JOURNAL OF NEUROTRAUMA 31:1721–1732 (October 15, 2014)

DOI: 10.1089/neu.2014.3361

Predicting Outcome after Traumatic Brain Injury: Development of Prognostic Scores Based on the IMPACT and the APACHE II

Rahul Raj,[1] Jari Siironen,[1] Riku Kivisaari,[1] Juha Hernesniemi,[1] and Markus B. Skrifvars[2]

Abstract

Prediction models are important tools for heterogeneity adjustment in clinical trials and for the evaluation of quality of delivered care to patients with traumatic brain injury (TBI). We sought to improve the predictive performance of the IMPACT (International Mission for Prognosis and Analysis of Clinical Trials) prognostic model by combining it with the APACHE II (Acute Physiology and Chronic Health Evaluation II) for 6-month outcome prediction in patients with TBI treated in the intensive care unit. A total of 890 patients with TBI admitted to a large urban level 1 trauma center in 2009–2012 comprised the study population. The IMPACT and the APACHE II scores were combined using binary logistic regression. A randomized, split-sample technique with secondary bootstrapping was used for model development and internal validation. Model performance was assessed by discrimination (by area under the curve [AUC]), calibration, precision, and net reclassification improvement (NRI). Overall 6-month mortality was 22% and unfavorable neurological outcome 47%. The predictive power of the new combined IMPACT–APACHE II models was significantly superior, compared to the original IMPACT models (AUC, 0.81–0.82 vs. 0.84–0.85; $p < 0.05$) for 6-month mortality prediction, but not for unfavorable outcome prediction (AUC, 0.81–0.82 vs. 0.83; $p > 0.05$). However, NRI showed a significant improvement in risk stratification of patients with unfavorable outcome by the IMPACT–APACHE II models, compared to the original models (NRI, 5.4–23.2%; $p < 0.05$). Internal validation using split-sample and resample bootstrap techniques yielded equivalent results, indicating low grade of overestimation. Our findings show that by combining the APACHE II with the IMPACT, improved 6-month outcome predictive performance is achieved. This may be applicable for heterogeneity adjustment in forthcoming TBI studies.

Key words: APACHE II; external validation; IMPACT; outcome; prognostic models; traumatic brain injury

Introduction

Traumatic brain injury (TBI) is one of the leading causes of morbidity and mortality in the world, and no major improvements in prognosis has been noted in recent decades.[1,2] Clinical trials are hard to conduct in TBI because of broad disorder heterogeneity and patient case mix, thus numerous trials have failed to show any improvements in patient outcome.[3] Further, because of the broad heterogeneity of TBI, clinical trials require large study populations, often from multiple centers. This creates further challenges, because one should now standardize for differences in quality of care provided by the participating institutions.[4] Benchmarking, by comparing the predicted and observed outcome, is an important tool for trauma health care quality evaluation.[5] Such benchmarking has been shown to reduce mortality rates after cardiac and noncardiac surgery.[6] It has been suggested that by using newly developed robust prediction models in TBI, similar improvements in quality of care and outcome could be achieved.[7]

Patients with TBI differ substantially in terms of prognosis from other critically ill patients, and several prognostic models specifically aimed for TBI have been developed.[8,9] The most robust and clinical applicable is the IMPACT (International Mission for Prognosis and Analysis of Clinical Trials) model, which uses admission characteristics to predict risk of 6-month outcome.[7,10] However, the IMPACT model is fairly new (introduced in 2008) and not as well established in centers around the world as the ''traditional'' intensive care unit (ICU) scoring systems, such as the APACHE II (Acute Physiology and Chronic Health Evaluation II).[11] The APACHE II (introduced in 1985) is the updated version of the original APACHE (introduced in 1981) and currently one of the

Departments of [1]Neurosurgery and [2]Intensive Care, Helsinki University Hospital, Helsinki, Finland.

world's most widely used ICU scoring systems.[12] In contrast to the IMPACT, the APACHE II does not consider admission characteristics, but instead uses 12 physiological variables measured in the first 24 h in the ICU.[11]

The IMPACT has been used for baseline risk adjustment in clinical trials in TBI (e.g., the Dexanabinol trial[13]), whereas the APACHE II is a common tool to adjust for differences in early intensive care and illness of severity in patients treated in the ICU.[14,15] We recently found satisfactory performance of the APACHE II for predicting 6-month mortality in patients with TBI treated in the ICU.[16] This implies that the APACHE II could be used for case-mix adjustment in TBI, as it has in other critically ill patients.

The aim of the present study was to create a new set of prediction models for patients with TBI treated in the ICU. We hypothesized that by combining the IMPACT, as an estimate of the severity of TBI, and the APACHE II, as an estimate of general illness severity, improved long-term outcome prediction would be achieved.

Methods

Patients and data collection

The ethics committee of Helsinki University Hospital (Helsinki, Finland) approved the study and waived the need for informed consent. The study population constituted of patients with moderate-to-severe TBI (admission Glasgow Coma Scale [GCS] 3–12) or complicated mild TBI (mTBI; admission GCS 13–15) treated in the ICU of one of the largest level 1 trauma centers in Scandinavia (Töölö Hospital, Helsinki University Hospital, catchment area population approximately 2 million) during a 4-year period (January 2009–December 2012). Definition of TBI was an S06.1–S06.9 ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th Edition) diagnosis, caused by an external force.[17] Patients with a history of head trauma, but no intracranial pathological findings (by computed tomography [CT] imaging) during the hospital stay, and those with subacute head injuries (>24 h) were not considered. Further exclusion criteria were age <16 years, penetrating head injury, dead on arrival, and death before CT imaging and/or ICU admission.[10,11]

Admission characteristics were assessed by an emergency department physician and were retrieved from subsequent electronic records. Patient head CT scans were retrospectively classified by a radiologist-neurosurgeon (R.K.) according to the Marshall and Rotterdam CT classification systems.[18,19] The APACHE II variables were extracted from the ICU software (PICIS, Anesthesia Manager®) by 5-minute intervals to pinpoint the worst physiological and laboratory values measured in the first 24 h in the ICU.[11] Treatment standards in the hospital followed the Brain Trauma Foundation guidelines.[20]

Outcome was 6-month mortality and neurological outcome. Non-Finnish citizens were excluded from the study, because they are not routinely followed up. Data on mortality were retrieved from the Finnish population register center (available for 100% of Finnish patients). Neurological outcome was determined based on outpatient clinic follow-ups by a neurosurgeon or neurologist 6 months from injury according to the Glasgow Outcome Scale (GOS), independently by two authors (R.R., J.S.), and dichotomized to unfavorable (GOS 1–3: death, vegetative state, or severe disability) and favorable (GOS 4–5: moderate disability and good recovery) outcome.[21] GOS assessment agreement was good, with a kappa of 0.90 (95% confidence interval [CI], 0.86–0.95); discrepancies were resolved by verbal discussion.

Prediction models

The IMPACT and APACHE II scores were calculated using the original methods.[10,11] The APACHE II is based upon 12 physiological variables (the most abnormal value measured in the first 24 h in the ICU), age, and chronic health status, giving a single score of 0–71. The IMPACT consists of three different models with increasing complexity (core, extended, and lab). The simplest IMPACT core consists of only age, motor score, and pupillary light reactivity and gives a single maximum score of 15. Addition of CT (epidural hematoma, traumatic subarachnoid hemorrhage, and Marshall CT classification) and secondary insult (hypoxia and hypotension) variables results in the IMPACT extended (maximal score, 22), and further addition of glucose and hemoglobin concentrations gives the IMPACT lab (maximum score, 29). None of the variables from the IMPACT and APACHE II overlap because they are measured on two separate occasions (measured on admission for IMPACT and measured during the first 24 h in the ICU for APACHE II).

For the development and internal validation of the new models, a split-sample technique, where the study population was randomly divided into a development and validation cohort, was used to mimic randomization.[22] Three new models were created using logistic regression: 1) the IMPACTcore–APAHCE II; 2) the IMPACText–APACHE II; and 3) the IMPACTlab–APACHE II. Each model was designed for 6-month mortality and unfavorable neurological outcome prediction using the following equation: $1/(1+e^{-\text{logit}})$, where each model has its own defined logit. To validate our results, secondary internal validation, using a resample bootstrap technique, was performed.[23] The equations of the new prediction models and the bootstrapped coefficients are found in the Supplementary Equations (see online supplementary material at http://www.liebertpub.com).

Statistical analyses

Categorical data are presented as *n* (%) and continuous data as median (interquartile range; IQR), unless otherwise mentioned. Differences between patients with good and poor outcome and between the validation and development cohort were tested using the chi-square (χ^2; two-tailed) test for categorical variables. Continuous data were tested for skewness and appropriate statistical test chosen accordingly for univariate analyses: the Mann-Whitney's U test for skewed data and the Student's *t*-test for normal distributed data.

The performance of the models was evaluated by assessing discrimination (by area under the receiver operator characteristic curve [AUC]), calibration (by GiViTI [Gruppo Italiano per la Valutazione degli Interventi in Terapia Intensiva, or Italian Group for the Evaluation of Interventions in Intensive Care Medicine] calibration belt and the Hosmer-Lemeshow $\hat{C}$-test [H-L]), and precision (Brier score).[24]

Discrimination refers to the model's ability to distinguish patients with good and poor outcome and is measured by the AUC. The AUC ranges from 0.5 (worst) to 1.0 (perfect). Generally, AUCs >0.90 are considered excellent, >0.80 good, >0.70 satisfactory, and <0.70 poor.[25] The AUC of the new models were compared to the original models using the Venkatraman test.[26] We also compared the performance of the new models to the original models by the continuous net reclassification improvement (NRI) test.[27] The NRI calculates the proportion of patients with outcome correctly assigned a higher probability and patients without outcome correctly assigned a lower probability by the new model, compared to the original model.

Calibration refers to the model's ability to differ between good and poor outcome over the whole risk spectrum. The calibration is classically measured using the H-L (also called the goodness-of-fit test). The H-L test divides the patients into equal-sized deciles (compared to the $\hat{H}$ test, which divides patients into deciles based on risk rather than group size), for which it calculates a χ^2 between the observed and predicted risk.[28] *p* values over 0.05 indicate no statistical difference between the predicted and observed risk and

the calibration is considered good, whereas p values under 0.05 indicate a significant difference between predicted and observed risk and the model is consequently considered to be poorly calibrated.[28] However, the H-L test is largely sample-size dependent and has been criticized.[29] To meet the critique of the H-L test, the GiViTI calibration belt was developed.[30] The GiViTI test calculates the relationship between the observed and predicted risks by fitting a polynomial function between the two and calculates the 80% CI (light gray area) and 95% CI (dark gray area). The diagonal bisector line indicates perfect calibration, and a statistically significant deviation between the predicted and observed outcome (i.e., poor calibration) occurs when the 95% CIs do not encompass the bisector line. This makes it possible to visually identify risk intervals of model over- and underprediction.

Precision is an overall performance indicator, and measured by the Brier score is the average of the sum of the squared difference between observed and predicted risk.[31] The Brier score ranges from 0.0 (perfect) to 0.25 (worst), with a 50% incidence of the outcome. When the outcome incidence is lower, the worst Brier score is lower. Accordingly, we performed the scaled Brier test calculating specific cut-off points for our data.

SPSS (version 21.0; IBM Corp., Armonk, NY), R (A Language and Environment for Statistical Computing; R-Foundation for Statistical Computing, Vienna, Austria), and Analyze-it for Microsoft Excel (version 3.5; Microsoft Corporation, Redmond, WA) were used for the statistical analyses. The PredictABEL library and was used for the H-L and NRI tests, the GiViTI library for the calibration belts, and the pROC library for the Venkatraman test.[30,32,33]

Results

In total, 1000 patients presenting with intracranial injury requiring admission to the ICU were screened, of which 890 met the study inclusion criteria (Fig. 1). Overall 6-month mortality was 23% ($n = 206$ of 895), and 48% ($n = 99$ of 206) died within 14 days of injury. A total of 48 patients were lost to follow-up, that is, their GOS could not be assessed (development cohort, $n = 24$; validation cohort, $n = 24$). Thus, 47% ($n = 394$ of 842) of patients had unfavorable neurological outcome (GOS 1–3) 6 months from injury.

Patient median age was 58 years (IQR, 44–68), median Rotterdam CT score was 3 (IQR, 3–4), and 35% of patients underwent immediate mass lesion evacuation. Of all patients, 34% had a hospital admission GCS of 13–15 and 66% a hospital admission GCS of 3–12. Median Rotterdam score for patients with admission GCS 13–15 was 3 (IQR, 2–3), compared to 4 (IQR, 3–5), for those with an admission GCS of 3–12 ($p < 0.001$). Of patients with an admission GCS of 13–15, 24% ($n = 71$ of 301) underwent acute mass lesion evacuation after admission, compared to 41% ($n = 239$ of 589) of those with an admission GCS of 3–12 ($p < 0.001$).

There were some significant ($p < 0.05$) differences in baseline characteristics between patients with good and poor outcome: Patients with poor outcome were younger, more often immunocompromised, had lower admission GCS, motor score, hemoglobin concentrations, platelet count and base excess, suffered more frequently from more hypoxia, had higher admission glucose concentrations, INR, and Rotterdam CT score, and fewer epidural

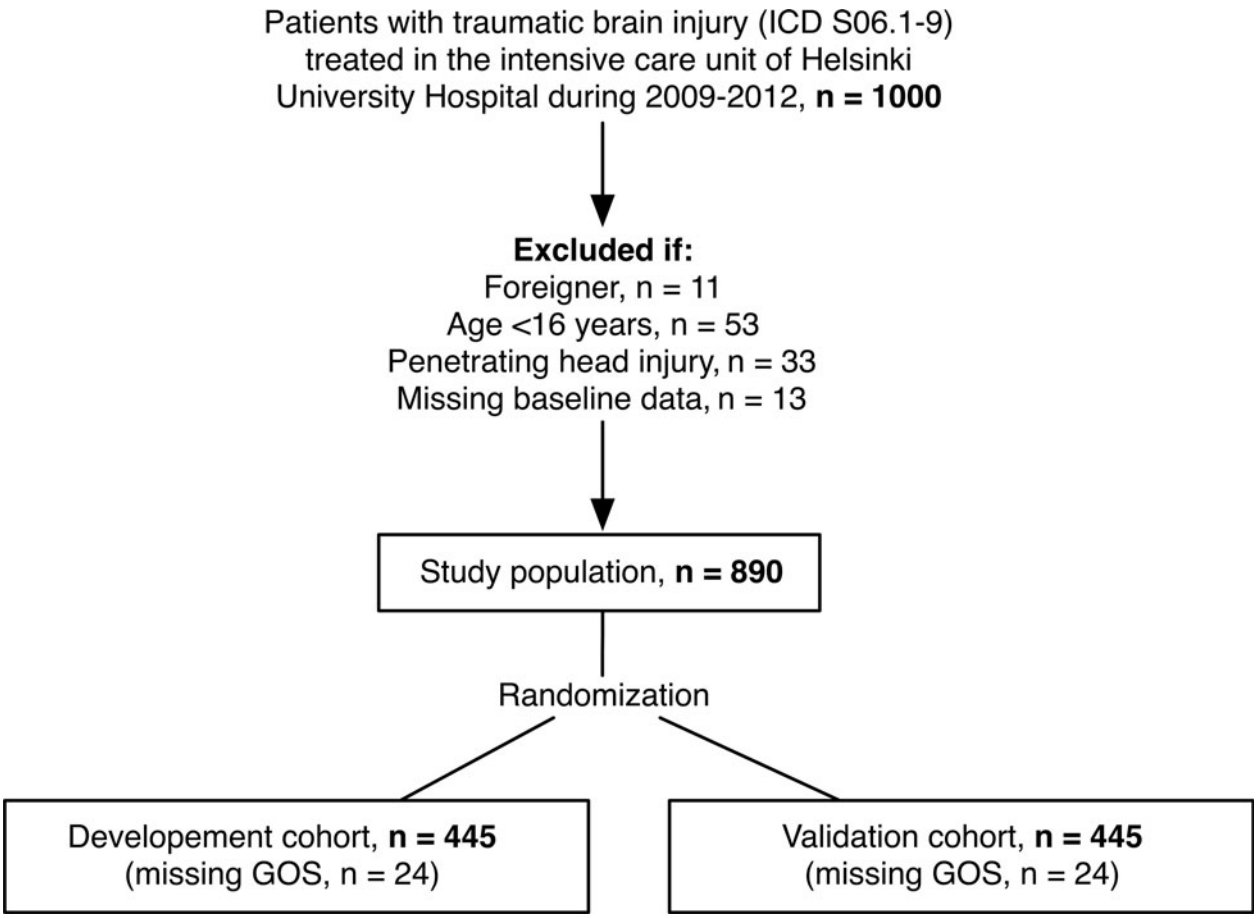

FIG. 1. Study population. ICD, International Statistical Classification of Diseases; GOS, Glasgow Outcome Scale.

Table 1. Study Population Characteristics

Variable	*All patients* (n = *895*)	*Six-month mortality*			*Six-month neurological outcome*[b]		
		Survivors (n = *684*)	*Nonsurvivors* (n = *206*)	p *value*[a]	*Favorable outcome* (n = *448*)	*Unfavorable outcome* (n = *394*)	p *value*[c]
Baseline characteristics							
Age	58 (44–68)	55 (41–65)	65 (56–73)	<0.001	51 (35–63)	63 (54–72)	<0.001
Chronic comorbidity							
Cardiovascular	8 (1)	5 (1)	3 (2)	0.396[d]	2 (0)	6 (2)	0.156[d]
Respiratory	10 (1)	7 (1)	3 (2)	0.705[d]	4 (1)	5 (1)	0.741[d]
Hepatic	14 (2)	8 (1)	6 (3)	0.105[d]	6 (1)	8 (2)	0.591[d]
Renal	2 (0)	1 (0)	1 (1)	0.410[d]	0 (0)	2 (1)	0.219[d]
Immunosuppression	28 (3)	14 (2)	13 (6)	0.002	9 (2)	18 (5)	0.035
Admission GCS							
3–8	379 (43)	243 (35)	136 (66)	<0.001	135 (30)	228 (58)	<0.001
9–12	210 (24)	168 (25)	42 (20)		111 (25)	88 (22)	
13–15	301 (33)	273 (40)	28 (14)		202 (45)	78 (20)	
Admission motor score							
Obeys/localizes	584 (66)	502 (73)	82 (40)	<0.001	347 (78)	197 (50)	<0.001
Normal/abnormal flexion	126 (14)	89 (13)	37 (18)		55 (12)	68 (17)	
None/extension	180 (20)	93 (14)	87 (42)		46 (10)	129 (33)	
Admission GCS, median (IQR)	10 (6–13)	12 (7–14)	6 (3–11)	<0.001	12 (8–14)	7 (4–12)	<0.001
Admission motor score, median (IQR)	5 (4–6)	5 (4–6)	4 (1–5)	<0.001	6 (5–6)	5 (2–6)	<0.001
Worst 24-h GCS, median (IQR)	9 (4–12)	10 (7–12)	4 (3–9)	<0.001	11 (8–13)	7 (3–10)	<0.001
Worst 24-h motor score, median (IQR)	5 (2–6)	5 (4–6)	2 (1–5)	<0.001	5 (4–6)	4 (1–5)	<0.001
Pupils							
Both react	668 (74)	572 (83)	96 (47)	<0.001	393 (88)	235 (60)	<0.001
One reacts	102 (12)	67 (10)	35 (17)		31 (7)	64 (16)	
None react	120 (14)	45 (7)	75 (36)		42 (5)	95 (24)	
Hypotension	61 (7)	46 (7)	15 (7)	0.782	33 (7)	26 (7)	0.663
Hypoxia	131 (15)	86 (13)	45 (22)	0.001	49 (11)	76 (19)	0.001
Glucose (mmol/L)	7.3 (6.1–8.8)	7.0 (6.0–8.5)	8.1 (6.8–9.5)	<0.001	7.0 (6.0–8.4)	7.6 (6.3–9.3)	<0.001
Hemoglobin (g/dL)	12.5 (11.1–13.8)	12.8 (11.5–14.1)	11.6 (10.4–12.9)	<0.001	13.0 (11.7–14.3)	11.9 (10.8–13.2)	<0.001
Platelet count (10^9)	187 (136–235)	192 (142–236)	164 (116–231)	0.003	198 (147–242)	171 (129–228)	<0.001
Base excess (mmol/L)	–1.6 (–4.4–0.6)	–1.4 (–4.1–0.7)	–2.4 (–5.2–0.5)	0.017	–1.5 (–4.1–0.5)	–1.8 (–4.7–0.9)	0.431
INR	1.1 (1.0–1.2)	1.1 (1.0–1.2)	1.1 (1.0–1.3)	<0.001	1.0 (1.0–1.1)	1.1 (1.0–1.3)	<0.001
Marshall CT							
DI I	15 (2)	15 (2)	0 (0)	<0.001	12 (3)	2 (1)	<0.001
DI II	275 (30)	249 (36)	26 (12)		184 (41)	74 (19)	
DI III–IV	67 (8)	49 (7)	18 (9)		32 (7)	33 (8)	
EML/NEML	533 (60)	371 (55)	162 (79)		220 (49)	285 (72)	
Rotterdam CT							
1–2	206 (23)	178 (26)	28 (14)	<0.001	130 (29)	59 (15)	<0.001
3–4	478 (54)	393 (58)	84 (41)		266 (59)	186 (47)	
5–6	206 (23)	112 (16)	94 (45)		52 (12)	149 (38)	

(*continued*)

Table 1. (Continued)

Variable	*All patients* (n = *895*)	*Six-month mortality*			*Six-month neurological outcome*[b]		
		Survivors (n = *684*)	*Nonsurvivors* (n = *206*)	p *value*[a]	*Favorable outcome* (n = *448*)	*Unfavorable outcome* (n = *394*)	p *value*[c]
Epidural hematoma	92 (10)	82 (12)	10 (5)	0.003	67 (15)	19 (5)	< 0.001
Traumatic SAH	504 (57)	379 (55)	125 (61)	0.181	237 (53)	243 (62)	0.010
Acute mass lesion evacuation	310 (35)	222 (33)	88 (43)	0.007	137 (31)	162 (41)	0.001
Length of stay (days)							
ICU	2 (1–6)	2 (1–6)	2 (1–5)	0.313	2 (1–5)	3 (1–7)	0.002
Hospital	8 (4–5)	8 (5–15)	6 (2–14)	< 0.001	8 (5–13)	8 (3–17)	0.775
Predicted risk of death							
IMPACT							
Core model	24 (15–35)	19 (12–24)	48 (24–68)	< 0.001	15 (12–24)	29 (19–55)	< 0.001
Extended model	23 (15–41)	19 (12–34)	49 (28–75)	< 0.001	39 (27–53)	59 (46–85)	< 0.001
Lab model	22 (13–35)	19 (11–26)	46 (26–66)	< 0.001	32 (18–44)	57 (38–82)	< 0.001
APACHE II	22 (12–34)	18 (12–28)	37 (28–52)	< 0.001	15 (9–25)	31 (20–45)	< 0.001
Observed outome							
14-day mortality	99 (11)	0 (0)	99 (48)	—	0 (0)	99 (25)	—
6-month mortality	206 (23)	0 (0)	206 (100)	—	0 (0)	206 (52)	—
6-month unfavorable outcome[b]	394 (47)	188 (30)	206 (100)	—	0 (0)	394 (100)	—

Continuous data presented as median (IQR), categorical data presented as *n* (%).

[a]Between 6-month survivors and nonsurvivors.

[b]Total of 842 patients.

[c]Between 6-month favorable and unfavorable neurological outcome.

[d]Fischer's exact test.

APACHE II, Acute Physiology and Chronic Health Evaluation II; CT, computerized tomography; DI, diffuse injury; GCS, Glasgow Coma Scale; INR, international normalized ratio; IMPACT, International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury; SAH, subarachnoid hemorrhage.

Table 2. Development and Validation Cohort Characteristics

Variable	*All* (n = *890*)	*Development* (n = *445*)	*Validation* (n = *445*)	p *value*
Prediction model scores				
IMPACT core sumscore	5 (3–7)	5 (3–7)	5 (3–7)	0.921
IMPACT extended sumscore	7 (5–10)	7 (5–10)	7 (5–11)	0.998
IMPACT lab sumscore	10 (7–13)	10 (7–13)	9 (7–13)	0.932
APACHE II total score	19 (14–23)	18 (14–23)	19 (14–23)	0.287
APS subscore	15 (11–20)	15 (11–19)	15 (12–20)	0.285
Chronic health subscore	0 (0–0)	0 (0–0)	0 (0–0)	0.999
Age subscore	3 (0–5)	3 (0–5)	3 (2–5)	0.346
GCS subscore	6 (3–11)	6 (3–11)	6 (3–11)	0.722
Outcome				
14-day mortality	99 (11)	49 (11)	50 (11)	0.915
1-month mortality	137 (15)	63 (14)	74 (17)	0.307
3-month mortality	178 (20)	84 (19)	94 (21)	0.402
6-month mortality	206 (23)	98 (22)	108 (24)	0.427
6-month unfavorable outcome[a,b]	394 (47)	191 (45)	203 (48)	0.407

Continuous data presented as median (IQR), categorical data presented as *n* (%).
[a]Total of 842 patients: 421 in the development and 421 in the validation cohort.
[b]Unfavorable outcome defined as GCS 1 (dead), 2 (vegetative state), and 3 (severe disability).
APACHE II, Acute Physiology and Chronic Health Evaluation II; APS, Acute Physiology Score; GCS, Glasgow Coma Scale; IMPACT, International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury.

hematomas (Table 1). Further, nonsurvivors had a shorter hospital length of stay (LOS), compared to survivors ($p < 0.001$), and patients with unfavorable neurological outcome had longer ICU LOS, compared to patients with favorable outcome ($p = 0.002$).

After randomization, a total of 445 patients (50%) were allocated to the development cohort and 445 (50%) to the validation cohort. There were 24 patients with missing GOS in the respective cohort, leaving a total of 842 (development, 421; validation, 421) for the development and validation for the neurological outcome prediction models. There were no significant differences in IMPACT or APACHE II score variables between the development and validation cohorts (Table 2).

Table 3. Prediction Model Performance for 6-Month Mortality

Prediction model	*Discrimination AUC (95% CI)*	*Calibration H-L* p *value*	*Precision Brier score*
All patients ($n = 890$)			
APACHE II[a]	0.80 (0.77–0.84)	0.099	0.139
IMPACT core[a]	0.80 (0.77–0.83)	< 0.001	0.140
IMPACT extended[a]	0.80 (0.77–0.83)	< 0.001	0.144
IMPACT lab[a]	0.81 (0.78–0.84)	0.054	0.136
Development sample (n = 450)			
APACHE II[a]	0.80 (0.75–0.84)	0.364	0.137
IMAPCT core[a]	0.78 (0.73–0.84)	0.030	0.137
IMPACT extended[a]	0.79 (0.74–0.84)	0.001	0.143
IMPACT lab[a]	0.80 (0.76–0.85)	0.103	0.133
IMPACTcore-APACHE II	0.84 (0.79–0.88)	0.915	0.117
IMPACText-APACHE II	0.84 (0.79–0.88)	0.366	0.119
IMPACTlab-APACHE II	0.84 (0.80–0.88)	0.727	0.119
Validation sample ($n = 450$)			
APACHE II[a]	0.81 (0.76–0.86)	0.315	0.142
IMAPCT core[a]	0.81 (0.77–0.86)	0.063	0.143
IMPACT extended[a]	0.81 (0.76–0.86)	0.026	0.144
IMPACT lab[a]	0.82 (0.77–0.86)	0.441	0.138
IMPACTcore-APACHE II	0.84 (0.80–0.89)	0.690	0.125
IMPACText-APACHE II	0.84 (0.80–0.89)	0.515	0.123
IMPACTlab-APACHE II	0.85 (0.81–0.89)	0.198	0.123

The predicted risk of 6-month mortality using the new models can be calculated by the equation: $1/(1+e^{-\text{logit}})$, where the logit is defined as:
$\text{logit}_{\text{IMPACTcore-APACHE II}} = -6.004 + 0.234$ * IMPACTcore sumscore + 0.160 * APACHE II score
$\text{logit}_{\text{IMPACText-APACHE II}} = -6.265 + 0.193$ * IMPACText sumscore + 0.158 * APACHE II score
$\text{logit}_{\text{IMPACTlab-APACHE II}} = -6.516 + 0.187$ * IMPACTlab sumscore + 0.149 * APACHE II score
[a]Original models.
AUC, area under the curve; CI, confidence interval; H-L, Hosmer-Lemeshow Ĉ-test (goodness of fit); APACHE, Acute Physiology and Chronic Health Evaluation; IMPACT, International Mission for Prognosis and Analysis of Clinical Trials in TBI.

Table 4. Prediction Model Performance for 6-Month Unfavorable Neurological Outcome

Prediction model	*Discrimination AUC (95% CI)*	*Calibration H-L* p *value*	*Precision Brier score*
All patients ($n=842$)			
APACHE II[a]	0.76 (0.73–0.79)	<0.001	0.248
IMPACT core[a]	0.78 (0.75–0.81)	<0.001	0.193
IMPACT extended[a]	0.79 (0.76–0.82)	0.006	0.190
IMPACT lab[a]	0.79 (0.76–0.82)	0.078	0.188
Development sample ($n=421$)			
APACHE II[a]	0.74 (0.69–0.79)	<0.001	0.250
IMAPCT core[a]	0.75 (0.71–0.79)	<0.001	0.234
IMPACT extended[a]	0.76 (0.72–0.80)	0.008	0.202
IMPACT lab[a]	0.76 (0.72–0.81)	0.008	0.199
IMPACTcore-APACHE II	0.78 (0.74–0.82)	0.283	0.190
IMPACText-APACHE II	0.79 (0.74–0.83)	0.327	0.188
IMPACTlab-APACHE II	0.79 (0.74–0.83)	0.318	0.188
Validation sample ($n=421$)			
APACHE II[a]	0.78 (0.74–0.82)	<0.001	0.245
IMAPCT core[a]	0.81 (0.77–0.85)	<0.001	0.227
IMPACT extended[a]	0.82 (0.78–0.85)	0.221	0.177
IMPACT lab[a]	0.82 (0.78–0.86)	0.712	0.177
IMPACTcore-APACHE II	0.83 (0.79–0.87)	0.476	0.170
IMPACText-APACHE II	0.83 (0.80–0.87)	0.037	0.167
IMPACTlab-APACHE II	0.83 (0.80–0.87)	0.439	0.168

The predicted risk of six-month unfavorable outcome using the new models can be calculated by the equation: $1/(1+e^{-logit})$, where the logit is defined as:
$logit_{IMPACTcore\text{-}APACHE\ II} = -3.363 + 0.198$ * IMPACTcore sumscore $+ 0.113$ * APACHE II score
$logit_{IMPACText\text{-}APACHE\ II} = -3.551 + 0.175$ * IMPACText sumscore $+ 0.105$ * APACHE II score
$logit_{IMPACTlab\text{-}APACHE\ II} = -3.732 + 0.162$ * IMPACTlab sumscore $+ 0.099$ * APACHE II score
[a]Original models.

AUC, area under the curve; CI, confidence intervals; H-L, Hosmer-Lemeshow Ĉ-test (goodness of fit); APACHE, Acute Physiology and Chronic Health Evaluation; IMPACT, International Mission for Prognosis and Analysis of Clinical Trials in TBI.

IMPACT versus APACHE II

The original models showed satisfactory to good discriminative power (AUC, 0.76–0.81) and poor to good calibration for 6-month mortality and neurological outcome prediction. There were no significant differences in AUC between the IMPACT models and the APACHE II for prediction of 6-month mortality ($p>0.05$ between all models; Table 3). However, the IMPACT models showed significantly higher AUCs for predicting 6-month neurological outcome, compared to the APACHE II ($p<0.05$ between all models; Table 4). The APACHE II showed good calibration for predicting mortality ($p=0.099$), but not neurological outcome ($p<0.001$). The IMPACT lab was the only model that showed good calibration ($p>0.05$) for both mortality and neurological outcome prediction by the H-L test (Fig. 2).

IMPACT plus APACHE II

Performance measures of the new combined IMPACT-APACHE II models are shown in Tables 3 (mortality) and 4 (neurological outcome). There were no significant differences in AUC between the development and validation cohort. The new combined models showed significantly higher AUCs, compared to the original IMPACT core ($p=0.026$), extended ($p=0.039$), and lab models ($p=0.043$) for predicting 6-month mortality (Fig. 3). Reclassification testing showed 54.3–81.6% of patients dying to be better classified by the new models, compared to the original IMPACT models ($p<0.001$) and 84.5–87.1% of patients dying to be better classified, as compared to the APACHE II ($p<0.001$; Table 5).

For the prediction of 6-month unfavorable outcome, the new combined models showed significantly higher AUCs, compared to the APACHE II ($p<0.05$), but not compared to the original IMPACT models ($p>0.05$; Fig. 3). However, reclassification testing showed 14.3% (IMPACTcore–APACHE II), 23.2% (IMPACText–APACHE II), and 16.3% (IMPACTlab–APACHE II) improvements in risk stratification by the new combined models, compared to the original IMPACT models, for unfavorable outcome prediction ($p<0.05$; Table 5).

The new models showed good calibration with no significant over- or underprediction intervals (Fig. 4). In the validation cohort, 6-month mortality was 26% and unfavorable outcome 48%, accordingl; for mortality, Brier scores ≤ 0.046 were considered excellent, 0.046–0.091 good, 0.091–0.137 satisfactory, and >0.137 poor; for neurological outcome, Brier scores ≤ 0.062 were considered excellent, 0.064–0.125 good, 0.125–0.187 satisfactory, and >0.187 poor. All new models displayed satisfactory precision, both regarding mortality (Brier score, 0.121–0.123) and neurological outcome (Brier score, 0.167–170).

Secondary internal bootstrapping gave similar results to that of the primary internal validation using the split-sample technique, indicating low grade of model overfitting (see Supplementary Equations) (see online supplementary material at http://www.liebertpub.com).

Post-hoc analysis

Because the IMPACT model was specifically developed for patients with an admission GCS of 3–12, we conducted post-hoc analysis testing this subgroup separately from those with an admission GCS of 13–15. In those with an admission GCS 3–12, AUC was 0.78–0.79 for mortality prediction and 0.76 (for all three

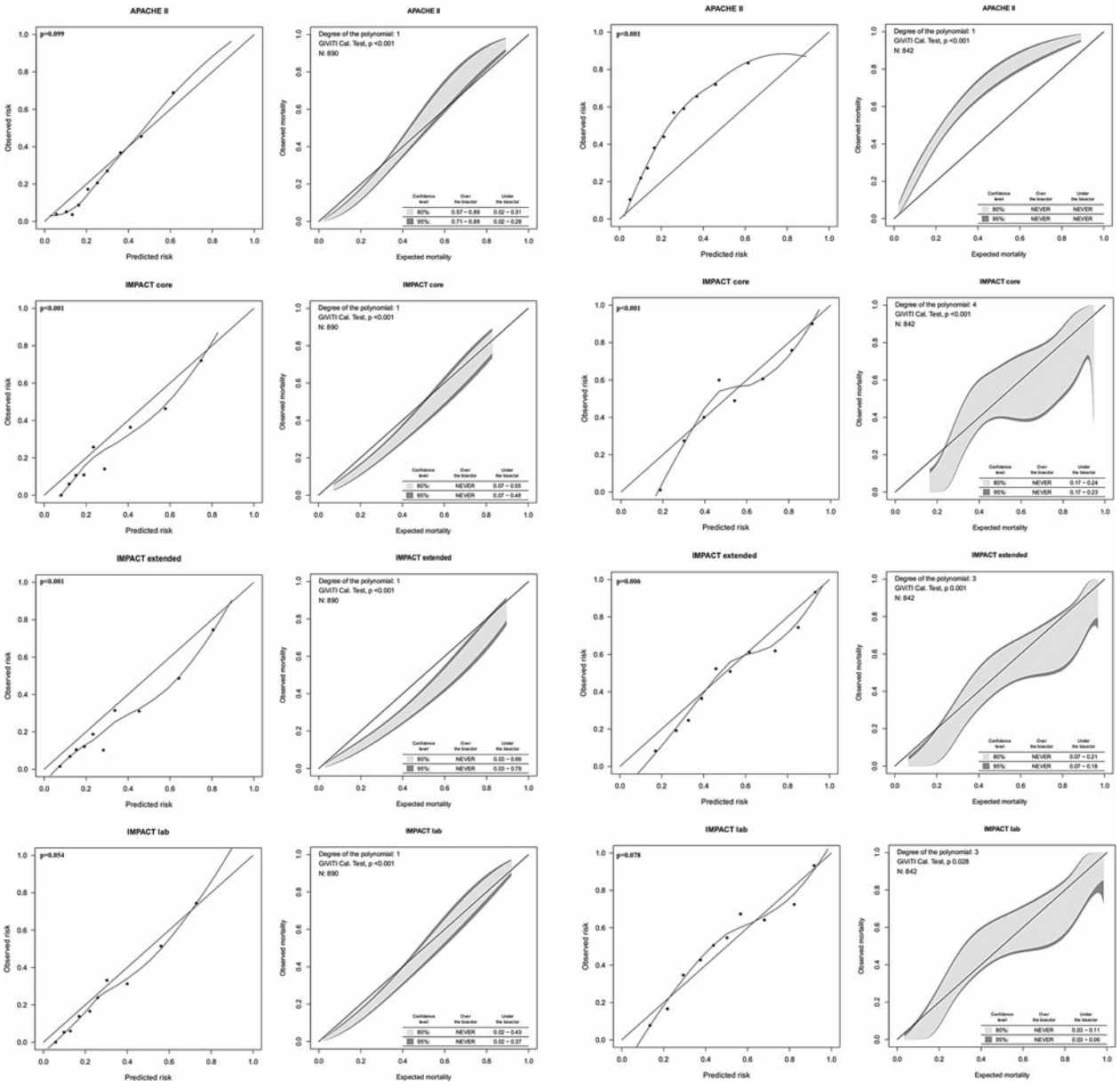

FIG. 2. Calibration tests for the original IMPACT and APACHE II models for 6-month mortality (left) and unfavorable outcome (right) prediction. The H-L calibration plots to the left (with a loess smoother curve fitted between the groups) and the GiViTI calibration belts to the right for 6-month mortality and neurological outcome, respectively. The GiViTI belt shows risk intervals of significant under- and overprediction when the 95% confidence interval does not encompass the diagonal bisector line (black line, indicating perfect calibration). The APACHE II model showed good calibration for mortality ($p = 0.099$), but not neurological outcome ($p < 0.001$) prediction. Both the IMPACT core and extended models showed poor calibration by both tests ($p < 0.05$). The IMPACT lab was the only model showing good calibration for both mortality ($p = 0.054$) and neurological outcome ($p = 0.078$) prediction. IMPACT, International Mission for Prognosis and Analysis of Clinical Trials; APACHE II, Acute Physiology and Chronic Health Evaluation II; H-L, the Hosmer-Lemeshow Ĉ-test; GiViTI, Italian Group for the Evaluation of Interventions in Intensive Care Medicine.

models) for neurological outcome prediction. In comparison, in patients with an admission GCS of 13–15, the AUC was 0.73–0.75 for mortality prediction and 0.77 (for all three models) for neurological outcome prediction.

Discussion

In this article, we describe the development of a series of prediction models, combining the IMPACT and the APACHE II, to predict risk of 6-month outcome in patients with TBI treated in the ICU. Further, this article shows an external validation of the IMPACT and the APACHE II models in such patients. We found the predictive performance of the IMPACT models to significantly increase after the addition of the APACHE II. Our results remained robust after internal validation, using both a split-sample technique and a bootstrap technique, indicating low grade of overestimation. Our findings may be applicable for heterogeneity adjustment in forthcoming epidemiological studies and clinical trials in TBI.

Further, this is, to our knowledge, the first study showing that the IMPACT model discriminates well for patients presenting with

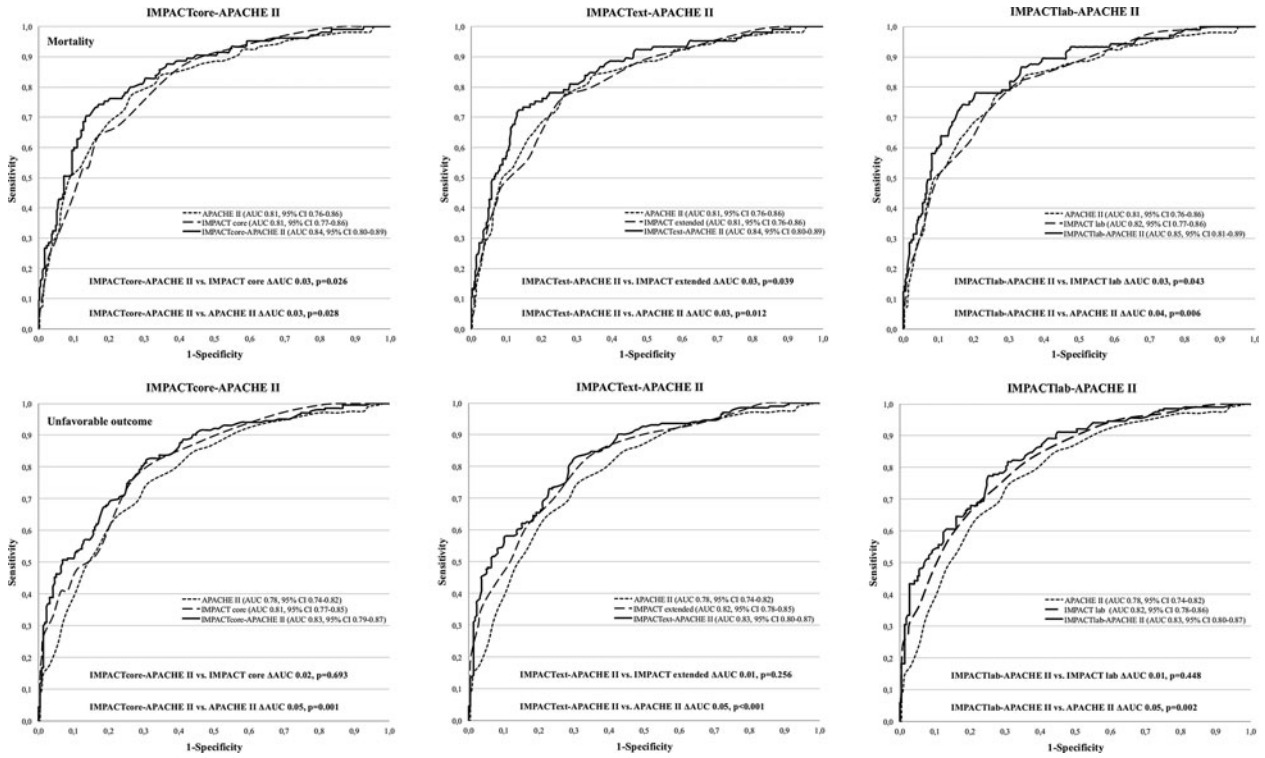

FIG. 3. Area under the receiver operator characteristic curve (AUC) for mortality (top) and neurological outcome (bottom) prediction. The IMPACTcore–APACHE II (left), the IMPACText–APACHE II (middle), and the IMPACTlab-APACHE II (right). The AUCs are compared between the models with a concomitant p value ($p<0.05$ indicates a significant difference). All new models showed significantly higher AUCs, compared to the original IMPACT and APACHE II models, for mortality prediction. The new models showed significantly higher AUCs, compared to the APACHE II, but not compared to the original IMPACT models, for neurological outcome prediction. IMPACT, International Mission for Prognosis and Analysis of Clinical Trials; APACHE II, Acute Physiology and Chronic Health Evaluation II; CI, confidence interval.

mTBI,before deterioration and ICU admission (i.e., complicated mTBI). Currently, there are no established prediction models for mTBI (uncomplicated or complicated) as there are for moderate-to-severe TBI. Future studies should look into using the IMPACT model as a framework for further development of mTBI prediction models.

IMPACT versus APACHE II

The APACHE II is one of the world's most-used ICU scoring systems and is frequently used as a measure of quality of care by calculating the ratio between predicted and observed outcome (i.e., standardized mortality ratio; SMR).[14] However, the APACHE II was developed for general ICU populations and its applicability to patients with TBI has previously been uncertain.[34] Further, calculating SMR based on hospital mortality prediction may cause biased results.[35] This is of special concern in patients with TBI, where a significant number of patients die subsequent to hospital discharge (approximately half of patients died within the first 14 days of injury in the present study).[36] Another major concern of the APACHE II is that it uses data from the first day in the ICU. Any score that uses data collected over this time is affected by the quality of care given—the very same thing you are trying to assess.[37] Patients initially receiving poor care will have higher

Table 5. Percent of Patients with Improved Classification by the New Models

		Mortality		*Unfavorable outcome*	
Updated model	*Reference model*	*NRI (95% CI)*	p *value*	*NRI (95% CI)*	p *value*
IMPACTcore-APACHE II	IMPACT core	77.6 (57.6–98.3)	<0.001	14.3 (5.0–23.5)	0.003
	APACHE II	84.5 (64.3–104.6)	<0.001	5.4 (0.4–10.5)	0.035
IMPACText-APACHE II	IMPACT extended	54.3 (34.2–74.3)	<0.001	23.2 (5.9–40.6)	0.009
	APACHE II	86.3 (66.2–106-4)	<0.001	12.8 (5.8–19.8)	<0.001
IMPACTlab-APACHE II	IMPACT lab	81.6 (61.3–101.9)	<0.001	16.3 ([−2.7] −35.3)	0.093
	APACHE II	87.1 (67.0–107.2)	<0.001	10.9 (3.8–17.9)	0.003

NRI values (with 95% CI) shown in percent (%).

NRI, Net Reclassification Improvement; CI, confidence interval, APACHE II, Acute Physiology and Chronic Health Evaluation II; IMPACT, International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury.

APACHE II scores resulting in overestimation of mortality and falsely low SMR values. The IMPACT model avoids this because it only uses characteristics assessed upon hospital admission, and thus the IMPACT is not affected by later care.

Although the IMPACT and APACHE II were designed for different purposes and differ substantially from each other, we found no significant differences in 6-month mortality predictive performance between the models (AUC, 0.80 vs. 0.78–0.80). Likewise, we recently showed satisfactory discrimination (AUC, 0.79) of the APACHE II for the prediction of 6-month mortality in patients with TBI treated in the ICU.[16] We believe this to be an effect of the APACHE II being originally designed to predict in-hospital mortality, but was, in the current study, used to predict 6-month mortality. This causes the improved outcomes that are achieved over time to be cancelled out by excess mortality after hospital discharge. This is supported by the fact that the APACHE II showed inferior performance for predicting neurological outcome, compared to mortality. It may be argued that, for future epidemiological TBI studies, the APACHE II might suffice for mortality analyses, whereas the IMPACT is better calibrated for neurological outcome analyses.

IMPACT plus APACHE II

Most validated prediction models in TBI have been mainly based on admission characteristics. Although substantial insights have been gained into the prognostic value of variables obtained during the subsequent treatment course, these have not, before the present study, been implemented in prognostic models.[9] Applying the new IMPACT–APACHE II models may offer several advantages over using the single admission characteristics model. First, in clinical trials, it would be possible not only to adjust for baseline risk differences, but also for differences in treatment standards between participating centers. This might increase statistical power of future studies. Second, for the evaluation of quality of delivered care in ICUs, it would be possible to more accurately differ between high- and low-performing units. For units reporting high mortality or unfavorable outcome rates, but low baseline risks, quality of delivered treatment might be questioned. In contrast, units reporting high baseline risks, but low mortality or unfavorable outcome rates, would be considered high performing. Third, the new models offer us important tools in proper case-mix adjustment for comparative

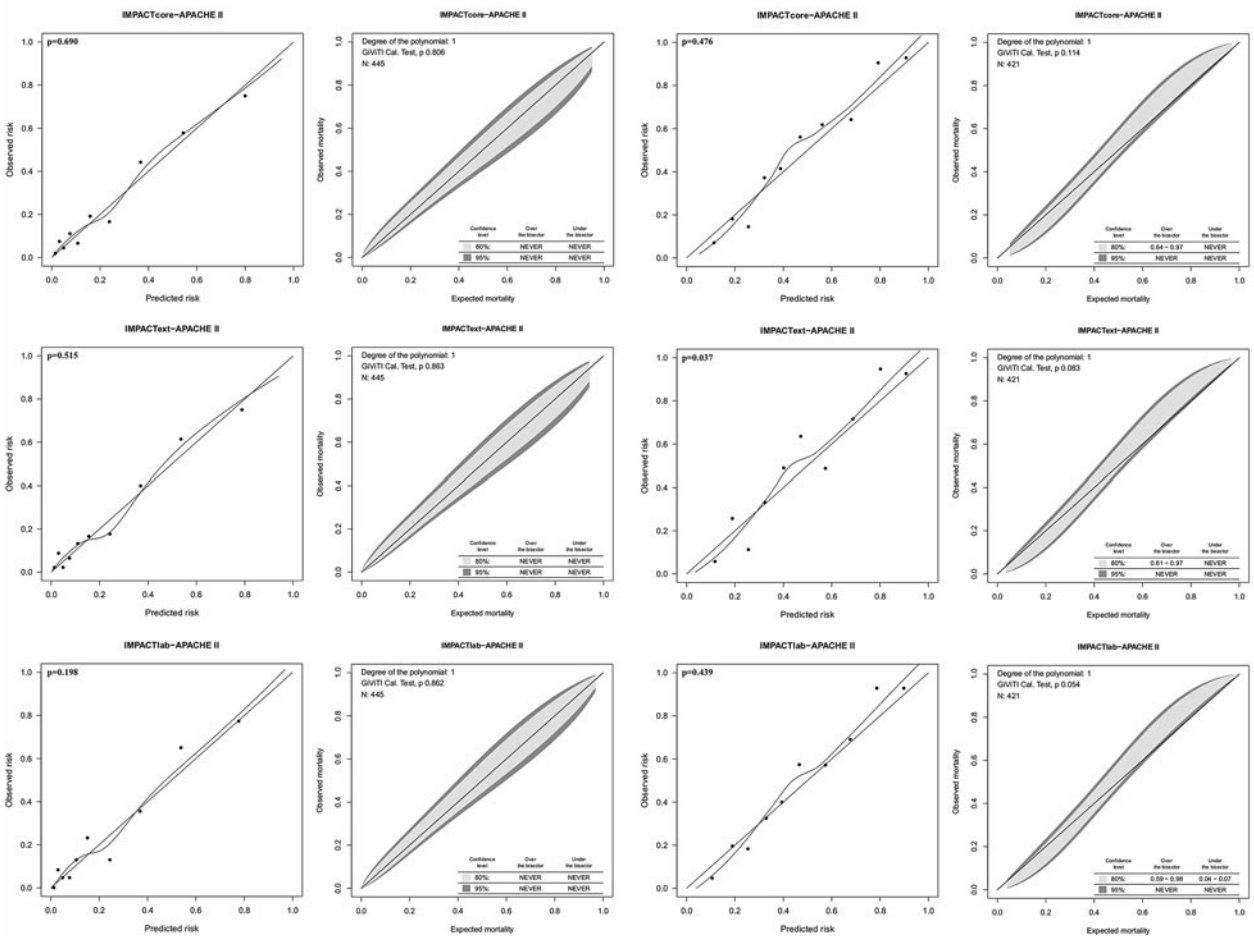

FIG. 4. Calibration tests for the newly developed IMPACT–APACHE II models in the validation cohort. Calibration for mortality prediction (left) and for neurological outcome prediction (right). The H-L calibration plots (left; with a loess smoother curve fitted between the groups) and the GiViTI calibration belts (right) for mortality and neurological outcome, respectively. All new models showed good calibration by the H-L test ($p > 0.05$). Only the IMPACText–APACHE II showed poor calibration by the H-L test ($p = 0.037$). Accordingly, the GiViTI calibration belt reveals significant under prediction (95% confidence interval over the diagonal bisector line) between for a risk interval between 0.61 and 0.97. IMPACT, International Mission for Prognosis and Analysis of Clinical Trials; APACHE II, Acute Physiology and Chronic Health Evaluation II; H-L, the Hosmer-Lemeshow Ĉ-test; GiViTI, Italian Group for the Evaluation of Interventions in Intensive Care Medicine.

effectiveness research between countries, centers, ICUs, and patients to identify best practices in the heterogeneous field of TBI research.[38]

Although we found the newly developed IMPACT–APACHE II models to significantly improve the individual models for mortality prediction, the improvements were less robust for neurological outcome prediction. Future studies should look at the possibility of increasing neurological outcome prediction accuracy. This may be achieved by including more TBI-specific intensive care variables, such as measures of intracranial pressure, cerebral perfusion pressure, partial brain tissue oxygenation, microdialysate monitoring, and different biomarkers.[39–41] Further, although the IMPACT models were introduced in 2008, they were developed upon studies conducted in 1984–1997.[42] Since then, advances in ICU and TBI treatment has been made, for example, the release of standardized international guidelines for the treatment of patients with TBI, cerebral perfusion pressure-targeted therapies, and advances in radiological techniques.[20] These advances may have improved patient outcome, which could explain why the IMPACT model was found to consistently overpredict risk of poor outcome in the present study (explaining the poor calibration noted for the original IMPACT models in our study). Thus, recalibration of the IMPACT models should be considered in the future to fit the underlying population and current practices.

Limitations

There are some limitations of the present study that have to be considered. First, because of the retrospective design of the study, we were limited to the simple GOS and could not assess the more-sensitive extended GOS.[43] Second, this was a single-center retrospective study, and thus these findings should be replicated in other settings. Third, data on extracranial injury severity were not available for all patients, but we have showed, in a previous study using a similar study sample, that the inclusion of extracranial injuries (by injury severity score) did not add any significant predictive ability to the IMPACT.[44]

Conclusion

The IMPACT and APACHE II models showed equal performance for 6-month mortality prediction in patients with TBI treated in the ICU. However, for neurological outcome prediction, the IMPACT was superior to the APACHE II. Combining the IMPACT and APACHE II resulted in superior 6-month mortality and neurological outcome predictive performance, compared to the individual models.

Acknowledgments

The study was funded by a Helsinki University Hospital EVO grant (TYH2012142), Medicinska Understödsföreningen Liv och Hälsa and by the Maire Taponen foundation. The authors express their gratitude to Jaakko Lappalainen (MD, PhD) and Professor Per Rosenberg (MD, PhD) for their valuable comments on the manuscript and to Giovanni Nattino (IRCCS-Istituto di Ricerche Farmacologiche ''Mario Negri'' Laboratory of Clinical Epidemiology) and Tuomas Selander (Science Service Center, Kuopio University Hospital) for their statistical assistance.

Author Disclosure Statement

No competing financial interests exist.

References

1. Stein, S.C., Georgoff, P., Meghan, S., Mizra, K., and Sonnad, S.S. (2010). 150 years of treating severe traumatic brain injury: a systematic review of progress in mortality. J. Neurotrauma 27, 1343–1353.
2. Roozenbeek, B., Maas, A.I.R., and Menon, D.K. (2013). Changing patterns in the epidemiology of traumatic brain injury. Nat. Rev. Neurol. 9, 231–236.
3. Maas, A.I.R., Steyerberg, E.W., Marmarou, A., McHugh, G.S., Lingsma, H.F., Butcher, I., Lu, J., Weir, J., Roozenbeek, B., and Murray, G.D. (2010). IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury. Neurotherapeutics 7, 127–134.
4. Lingsma, H.F., Roozenbeek, B., Li, B., Lu, J., Weir, J., Butcher, I., Marmarou, A., Murray, G.D., Maas, A.I.R., and Steyerberg, E.W. (2011). Large between-center differences in outcome after moderate and severe traumatic brain injury in the international mission on prognosis and clinical trial design in traumatic brain injury (IMPACT) study. Neurosurgery 68, 601–607; discussion, 607–608.
5. Hemmila, M.R., Nathens, A.B., Shafi, S., Calland, J.F., Clark, D.E., Cryer, H.G., Goble, S., Hoeft, C.J., Meredith, J.W., Neal, M.L., Pasquale, M.D., Pomphrey, M.D., and Fildes, J.J. (2010). The Trauma Quality Improvement Program: pilot study and initial demonstration of feasibility. J. Trauma 68, 253–262.
6. Khuri, S.F., Daley, J., and Henderson, W.G. (2002). The comparative assessment and improvement of quality of surgical care in the Department of Veterans Affairs. Arch. Surg. 137, 20–27.
7. Maas, A.I.R., Murray, G.D., Roozenbeek, B., Lingsma, H.F., Butcher, I., McHugh, G.S., Weir, J., Lu, J., and Steyerberg, E.W.; International Mission on Prognosis Analysis of Clinical Trials in Traumatic Brain Injury (IMPACT) Study Group. (2013). Advancing care for traumatic brain injury: findings from the IMPACT studies and perspectives on future research. Lancet Neurol. 12, 1200–1210.
8. Stevens, R.D., and Sutter, R. (2013). Prognosis in severe brain injury. Crit. Care Med. 41, 1104–1123.
9. Lingsma, H.F., Roozenbeek, B., Steyerberg, E.W., Murray, G.D., and Maas, A.I.R. (2010). Early prognosis in traumatic brain injury: from prophecies to predictions. Lancet Neurol. 9, 543–554.
10. Steyerberg, E.W., Mushkudiani, N., Perel, P., Butcher, I., Lu, J., McHugh, G.S., Murray, G.D., Marmarou, A., Roberts, I., Habbema, J.D.F., and Maas, A.I.R. (2008). Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med. 5, e165.
11. Knaus, W.A., Draper, E.A., Wagner, D.P., and Zimmerman, J.E. (1985). APACHE II: a severity of disease classification system. Crit. Care Med. 13, 818–829.
12. Vincent, J.-L., and Moreno, R. (2010). Clinical review: scoring systems in the critically ill. Crit. Care 14, 207.
13. Maas, A.I., Murray, G., Henney, H. 3rd, Kassem, N., LeGrand, V., Mangelus, M., Muizelaar, J.P., Stocchetti, N., Knoller, N. (2006). Efficacy and safety of randomised dexanabinol in severe traumatic brain injury: Results of a phase III randomised, placebo-controlled, clinical trial. Lancet Neurol. 5, 38–45.
14. Le Gall, J.-R. (2005). The use of severity scores in the intensive care unit. Intensive Care Med. 31, 1618–1623.
15. Kaukonen, K.-M., Bailey, M., Suzuki, S., Pilcher, D., and Bellomo, R. (2014). Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000–2012. JAMA 311, 1308–1316.
16. Raj, R., Skrifvars, M.B., Bendel, S., Selander, T., Kivisaari, R., Siironen, J., and Reinikainen, M. (2014). Predicting six-month mortality of patients with traumatic brain injury: usefulness of common intensive care severity scores. Crit. Care 18, R60.
17. Menon, D.K., Schwab, K., Wright, D.W., and Maas, A.I.; Demographics and Clinical Assessment Working Group of the International and Interagency Initiative toward Common Data Elements for Research on Traumatic Brain Injury and Psychological Health. (2010). Position statement: definition of traumatic brain injury. Arch. Phys. Med. Rehabil. 91, 1637–1640.
18. Marshall, L.F., Marshall, S.B., Klauber, M.R., and Clark, M.B. (1991). A new classification of head injury based on computerized tomography. J. Neurosurg. 75, S14–S22.
19. Maas, A.I.R., Hukkelhoven, C.W.P.M., Marshall, L.F., and Steyerberg, E.W. (2005). Prediction of outcome in traumatic brain injury

with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors. Neurosurgery 57, 1173–1182; discussion, 1173–1182.
20. Brain Trauma Foundation, American Association of Neurological Surgeons, Congress of Neurological Surgeons. (2007). Guidelines for the management of severe traumatic brain injury. J. Neurotrauma 24 Suppl. 1, S1–S106.
21. Jennett, B., and Bond, M. (1975). Assessment of outcome after severe brain damage. Lancet 1, 480–484.
22. Picard, R.R., and Berk, K.N. (1990). Data splitting. Am. Stat. 44, 140–147.
23. Steyerberg, E.W., Harrell, F.E., Borsboom, G.J., Eijkemans, M.J., Vergouwe, Y., and Habbema, J.D. (2001). Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J. Clin. Epidemiol. 54, 774–781.
24. Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M.J., and Kattan, M.W. (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 21, 128–138.
25. David W Hosmer, Jr., Lemeshow, S., and Sturdivant, R.X. (2013). *Applied Logistic Regression*. John Wiley & Sons: New York.
26. Seshan, V.E., Gonen, M., and Begg, C.B. (2013). Comparing ROC curves derived from regression models. Stat. Med. 32, 1483–1493.
27. Pencina, M.J., D'Agostino, R.B., D'Agostino, R.B., and Vasan, R.S. (2008). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat. Med. 27, 157–172; discussion, 207–212.
28. Lemeshow, S., and Hosmer, D.W. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. Am. J. Epidemiol. 115, 92–106.
29. Kramer, A.A., and Zimmerman, J.E. (2007). Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit. Care Med. 35, 2052–2056.
30. Finazzi, S., Poole, D., Luciani, D., Cogo, P.E., and Bertolini, G. (2011). Calibration belt for quality-of-care assessment based on dichotomous outcomes. PLoS One 6, e16110.
31. Hilden, J., Habbema, J.D., and Bjerregaard, B. (1978). The measurement of performance in probabilistic diagnosis. III. Methods based on continuous functions of the diagnostic probabilities. Methods Inf. Med. 17, 238–246.
32. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12, 77.
33. Kundu, S., Aulchenko, Y.S., van Duijn, C.M., and Janssens, A.C.J.W. (2011). PredictABEL: an R package for the assessment of risk prediction models. Eur. J. Epidemiol. 26, 261–264.
34. Vassar, M.J., Lewis, F.R., Chambers, J.A., Mullins, R.J., O'Brien, P.E., Weigelt, J.A., Hoang, M.T., and Holcroft, J.W. (1999). Prediction of outcome in intensive care unit trauma patients: a multicenter study of Acute Physiology and Chronic Health Evaluation (APACHE), Trauma and Injury Severity Score (TRISS), and a 24-hour intensive care unit (ICU) point system. J. Trauma 47, 324–329.
35. Pouw, M.E., Peelen, L.M., Moons, K.G.M., Kalkman, C.J., and Lingsma, H.F. (2013). Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics. BMJ 347, f5913.
36. Myburgh, J.A., Cooper, D.J., Finfer, S.R., Venkatesh, B., Jones, D., Higgins, A., Bishop, N., and Higlett, T.; Australasian Traumatic Brain Injury Study (ATBIS) Investigators for the Australian, New Zealand Intensive Care Society Clinical Trials Group. (2008). Epidemiology and 12-month outcomes from traumatic brain injury in Australia and New Zealand. J. Trauma 64, 854–862.
37. Shann, F. (2000). Mortality prediction model is preferable to APACHE. BMJ 320, 714.
38. Maas, A.I.R., Menon, D.K., Lingsma, H.F., Pineda, J.A., Sandel, M.E., and Manley, G.T. (2012). Re-orientation of clinical research in traumatic brain injury: report of an international workshop on comparative effectiveness research. J. Neurotrauma 29, 32–46.
39. Beynon, C., Kiening, K.L., Orakcioglu, B., Unterberg, A.W., and Sakowitz, O.W. (2012). Brain tissue oxygen monitoring and hyperoxic treatment in patients with traumatic brain injury. J. Neurotrauma 29, 2109–2123.
40. Sanchez, J.J., Bidot, C.J., O'Phelan, K., Gajavelli, S., Yokobori, S., Olvey, S., Jagid, J., Garcia, J.A., Nemeth, Z., and Bullock, R. (2013). Neuromonitoring with microdialysis in severe traumatic brain injury patients. Acta Neurochir. Suppl. 118, 223–227.
41. Papa, L., Ramia, M.M., Kelly, J.M., Burks, S.S., Pawlowicz, A., and Berger, R.P. (2013). Systematic review of clinical research on biomarkers for pediatric traumatic brain injury. J. Neurotrauma 30, 324–338.
42. Marmarou, A., Lu, J., Butcher, I., McHugh, G.S., Mushkudiani, N.A., Murray, G.D., Steyerberg, E.W., and Maas, A.I.R. (2007). IMPACT database of traumatic brain injury: design and description. J. Neurotrauma 24, 239–250.
43. Weir, J., Steyerberg, E.W., Butcher, I., Lu, J., Lingsma, H.F., McHugh, G.S., Roozenbeek, B., Maas, A.I.R., and Murray, G.D. (2012). Does the extended Glasgow Outcome Scale add value to the conventional Glasgow Outcome Scale? J. Neurotrauma 29, 53–58.
44. Raj, R., Siironen, J., Kivisaari, R., Hernesniemi, J., Tanskanen, P., Handolin, L., and Skrifvars, M.B. (2013). External validation of the IMPACT Model and the role of markers of coagulation. Neurosurgery 73, 305–311.

Address correspondence to:
Rahul Raj, PhD(c)
Department of Neurosurgery
Helsinki University Hospital
Topeliuksenkatu 5
PB-266, 00029 HUS
Helsinki, Finland

E-mail: rahul.br.raj@icloud.com