include a wide range of injury severities from a minor spinous process fracture requiring no treatment to an unstable multicolmn fracture requiring surgery. Authors using ICDPIC should look for the prevalence of these codes in their data sets and consider them potential markers for inaccuracy if they are not accompanied by more specific codes for the same body region.

Although this study assumed hand-coded ISS to be the gold standard and did not directly test the accuracy of coding by re-abstraction of the charts or against another external standard, the examination of cases with maximal discrepancy between hand-coded and ICD-9-derived ISS suggests two areas that could be the focus of additional training and quality assurance for coders. First, the discrepancies between hand-coded and ICD-9-derived scores in cases with an AIS score of 6 suggest that these cases could be the source of the greatest variability in coding. In complying with the American College of Surgeons recommendation that 5%–10% of charts be re-abstracted, it would be prudent that charts with AIS scores of 5 and 6 be preferentially chosen for rereview. Second, the higher number of cases with head injuries defined by ICD 850.11 “concussion with loss of consciousness 30 min or less” suggests that these cases of mild to moderate traumatic brain injury should also be the focus of additional scrutiny.

Performance was similar across age groups, with the Bland–Altman limits of agreement being slightly narrower and the weighted kappas the same or higher in the 18 years or younger and 65 years or older age groups than in the middle-aged group or overall cohort. The only inferior performance in the 18 years or younger age group was seen in the secondary metric of the median ISS derived by ICDPIC or ICDMAP versus ISS-registry (Table 2). This difference was not seen in the primary outcome of Bland–Altman mean difference or limits of agreement, nor was there a difference between the median of the differences between ICD-9-derived ISS and ISS-registry in each patient. One possible explanation for the lower accuracy in the pediatric age group was the overuse of the code 850.11. In contrast, this code was much less frequently seen in the elderly, in which accuracy was excellent. The results were the same in the 5- to 14- and 15- to 18-year-old groups. Infants and children younger than 4 years showed the strongest agreement between measures, with the same median ISS for both the ICD-9-derived ISS and ISS-registry.

Previous studies have shown hand-coded ISS to be superior to AIS scores derived from ICD-9 codes at predicting mortality (Di Bartolomeo et al., 2010; Meredith et al., 2002). If hand-coded ISS can be obtained for a sample, we would still recommend that those be used. However, this requires both the availability of a trained coder and a data set small enough to be manually coded. Even if charts were being reviewed for abstraction of other data elements, this might not be practical for large data sets, in which case our results show ICDPIC and ICDMAP to be acceptable alternatives. As neither ICDPIC nor ICDMAP was clearly superior, the ICDPIC program can be used in place of ICDMAP, which is no longer distributed.

Programs for generating ISS are useful for a number of reasons. The ISS remains the most widely used anatomical injury severity scale for trauma quality assurance programs, evaluating the effectiveness of field triage practices, trauma research and identifying the group of patients most likely to benefit from trauma center care. The ISS also remains the primary measure of injury severity described in the American College of Surgeons Committee on Trauma “Resources for Optimal Care of the Injured Patient” (Rotondo, Cribari, & Smith, 2014). Hospitals without trauma registries do not code ISS, which leaves an important gap in evaluating injured patients transported to nontrauma hospitals for such issues as under- and overtriage rates for field triage, regionalization of trauma care,

interhospital transfer patterns, and patterns of injury at different types of hospitals. ICD-9-based risk adjustment tools such as the ICD-Based Injury Severity Score (ICISS) are useful, but lack appropriate and recognized cutoff points for splitting patients based on minor, moderate, and serious injury (Osler, Rutledge, Deis, & Bedrick, 1996).

Our study was similar to Di Bartolomeo et al.'s (2010) single-center Italian study comparing ICDPIC with hand-coded ISS. Our study demonstrated much better agreement between the ICD-9-derived ISS and the hand-coded values. Only 12.1% of their 272 cases had the same ISS-ICDPIC and ISS-registry versus 59.2% in ours. In that study, ICDPIC underestimated ISS more significantly, with a mean difference of 6.4 versus 1.7 that we observed. Di Bartolomeo et al. (2010) note many reasons for differences between European and U.S. results. Their case mix was more seriously injured, including only patients admitted to an intensive care unit or with an ISS of more than 15, which is standard practice in European trauma registries (Probst, Paffrath, Krettek, & Pape, 2006). Only 13% of the patients in Di Bartolomeo et al.'s (2010) study versus 76% in ours had an ISS of 15 or less. A sensitivity analysis on our data including only patients with ISS-registry of more than 15 lowered the percentage of cases with exact agreement between ISS-ICDPIC and ISS-registry to 36.0%, which was still higher than Di Bartolomeo et al.'s (2010) finding. Another explanation for differences between European and U.S. results is the quality of the coding of both the trauma registry ISS and administrative ICD-9 codes, which would have widened the difference between the two scores. Therefore, Di Bartolomeo et al.'s (2010) finding of poor accuracy for ICDPIC does not challenge the reasonable accuracy and usefulness of the tool for research on U.S. data sets. Another study of 76,781 National Trauma Data Bank records compared ICDMAP-derived scores with those coded in the registry (Meredith et al., 2002). Similar to our results, the median ISS-ICDMAP was slightly lower than that coded in the registry (8 vs. 9). Hand-coded ISS slightly better discriminated survivors from nonsurvivors (area under the ROC = 0.88 vs. 0.84).

The change from ICD-9 to ICD-10 will require the development of new tools, including an updated ICDPIC procedure (Office of the Federal Register, 2015). One such algorithm that maps ICD-10 to AIS 1998 revision codes by expert consensus has been developed and validated by Haas, Xiong, Brennan-Barnes, Gomez, and Nathens (2012). However, with ICD-9 codes still being recorded through 2015, this work will be relevant for retrospective research for a decade or more to come. As we found that many of the cases with the highest discrepancies were related to a few codes that could not accurately map to an AIS score, the far greater number of codes in ICD-10 should allow more precise and accurate mapping.

One caution for researchers using ICD-9-derived ISS is that they should not be mixed with hand-coded ISS. Our study showed that both tools for deriving ISS systematically underestimated injury severity. Mixing ICD-9-derived ISS with hand-coded scores from registries in the same sample of patients could create an unpredictable bias in the scoring. As an example, trauma centers may have patients with higher ISS scores, in part, due to hand-coding compared with nontrauma hospitals using ICD-9-derived scores. Researchers should use one method or the other in a study. Using purely ICD-9-derived injury severity scores will underestimate the percentage of patients with serious injuries. The resulting regionalization metrics and under- and overtriage rates would be biased as a result. This

issue will become more complex in the future, as longitudinal studies span eras of multiple AIS versions, ICD-9 and ICD-10.

CONCLUSIONS

Deriving ISS from ICD-9 codes slightly underestimated ISS when compared with hand-coded scores. Scores derived from ICDPIC and ICDMAP were very similar, making the use of ICDPIC reasonable, given its continued availability and transparency of its coding algorithms. Although hand-coded ISS should remain the gold standard, ICD-9-derived scores are useful in research when hand-coded scores are unavailable, particularly in U.S. data sets. Patients with AIS scores of 5 and 6, as well as those with mild-moderate traumatic brain injuries, represent areas of potential extra focus in training and quality assurance re-abstractation.

LIMITATIONS

This study assumed hand-coded ISS to be the gold standard. We did not have access to the original charts for rereview to determine the accuracy of that coding and the amount of variability induced by both subjectivity of coding and outright error. However, these data came from established trauma systems in compliance with the standards for quality assurance set by the American College of Surgeons. Similarly, the accuracy of ICD-9-derived ISS will vary on the basis of the quality and comprehensiveness of ICD-9 coding. As diagnoses are typically assigned by the providers at the time of care and then may be modified or added to by coders for billing purposes, there are multiple steps at which variability in coding could be introduced. This study did not have access to the original charts to validate the ICD-9 coding and did not look at variability in the type and number of diagnoses by hospital site.

The scope of our patient population imposed some other limitations. Our data set only included patients cared for in trauma centers. Whether ICD-9 diagnoses are assigned differently in trauma versus nontrauma centers is unknown. In the United States, financial incentives influencing how comprehensively ICD-9 diagnosis codes are given may vary between academic or public hospitals (i.e., Level I and II trauma centers) and other hospitals. Our data set did not identify trauma center level, so we could not investigate this.

Another limitation is that ICDPIC could not code for burn diagnoses, requiring that patients with only burn diagnoses be excluded. Patients with both burn and other trauma diagnoses were not likely to be a significant confounder, as they constituted less than 0.4% of our cases. Other authors have pointed out that ISS is poor at differentiating moderate from severe burns and predicting mortality in those cases (Cassidy et al., 2014). Therefore, ISS is probably not a good choice in predicting mortality in cohorts with large populations of burn patients, regardless of the method by which they are derived. Finally, this data set comprised patients who were transported by EMS and who could be probabilistically matched to a trauma registry record. Therefore, less severely injured patients who arrived at the hospital by private transportation or who were not entered in a trauma registry would not be included. We conducted a sensitivity analysis on the patients with ISS-registry of 15 or less. Accuracy

in this group was superior to the overall cohort, with absolute agreement between ISS-ICDPIC and ISS-registry of 66.7% (vs. 59.3% in the overall cohort), with similarly higher accuracies in the ISS-ICDMAP to ISS-registry and ISS-ICDPIC to ISS-ICDMAP comparisons. Therefore, it would be safe to conclude that the overall accuracy of these tools would be at least as great in a population that included non-EMS arrivals, although there may be particular reasons why certain types of populations might not come by EMS, which would bias the results in those subgroups

One difficulty in comparing the older ICDMAP programs with the newer ICDPIC program comes from the change to the AIS score that has occurred over time. As our data were collected from 2006 to 2008, this was subsequent to the major revision in 2005. The mapping tables used by ICDPIC were updated in 2011, so also would have been based on the post-2005 definitions. As ICDMAP ceased production in 2001, its mapping scheme would have been based on an older AIS version. A study of 109 patients showed significant differences in AIS scores and overall ISS when they were coded according to 1998 versus 2005 schemes (Palmer, Niggemeyer, & Charman, 2010). Although our study does not directly address the significance of changing coding schemes over time, it suggests the superiority of using the more current ICDPIC program because of transparency in the source of its lookup tables.

Acknowledgments

This project was supported by the Robert Wood Johnson Foundation Physician Faculty Scholars Program; the Oregon Clinical and Translational Research Institute (grant UL1 RR024140); UC Davis Clinical and Translational Science Center (grant UL1 RR024146); Stanford Center for Clinical and Translational Education and Research (grant UL1 RR025744); University of Utah Center for Clinical and Translational Science (grant UL1 RR025764 and C06 RR11234); and UCSF Clinical and Translational Science Institute (grant UL1 RR024131). All Clinical and Translational Science Awards are from the National Center for Research Resources, a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research.

References

- Association for the Advancement of Automotive Medicine. Abbreviated Injury Scale (AIS) 2005 manual, update 2008. Barrington, IL: Author; 2008.
- Baker SP, O'Neill B, Haddon W, Long W. The Injury Severity Score: A method for describing patients with multiple injuries and evaluating emergency care. *Journal of Trauma*. 1974; 14(3):187–196. [PubMed: 4814394]
- Barnard RT, Loftis KL, Martin RS, Stitzel JD. Development of a robust mapping between AIS 2+ and ICD-9 injury codes. *Accident Analysis and Prevention*. 2013; 52:133–143. [PubMed: 23333320]
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*. 1986; 327(8476):307–310.
- Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: The TRISS method. Trauma Score and the Injury Severity Score. *Journal of Trauma*. 1987; 27(4):370–380. [PubMed: 3106646]
- Cassidy JT, Phillips M, Fatovich D, Duke J, Edgar D, Wood F. Developing a burn injury severity score (BISS): Adding age and total body surface area burned to the injury severity score (ISS) improved mortality concordance. *Burns*. 2014; 40(5):805–813. [PubMed: 24315362]
- Chawda MN, Hildebrand F, Pape HC, Giannoudis PV. Predicting outcome after multiple trauma: Which scoring system? *Injury*. 2004; 35(4):347–358. [PubMed: 15037369]
- Clark, DE., Osler, TM., Hahn, DR. ICDPIC: STATA module to provide methods for translating International Classification of Diseases (Ninth Revision) diagnosis codes into standard injury

categories and/or scores. 2015 Apr 29. Retrieved from <http://econpapers.repec.org/software/bocbocode/s457028.htm>

- Copes WS, Champion HR, Sacco WJ, Lawnick MM, Keast SL, Bain LW. The Injury Severity Score revisited. *Journal of Trauma*. 1988; 28(1):69–77. [PubMed: 3123707]
- Di Bartolomeo S, Tillati S, Valent F, Zanier L, Barbone F. ISS mapped from ICD-9-CM by a novel freeware versus traditional coding: A comparative study. *Scandinavian Journal of Trauma, Resuscitation, and Emergency Medicine*. 2010; 18(17):1–7.
- Haas B, Xiong W, Brennan-Barnes M, Gomez D, Nathens AB. Overcoming barriers to population-based research: Development and validation of an ICD10-AIS algorithm. *Canadian Journal of Surgery*. 2012; 55(1):11–16.
- Help for ICDPIC. 2014 Jul 31. Retrieved from <http://fmwww.bc.edu/repec/bocode/i/icdpic.html>
- Kingma J, TenVergert E, Werkman HA, Ten Duis HJ, Klasen HJ. A Turbo Pascal program to convert ICD-9CM coded injury diagnoses into Injury Severity Scores: ICDTOAIS. *Perceptual and Motor Skills*. 1994; 78(3):915–936. [PubMed: 8084713]
- MacKenzie EJ, Rivara FP, Jurkovich GJ, Nathen AB, Frey KP, Egleston BL, ... Scharfstein DO. A national evaluation of the effect of trauma-center care on mortality. *The New England Journal of Medicine*. 2006; 354(4):366–378. [PubMed: 16436768]
- MacKenzie EJ, Steinwachs DM, Shankar BS, Turney SZ. An ICD-9CM to AIS conversion table: Development and application. *Proceedings of the Association for Advancement of Automotive Medicine*. 1986; 30:135–151.
- MacKenzie EJ, Steinwachs DM, Shankar B. Classifying trauma severity based on hospital discharge diagnoses. Validation of an ICD-9CM to AIS-85 conversion table. *Medical Care*. 1989; 27(4):412–422. [PubMed: 2649755]
- Meredith JW, Evans G, Kilgo PD, MacKenzie E, Osler T, McGwin G, ... Rue L. A comparison of the abilities of nine scoring algorithms in predicting mortality. *Journal of Trauma*. 2002; 53(4):621–629. [PubMed: 12394857]
- Mullins RJ, Veum-Stone J, Helfand M, Zimmer-Gembeck M, Hedges JR, Southard PA, Trunkey DD. Outcome of hospitalized injured patients after institution of a trauma system in an urban area. *JAMA*. 1994; 271(24):1919–1924. [PubMed: 8201736]
- Newgard CD, Malveau S, Staudenmayer K, Wang NE, Hsia RY, Mann NC, ... Cook LJ. Evaluating the use of existing data sources, probabilistic linkage and multiple imputation to build population-based injury databases across phases of trauma care. *Academic Emergency Medicine*. 2012; 19(4):469–480. [PubMed: 22506952]
- Newgard CD, Zive D, Holmes JF, Bulger EM, Staudenmayer K, Liao M, ... Hedges JA. A multisite assessment of the American College of Surgeons Committee on Trauma field triage decision scheme for identifying seriously injured children and adults. *Journal of the American College of Surgeons*. 2011; 213(6):709–721. [PubMed: 22107917]
- Office of the Federal Register. CD-10-CM/ AIS mapping software: A notice by the National Highway Traffic Safety Administration. 2015 Jan 29. Retrieved from <https://www.federalregister.gov/articles/2014/06/12/2014-13727/icd-10-cmais-mapping-software>
- Osler T, Rutledge R, Deis J, Bedrick E. ICISS: An International Classification of Disease-9 based Injury Severity Score. *Journal of Trauma*. 1996; 41:380–386. [PubMed: 8810953]
- Palmer CS, Niggemeyer LE, Charman D. Double coding and mapping using Abbreviated Injury Scale 1998 and 2005: Identifying issues for trauma data. *Injury*. 2010; 41(3):948–954. [PubMed: 20074729]
- Probst C, Paffrath T, Krettek C, Pape HC. Comparative update on documentation of trauma in seven national registries. *European Journal of Trauma*. 2006; 32(4):357–364.
- Rotondo, MF., Cribari, C., Smith, RS. Resources for optimal care of the injured patient. Chicago, IL: American College of Surgeons; 2014.
- Sampalis JS, Denis R, Lavoie A, Fréchette P, Boukas S, Nikolis A, ... Mulder D. Trauma care regionalization: A process-outcome evaluation. *Journal of Trauma*. 1999; 46(4):565–579. [PubMed: 10217218]

U.S. Centers for Disease Control and Prevention. 10 Leading causes of death by age group, United States—2014. 2016 Mar 14. Retrieved from http://www.cdc.gov/injury/images/lc-charts/leading_causes_of_death_age_group_2014_1050w760h.gif

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

KEY POINTS

- ICD-9-derived ISS showed moderate to substantial agreement to hand-coded scores but slightly underestimated them.
- Although hand-coding should remain the standard by which ISS values are generated, ICD-9-derived scores may be useful in quality assurance and research when hand-coded scores are unavailable.
- Codes for concussion with loss of consciousness, skull fracture, and lumbar fracture were associated with the greatest discrepancy between ICD-9-derived and hand-coded values. Cases given these codes should be the focus of additional review.

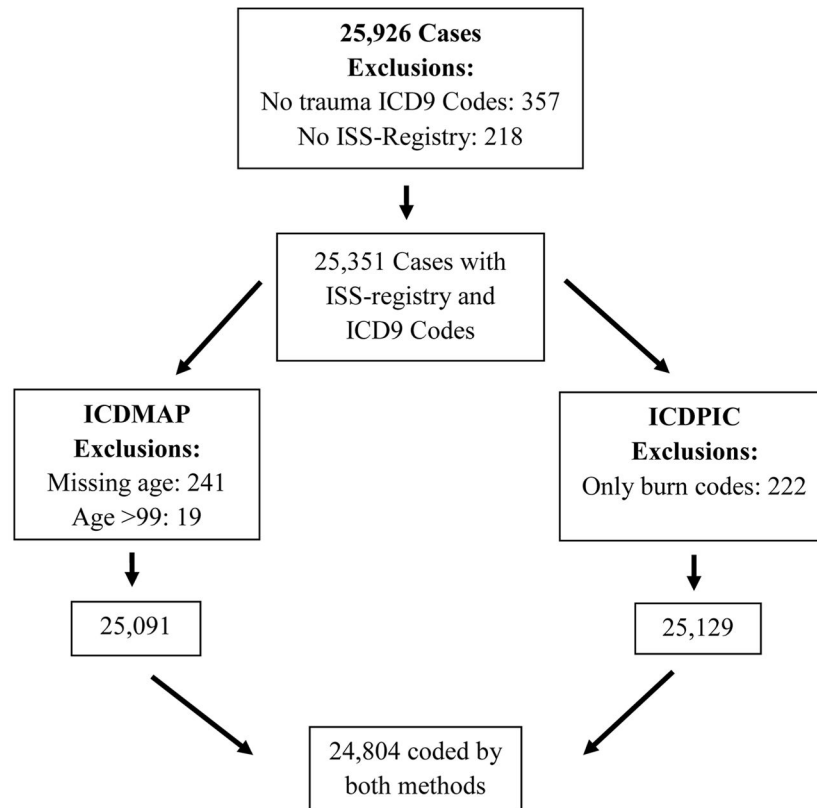


Figure 1. Flow diagram of excluded cases. Only patients for whom ISS-ICDPIC, ISS-ICDMAP, and ISS-registry could all be calculated were included in the final analysis.

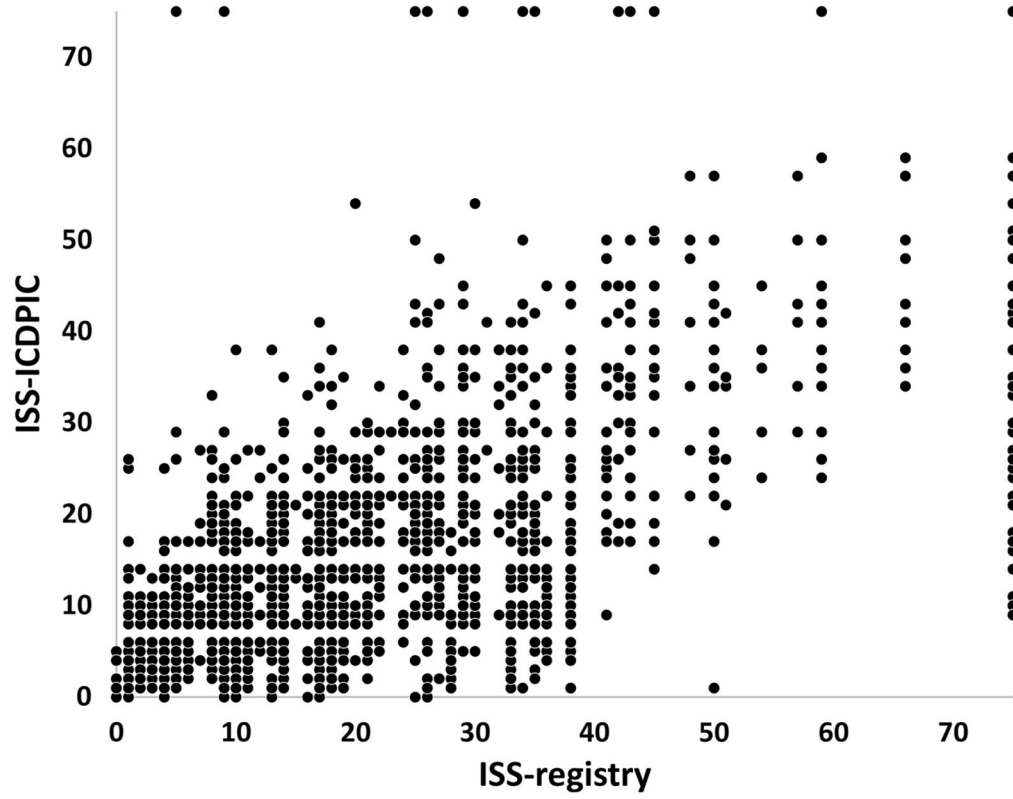


Figure 2. Scatter plot of ISS-ICDMAP versus ISS-registry. The plot of ISS-ICDMAP versus ISS-registry (not shown) had a very similar appearance.

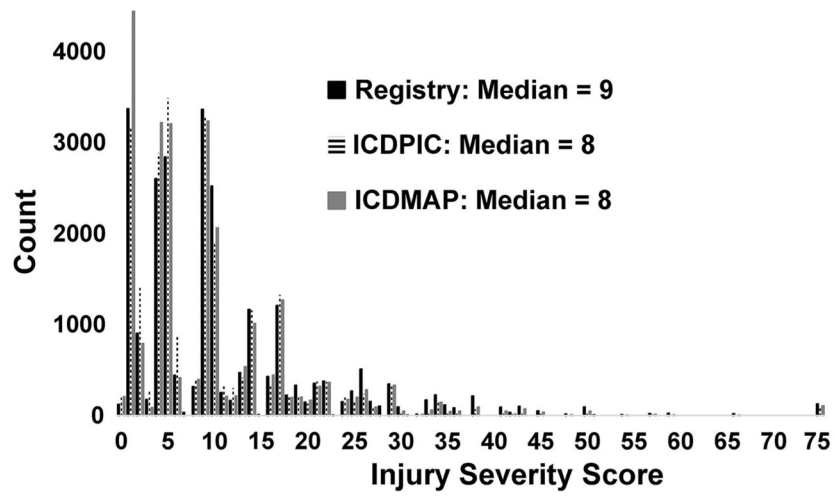


Figure 3. Distribution of Injury Severity Score coding by three methods ($N = 24,804$).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

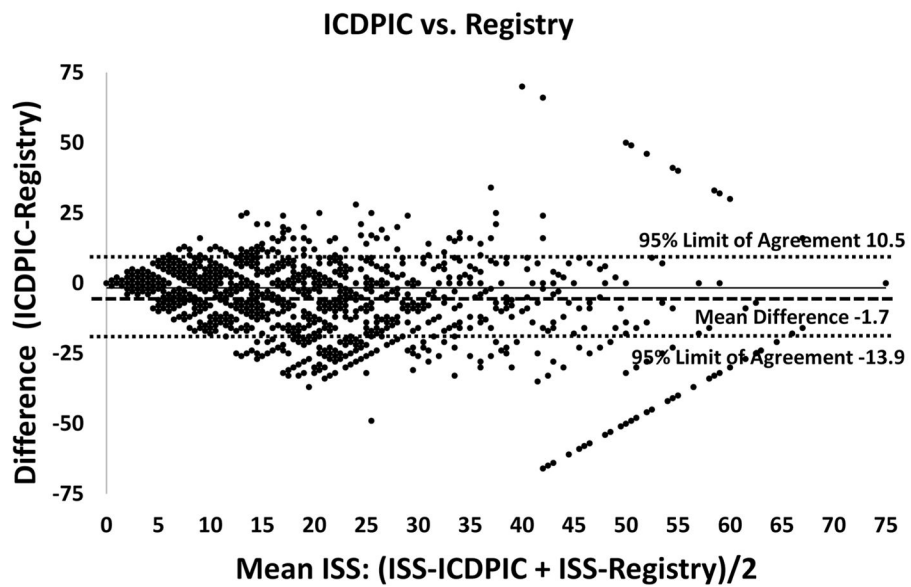


Figure 4. Bland–Altman plot of the mean of ISS-ICDPIC and ISS-registry versus the difference between ISS-ICDPIC and ISS-registry. Bland–Altman plot for mean Injury Severity Score versus ISS-ICDPIC minus ISS-registry. Mean difference -1.7 (horizontal dashed line, 95% CI $[-1.8, -1.6]$). Limits of agreement -13.9 to 10.5 (horizontal dotted lines).

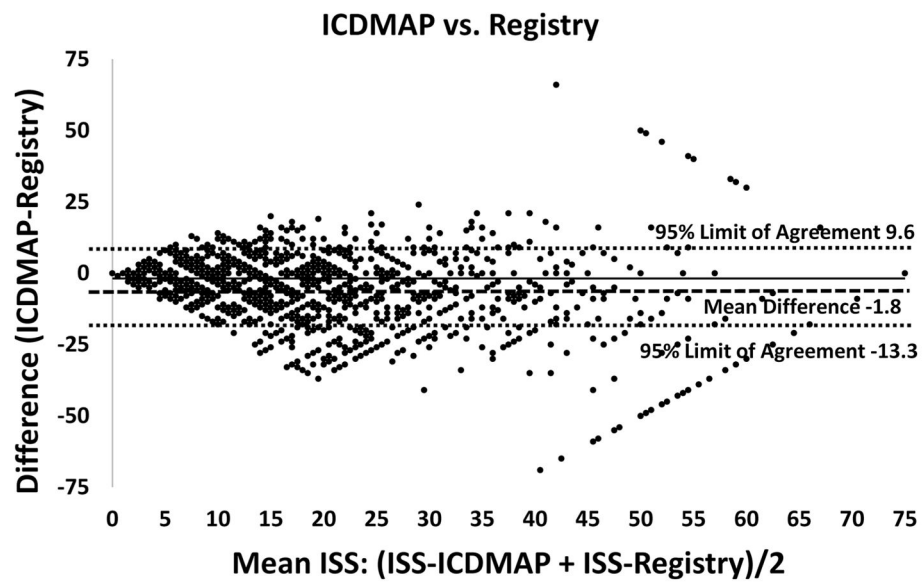


Figure 5. Bland–Altman plot of the mean of ISS-ICDMAP and ISS-registry versus the difference between ISS-ICDMAP and ISS-registry. Bland–Altman plot for mean Injury Severity Score versus ISS-ICDMAP minus ISS-registry. Mean difference -1.8 (horizontal dashed line, 95% CI $[-1.9, -1.7]$). Limits of agreement -13.3 to 9.6 (horizontal dotted lines).

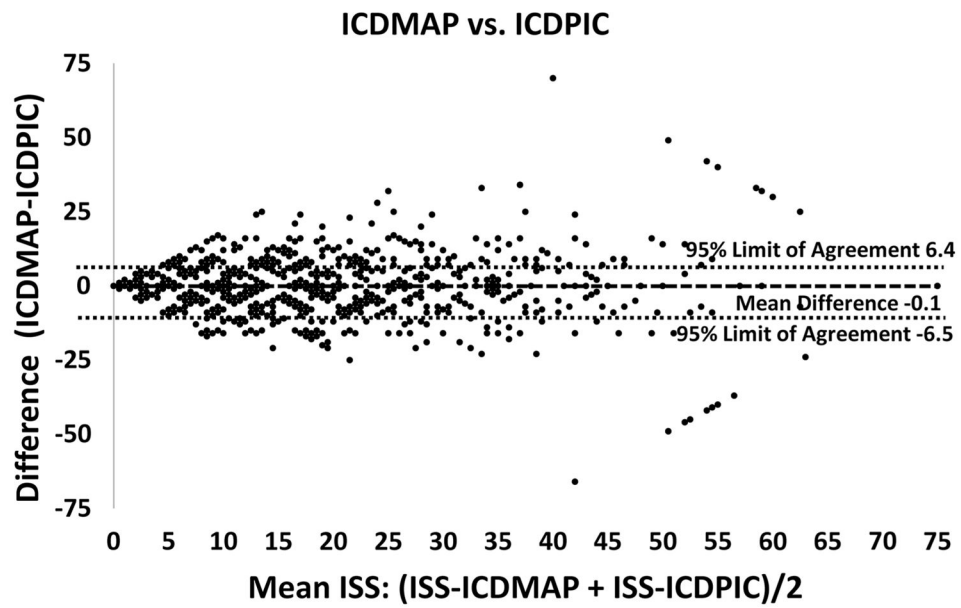


Figure 6. Bland–Altman plot of the mean of ISS-ICDMAP and ISS-ICDPIC versus the difference between ISS-ICDMAP and ISS-ICDPIC. Bland–Altman plot for mean Injury Severity Score versus ISS-ICDMAP minus ISS-ICDPIC. Mean difference -0.1 (horizontal dashed line, 95% CI $[-0.1, 0]$). Limits of agreement -6.5 to 6.4 (horizontal dotted lines).

TABLE 1Characteristics of Sample ($N = 24,804$)

		<i>n</i>
<i>Patient characteristics</i>		
Age, median (IQR), years	40 (23–58)	
Ages 0–18 years, [95% CI]	14.6% [14.1–15.0]	3,610
Ages 0–4 years, [95% CI]	2.3% [2.1–2.5]	570
Ages 5–14 years, [95% CI]	5.9% [5.6–6.2]	1,460
Ages 15–18 years, [95% CI]	6.4% [6.1–6.7]	1,580
Ages ≥ 65 years, [95% CI]	19.6% [19.1–20.1]	4,869
Female sex, ^a [95% CI]	36.8% [32.2–37.4]	8,302
Total ICD-9 codes, median (IQR)	3 (2–5)	
Trauma ICD-9 codes, ^b median (IQR)	3 (2–5)	
<i>Outcomes</i>		
In-hospital mortality, [95% CI]	3.9% [3.7–4.2]	970
Admission or death, [95% CI]	85.0% [84.6–85.5]	21,088
Length of stay, median days, (IQR)	3 (1–5)	
Any surgery, [95% CI]	35.7% [35.1–36.3]	8,863
Nonorthopedic surgery, [95% CI]	13.0% [12.6–13.5]	3,234

Note. ICD-9 = International Classification of Diseases, Ninth Revision; IQR = interquartile range.

^a2,246 patients in the trauma registries did not have sex recorded.

^bICD-9 codes 800–960 and 990–999.

TABLE 2

Injury Severity Scores Derived by Three Methods

	Trauma Registry		ICDPIC		ICDMAP	
	Median	IQR	Median	IQR	Median	IQR
<i>ISS</i>						
All ages	9	4-14	8	4-13	8	4-13
Ages 0-18 years	9	4-14	5	3-10	5	2-10
Ages 0-4 years	5	2-10	5	2-10	5	1-9
Ages 5-14 years	6	4-10	5	3-10	5	2-10
Ages 15-18 years	9	4-16	8	4-14	8	2-13
Ages 19-64 years	9	4-16	8	4-13	6	4-13
Age 65 years	9	5-14	9	4-11	9	4-11
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<i>ISS > 15</i>						
All ages	6,061	24.4	4,706	19.0	4,790	19.3

Note. IQR = interquartile range; ISS = Injury Severity Score.

TABLE 3

Mean Differences, Bland–Altman Limits of Agreement, Median Differences, and Weighted Kappas Between Methods of Calculating ISS by Age Groups

	Mean Difference [95% CI]	Limits of Agreement	Median Difference (IQR)	Weighted κ
All ages				
ICDPIC – Registry	-1.7 [-1.7, -1.8]	-13.9 to 10.5	0 [-3, 0]	0.72
ICDMAP – Registry	-1.8 [-1.9, -1.7]	-13.3 to 9.6	0 [-1, 9]	0.76
ICDMAP – ICDPIC	-0.1 [-0.1, 0]	-6.6 to 6.4	0 [0, 0]	0.86
Ages 0–18				
ICDPIC – Registry	-1.4 [-1.6, -1.2]	-12.7 to 9.9	0 [-1, 0]	0.74
ICDMAP – Registry	-1.6 [-1.7, -1.4]	-12.4 to 9.3	0 [-1, 0]	0.77
ICDMAP – ICDPIC	-0.2 [-0.3, -0.1]	-5.7 to 5.3	0 [0, 0]	0.87
Ages 0–4				
ICDPIC – Registry	-1.5 [-2.0, -1.0]	-13.4 to 10.4	0 [-1, 0]	0.70
ICDMAP – Registry	-1.6 [-2.1, -1.2]	-12.8 to 9.6	0 [-1, 0]	0.75
ICDMAP – ICDPIC	-0.1 [-0.3, 0.1]	-5.6 to 5.4	0 [0, 0]	0.85
Ages 5–14				
ICDPIC – Registry	-1.4 [-1.6, -1.1]	-11.9 to 9.1	0 [-1, 1]	0.71
ICDMAP – Registry	-1.6 [-1.8, -1.3]	-11.8 to 8.7	0 [-1, 0]	0.74
ICDMAP – ICDPIC	-0.2 [-0.3, -0.1]	-4.9 to 4.5	0 [0, 0]	0.87
Ages 5–18				
ICDPIC – Registry	-1.4 [-1.7, -1.1]	-13.2 to 10.5	0 [-1, 0]	0.76
ICDMAP – Registry	-1.6 [-1.8, -1.3]	-12.9 to 9.8	0 [-1, 0]	0.79
ICDMAP – ICDPIC	-0.2 [-0.4, 0.0]	-6.4 to 6.0	0 [0, 0]	0.87
Ages 19–64				
ICDPIC – Registry	-1.8 [-1.9, -1.7]	-20.0 to 14.7	0 [-3, 0]	0.72
ICDMAP – Registry	-2.0 [-2.0, -1.9]	-14.0 to 10.1	0 [-3, 0]	0.76
ICDMAP – ICDPIC	-0.1 [-0.2, -0.1]	-7.0 to 6.8	0 [0, 0]	0.85
Ages 65				
ICDPIC – Registry	-1.6 [-1.7, -1.4]	-12.2 to 9.1	0 [-1, 0]	0.72
ICDMAP – Registry	-1.5 [-1.6, 1.5]	-11.1 to 8.2	0 [0, 0]	0.78
ICDMAP – ICDPIC	-0.1 [-0.2, 0]	-5.8 to 5.6	0 [0, 0]	0.86

Note. IQR = interquartile range.

TABLE 4

Measures of Agreement Between Methods of Calculating ISS

	Agreement	Weighted κ
Same ISS		
Registry vs. ICDPIC	59.2%	0.73
Registry vs. ICDMAP	66.3%	0.76
ICDPIC vs. ICDMAP	69.9%	0.86
Same ISS range ^a		
Registry vs. ICDPIC	76.4%	0.72
Registry vs. ICDMAP	79.4%	0.76
ICDPIC vs. ICDMAP	88.9%	8.87
Same or adjacent ISS range		
Registry vs. ICDPIC	95.6%	
Registry vs. ICDMAP	96.2%	
ICDPIC vs. ICDMAP	98.9%	
ISS >15		
Registry vs. ICDPIC	91.1%	0.75
Registry vs. ICDMAP	92.3%	0.78
ICDPIC vs. ICDMAP	96.2%	0.88

Note. The agreement is the actual proportion of cases in which the two methods had the same outcome, not the agreement calculated by the weighted kappa method. ISS = Injury Severity Score.

^aISS values were collapsed into ranges 0–8, 9–15, 16–24, 25–40, 41–49, and 50–75.