Original article

Comparison of two prognostic models in trauma outcome

A. Cook[1,2], T. Osler[3], L. Glance[5], F. Lecky[9], O. Bouamra[10], J. Weddle[6], B. Gross[4], J. Ward[1], F. O. Moore III[1], F. Rogers[7] and D. Hosmer[8]

[1]Department of Surgery, Chandler Regional Medical Center, Chandler, and [2]Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona, [3]Department of Surgery and [4]College of Medicine, University of Vermont, Burlington, Vermont, [5]Department of Anesthesiology, University of Rochester, Rochester, New York, [6]Department of Surgery, Baylor University Medical Center at Dallas, Dallas, Texas, [7]Department of Surgery, Lancaster General Hospital, Lancaster, Pennsylvania, and [8]Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, Massachusetts, USA, and [9]Department of Emergency Medicine, University of Sheffield, Sheffield, and [10]Institute of Population Health, University of Manchester, Manchester, UK

Correspondence to: Dr A. Cook, Department of Surgery, Chandler Regional Medical Center, 485 South Dobson Road, Suite 201, Chandler, Arizona 85224, USA (e-mail: adcookmd@gmail.com)

Background: **The Trauma Audit and Research Network (TARN) in the UK publicly reports hospital performance in the management of trauma. The TARN risk adjustment model uses a fractional polynomial transformation of the Injury Severity Score (ISS) as the measure of anatomical injury severity. The Trauma Mortality Prediction Model (TMPM) is an alternative to ISS; this study compared the anatomical injury components of the TARN model with the TMPM.**

Methods: **Data from the National Trauma Data Bank for 2011–2015 were analysed. Probability of death was estimated for the TARN fractional polynomial transformation of ISS and compared with the TMPM. The coefficients for each model were estimated using 80 per cent of the data set, selected randomly. The remaining 20 per cent of the data were used for model validation. TMPM and TARN were compared using calibration curves, measures of discrimination (area under receiver operating characteristic curves; AUROC), proximity to the true model (Akaike information criterion; AIC) and goodness of model fit (Hosmer–Lemeshow test).**

Results: **Some 438 058 patient records were analysed. TMPM demonstrated preferable AUROC (0·882 for TMPM *versus* 0·845 for TARN), AIC (18 204 *versus* 21 163) and better fit to the data (32·4 *versus* 153·0) compared with TARN.**

Conclusion: **TMPM had greater discrimination, proximity to the true model and goodness-of-fit than the anatomical injury component of TARN. TMPM should be considered for the injury severity measure for the comparative assessment of trauma centres.**

Presented to the 12th World Conference on Injury Prevention and Safety Promotion, Tampere, Finland, September 2016

Paper accepted 22 October 2017

Published online in Wiley Online Library (www.bjs.co.uk). **DOI:** 10.1002/bjs.10764

Introduction

The need for valid and reliable performance measures is vital as healthcare commissioners, patients and accrediting bodies compare outcomes among hospitals or physicians. For trauma, outcome assessment is mandated by the American College of Surgeons in the USA and the National Health Service (NHS) in England[1,2]. Both the Trauma Audit and Research Network (TARN) in the UK and the Trauma Quality Improvement Program (TQIP) in the USA calculate a patient's baseline predicted probability of dying from their traumatic injuries[3,4]. Central to the predicted probability of death is the extent of the anatomical injury from the traumatic event.

Over the past 40 years, several methods of injury severity measurement have been proposed[5–9], although the first such method, the Injury Severity Score (ISS)[10], remains the most widely applied measure. The familiarity of the ISS to clinicians and researchers is probably the greatest contributor to its longevity. The ISS has functioned as a stand-alone severity measure, and has also been incorporated into the leading models of trauma mortality, such as the Trauma Injury Severity Score (TRISS) method[11] and TARN. Although it has remained the leading measure of

Table 1 Characteristics of the 438 058 patients in the study

	All patients	Survivors	Non-survivors	*P*§
Total	438 058 (100)	422 128 (96·4)	15 930 (3·6)	
Age (years)*	49·8 (49·8, 49·9)	49·5 (49·4, 49·6)	58·3 (58·0, 58·7)	< 0·001¶
Men	284 764 (65·0)	273 655 (96·1)	11 109 (3·9)	< 0·001
Race/ethnicity				< 0·001
White, non-Hispanic	277 794 (63·4)	267 350 (96·2)	10 444 (3·8)	
Black, non-Hispanic	52 946 (12·1)	51 336 (97·0)	1610 (3·0)	
Hispanic/Latino	66 976 (15·3)	65 004 (97·1)	1972 (2·9)	
Race/other	40 342 (9·2)	38 438 (95·3)	1904 (4·7)	
Alcohol present	56 203 (12·8)	54 437 (96·9)	1766 (3·1)	< 0·001
Mechanism of injury‡				< 0·001
Blunt	268 605 (85·6)	258 574 (96·3)	10 031 (3·7)	
Fall	159 053 (59·2)	152 407 (95·8)	6646 (4·2)	
Motor vehicle crash	68 956 (25·7)	66 217 (96·0)	2739 (4·0)	
Penetrating	32 867 (10·5)	31 761 (96·6)	1106 (3·4)	
Firearm-related	9948 (30·3)	9117 (91·6)	831 (8·4)	
Other mechanism	12 365 (3·9)	11 998 (97·0)	367 (3·0)	
Payer status at discharge				< 0·001
Medicare/Medicaid	142 424 (32·5)	135 832 (95·4)	6592 (4·6)	
Private insurance	118 565 (27·1)	115 623 (97·5)	2942 (2·5)	
Other insurance	33 850 (7·7)	32 422 (95·8)	1428 (4·2)	
Uninsured	77 266 (17·6)	74 608 (96·6)	2658 (3·4)	
Insurance status unknown	65 953 (15·1)	63 643 (96·5)	2310 (3·5)	
Length of hospital stay (days)†	4 (2–7)	4 (2–7)	3 (1–8)	< 0·001#
ICU admission	138 837 (31·7)	125 697 (90·5)	13 140 (9·5)	< 0·001
ICU length of stay (days)†	3 (2–6)	3 (2–6)	3 (1–7)	< 0·001#
Mechanical ventilation (days)†	3 (1–8)	3 (1–9)	2 (1–6)	< 0·001#
Injury Severity Score†	9 (5–17)	9 (5–16)	26 (17–34)	
Probability of death from mortality prediction model†				
TMPM	0·010 (0·005–0·024)	0·010 (0·005–0·023)	0·176 (0·041–0·453)	< 0·001#
TARN	0·011 (0·005–0·042)	0·011 (0·004–0·036)	0·106 (0·042–0·187)	< 0·001#

Values in parentheses are percentages unless indicated otherwise; values are *mean (95 per cent c.i.) and †median (i.q.r.). ‡Data were missing for 124 221 subjects and available for 313 837; of the 313 837 patients, 11 504 died. TMPM, Trauma Mortality Prediction Model; TARN, Trauma Audit and Research Network. §Survivors *versus* non-survivors (χ^2 test, except ¶Student's *t* test and #Kruskal–Wallis test).

injury severity since it was introduced, ISS has four major limitations. First, it is based on the Abbreviated Injury Scale (AIS) severity values, which are determined by expert consensus rather than being derived empirically. Second, ISS accommodates only the worst injury from three separate body regions. As such, it cannot account for two or more serious injuries in the same body region. Third, many hospitals do not assign AIS codes to their patients' injuries *de novo*. Instead, they convert the Clinical Modification of ICD-9 codes to AIS codes, thereby reducing the accuracy of injury descriptions[12]. Finally, Kilgo and colleagues[13] have described the ISS as 'choppy', owing to sharp increases or decreases with respect to mortality over incremental increases in ISS values. Of note, TARN uses a mathematical transformation of ISS to achieve better performance of the model.

The Trauma Mortality Prediction Model (TMPM) was developed as an empirically based alternative to ISS without the limitations of ISS. TMPM incorporates the patient's five worst injuries as predictors of their probability of death using a logistic regression model. Previous work[14–16] comparing TMPM with ISS and ISS-based injury severity models showed that TMPM was better able to predict mortality. The TMPM was hypothesized to be a better predictor of survivors from fatalities, given the limitations of the ISS.

Methods

After obtaining approval from the institutional review board of Chandler Regional Medical Center, data from the National Trauma Data Bank (NTDB) for the years 2011–2015 were used in a retrospective cohort study. The NTDB contains data for nearly 7 million patient visits in more than 900 hospitals. The NTDB data are anonymized to ensure confidentiality of patients, physicians and participating hospitals[17]. The outcome of interest was in-hospital death.

Patients were excluded if they were younger than 18 years old, were burn victims, their discharge status (alive or

Table 2 Area under the receiver operating characteristic (ROC) curve and Akaike information criterion values for the Trauma Audit and Research Network and Trauma Mortality Prediction Model

	AUROC value	AIC value	Hosmer–Lemeshow statistic
Derivation			
TARN	0·841 (0·840, 0·841)	86 980 (86 942, 87 019)	2684·6 (681·3, 687·8)
TMPM	0·880 (0·880, 0·880)	74 509 (74 773, 74 545)	390·6 (89·4, 91·7)
Validation			
TARN	0·845 (0·845, 0·846)	21 163 (21 143, 21 182)	153·0 (151·6, 154·3)
TMPM	0·882 (0·882, 0·882)	18 204 (18 186, 18 223)	32·4 (31·7, 33·1)

Values in parentheses are 95 per cent confidence intervals. AUROC, area under the receiver operating characteristic (ROC) curve; AIC, Akaike information criterion; TARN, Trauma Audit and Research Network; TMPM, Trauma Mortality Prediction Model.

dead) was unknown, they were dead on arrival, or were missing ISS values or AIS codes. Patients for whom the TMPM probability of death could not be calculated were also excluded. The reliability and quality of injury documentation in the AIS lexicon was essential to this study. To this end, AIS codes were taken from the RDS_AISPCODE files, as these were submitted by each hospital and least likely to contain codes mapped from ICDMAP-90. Two additional exclusion criteria were applied using methods described previously[14]. Patients were excluded if they were from hospitals that admitted fewer than 300 patients per year, or if their hospital used less than 20 per cent of the available AIS codes.

Comparisons of surviving patients with those who died were performed using χ^2 and Kruskal–Wallis test statistics, as appropriate. Summary measures are presented as means with 95 per cent confidence intervals where appropriate. Median (i.q.r.) values were used for some non-categorical data.

The anatomical injury component of the TARN model was compared with TMPM. The TARN 2014 model (Ps14) is based on the ISS transformed by applying fractional polynomials. However, given the TARN Ps14 model was developed for data defined by the TARN inclusion criteria[18], coefficients for the fractional polynomial transformation of ISS were calculated *de novo* using NTDB data for the present study. This model is as follows:

$$\left[\mathrm{ISS}_1 = \left(\log_e(\mathrm{ISS}/10) - 0{\cdot}19499\right),\right.$$
$$\left.\mathrm{ISS}_2 = \left(\log_e(\mathrm{ISS}/10)^2 - 0{\cdot}03802\right)\right] \text{ probit}$$
$$\text{death } \mathrm{ISS}_1\ \mathrm{ISS}_2 = \beta_1\mathrm{ISS}_1 + \beta_2\mathrm{ISS}_2$$
$$+ \text{constant} = \mathrm{ISS}_1 \times 0{\cdot}762669 + \mathrm{ISS}_2$$
$$\times 0{\cdot}2587654 - 2{\cdot}052421$$

TMPM uses the worst five anatomical injuries coded as categorical variables along with two interaction terms[15]. Coefficients for these seven TMPM terms were calculated for the study data. For the present study, the AIS lexicon was applied as the descriptor of anatomical injury. Of note, however, the TMPM uses only the six-digit 'predot' code to define injuries, and does not incorporate the AIS severity value in its calculation of the probability of death. In contrast, the ISS metric is based exclusively on the expert consensus-based severity value[10].

Using 80 per cent of the data, the fractional polynomial transformation of the ISS was calculated for the TARN model, and coefficients for TARN and TMPM models were estimated. The models were validated using the remaining 20 per cent of the data set.

Measures of model performance included area under the receiver operating characteristic curve (AUROC) (C statistic)[19] and the Akaike information criterion (AIC)[20]. The AUROC is a measure of sensitivity over 1 – specificity. This measures the ability of a model to discriminate subjects having the outcome of interest (here mortality) from those who do not. As AUROC values approach 1·0 the model's discrimination improves[21]. The AIC provides a means of ranking competing models and of estimating which model most closely approximates the hypothetical 'true' model of the phenomenon at hand. Generally, the model with the lowest associated AIC value is preferred. The Hosmer–Lemeshow goodness-of-fit test is a means of assessing how well the model describes the data under analysis and informing the plausibility of the inferences drawn from the model[19]. The Hosmer–Lemeshow test statistic was calculated for each score with ten degrees of freedom specified in the development sample and eight in the validation sample. Calibration curves were constructed for each model to assess monotonicity of the severity measures. Monotonicity describes incremental increases in the observed outcome (here mortality) as a consistent function of incremental increases in the injury severity values.

Observed mortality was plotted against expected mortality for TARN and TMPM. Of note, ISS has 44 unique values for predicted probability of mortality, so the anatomical injury component of the TARN model also has 44 distinct values. The TMPM produced 50 596 unique predicted

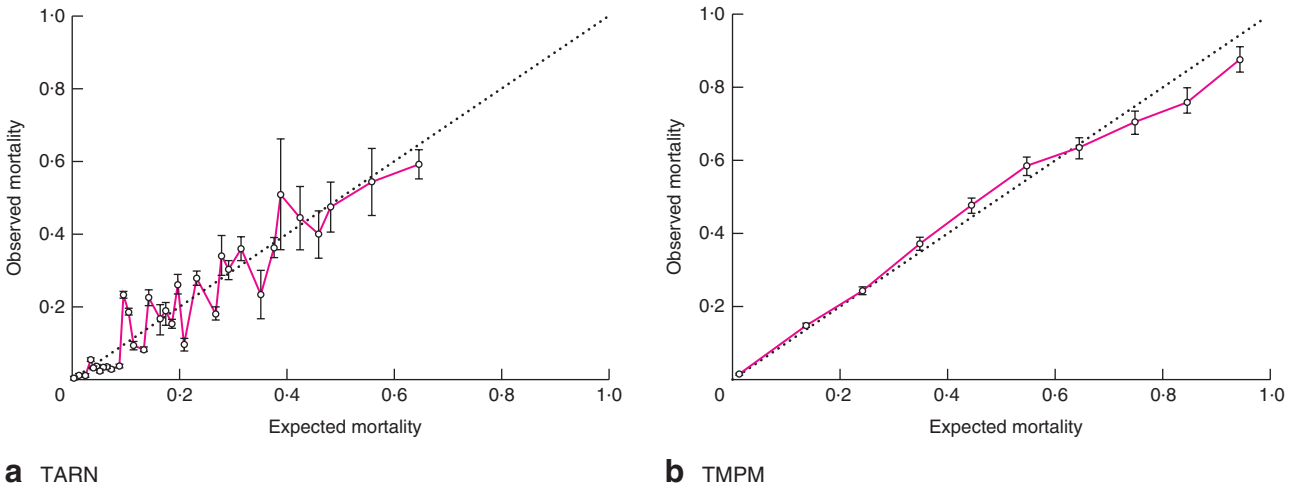

Fig. 1 Derivation of the calibration curve for the anatomical components of **a** the Trauma Audit and Research Network (TARN) and **b** the Trauma Mortality Prediction Model (TMPM)

probability values in the derivation group and 174 974 such values in the validation group. Given this level of granularity in the TMPM, the probabilities were grouped into deciles and plotted against the observed mortality for each decile. The 95 per cent confidence intervals for the AUROC, AIC and Hosmer–Lemeshow test for each severity score were calculated using 1000 bootstrap samples of the data set. There were no missing values for any model in the study.

All statistical analyses were performed using Stata/MP™ version 14.2 (Stata Corporation, College Station, Texas, USA).

Results

After applying the exclusion criteria, some 438 058 patients were included in the study (*Fig. S1*, supporting information). A total of 15 930 patients died (3·6 (95 per cent c.i. 3·6 to 3·7) per cent) during their hospitalization. The mean age was 49·8 years, and non-survivors were older (58·3 (58·0 to 58·7) years). White men represented the largest race–gender demographic group in the cohort (38·6 per cent). Blunt mechanisms of injury were predominant (85·6 per cent). Of these, the majority were due to falls (59·2 (59·0 to 59·4) per cent), followed by motor vehicle crashes (25·7 (25·5 to 25·8) per cent). Firearm-related injuries were associated with the highest mortality (8·4 (7·8 to 8·9) per cent). Median (i.q.r.) ISS and TMPM values for probability of death were 9 (5–17) and 0·010 (0·005–0·024) respectively (*Table 1*). The range of predicted probabilities of death were $1{\cdot}34 \times 10^{-6}$ – 1·0 and 0·002 – 0·666 for TMPM and TARN respectively.

There were 350 325 (80·0 per cent) and 87 733 (20·0 per cent) patients in the derivation and validation groups respectively. There were no significant differences between these two groups with respect to age, sex, race/ethnicity, mechanisms of injury, median ISS, duration of stay in hospital or the ICU, and rates of ICU admission and death.

AUROC values were used to compare discrimination between survivors and non-survivors for the anatomical injury scores. TMPM exhibited better discrimination compared with TARN in the derivation group (0·880 (95 per cent c.i. 0·880 to 0·880) *versus* 0·841 (0·840 to 0·841) respectively). Similarly, AUROC was higher for TMPM in the validation group (0·882 (0·882 to 0·882), compared with 0·845 (0·845 to 0·846)) for the fractional polynomial transformation of ISS used by TARN (*Table 2*).

The AIC was used to estimate each model's proximity to a theoretical and unknown 'ideal' probability of mortality model. Given a group of prediction models, the best model is the one with the smallest AIC value. TMPM performed better in comparison to the fractional polynomial transformation of ISS used by TARN. Similarly, TMPM demonstrated a lower value in the Hosmer–Lemeshow test and is thus a better fit to the data in the present study (*Table 2*).

Calibration curves were plotted for TMPM and TARN. When the observed percentage mortality was plotted over the proportion of mortality predicted by each of the 44 the discrete TARN ISS values, a non-monotonic pattern was observed in both the derivation and validation groups. It is noteworthy that the maximum probability of mortality predicted by the fractional polynomial transformation of ISS

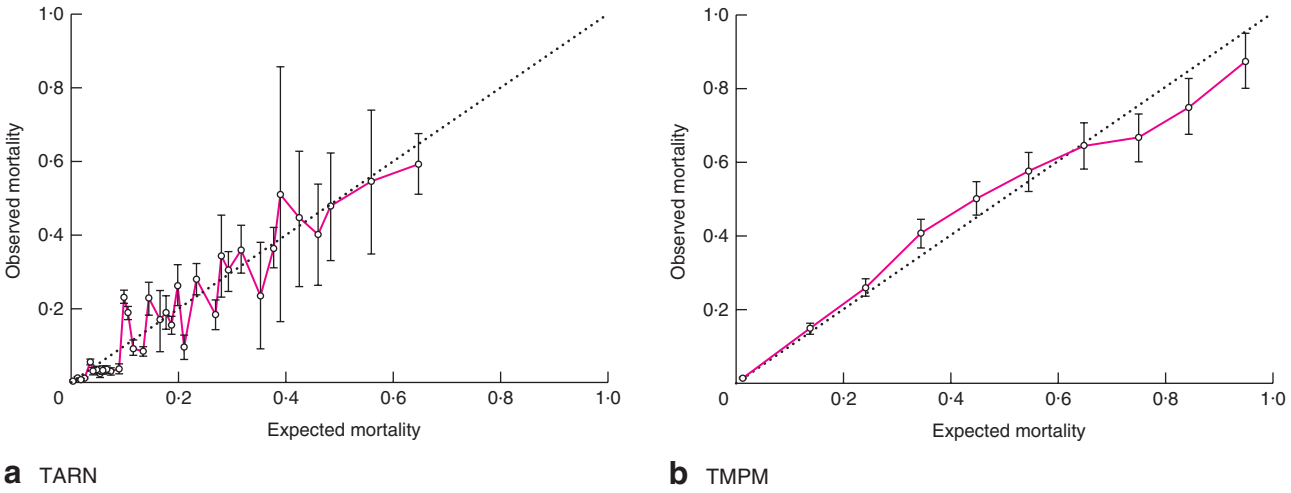

Fig. 2 Validation of the calibration curve for the anatomical components of **a** the Trauma Audit and Research Network (TARN) and **b** the Trauma Mortality Prediction Model (TMPM)

was approximately 0·6. This is likely due to the inherent limitations of ISS as the mortality rate among patients with ISS values of 75 in this study was 59·2 per cent (*Figs 1* and *2*).

Discussion

TMPM exhibited greater discrimination over a broader range of injury severity than the ISS-based anatomical injury component of TARN. Additionally, TMPM had a lower AIC value, indicating that it approximates a hypothetical 'ideal' model more closely. Moreover, the goodness-of-fit for TMPM surpassed that of TARN. In sum, the inherent limitations of the ISS, which have been well documented in the literature[13,14,22], are not overcome by mathematical transformation.

Illness severity in injured patients is measured using the probability of injury-related death. Unlike TMPM, in which injury severity is estimated empirically using the AIS so-called 'predot' codes, the basis for the injury severity component of the TARN model is the AIS injury severity based on expert consensus. Robust measures of injury severity are essential for performance benchmarking in order to improve outcomes in injured patients, and for incentivizing higher-quality care in the NHS National Tariff Payment System and the Merit-Based Incentive Payment System and Value Based Purchasing from the Centers for Medicare and Medicaid Services in the USA. As anatomical injury is the fundamental element of trauma, the credible measurement of injury severity is essential to the evaluation of trauma care quality. In the present study cohort, the AUROC was four percentage points greater for TMPM. Thus, TMPM accurately predicted mortality in 589 more patients than did the injury severity component of TARN. The cumulative weight of the findings suggest that TARN should consider using TMPM to quantify injury severity instead of relying on an ISS-based severity measure.

Developed over 40 years ago, ISS was adopted as the standard for measuring anatomical injury severity[10]. It is the sum of the squared AIS severity values for the worst injury in each of three separate body regions. AIS severity values are based on expert consensus, rather than being derived empirically. This algorithm results in 56 combinations of squared AIS severity scores that yield 44 possible unique ISS values. The ISS is relatively simple to calculate, and it is this simplicity that is probably the key to its longevity. However, the ISS is also known for its non-monotonic nature due to steep peaks and valleys in predicted mortality as the ISS value gets larger[13]. More importantly, trauma care has advanced considerably in the 43 years since its publication, as evidenced by the limitation of ISS in predicting death in the most severely injured patients. The TARN mortality model applies a sophisticated polynomial transformation of the ISS as the measure of anatomical severity. This improved the performance of the TARN model, as demonstrated by Bouamara and colleagues in 2006[23].

The TMPM was developed in 2008 using an empirical regression-based approach to estimate injury severity, instead of relying on the expert-based estimates of injury severity used in ISS-based severity measures. TMPM provides empirically derived probabilities of death using the AIS lexicon. Previous work[14] comparing

TMPM to ISS, maximum AIS score, New Injury Severity Score and the ICD-9-based injury severity score (ICISS) demonstrated that TMPM predicts trauma mortality more accurately. Although somewhat technical in approach, it is important to assess the adequacy of the models before interpreting their results[24]. The present analysis is the first study to compare the anatomical component of TARN with TMPM, and find TMPM performed best.

This study has several limitations. First, the NTDB is not population-based but is built on a self-selected group of trauma centres. Thus, the findings may not be generalizable to all trauma and non-trauma centres. However, *a priori*, there is no reason to believe that the present finding of empirical measures of injury severity outperforming measures of injury severity based on expert consensus would be limited to this sample of injured patients. Second, the complete TARN model includes variables for age, Glasgow Coma Scale score and terms for the Charlson Co-morbidity Index[18,25,26]. Thus, the present results represent the anatomical injury components only, and no inference should be made regarding the predictive capabilities of the complete TARN model. However, the basis for accurate risk adjustment in trauma mandates accurate specification of the anatomical component of injury severity, before including measures of physiological derangement and co-morbidity.

Disclosure

The authors declare no conflict of interest.

References

1 Committee on Trauma. *Resources for Optimal Care of the Injured Patient* 2014. American College of Surgeons: Chicago, 2014.

2 Intercollegiate Group on Trauma Standards. *Regional Trauma Systems: Interim Guidance for Commissioners*. Royal College of Surgeons of England: London, 2009.

3 Hemmila MR, Nathens AB, Shafi S, Calland JF, Clark DE, Cryer HG *et al.* The Trauma Quality Improvement Program: pilot study and initial demonstration of feasibility. *J Trauma* 2010; **68**: 253–262.

4 Lecky F, Woodford M, Yates DW. Trends in trauma care in England and Wales 1989–97. *Lancet* 2000; **355**: 1771–1775.

5 Champion H, Copes W, Sacco W *et al.* A new characterization of injury severity. *J Trauma* 1990; **30**: 539–546.

6 Champion HR, Sacco WJ, Carnazzo AJ, Copes W, Fouty WJ. Trauma score. *Crit Care Med* 1981; **9**: 672–676.

7 Champion HR, Sacco WJ, Lepper RL, Atzinger EM, Copes WS, Prall RH. An anatomic index of injury severity. *J Trauma* 1980; **20**: 197–202.

8 Levy PS, Mullner R, Goldberg J, Gelfand H. The estimated survival probability index of trauma severity. *Health Serv Res* 1978; **13**: 28–35.

9 Osler T, Rutledge R, Deis J, Bedrick E. ICISS: an International Classification of Disease-9 based injury severity score. *J Trauma* 1996; **41**: 380–386.

10 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.

11 Boyd C, Tolson M. Evaluating trauma care: the TRISS method. *J Trauma* 1987; **27**: 370–378.

12 MacKenzie E, Steinwachs D, Shankar B. An ICD-9-CM to AIS conversion table: development and application. *Annual Conference of the American Association for Automotive Medicine*, 6–8 October 1986, Montreal.

13 Kilgo PD, Meredith JW, Hensberry R, Osler TM. A note on the disjointed nature of the Injury Severity Score. *J Trauma* 2004; **57**: 479–485.

14 Cook A, Weddle J, Baker S, Hosmer D, Glance L, Friedman L *et al.* A comparison of the Injury Severity Score and the Trauma Mortality Prediction Model. *J Trauma Acute Care Surg* 2014; **76**: 47–52.

15 Osler T, Glance L, Buzas JS, Mukamel D, Wagner J, Dick A. A trauma mortality prediction model based on the Anatomic Injury Scale. *Ann Surg* 2008; **247**: 1041–1048.

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

17 American College of Surgeons. *National Trauma Data Bank Annual Report 2014*; 2014. https://www.facs.org/quality%20programs/trauma/ntdb/docpub [accessed 10 November 2015].

18 Bouamra O, Jacques R, Edwards A *et al.* Prediction modelling for trauma using comorbidity and 'true' 30-day outcome. *Emerg Med J* 2015; **32**: 933–938.

19 Hosmer DW Jr, Lemeshow S, Sturdivant RX. Assessing the fit of the model. In *Applied Logistic Regression* (3rd edn). John Wiley: Hoboken, 2013; 153–226.

20 Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974; **19**: 716–723.

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

22 Haider AH, Villegas CV, Saleem T, Efron DT, Stevens KA, Oyetunji TA *et al.* Should the IDC-9 Trauma Mortality Prediction Model become the new paradigm for benchmarking trauma outcomes? *J Trauma Acute Care Surg* 2012; **72**: 1695–1701.

23 Bouamra O, Wrotchford A, Hollis S, Vail A, Woodford M, Lecky F. A new approach to outcome prediction in trauma: a

comparison with the TRISS model. *J Trauma* 2006; **61**: 701–710.
24 Hosmer DW Jr, Lemeshow S, Sturdivant RX. Interpretation of the fitted logistic regression model. In *Applied Logistic Regression* (3rd edn). John Wiley: Hoboken, 2013; 49–88.
25 Charlson M, Pompei P, Ales K, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; **40**: 373–383.
26 Dr Foster Intelligence. *Understanding HSMRs: a Toolkit on Hospital Standardised Mortality Ratios*; 2014. http://www.drfoster.com/wp-content/uploads/2014/09/HSMR_Toolkit_Version_9_July_2014.pdf [accessed 1 December 2017].

Supporting information

Additional supporting information can be found online in the supporting information tab for this article.