Development and Validation of the Revised Injury Severity Classification Score for Severely Injured Patients

Rolf Lefering[1]

Abstract

Introduction: Trauma scores are often used for prognostication and the adjustment of mortality data. The appropriate consideration of identified prognostic factors is mandatory for a valid score with good outcome prediction properties. The Trauma Registry of the German Society for Trauma Surgery (TR-DGU) initially used the Trauma and Injury Severity Score (TRISS) but various reasons led to the development of a new scoring system, the Revised Injury Severity Classification (RISC).

Patients and Methods: A total of 2,008 severely injured patients with complete data documented in the TR-DGU during the period 1993–2000 were used to develop a new score. Patients were split into a development sample (n = 1,206) and a validation sample (n = 802). Multivariate logistic regression analysis was applied, and the results were compared with existing score systems. The quality of prediction was determined regarding discrimination (disparity, sensitivity, specificity, receiver operating characteristic [ROC] curve), precision (predicted versus observed mortality), and calibration (Hosmer–Lemeshow goodness-of-fit).

Results: Existing score systems (ISS, NISS, RTS, ASCOT, TRISS, Rixen) revealed areas under the ROC curve ranging from 0.767 to 0.877. The RISC combines 11 different components: age, NISS, head injury, severe pelvic injury, Glasgow Coma Scale, partial thromboplastin time (PTT), base excess, cardiac arrest, and indirect signs of bleeding (shock, mass transfusion, and low hemoglobin). The new RISC score reached significantly higher values of above 0.90 for the area under the ROC curve in both development and validation samples. Application to data from 2001 confirmed these results.

Conclusion: Outcome prediction including initial laboratory values was able to significantly improve the ability to discriminate between survivors and non-survivors. The adjustment of mortality rates should be based on the best available prediction model.

Key Words
Severe injuries · Polytrauma · Scoring systems · Prognosis · Outcome prediction · Mortality · Trauma registries

Eur J Trauma Emerg Surg 2009;35:437–47

DOI 10.1007/s00068-009-9122-0

Introduction

The description and comparison of multiple-injured patients in clinical trials and registries require tools to describe the severity of injuries. Outcome comparisons without considering the different levels of injury severity are worthless and misleading. Therefore, as one of the first systems of its kind, the Injury Severity Score (ISS) was introduced in the early 1970s [1]. However, this anatomical description of injury severity does not consider the physiological response to injury which also determines outcome. Trauma Score and the Revised Trauma Score have been developed to cover this part of severity classification [2, 3]. Based on the results of the Major Trauma Outcome Study (MTOS), the combination of these two aspects enhanced by the information about the patient's age led to more sophisticated scores like the Trauma and Injury Severity Score (TRISS) and A Severity Classification of Trauma (ASCOT) [4, 5]. Especially, the TRISS has subsequently been broadly accepted and applied

[1]Institute for Research in Operative Medicine (IFOM), Medical Faculty, University Witten/Herdecke, Cologne, Germany.

Received: June 28, 2009; revision accepted: August 9, 2009;
Published Online: September 18, 2009

worldwide in trauma registries and clinical studies. Observed mortality rates are compared with prognostic estimates derived from severity scores. Thus, the comparison of outcome in patient groups with varying severity became possible.

Also the Trauma Registry of the German Society for Trauma Surgery (TR-DGU), founded in 1993, inspired by the MTOS results, used the TRISS as a tool to compare outcomes across hospitals.

However, the use of the TRISS for the estimation of prognosis is also discussed critically [6]. Using the TRISS as a standard means, our own results are compared to the outcome of mostly American hospitals in the late 1980s. Furthermore, the German preclinical emergency system differs substantially from the American one, where paramedics minimize preclinical time ('load and go' vs. 'stay and play') [7, 8]. Germany provides a physician-based preclinical system with the initiation of first therapeutic interventions before transportation. Thus, hospital admission data are difficult to compare between these two countries.

Other authors criticize the TRISS for not adequately considering the importance of age or head injury. Our own investigations based on the TR-DGU have shown that additional prognostic factors like coagulation and base deficit could improve the outcome estimation [9, 10]. Furthermore, TRISS calculation was possible only for a limited number of patients, mainly due to missing respiratory rate data. Therefore, in 2001, it was decided to initiate an update of the TRISS and to include additional prognostic factors in a new prognostic system in order to optimize the adjustment for injury severity in outcome comparisons.

Patients and Methods

Trauma Registry

The Trauma Registry of the German Society for Trauma Surgery (TR-DGU, DGU stands for Deutsche Gesellschaft für Unfallchirurgie) was founded in 1993 as a nationwide initiative for prospective, standardized, and anonymous documentation of multiple-injured patients. The primary aim of the TR-DGU was to improve the quality of trauma care by providing a tool for inter-hospital benchmarking. Participation is voluntary and free of costs for the hospitals (www.traumaregister.de).

The data collected for each patient are structured in four groups corresponding to the consecutive phases of acute trauma care: (A) preclinical phase: mechanism of injury, initial physiology, first therapy, neurological signs, rescue time; (B) emergency room: physiology, laboratory findings, suspected pattern of injury, therapy, time sequence of diagnostics; (C) intensive care unit: status on admission, organ failure, sepsis, duration of ventilation; and (D) final outcome: hospital stay, survival, complete list of injuries and operative procedures. Injuries are described in the text and are coded according to the Abbreviated Injury Scale (AIS), version 1990/update 1998 [11]. The documentation is continued until the patient's death or discharge from hospital.

Documentation was done on paper forms by the responsible physician in each participating hospital. Extended manuals are provided, and instructions for the documentation is included as part of the documentation sheet. Data were checked, coded, and entered into a database by experienced personnel in three local documentation centers (Cologne, Essen, Hannover). Since 2002, data collection was done with a central password-protected Internet-based documentation system with automatic checks for the completeness of selected variables and plausibility checks.

Only anonymized data were collected as part of the legally required activities of hospitals for external quality assessment, thus, no informed consent was required from the patients. Once a year, all of the participating hospitals receive an extended individual audit report containing more than 2,000 specific data during a 1-day workshop with scientific presentations.

Patients

From 1993 to 2000, 8,056 patients from 88 hospitals in Germany, Austria, and Switzerland were documented in the TR-DGU. The criteria for inclusion in the TR-DGU are as follows: the patient has to be alive upon admission to hospital and is in potential need for intensive care. Cases with burns, poisoning, or drowning were excluded, as well as isolated proximal femoral fractures in elderly patients. There is no minimum ISS required, however, the average ISS is high with 23.1 points, and 69% of cases had an ISS of 16 or more. Intensive care was required for 96% of cases. Seventy-six percent of patients were directly admitted from the scene of the accident, and the remaining patients were initially treated at another facility and subsequently transferred to the participating hospital.

For the present analysis, only primary admitted patients with available data for the calculation of the various scoring systems (ISS, RTS, TRISS, ASCOT) were considered. The availability of potential new prognostic factors like base excess and coagulation was

also required (for details, see the Development of the New Score section below).

Trauma Scores

The AIS is a hierarchical system of about 2,000 different injury codes [11]. As part of this code, the last digit represents the severity of injury, ranging from 1 (minor) to 6 (actually untreatable).

The AIS is used to calculate the ISS and the New Injury Severity Score (NISS) [1, 12]. For the ISS, the maximum AIS severity in six body regions (head and neck, face, thorax, abdomen, extremities, soft tissue) has to be derived. The severity scores of the three most severely injured body regions are then squared and added to yield the ISS. The NISS is calculated in a similar fashion, but considers the three worst injuries, irrespective of their location in the body.

The Glasgow Coma Scale (GCS) is a score for describing unconsciousness ranging from 3 (worst) to 15 (best). It combines three different aspects of coma: eye opening (1–4 points), verbal response (1–5 points), and motor function (1–6 points) [13].

The Revised Trauma Score (RTS) is based on blood pressure, GCS, and respiratory rate. These variables are categorized in a five-point scheme, which could also be used for triage purposes (T-RTS). For prognostic formulas, the RTS is calculated as a linear combination of these three variables, weighted with coefficients [3].

The TRISS combines ISS, RTS, and age, where age was included only as 'below/above 55 years of age.' There are different formulas for blunt and penetrating trauma. Various coefficients have been published for the TRISS; in this investigation, the coefficients from Champion's paper about national norms were used [5].

ASCOT is a further score derived from the MTOS data by the same authors as the TRISS, with a more sophisticated description of the anatomical injuries [4].

Finally, a score developed within the TR-DGU by Rixen et al. was also used here. This score is a weighted linear combination of the following continuous measurements: base excess, ISS, age, GCS, and Quick's value [9, 10, 14].

Details and formulas of the above-mentioned scores can be derived from the original publications or from recent reviews [15, 16].

Development of the New Score

Patients eligible for analysis (i.e., with complete data for all variables required) were split into a development sample and a validation sample by random allocation. After univariate analysis, multiple logistic regression analysis was applied, with hospital mortality as the dependent variable. Model building was performed by stepwise forward selection of the most discriminating variables. Continuous variables like age were also offered as categorical variables (with at least five categories) during model building in order to compensate for the often observed non-linear relationship with outcome. If a certain category of a categorical variable in the model was found not to be associated with mortality, this category was merged with the standard category of that variable. Adjacent categories with similar coefficients were merged as well.

In a first analysis based on 2,030 cases, a first model was derived. Candidate variables for this first model were age, gender, type of injury (blunt/penetrating), ISS, maximum AIS in various body regions, first GCS, first systolic blood pressure, first respiratory rate, first heart rate, several preclinical interventions, coagulation (partial thromboplastin time [PTT] and Quick's value), base excess, hemoglobin, preclinical volume, and mass transfusion. This model was extensively discussed in the TR-DGU Steering Group, and based on these results, a second model was calculated.

The second and final analysis was based on 2,008 primary admitted cases with complete data for the following variables: ISS, age, gender, type of injury, first systolic blood pressure, first GCS, preclinical cardiac arrest, base excess, coagulation (PTT and Quick's value), and indirect signs of relevant bleeding (shock, hemoglobin, and mass transfusion). For comparison, it was also required to have a valid TRISS score.

The patient group was split into a development sample (n = 1,206) and a validation sample (n = 802) by random allocation. Validation of the newly developed RISC score was also performed on TR-DGU data from 2001.

Statistics

Similarly to the MTOS, the data were described with the M-statistic of Boyd et al. [17]. In each database, the patients are divided into six risk bands based on their prognosis, and the relative percentage of the risk bands in each dataset is calculated. The sum of the respectively smaller percentages for each of the six risk bands gives the M-statistic. According to Hollis et al. [18], an adequate comparability is given if the M-statistic exceeds 0.88.

The quality of the new scoring systems in predicting mortality was analyzed and presented in terms of discrimination, precision, and calibration.

Discrimination measures the ability of a scoring system to separate survivors from non-survivors. This

is measured by sensitivity (percentage of correct prognoses in non-survivors) and specificity (percentage of correct prognoses in survivors), based on a certain cut-off value. This cut-off point is the 50% risk of death prediction value in the case of a probability score. In the case of a simple point score (e.g., ISS), the value which closest produced the observed mortality rate was chosen as the cut-off point. The total number of correct predictions (accuracy) was also calculated. No cut-off value is required for calculating the mean score values for survivors and non-survivors. The difference between these two values is called disparity.

Independently from a cut-off point, the mean difference in score values for survivors and non-survivors (disparity) was presented. This allows the easy comparison of discrimination if the same scale was used (e.g., scores producing a probability of death). As a summary measure for discrimination not depending on a certain cut-off point, the area under the receiver operating characteristic (ROC) curve was determined, together with its 95% confidence interval.

Precision describes how well a score-based prognosis is able to meet the overall observed mortality rate. This could, of course, only be calculated for probability scores.

Calibration describes the ability to predict low and high values equally well. As a standard measure for calibration, the goodness-of-fit statistic of Hosmer and Lemeshow (H–L statistic) was used here [19]. This H–L statistic measures and combines how close the predicted and observed mortality coincide in ten equally sized subgroups of increasing severity. A well calibrated score receives low values in the H–L statistic, which is not significantly different from zero.

Differences between survivors and non-survivors in the univariate analysis were evaluated with the Chi-square test or Student's *t*-test in case of prevalence rates or measurements, respectively. A p-value < 0.05 was considered to be statistically significant.

All analyses were performed with SPSS statistical software (SPSS Inc., Chicago, IL, USA).

Imputation of Missing Values

In order to achieve a reasonable and valid score result for as many patients as possible, the following concept was followed for the imputation of missing values:

1. Variables with absolute importance for scoring were not replaced; this regards age and ISS with its derived variables (e.g., head injury severity)
2. Cases with missing values in more than half of the required variables were excluded from replacement; because the information content was deemed too small
3. For a variable with missing values, a proxy variable was sought initially; if categorization was required, this was done based on univariate analysis (similar mortality rates)
4. If no proxy variable was available, it was checked whether a standard category could be used for the imputation of missing values (i.e., no points in the score)
5. If none of the above mentioned procedures was successful, the score was recalculated with a reduced set of variables where the interesting variable was missing

For the final comparisons, it was decided to use the TRISS prognosis in cases where the TRISS was available but not the RISC (in spite of imputation). This allows to calculate a prognosis at least for all cases with TRISS.

Results

The comparability of patients in the development sample (n = 1,206) and the validation sample (n = 802) is presented in table 1. Primary admitted patients not considered for model building due to at least one missing data point are also presented in table 1.

The discrimination ability of the six existing trauma score systems in the development dataset is presented in table 2 and figure 1. The highest area under the curve (AUC) of the ROC curve was observed for the Rixen score (0.877), but the TRISS and ASCOT follow closely. The precision (overall predicted mortality rate) of the Rixen score, TRISS, and ASCOT is moderate when compared to the observed mortality rate of 16.6% in this group. The results of the TRISS, especially the precision, could be improved by calculating new coefficients adapted to the given dataset. This modification was able to improve the AUC of the ROC curve for the TRISS by 0.01. The good precision of ISS, NISS, and RTS is based on the appropriate selection of the cut-off point.

The univariate analysis of prognostic factors is presented in table 3. All variables except gender, type of injury, and extremity injuries (AIS ≥ 3) were significantly associated with hospital mortality. Repeated stepwise logistic regression analysis was applied to reduce the number of variables, and also to reduce the number of categories per variable. The final Revised Injury Severity Classification (RISC) score is presented in table 4. The only continuous

Table 1. Demographic data of the development and the validation sample, also showing the excluded cases (only primary admissions). The second p-value compares the combined development and validation sample (n = 2,008) with the excluded patients.

	Development sample (n = 1,206)	Validation sample (n = 802)	p-value	Excluded patients (n = 4,079)	p-value
Age (mean)	38.2	38.8	0.36	39.1	0.019
Male gender (%)	74	73	0.55	72	0.20
Blunt trauma (%)	95	96	0.70	95	0.52
ISS (mean)	25.4	25.2	0.75	22.0	< 0.001
Helicopter transfer (%)	44	44	0.86	40	0.006
Traffic related (%)	67	68	0.82	61	< 0.001
Hospital mortality (%)	16.6	16.2	0.83	18.9	0.018
Length of stay (days)	29.0	30.0	0.48	25.2	< 0.001

Table 2. Discriminative ability of seven different trauma score systems based on 1,206 primary admitted trauma patients (development sample). The observed mortality in this group was 16.6%. TRISS* is TRISS with adjusted coefficients.

Trauma scores	ISS	NISS	RTS	TRISS*	ASCOT	Rixen	TRISS
Cut-off point	40	49	4.1	0.5	0.5	0.5	0.5
Sensitivity (%)	45	49	44	56	51	56	41
Specificity (%)	89	90	89	90	92	91	96
Accuracy (%)	81.9	83.2	81.1	86.1	85.2	85.4	86.7
Mean for survivors	22.8	27.9	6.69	14.3	12.6	14.8	11.3
Mean for non-survivors	38.8	46.9	4.75	55.5	48.8	54.3	42.8
AUC of ROC curve	0.786	0.804	0.767	0.852	0.849	0.877	0.862
95% CI low	0.753	0.770	0.733	0.824	0.822	0.852	0.834
95% CI high	0.820	0.838	0.803	0.880	0.875	0.902	0.890
Prediction (%)	16.3	16.5	16.9	21.1	18.6	21.3	16.5

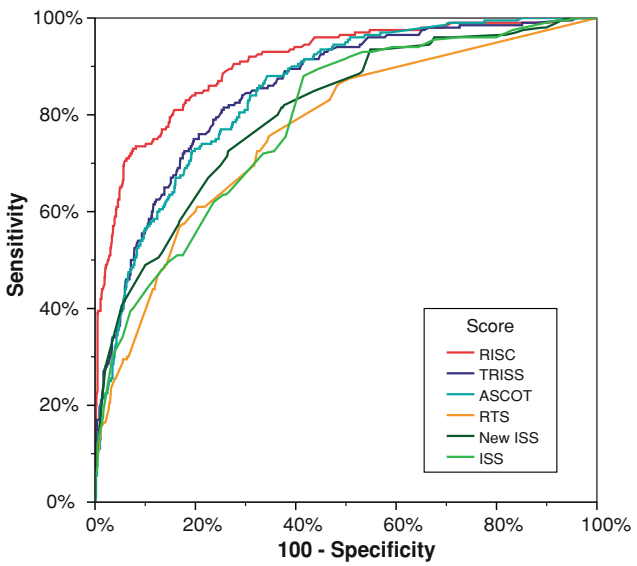

Figure 1. Receiver operating characteristic (ROC) curves for six prognostic score systems. The area under the ROC curve ranges from 0.767 (RTS) to 0.907 (RISC), see Tables 2 and 5. The analysis is based on 1,206 cases from the development sample.

variable in the final model is NISS. Relevant injury to the extremities which was not significantly associated with outcome in the univariate analysis was included but only with an AIS severity level of 5, which corresponds to an unstable pelvic fracture with relevant blood loss.

The score points for each category (except the coefficient for NISS) were rounded to one decimal place. This changed the individual prognosis by ±2% at the most. For categorical variables, the standard category was chosen to be the one with the lowest mortality rate, so that all categories received negative values. For an individual patient point weights for each variable were subtracted from the constant term of 5.0 to give the final score X, which is then transformed into a probability of survival $P(s)$ with the logistic function:

$$P(s) = 1/(1 + e^{-X}) = e^{X}/(1 + e^{X})$$

Table 5 presents the quality criteria of the RISC and TRISS in the development sample and in the validation sample. The AUC of the ROC curve for the

Table 3. Univariate analysis of potential prognostic variables in 1,206 cases from the development sample with respect to hospital outcome. The subscript A refers to preclinical values and the subscript B to first clinical data.

		Survivor (n = 1,006)	Non-survivor (n = 200)	p-value
Age	Mean	37.0	44.2	< 0.001
Male gender	%	74.7	69.0	0.21
Blunt trauma	%	95.3	95.0	0.84
ISS	Mean	22.8	38.8	< 0.001
NISS	Mean	27.9	46.9	< 0.001
Head injury (AIS ≥ 3)	%	39.9	75.0	< 0.001
Thorax injury (AIS ≥ 3)	%	46.6	58.0	0.003
Abdominal injury (AIS ≥ 3)	%	16.3	32.5	< 0.001
Extremities (AIS ≥ 3)	%	45.1	46.0	0.82
GCS_A 3–5	%	14.5	52.5	< 0.001
$Shock_A$ (< 90 mmHg)	%	11.4	35.0	< 0.001
Cardiac $arrest_A$	%	0.1	8.0	< 0.001
Base $excess_B$ (mmol/l)	Mean	–2.5	–6.9	< 0.001
PTT_B (s)	Mean	34.8	62.0	< 0.001
Quick's $value_B$ (%)	Mean	78.8	56.0	< 0.001
$Hemoglobin_B$ (< 9 mg/dl)	%	17.3	51.0	< 0.001
Mass $transfusion_B$	%	10.9	38.5	< 0.001
Indirect bleeding signs	Mean	0.4	1.2	< 0.001

RISC is about 0.05 points better than the respective AUC for the TRISS in both datasets. This difference is statistically significant (p = 0.003 and p = 0.013 in the development and validation samples, respectively). The calibration and precision of the RISC is satisfactory, while the prognosis of the TRISS exceeds the observed mortality by 4–5%.

The RISC score requires 11 different data points from each patient, where three of them were derived from injury coding. Only 40% of all primary admitted patients (2,437 of 6,087) had all 12 data points available without any missing. Another 37% of cases had missing data for only one data point. Patients without data about the injury coding (n = 263, 4.3%) or with missing age (n = 12; 0.2%) were not considered for the imputation of missing values. Also, patients with missing data in five or more of the required data points (n = 167, 2.7%) were excluded. Together, 356 patients (5.8%) were, thus, excluded from the imputation of missing values.

The replacement strategies for each variable in the RISC score is presented in table 6. Age and ISS were compulsory; for all other variables, specific strategies were derived. When all of these strategies were applied, calculation of the RISC prognosis became possible for an additional 3,155 patients (51.8%). In 39 cases (0.6%), the TRISS was available but not the RISC, despite the replacement attempts. In these patients, the RISC prognosis was set equal to the TRISS prognosis. Finally, 100 patients (1.6%) did not receive an RISC prognosis because of remaining missing data. Overall, the RISC could, thus, be calculated for 92.5% of patients.

Assessment of the quality criteria for the RISC in patients with complete data and those where some variables were imputed is given in table 7. The AUC of the ROC curve was comparable, but the precision and calibration decreased slightly in cases with replacements.

Validation of the RISC in the TR-DGU data from 2001 is also presented in table 7. Patients with complete data for the RISC and those with partial imputation were considered together (n = 2,070). The mean age in this group was 39.9 years (standard deviation [SD] 19.8), and 72% were male. The mean ISS was 24.6 points (SD 14.4), and 95.7% of patients sustained a blunt trauma. Discrimination, precision, and calibration were satisfactory.

Discussion

The outcome data of severely injured patients from different hospitals could only be interpreted directly if the patient populations are comparable. But in most instances, this comparability is not given. Level one trauma centers usually treat the more severe cases, and, thus, observe higher mortality rates than smaller hospitals with less severe cases. Therefore, mortality rates should be analyzed together with the injury severity of the treated patients.

Table 4. The Revised Injury Severity Classification (RISC) score. The category without a coefficient (marked with '–') is the standard category which receives no points when the score is calculated. The column 'Patients' describes the distribution in the development data set. ED: emergency department; BP: blood pressure; pRBC: packed red blood cells; AIS: Abbreviated Injury Scale.

Variable	Unit	Value	Coefficient	Patients (n, %)
Age	Years	< 55	–	943 (78.1)
Age at injury		55–64	–1.0	152 (12.6)
		65–74	–2.0	63 (5.2%)
		75+	–2.3	48 (4.0%)
NISS	Score	1–75	–0.03	All
New injury severity score				
Head injury	AIS	0–3	–	826 (68.5%)
Head injury severity according to AIS		4	–0.5	202 (16.7%)
		5/6	–1.8	178 (14.8%)
Extremities	AIS	0–4	–	1,165 (96.6%)
Severity of injury to the extremities according to AIS		5	–1.0	41 (3.4%)
GCS	Points	6–15	–	955 (79.2%)
Glasgow Coma Scale; first preclinical assessment		3–5	–0.9	251 (20.8%)
PTT	s	< 40	–	888 (73.6%)
Partial thromboplastin time; first measurement in the ED		40–49	–0.8	163 (13.5%)
		50–79	–1.0	105 (8.7%)
		80+	–1.2	50 (4.1%)
Base excess	mmol/L		–	1,073 (89.0%)
Base excess. first measurement in the ED		–9.0 to –19.9	–0.8	112 (9.3%)
		Under –20	–2.7	21 (1.7%)
Relevant bleeding signs	Number	None	–	790 (65.5%)
Number of indirect signs of relevant bleeding: systolic BP < 90 mmHg		1	–0.4	232 (19.2%)
preclinical/hemoglobin < 9 mg/dl in ED/mass		2	–0.8	136 (11.3%)
transfusion > 9 units of pRBC in the ED		3	–1.6	48 (4.9%)
Cardiac arrest		No	–	1189 (98.6%)
Preclinical cardiac arrest with reanimation/defibrillation		Yes	–2.5	17 (1.4%)
Constant			5.0	All

Table 5. Quality criteria for the RISC and TRISS in both datasets, development sample (n = 1,206) and validation sample (n = 802).

	Development sample		Validation sample	
	TRISS	RISC	TRISS	RISC
Discrimination				
Mean for survivors	14.3	8.9	13.6	8.7
Mean for non-survivors	55.5	54.4	55.7	51.6
Disparity	41.2	45.5	42.1	42.9
Sensitivity (%)	56.0	55.0	59.2	50.8
Specificity (%)	90.1	96.5	90.9	97.5
Accuracy (%)	84.4	89.6	85.8	89.9
AUC of ROC curve	0.852	0.907	0.860	0.909
95% confidence interval	0.824–0.880	0.883–0.931	0.827–0.892	0.879–0.938
Precision				
Observed mortality (%)	16.6	16.6	16.2	16.2
Score prognosis (%)	21.1	16.4	20.9	16.1
Calibration				
H-L statistic	172.0	6.2	101.0	6.7
p-value	0.001	0.63	< 0.001	0.57

Table 6. Strategy for the imputation of missing values for the RISC score. The rate of missing values refers to all primary patients (n = 6,087). The subscript A refers to preclinical values, and the subscript B to the first assessment in the hospital.

Variable	Missing values (%)	Replacement strategy	Comment
Age	0.2	None	Compulsory variable
NISS	4.3	None	Compulsory variable
Head injury	4.3	None	Compulsory variable
Extremity injury	4.3	None	Compulsory variable
Cardiac arrest$_A$	3.2	Blood pressure$_A$ = 0 or cardiac arrest$_B$	Use standard category if no data available
GCS$_A$	5.8	GCS$_B$	Use standard category if no GCS available
Hemoglobin$_B$	8.2	Blood pressure$_B$	No replacement if both values were missing
Blood pressure$_A$	11.2	Blood pressure$_B$	No replacement if neither blood pressure is available
Blood transfusion$_B$	17.0	Standard category	Hemodynamic data suggest that cases with missing data were not transfused
PTT$_B$	19.0	Quick's value	If PTT and Quick were missing, double the points
40–49		50–79%	for indirect bleeding signs; no replacement if
50–79		30–49%	bleeding signs were missing
80 or above		Below 30%	
Base excess$_B$	40.9%	Choose the worst:	Use standard category if no data available
–9.0 to –19.9		platelets < 100,000	
–20 or below		cardiac arrest$_B$	

Table 7. Quality criteria for the RISC score in primary admitted patients with complete data (n = 2,437), in patients with replacement of at least one data point (n = 3,194), and in primary admitted patients from the 2001 dataset of the TR-DGU (n = 2,070; complete cases and cases with replacement combined).

	Complete cases	Cases with imputation	TR-DGU 2001
Discrimination			
Mean for survivors	9.2	8.3	9.1
Mean for non-survivors	52.3	54.5	52.9
Disparity	43.1	46.2	43.8
Sensitivity (%)	51.9	52.6	52.5
Specificity (%)	96.7	96.6	95.8
Accuracy (%)	89.1	88.3	89.1
AUC of ROC curve	0.902	0.919	0.907
95% confidence interval	0.885–0.918	0.908–0.931	0.890–0.924
Precision			
Observed mortality (%)	16.9	18.9	15.6
Score prognosis (%)	16.5	17.1	15.9
Calibration			
H–L statistic	11.2	30.3	12.8
p-value	0.19	< 0.001	0.12

Injury severity, of course, depends on the number, the extent, and the pattern of injuries. For this anatomical description of injury severity, the ISS is by far the most frequently used index. The MTOS results have shown, however, that the physiological response to trauma is at least as important. Together with age and type of injury, these prognostic factors have been combined in the TRISS score. Using the TRISS for prognostication in your own patients means that you compare your outcome with what would have been expected if these patients were treated in an average MTOS hospital about 20 years ago.

Like many other trauma registries, the TR-DGU used the TRISS to adjust for variation among hospitals when comparing mortality rates. However, various critiques have been brought forward for the TRISS. Age was not considered sufficiently and head injury had a worse outcome than predicted. The ISS, as part

of the TRISS, was said to not adequately consider multiple injuries to the same body region, which led to the development of the NISS by Osler et al. [12]. Also, the application of the TRISS in a prehospital emergency system where physicians had already started to treat the patient at the scene can cause further difficulties. Furthermore, MTOS data were more than 15 years older than the actual data, leading to an overestimation of the risk of death. Previous research in the TR-DGU has shown that additional prognostic factors like base deficit or coagulation exist if first laboratory values in the hospital were also considered. Finally, the completeness of data for the calculation of the TRISS is another problem not only in the TR-DGU where missing preclinical respiratory rates (27% of cases) was the primary source. These arguments led us to the development of a new prognostic score system presented here.

What are the major differences to the TRISS? There are no different formulas for blunt and penetrating trauma cases, which could partly be explained by the limited number of cases with penetrating trauma in the TR-DGU (about 5%). Like in the ASCOT score, age is now considered in four categories. However, the cut-off point of 55 years used in the TRISS could be verified here. Detailed analysis showed that the 55–59 years subgroup was the first with an increased mortality rate when compared to younger patients.

Injury severity is now reflected by three variables; NISS (instead of ISS) plus additional points for head injury and for unstable pelvic fractures with relevant blood loss. This confirms the specific importance of these injuries for survival [20, 21].

The major difference to the TRISS, however, is the inclusion of new parameters from initial laboratory assessment of the patients. Specifically, coagulation disorders are reflected by PTT (or Quick's value) and base excess. These parameters have already been identified in previous analyses from the TR-DGU and by other authors [9, 10, 22].

Relevant bleeding was felt to be one of the major determinants of outcome. However, the possibility to measure blood loss is rather limited. It was, therefore, attempted to have a number of indirect signs for relevant blood loss both from the preclinical setting and early emergency room phase. Initially, there was a fourth indirect sign, preclinical volume administration of at least 3,000 ml, which was discarded after internal discussion. Preclinical volume was used very deliberately in Germany in the 1990s; the average amount in severely injured patients was about 2,000 ml. But in recent years, a trend to smaller amounts could be observed. Furthermore, the international applicability of the RISC would be limited if this parameter was used. However, there is still one indirect sign for relevant bleeding (mass transfusion) which could not be determined shortly after admission, as with all of the other components of the RISC. Nonetheless, it was decided to keep this item because of the importance of bleeding and transfusion for outcome prediction [23]. In future revisions of the RISC, however, it is intended to replace this item by data which are available initially after admission. This could be, for example, the Trauma Associated Severe Hemorrhage (TASH) score that predicts mass transfusion [24].

Regarding the quality criteria of the RISC, it could be shown that there is a substantial advantage of the RISC over the TRISS. Calculating new coefficients for the TRISS would result in a correct precision (overall predicted mortality rate) but discrimination could only slightly be improved. The AUC of the ROC curve for the RISC improved by about 0.05 to values above 0.90, which is rather high. It could not be expected to improve these values further and further, since there is a natural limit. A value very close to 1 would mean that the outcome of a patient is already nearly completely determined by his/her condition on hospital admission. When comparing our results for the AUC of the ROC curve with other published values, it is mandatory to check for comparable patient groups. Outcome prediction in patients with minor injuries is not difficult and could easily raise the AUC for any score.

The replacement of missing data is another critical point which has to be discussed. On the one hand, prognostic estimates derived from complete data are clearly preferable. On the other hand, if the intended use of the RISC is considered, the exclusion of all cases with missing data could severely invalidate those results by selection bias [25]. The more data points are used for the calculation of a score, the higher the portion of cases with at least partial missing data. Only 40% of cases had complete data for RISC calculation. Therefore, a compromise was sought which allows for a reasonable outcome prediction despite partial missing data, but which also excludes cases from RISC calculation if too much of the information was missing. Finally, cases with no data about injury pattern or age and cases who miss more than half of the required data points were excluded from the RISC calculation. Fortunately, this group of patients was small (6%). After application of the replacement strategies, more than 90% of patients received a valid RISC score. Quality criteria for the RISC in patients with replacements are

slightly worse than in complete cases, as expected. However, the results of the 2001 dataset (complete and replaced cases together, see Table 7) are as good as in the development dataset. Repeated validations performed in the 2002 and 2003 data of the TR-DGU (results not shown here) revealed similar results.

Since 2004, the TR-DGU has used the RISC score instead of the TRISS for outcome adjustment in the annual audit reports for the participating hospitals. Although developed in 1,200 cases from the last decade only, the RISC is able to predict the observed hospital mortality in the whole database (> 30,000 cases) with a deviation of less than 1%. However, in the most recent years, the outcome remained significantly below the estimated level, which means that hospital mortality today is lower than in the 1990s [26]. The reasons for this improvement are multiple, but with a valid outcome prediction tool like the RISC, it is possible to evaluate different strategies like whole-body computed tomography (CT) scanning more accurately [27].

Conclusion

The introduction of new prognostic factors in a novel outcome prediction model (Revised Injury Severity Classification [RISC]) could significantly improve the prognostic quality. Comparison of mortality rates in severely injured patients, such as in hospital audit reports, should only be made with an adequate consideration of the initial situation of the patients, where anatomical injury severity is only one aspect.

Acknowledgments

The present paper is a short version of my PhD thesis. I would like to thank very much the head of our department, Edmund A. M. Neugebauer, who succeeded in generating a highly scientific atmosphere in his institute, which promotes the close cooperation of surgeons and other scientists. I would also like to thank all members of the Trauma Registry Steering Group, especially Dieter Rixen, Thomas Paffrath, Bertil Bouillon, and Steffen Ruchholtz, who gave a large amount of stimulating and critical input during the development process. I learned so much about traumatology from this project. Finally, I would like to thank the large group of enthusiastic trauma surgeons from more than 100 hospitals in Europe who kept the Registry running over the years, and who made it a unique source for scientific research.

Conflict of interest statement

The author declares that there is no actual or potential conflict of interest in relation to this article.

References

1. 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;14:187–96.
2. Champion HR, Sacco WJ, Carnazzo AJ, Copes W, Fouty WJ. Trauma score. Crit Care Med 1981;9:672–6.
3. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the trauma score. J Trauma 1989;29:623–9.
4. Champion HR, Copes WS, Sacco WJ, Lawnick MM, Bain LW, Gann DS, Gennarelli T, Mackenzie E, Schwaitzberg S. A new characterization of injury severity. J Trauma 1990; 30:539–45.
5. Champion HR, Copes WS, Sacco WJ, Lawnick MM, Keast SL, Bain LW Jr, Flanagan ME, Frey CF. The major trauma outcome study: establishing national norms for trauma care. J Trauma 1990;30:1356–65.
6. Gabbe BJ, Cameron PA, Wolfe R. TRISS: does it get better than this? Acad Emerg Med 2004;11:181–6.
7. Westhoff J, Hildebrand F, Grotz M, Richter M, Pape HC, Krettek C. Trauma care in Germany. Injury 2003;34:674–83.
8. Roudsari BS, Nathens AB, Arreola-Risa C, Cameron P, Civil I, Grigoriou G, Gruen RL, Koepsell TD, Lecky FE, Lefering RL, Liberman M, Mock CN, Oestern HJ, Petridou E, Schildhauer TA, Waydhas C, Zargar M, Rivara FP. Emergency medical service (EMS) systems in developed and developing countries. Injury 2007;38:1001–13.
9. Raum MR, Bouillon B, Rixen D, Lefering R, Tiling T, Neugebauer E; German Polytrauma Study Group of the German Trauma Society. The prognostic value of prothrombin time in predicting survival after major trauma: a prospective analysis of 1351 patients from the German trauma registry. Eur J Trauma 2001;27:110–6.
10. Rixen D, Raum M, Bouillon B, Lefering R, Neugebauer E; Atbeitsgemeinschaft "Polytrauma" of the Deutsche Gesellschaft für Unfallchirurgie. Base deficit development and its prognostic significance in posttrauma critical illness: an analysis by the trauma registry of the Deutsche Gesellschaft für Unfallchirurgie. Shock 2001;15:83–9.
11. Association for the Advancement of Automotive Medicine. The abbreviated injury scale 1990 revision – update 1998. AAAM: Barlington, 1998.
12. Osler T, Baker SP, Long W. A modification of the injury severity score that both improves accuracy and simplifies scoring. J Trauma 1997;43:922–6.
13. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet 1974;2:81–4.
14. Rixen D, Raum M, Bouillon B, Schlosser LE, Neugebauer EAM; AG Polytrauma der DGU. Prognoseabschätzung des Schwerverletzten – Eine Analyse von 2069 Patienten des Traumaregisters der DGU. Unfallchirurg 2001;104:230–9.
15. Lefering R. Scoring beim Polytrauma. In: Oestern HJ, ed. Das Polytrauma: Präklinisches und klinisches management. Urban & Fischer: München, 2007:41–54.
16. Senkowski CK, McKenney MG. Trauma score systems: a review. J Am Coll Surg 1999;189:491–503.
17. Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. J Trauma 1987;27:370–8.
18. Hollis S, Yates DW, Woodford M, Foster P. Standardized comparison of performance indicators in trauma: a new approach to case-mix variation. J Trauma 1995;38:763–6.

19. Hosmer DW, Lemeshow S. Goodness of fit tests for the multiple logistic regression model. Commun Stat 1980;A9:1043–69.
20. Lefering R, Paffrath T, Linker R, Bouillon B, Neugebauer E; German Society for Trauma Surgery. Head injury and outcome – what influence do concomitant injuries have? J Trauma 2008;65:1036–44.
21. McMahon CG, Yates DW, Campbell FM, Hollis S, Woodford M. Unexpected contribution of moderate traumatic brain injury to death after major trauma. J Trauma 1999;47:891–5.
22. Kroezen F, Bijlsma TS, Liem MSL, Meeuwis JD, Leenen LPH. Base deficit-based predictive modeling of outcome in trauma patients admitted to intensive care units in Dutch trauma centers. J Trauma 2007;63:908–13.
23. Malone DL, Dunne J, Tracy JK, Putnam T, Scalea TM, Napolitano LM. Blood transfusion, independent of shock severity, is associated with worse outcome in trauma. J Trauma 2003;54:898–907.
24. Yücel N, Lefering R, Maegele M, Vorweg M, Tjardes T, Ruchholtz S, Neugebauer EAM, Wappler F, Bouillon B, Rixen D; "Polytrauma Working Group" of the German Trauma Society. Trauma associated severe hemorrhage (TASH)-score: probability of mass transfusion as surrogate for life threatening hemorrhage after multiple trauma. J Trauma 2006;60:1228–37.
25. Newgard CD. The validity of using multiple imputation for missing out-of-hospital data in a state trauma registry. Acad Emerg Med 2006;13:314–24.
26. Ruchholtz S, Lefering R, Paffrath T, Oestern HJ, Neugebauer E, Nast-Kolb D, Pape HC, Bouillon B. Rückgang der Traumaletalität. Ergebnisse des Traumaregisters der Deutschen Gesellschaft für Unfallchirurgie. Dt Ärzteblatt 2008;105:225–31.
27. Huber-Wagner S, Lefering R, Qvick LM, Körner M, Kay MV, Pfeifer KJ, Reiser M, Mutschler W, Kanz KG; Working Group on Polytrauma of the German Trauma Society. Effect of whole-body CT during trauma resuscitation on survival: a retrospective, multicentre study. Lancet 2009;373:1455–61.

Address for Correspondence
Rolf Lefering, PhD
Institute for Research in Operative Medicine (IFOM)
Medical Faculty
University Witten/Herdecke
Ostmerheimer Str. 200
51109 Cologne
Germany
Phone (+49/221) 98957-16, Fax -30
e-mail: rolf.lefering@ifom-uni-wh.de