Injury Severity Score coding: Data analyst v. emerging m-health technology

R T Spence,[1,2] MB ChB, MPH; **E Zargaran**,[3] MD; **M Hameed**,[3] MD, FACS; **D Fong**;[3] **E Shangguan**;[3] **R Martinez**;[3]
P Navsaria,[2] MB ChB, MMed, FACS, FCS; **A Nicol**,[2] MB ChB, FCS, PhD

[1] *Codman Center, Massachusetts General Hospital, Boston, USA*
[2] *Department of General Surgery, Faculty of Health Sciences, University of Cape Town, South Africa*
[3] *Department of General Surgery, Vancouver General Hospital, British Columbia, Canada*

Corresponding author: R T Spence (spnric004@myuct.ac.za)

Background. The cost of Abbreviated Injury Scale (AIS) coding has limited its utility in areas of the world with the highest incidence of trauma. We hypothesised that emerging mobile health (m-health) technology could offer a cost-effective alternative to the current gold-standard AIS mechanism in a high-volume trauma centre in South Africa.

Methods. A prospectively collected sample of consecutive patients admitted following a traumatic injury that required an operation during a 1-month period was selected for the study. AISs and Injury Severity Scores (ISSs) were generated by clinician-entered data using an m-health application (ISS eTHR) as well as by a team of AIS coders at Vancouver General Hospital, Canada (ISS VGH). Rater agreements for ISSs were analysed using Bland-Altman plots with 95% limits of agreement (LoA) and kappa statistics of the ISSs grouped into ordinal categories. Reliability was analysed using a two-way mixed-model intraclass correlation coefficient (ICC). Calibration and discrimination of univariate logistic regression models built to predict in-hospital complications using ISSs coded by the two methods were also compared.

Results. Fifty-seven patients were managed operatively during the study period. The mean age of the cohort was 27.2 years (range 14 - 62), and 96.3% were male. The mechanism of injury was penetrating in 93.4% of cases, of which 52.8% were gunshot injuries. The LoA fell within –8.6 - 9.4. The mean ISS difference was 0.4 (95% CI –0.8 - 1.6). The kappa statistic was 0.53. The ICC of the individual ISS was 0.88 (95% CI 0.81 - 0.93) and the categorical ISS was 0.81 (95% CI 0.68 - 0.87). Model performance to predict in-hospital complications using either the ISS eTHR or the ISS VGH was equivalent.

Conclusions. ISSs calculated by the eTHR and gold-standard coding were comparable. Emerging m-health technology provides a cost-effective alternative for injury severity scoring.

In 2010, an estimated 16.9 million lives (32.9% of all deaths worldwide) were lost as a result of conditions needing surgical care, and of these deaths, 5.7 million were due to injury.[1] This well surpassed the number of deaths from HIV/AIDS (1.46 million), tuberculosis (1.20 million) and malaria (1.17 million) combined.[2] It is difficult to propose suggestions for national and international strategies to improve trauma care when the impact of any corrective action is currently almost immeasurable with the tools at hand in most resource-limited environments (the so-called 'closing the loop').[3]

The Abbreviated Injury Scale (AIS) has given rise to the most commonly used injury scoring systems for descriptive analyses, benchmarking and quality improvement initiatives in the trauma literature to date.[4] The worth of its mathematical functions (or scores) in predicting mortality has been well documented in various patient populations and across various databases.[5] AIS functions include the popular Injury Severity Score (ISS), the New Injury Severity Score (NISS) and the Anatomic Profile Score, all of which have varied qualitative and quantitative properties. However, the cost of AIS coding has limited its benefit in the developing world, where >90% of deaths due to injury occur. The average traumatic incident takes between 10 and 20 minutes to code, a significant financial commitment for any trauma centre in the world.[6]

We hypothesised that emerging mobile health (m-health) technology could offer a cost-effective alternative to current gold-standard AIS coding mechanisms in a high-volume trauma centre in South Africa (SA). The primary aim of this study was to compare ISSs coded by the data analysts and those coded using the m-health application, including a comparison of the performance of the resultant scores in predicting an adverse event.

Methods

Patient population
Patient records were selected from the newly implemented electronic Trauma Health Registry (eTHR) in the level 1 trauma unit at Groote Schuur Hospital (GSH), Cape Town, SA. This is one of the busiest trauma centres in SA, receiving over 12 000 patients annually. A sample of consecutive patients aged >13 years who were admitted following a traumatic injury that required an operation under general anaesthetic during the 1-month period 1 February 2015 - 1 March 2015 was selected for the study.

Data collection
Every patient had clinician-entered admission, operative and discharge records prospectively generated on the eTHR. Included in these records were the reports of any radiological investigations performed during the index hospital admission as well as any complications that occurred. Attending clinicians updated patient entries using the m-health application during patient admissions and consultant ward rounds, while writing any operative records and on hospital discharge. The admission records, operative records, radiology reports and discharge summaries of eligible patients were then retrospectively extracted from the eTHR and uploaded into a database generated for the purpose of this study using REDCap (Research Electronic Data Capture).[7]

The AIS classification system

The AIS classification system is a consensus-derived, anatomically based, seven-digit injury scoring system. The first six digits refer to a unique numerical identifier that designates the injured body region (out of nine regions), the type of anatomical structure and the specific anatomical structure; the seventh digit refers to an ordinal injury severity scale with categories ranging from 1 ('minor injury') to 6 ('maximal injury'). From the AIS scores, an ISS value, a pragmatic quantitative summary measure of the overall severity of anatomical and functional damage, is calculated by summing the squares of the highest AIS severity codes in each of the three most severely injured ISS body regions.[6]

AIS coding at GSH (ISS eTHR)

It has become protocol in the unit for the operating surgeon to classify all injuries postoperatively according to the AIS 2005 Update 2008. This is done by means of drop-down menus that include AIS injury descriptions that are built into the eTHR application under the nine anatomical regions that make up the scoring system. The application is then programmed to calculate the ISS using these clinician-entered AIS scores by summing the squares of the three most severely injured ISS body regions.

AIS coding at Vancouver General Hospital (ISS VGH)

VGH, an accredited adult level 1 trauma centre in Vancouver, Canada, has a working collaboration with GSH. Data analysts at VGH were invited to participate in this study. Three data analysts agreed, each with trauma coding experience of over 5 years and certified in coding AIS 2005 and Update 2008. Access to the REDCap database was then given to these three data analysts, who collectively proceeded to code every patient's ISS according to the Association for the Advancement of Automotive Medicine (AAAM) guidelines. Each injury was scored in AIS 2005 Update 2008 and the ISS was then manually calculated and entered into REDCap. After all patients were coded, the GSH collaborators were informed, the ISS VGH scores were extracted and the REDCap database was closed.

Classification and collection of endpoints

In keeping with the design and rationale of the American College of Surgeons Trauma Quality Improvement programme, our definition of in-hospital complication included the presence or absence of the following complications: abdominal compartment syndrome, acute lung injury/respiratory distress, acute renal failure, bleeding requiring transfusion, cardiac arrest requiring cardiopulmonary resuscitation with return of spontaneous circulation, catheter-related bloodstream infections, cellulitis, coagulopathy, decubitis ulcer, deep-vein thrombosis, surgical site infection, delayed haemo/pneumothorax, extremity compartment syndrome, myocardial infarction, pneumonia, pulmonary embolus, retained haemothorax, stroke/cerebrovascular accident, systemic sepsis, urinary tract infections, unplanned intubation, unplanned reoperation, unplanned return to the intensive care unit and death.[8] The eTHR is available on a mobile iPad or phone and is active during consultant-led ward rounds, where there is a daily consensus discussion of any complication that has been recorded.

Statistical analysis

The resultant ISS data for both scores were reasonably normal, and both parametric and non-parametric tests were therefore described to comprehensively illustrate the difference between the two sets of ISS scores. Inter-rater agreement was assessed using the Bland-Altman limits of agreement (LoA) method. This method compares the estimated variation in the data with a clinical evaluation of what is an acceptable variation in order for measurements to be considered 'not different'.[6] The inter-rater agreement between the VGH and the eTHR was further assessed by calculating the kappa statistic of the ISS grouped into validated ordinal categories. The ISS categories were 3 - 8, 9 - 15, 16 - 24 and 25 - 75, as proposed by Haider *et al*.[9] A kappa statistic is interpreted according to the following: 0.01 - 0.20 slight agreement, 0.21 - 0.40 fair agreement, 0.41 - 0.60 moderate agreement, 0.61 - 0.80 substantial agreement and 0.81 - 0.99 almost perfect agreement.[10] The purpose of the ISS is to rank-order severity of injuries for prognostication. For this reason, the Kruskal-Wallis rank test was also employed to compare the ranks of the ISSs coded by the eTHR and the VGH. Reliability is defined as the ratio of variation between measurements to the total variation of all the measurements it is intended to measure. Reliability was estimated by intraclass correlation coefficient (ICC) statistics and corresponding 95% confidence intervals (CIs) using a two-way mixed model with absolute agreement index.[6] ICC statistics give a number on a scale from 0 to 1, where 0 indicates agreement no better than chance and 1 indicates perfect agreement.

Furthermore, the univariate association between an ISS and the presence or absence of an in-hospital complication was tested using logistic regression for each set of ISSs. The discrimination and calibration of these two regression models were then compared by calculating the area under the receiver-operating curve (ROC) and the Hosmer-Lemeshow goodness-of-fit (GOF) statistic for both models, respectively. A large *p*-value generated from the GOF test suggests a well-fitted model. This step is repeated for the ordinal categories of ISS. All analyses were performed using STATA/SE version 14.0 (StataCorp LP, USA). The study was approved by the Human Research Ethics Committee of the University of Cape Town (UCT HREC REF: 338/2014).

Results

Descriptive statistics

During the study period, 57 patients were managed operatively, and the ISSs of 57 patients were therefore compared in this analysis. The mean age of the cohort was 27.2 years (range 14 - 62), and 96.5% were male. The mechanism of injury was penetrating in 93.0% of cases, of which 52.8% were firearm injuries. The operative cohort included 33 exploratory laparotomies, 15 cardiothoracic procedures, 6 vascular dissections, 2 videoscopy procedures and 1 neck dissection. An in-hospital complication occurred in 25.2% of cases (95% CI 17.76 - 33.98), including a mortality rate of 4.6% (95% CI 2.71 - 7.57). The summary descriptive statistics, including those of the ISSs, are presented in Table 1.

The median ISS VGH was 10 (interquartile range (IQR) 9 - 17.5) and the median ISS eTHR was 11 (IQR 9 - 18). The mean ISS VGH was 14.92 (95% CI 12.38 - 17.48) compared with the ISS eTHR mean of 14.51 (95% CI 12.13 - 16. 89). The variance of the ISS VGH was 90.87 compared with 78.98 for the ISS eTHR. Fig. 1 represents the box plots of both ISSs.

Agreement

The Bland-Altman comparison of the ISS eTHR and ISS VGH identified the LoA to fall within −8.669 - 9.490. The mean ISS difference was 0.411, and this difference was not statistically significant (95% CI −0.805 - 1.626). Pitman's test showed a difference in variance of *r*=0.147 (*p*=0.281). The resultant kappa statistic was 0.5303 (95% CI 0.496 - 0.618; *p*<0.0001), translating to moderate agreement. The Kruskal-Wallis rank test gave a rank sum of 3 270.50 in the ISS VGH compared with a rank sum of 3 057.50 in the ISS eTHR. The difference in ranks between the two groups was not significant (*p*=0.54).

Table 1. Summary statistics of the operative cohort (*N*=57)

Characteristic	
Gender, *n* (%)	
Female	2 (3.5)
Male	55 (96.5)
Age (yr), mean (range)	27.24 (14 - 62)
Mechanism, *n* (%)	
Blunt	4 (7.0)
Penetrating	53 (93.0)
Penetrating type, *n* (%)	
Gunshot wound	28 (52.8)
Knife	25 (47.2)
Other	0 (0)
Operation, *n* (%)	
Exploratory laparotomy	33 (57.9)
Cardiothoracic procedure	15 (26.3)
Vascular dissection	6 (10.5)
Videoscopy procedure	2 (3.5)
Neck dissection	1 (1.8)
ISS VGH	
Median (IQR)	10 (9 - 17.5)
Mean (95% CI)	14.92 (12.38 - 17.48)
Variance (SD)	90.87 (9.53)
ISS eTHR	
Median (IQR)	11 (9 - 18)
Mean (95% CI)	14.51 (12.13 - 16.89)
Variance (SD)	78.98 (8.89)

SD = standard deviation.

Reliability

The ICC of the individual ISSs was 0.8797 (95% CI 0.8034 - 0.9276), translating to excellent reliability between the ISS VGH and the ISS eTHR. The ICC of the categorical ISSs was 0.8016 (95% CI 0.6833 - 0.8788).

Model validation

The risk estimates when using both continuous and categorical ISSs for VGH and eTHR are presented in Table 2.

The ORs estimated using individual ISSs were very comparable. The OR using the ISS VGH was 1.24 (95% CI 1.10 - 1.39) compared with an OR of 1.19 (95% CI 1.07 - 1.34) when using the ISS eTHR in the logistic regression models predicting an in-hospital complication. Both of these models had good ability to discriminate, with an ROC using the ISS VGH of 0.8827 (95% CI 0.7677 - 0.9976) compared with an ROC
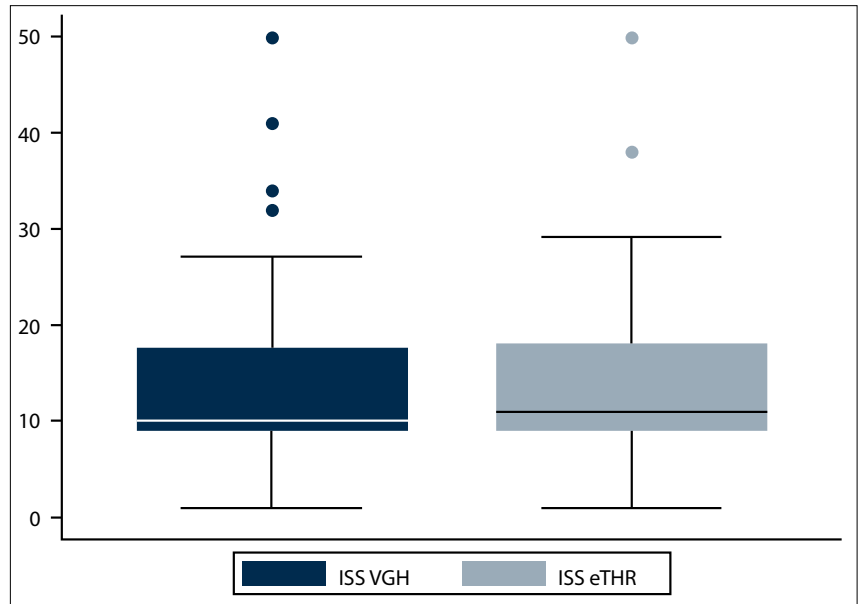

Fig. 1. Box plots comparing the ISS VGH and the ISS eTHR.

Table 2. Comparison of the univariate models predicting an adverse outcome using the ISS eTHR and the ISS VGH as continuous and categorical predictors

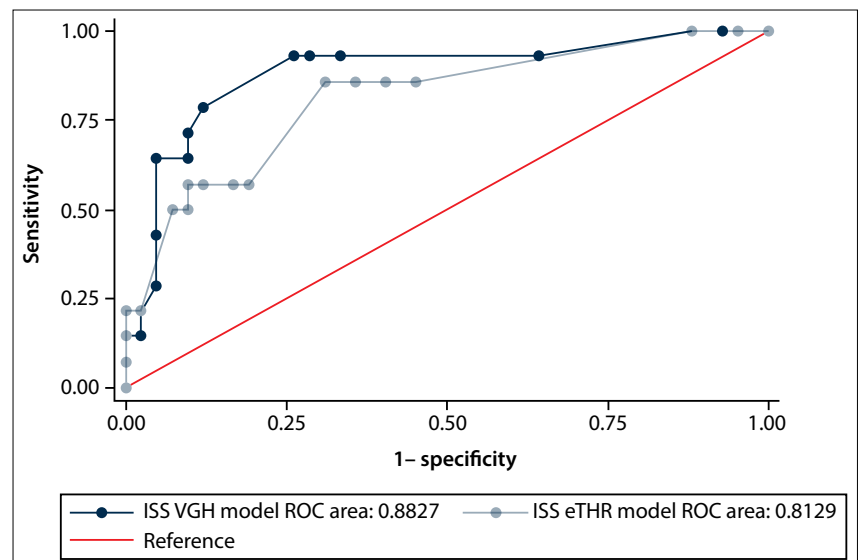| Method of coding	OR	95% CI	ROC	95% CI	GOF *p*-value*
Continuous ISS VGH	1.24	1.10 - 1.39	0.8827	0.7677 - 0.9976	0.21
Continuous ISS eTHR	1.19	1.07 - 1.34	0.8129	0.6802 - 0.9456	0.67
Categorical ISS VGH	5.29	2.01 - 13.98	0.8172	0.6947 - 0.9395	0.91
Categorical ISS eTHR	10.85	3.13 - 37.69	0.8912	0.7967 - 0.9855	0.84

*Hosmer Lemeshow GOF test.

Fig. 2. Comparison of coder and eTHR continuous models to discriminate adverse events.

of 0.8129 (95% CI 0.6802 - 0.9456) for the ISS eTHR, which was not significantly different (*p*=0.33). These curves are plotted against each other in Fig. 2.

Both models had acceptable ability to calibrate in-hospital complications, Hosmer-

Lemeshow tests being non-significant for both models (*p*=0.21 and *p*=0.67, respectively).

The models using categorical ISSs had much larger standard errors than those using individual ISSs, and these estimates

were less precise, as presented in Table 2. Using categories for the ISS VGH in the logistic regression resulted in an OR of 5.29 (95% CI 2.01 - 13.98) compared with 10.85 (95% CI 3.13 - 37.69) when using categories for the ISS eTHR. However, both the models maintained outstanding ability to discriminate, with an ROC using the categorical ISS VGH of 0.8172 (95% CI 0.6947 - 0.9395) compared with an ROC of 0.8912 (95% CI 0.7967 - 0.9855) for the ISS eTHR, which was not significantly different (p=0.23). These curves are plotted against each other in Fig. 3.

Both models had acceptable ability to calibrate in-hospital complications, Hosmer-Lemeshow tests being non-significant for both models (p=0.91 and p=0.84, respectively). The results of the Hosmer-Lemeshow GOF using quintiles of 10 for grouping the data of both sets of individual ISSs are presented graphically in Fig. 4. The probability of adverse events is very comparable at similar VGH and eTHR ISSs.

Discussion
We have used multiple methods to illustrate that ISSs for traumatic injuries managed operatively and coded by the m-health app were comparable to the current gold-standard AIS coding. As referenced in the literature, a clinically acceptable LoA is ±9 units, equivalent to the increase in the derived ISS value when the severity of a single injury is increased from AIS 4 to 5.[6] The Bland-Altman comparison of ISSs identified the LoA to fall within −8.669 - 9.490, very close to the acceptable LoA range of ±9 units. Although a kappa statistic of 0.53 translates to moderate agreement, 67.86% of the scores were grouped into the same ordinal categories by the two scoring methods compared with an expected agreement of 31.57% under chance alone. The reliability for both individual and categorical ISSs was excellent, with upper limits of both CIs almost approaching 1 (perfect reliability). This high degree of agreement and reliability translated into excellent model performance for predicting in-hospital complications using both scoring systems.

According to our knowledge, this is the first description of such a comparison between gold-standard AIS coding and a cost-effective alternative in a high-volume trauma centre with a high incidence of penetrating injury. The results of our study must be considered in the light of previous work that has been done comparing the inter-rater agreement and reliability of the ISS and the NISS, all coded using gold-
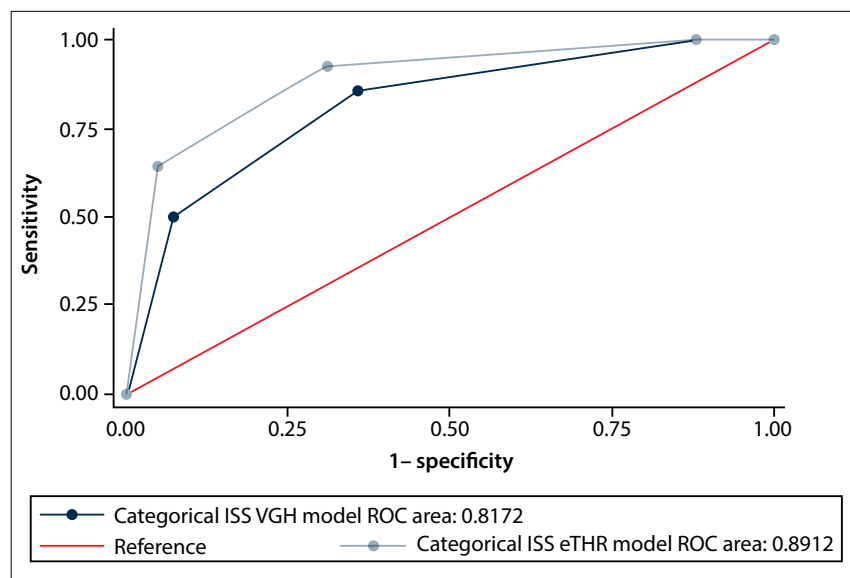

Fig. 3. Comparison of coder and eTHR categorical models to discriminate adverse events.
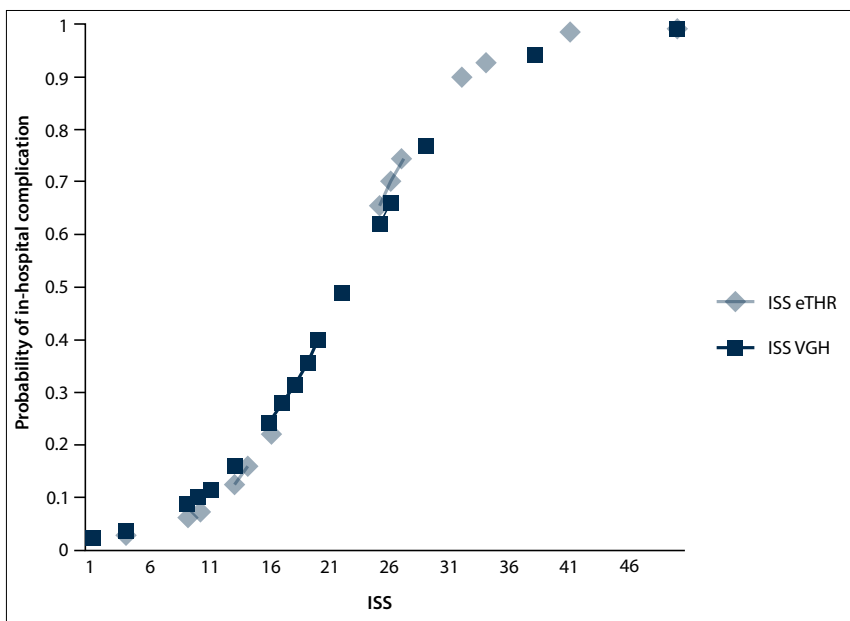

Fig. 4. Estimated probabilities of in-hospital adverse events at increasing ISS VGH and ISS eTHR scores.

standard mechanisms but by multiple raters. A study by Ringdal et al.[6] reported all LoAs were wider than the predefined, clinically acceptable limit of ±9, for both the ISS and the NISS. The joint ICC (range) for inter-rater reliability was 0.49 (0.19 - 0.85) for the ISS and 0.49 (0.16 - 0.82) for the NISS. MacKenzie et al.[5] reported that physicians and nurses had higher intra-rater reliability than medical technicians and non-clinical technicians. The inter-rater AIS score agreement was significantly higher for blunt than for penetrating injuries. Zoltie and De Dombal[11] found a large variation in ISS, with a mere 28% probability of agreement between two independent raters. A comprehensive assessment of the

multiple scoring systems that are available was presented by Meredith et al.,[12] who concluded that differences in performance were relatively small.

Another important, less resource-intensive alternative to AIS coding is the well-known International Classification of Diseases, 9th revision (ICD-9)-based ISS introduced by Osler et al.[13] in 1996. This still relies on ICD-9 coders and will need to be revised with every new edition of the ICD. However, introducing this system into m-health applications will provide another very useful tool for clinicians operating in resource-limited environments with invested interests in improving the outcomes of trauma victims.

Since the implementation of the eTHR app in our trauma centre in April 2014, front-room clinicians have entered over 15 000 patients and surgeons have coded over 3 600 ISSs, without the assistance of a single data analyst.[14] Evidence that this tool offers a cost-effective, validated alternative to AIS coding in a resource-limited environment has significant implications for our centre and potentially for centres around the world operating in similar settings. Validated injury severity scoring allows centres to take into account the heterogeneity of trauma victims presenting to their unit and enables risk-adjusted outcomes to be benchmarked over time. This creates the opportunity for the outcome of any implemented hospital-, national-, and even global-level quality improvement efforts to be measured and benchmarked.[15]

The findings of our study must be interpreted in the context of the study design. A sample size of 50 for ISS reliability studies has been suggested and used in the limited amount of work done in this field previously.[11] A larger sample size would have been preferable, but was not feasible owing to resource constraints and reliance on external data analysts. Only patients requiring an operation were included in this consecutive sample of 57 patients. The recently implemented wider quality improvement project that is taking place at our hospital has commenced by targeting postoperative patients, and scoring in operatively managed patients was therefore prioritised. The operating surgeon for each case performed the AIS coding, as he or she was thought to be in the best position to provide the most accurate assessment of all the injuries sustained. A similar study extending the methodology to patients treated non-operatively would be meaningful and increase the generalisability of this scoring method. Feedback from the VGH coders was that the information provided on each case was less than the usual amount of information used for coding purposes. Coders therefore proceeded with the AAAM guidelines of coding conservatively, not coding queried or unconfirmed diagnoses. The performance of the model using the ISS VGH was excellent (ROC 0.8827, 95% CI 0.7676 - 0.9976), suggesting that their coding was accurate and highly predictive of an in-hospital complication. The ROC of the model using the ISS eTHR was no different (p=0.33), suggesting comparable predictive ability of the two sets of scores.

Conclusions

ISSs calculated by the eTHR and gold-standard coding were comparable. Emerging m-health technology provides a cost-effective alternative for injury severity scoring.

1. Meara JG, Leather AJ, Hagander L, et al. Global Surgery 2030: Evidence and solutions for achieving health, welfare, and economic development. Lancet 2015;386(9993):569-624. DOI:10.1016/S0140-6736(15)60160-X
2. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380(9859):2095-2128. DOI:10.1016/S0140-6736(12)61728-0
3. Mock C. WHO releases: Guidelines for trauma quality improvement programmes. Inj Prev 2009;15(5):359. DOI:10.1136/ip.2009.024315
4. Chada MN, Hildebrand F, Pape HC, Giannoudis PV. Predicting outcome after multiple trauma: Which scoring system? Injury 2004;35(4):347-358. DOI:10.1016/S0020-1383(03)00140-2
5. MacKenzie EJ, Shapiro S, Eastham JN. The Abbreviated Injury Scale and Injury Severity Score: Levels of inter- and intrarater reliability. Med Care 1985;23(6):823-835.
6. Ringdal KG, Skaga NO, Hestnes M, et al. Abbreviated Injury Scale: Not a reliable basis for summation of injury severity in trauma facilities? Injury 2013;44(5):691-699. DOI:10.1016/j.injury.2012.06.032
7. Project REDCap. 2015. http://www.project-redcap.org/ (accessed 24 June 2015).
8. Newgard CD, Fildes JJ, Wu L, et al. Methodology and analytic rationale for the American College of Surgeons Trauma Quality Improvement Program. J Am Coll Surg 2013;216(1):147-157. DOI:10.1016/j.jamcollsurg.2012.08.017
9. Haider AH, Hashmi ZG, Gupta S, et al. Benchmarking of trauma care worldwide: The potential value of an International Trauma Data Bank (ITDB). World J Surg 2014;38(8):1882-1891. DOI:10.1007/s00268-014-2629-5
10. Neale R, Rokkas P, McClure RJ. Interrater reliability of injury coding in the Queensland Trauma Registry. Emerg Med (Fremantle) 2003;15(1):38-41.
11. Zoltie N, de Dombal FT. The hit and miss of ISS and TRISS: Yorkshire Trauma Audit Group. BMJ 1993;307(6909):906-909. DOI:307(6909):906-9
12. Meredith JW, Evans G, Kilgo PD, et al. A comparison of the abilities of nine scoring algorithms in predicting mortality. J Trauma 2002;53(4):621-618; discussion 628-629. DOI:10.1097/01.TA.0000032120.91608.52
13. Osler T, Baker SP, Long W. A modification of the injury severity score that both improves accuracy and simplifies scoring. J Trauma 1997;43(6):922-925; discussion 925-926.
14. Zargaran E, Schuurman N, Nicol AJ, et al. The electronic Trauma Health Record: Design and usability of a novel tablet-based tool for trauma care and injury surveillance in low resource settings. J Am Coll Surg 2014;218(1):41-50. DOI:10.1016/j.jamcollsurg.2013.10.001
15. Farmer PE, Kim JY. Surgery and global health: A view from beyond the OR. World J Surg 2008;32(4):533-536. DOI:10.1007/s00268-008-9525-9