THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Predicting outcome after traumatic brain injury

OPEN ACCESS

Predicting Outcome after Traumatic Brain Injury: Development and International Validation of Prognostic Scores Based on Admission Characteristics

Ewout W. Steyerberg[1*], Nino Mushkudiani[1], Pablo Perel[2], Isabella Butcher[3], Juan Lu[4], Gillian S. McHugh[3], Gordon D. Murray[3], Anthony Marmarou[4], Ian Roberts[2], J. Dik F. Habbema[1], Andrew I. R. Maas[5]

1 Center for Medical Decision Sciences, Department of Public Health, Erasmus MC, Rotterdam, The Netherlands, 2 London School of Hygiene and Tropical Medicine, Nutrition and Public Health Intervention Research Unit, London, United Kingdom, 3 Division of Community Health Sciences, University of Edinburgh, Scotland, 4 Department of Neurosurgery, Virginia Commonwealth University, Richmond, Virginia, United States of America, 5 Department of Neurosurgery, Erasmus MC, Rotterdam, The Netherlands

Abbreviations: AUC, area under the receiver operating characteristic curve; CRASH, Corticosteroid Randomisation after Significant Head Injury; CT, computed tomography; EDH, epidural hematoma; GCS, Glasgow Coma Scale; GOS, Glasgow Outcome Scale; IMPACT, International Mission for Prognosis and Analysis of Clinical Trials in TBI; OR, odds ratio; RCT, randomized controlled trial; TBI, traumatic brain injury; tSAH, traumatic subarachnoid hemorrhage

* To whom correspondence should be addressed. E-mail: E.Steyerberg@ErasmusMC.nl

ABSTRACT

Background

Traumatic brain injury (TBI) is a leading cause of death and disability. A reliable prediction of outcome on admission is of great clinical relevance. We aimed to develop prognostic models with readily available traditional and novel predictors.

Methods and Findings

Prospectively collected individual patient data were analyzed from 11 studies. We considered predictors available at admission in logistic regression models to predict mortality and unfavorable outcome according to the Glasgow Outcome Scale at 6 mo after injury. Prognostic models were developed in 8,509 patients with severe or moderate TBI, with cross-validation by omission of each of the 11 studies in turn. External validation was on 6,681 patients from the recent Medical Research Council Corticosteroid Randomisation after Significant Head Injury (MRC CRASH) trial. We found that the strongest predictors of outcome were age, motor score, pupillary reactivity, and CT characteristics, including the presence of traumatic subarachnoid hemorrhage. A prognostic model that combined age, motor score, and pupillary reactivity had an area under the receiver operating characteristic curve (AUC) between 0.66 and 0.84 at cross-validation. This performance could be improved (AUC increased by approximately 0.05) by considering CT characteristics, secondary insults (hypotension and hypoxia), and laboratory parameters (glucose and hemoglobin). External validation confirmed that the discriminative ability of the model was adequate (AUC 0.80). Outcomes were systematically worse than predicted, but less so in 1,588 patients who were from high-income countries in the CRASH trial.

Conclusions

Prognostic models using baseline characteristics provide adequate discrimination between patients with good and poor 6 mo outcomes after TBI, especially if CT and laboratory findings are considered in addition to traditional predictors. The model predictions may support clinical practice and research, including the design and analysis of randomized controlled trials.

The Editors' Summary of this article follows the references.

Introduction

Traumatic brain injury (TBI) is a leading cause of death and disability. Establishing a reliable prognosis early after injury is notoriously difficult, as is captured in the Hippocratic aphorism, "No head injury is too severe to despair of, nor too trivial to ignore." Following the development of the Glasgow Coma Scale (GCS) [1] and the Glasgow Outcome Scale (GOS) [2], it was found that confident predictions could be made after 24 h following the injury, but were difficult to establish on admission [3]. Prognostic models with admission data are essential to support early clinical decision-making, and to facilitate reliable comparison of outcomes between different patient series and variation in results over time. Furthermore, prognostic models have an important role in randomized controlled trials (RCTs), for stratification [4] and statistical analyses that explicitly consider prognostic information, such as covariate adjustment [5,6].

Many models include data obtained after admission, and most were developed on relatively small sample sizes originating from a single center or region [7,8]. Many models lack external validation, which is essential before the broad application of a model can be advised [9,10]. Furthermore, few models are presented in a clinically practical way.

We aimed to develop prognostic models based on admission characteristics, which would allow application of the model before in-hospital therapeutic interventions. We used several large patient series for model development as available in the International Mission for Prognosis and Analysis of Clinical Trials in TBI (IMPACT) project [11], as an extension of multivariable analyses reported before [12]. External validation was possible on data from a large, recently completed RCT [13]. This RCT was used to develop a series of prediction models with a specific focus on non-Western countries [14]. In parallel with this work and as part of a collaboration between CRASH and IMPACT investigators, we developed and describe here a basic model that includes easily accessible clinical features, and additional models that included findings from computed tomography (CT) scanning, and laboratory measurements.

Methods

Patients

The IMPACT database includes patients with moderate and severe TBI (GCS ≤ 12) from eight randomized controlled trials and three observational studies conducted between 1984 and 1997 [11]. Detailed characteristics of these 11 studies and data management have been described previously [15]. The endpoint for the prognostic analyses was the 6 mo GOS, which is an ordered outcome with five categories: 1, dead; 2, vegetative state; 3, severe disability; 4, moderate disability; and 5, good recovery. In patients whose 6 mo assessment was not available we used the 3 mo GOS ($n = 1,611$, 19% of the patients). We selected 8,509 patients aged ≥ 14 y [12].

We externally validated prognostic models using patients enrolled in the Medical Research Council Corticosteroid Randomisation after Significant Head Injury (MRC CRASH) trial (trial registration ISRCTN74459797, ISRCTN Register, http://www.controlled-trials.com/), who were recruited between 1999 and 2004 [13]. This was a large international double-blind, randomized placebo-controlled trial of the effect of early administration of a 48-h infusion of methylprednisolone on outcome after head injury. It was found that the risks of death and disability were higher in the corticosteroid group than in the placebo group. The trial included 10,008 adults with GCS ≤ 14, who were enrolled within 8 h after injury. We selected 6,681 patients with a GCS ≤ 12 and with complete 6 mo GOS. Secondary analyses considered only placebo patients ($n = 3,287$) and patients from high-income countries ($n = 1,588$). For the validation we focused on prediction of mortality (GOS 1) versus survival (GOS 2–5) and of unfavorable (GOS 1–3) versus favorable outcome (GOS 4–5).

Predictors and Model Development

We considered patient characteristics that could be determined easily and reliably within the first few hours after injury. We initially examined a set of 26 potential predictors [12]. These included demographics (age, sex, race, education), indicators of clinical severity (cause of injury, GCS components, pupillary reactivity), secondary insults (hypoxia, hypotension, hypothermia), blood pressure (systolic, diastolic), various CT characteristics and various biochemical variables. For the present analyses, we selected predictors that were important in predicting outcome (according to the Nagelkerke R^2 in multivariable analyses), and available for a substantial numbers of patients in the development cohort [12]. Three prognostic models were defined: (1) The core model included age, the motor score component from the GCS, and pupillary reactivity; (2) the extended model included the three predictors from the core model plus information on secondary insults (hypoxia, hypotension), CT characteristics (Marshall CT classification [16]), traumatic subarachnoid hemorrhage (tSAH), and epidural hematoma (EDH); and (3) the lab model included the characteristics from the extended model and additional information on glucose and hemoglobin (Hb). Definitions of predictors have been described in detail [15].

Age and motor score were available for all patients. Missing values occurred for several other predictors, especially because some predictors were not recorded in some studies. Within studies, predictor values were generally over 90% complete if the predictor was recorded [15]. Pupillary reactivity was not recorded in two trials ($n = 1,045$), but were nearly complete in the other studies (338 missing values among 7,474 patients).

For the extended model we excluded one trial, since hypoxia, hypotension, and the CT classification were not recorded, leaving 6,999 patients. For the development of the lab model, we were limited to four studies in which glucose and Hb had been recorded ($n = 3,554$). Missing values occurred for 167 glucose values (5%), and 132 Hb values (4%).

We multiply imputed ten sets of data that were identical in known information, but could differ on imputed values for missing information. We used the method of chained equations, sampling imputed values from the posterior predictive distribution of the missing data [17–20]. We used the MICE algorithm [21], which works with R software [22]. The imputation models used all the variables that we considered as potential predictors as well as the 6 mo GOS. In total, 1,383 of the required 25,527 values (5%) in the core model were imputed, 7,477 of the 55,992 required values in

the extended model (13%), and 2,965 of the 35,540 required values in the lab model (8%).

Statistical Analysis

Proportional odds logistic regression analysis was performed with the 6 mo GOS as an ordinal outcome [23]. This analysis efficiently summarizes predictive relationships with an ordinal outcome such as the GOS. The proportionality assumption was checked for each selected predictor and found to be reasonable [12]. Interaction terms between predictors were examined with likelihood ratio tests, but none was of sufficient relevance to extend the models beyond the main effects for each predictor. Similarly, study-specific effects were assessed with interaction terms between study and each predictor. Final prognostic models were developed with logistic regression analysis for dichotomized versions for the GOS: mortality (versus survival) and unfavorable outcome (versus favorable outcome). All analyses were stratified by study.

For the continuous predictors age, glucose, and Hb, a linear relationship with outcome was found to be a good approximation after assessment of nonlinearity using restricted cubic splines [24]. The odds ratios (ORs) were scaled so that they corresponded to a change from the 25th percentile to the 75th percentile of the predictor distribution. This scaling allowed for a direct comparison of the prognostic value of predictors that had been recorded in different units or on different scales. Pooled ORs were estimated over the imputed datasets (`fit.mult.impute` function from the Harrell *Design* library [25]). All analyses were repeated using only complete data, which gave similar results (unpublished data).

Internal Validation

The discriminatory power of the three models was indicated by the area under the receiver operating characteristic curve (AUC). The AUC varies between 0.5 (a noninformative model) and 1.0 (a perfect model). AUC was calculated in a cross-validation procedure, where each study was omitted in turn. Results were pooled over the ten imputed datasets for eight studies with sufficient numbers for reliable validation ($n > 500$) [26].

External Validation

We aimed to validate all models externally using data from selected patients in the CRASH trial. However, lab values were not recorded in this trial, nor were hypoxia, hypotension, and EDH. We therefore validated the core model, and a variant of the extended model, in which only the Marshall CT classification and presence of tSAH were added to the core model (i.e., the core + CT model). Results are shown for patients with complete data (core model: $n = 6,272$; extended model variant, $n = 5,309$). Imputation of missing values was performed as for the IMPACT studies, leading to similar results (unpublished data). Performance criteria comprised discrimination (measured using the AUC) and calibration (agreement of observed outcomes with predicted risk). Calibration was assessed with the Hosmer-Lemeshow test and graphically using a calibration plot [24].

Model Presentation

The final models were presented in a score chart, with scores based on the regression coefficients in the propor-

tional odds models [27]. Coefficients were scaled such that the same rounded score was obtained for predictors that were used across the different models (e.g., age, motor score, pupils). Logistic regression was subsequently used to calibrate the risks of mortality and unfavorable outcome according to the scores, with the model intercept referring to the Tirilazad international trial [15]. This intercept was chosen since it represented typical proportions of mortality (278/1,118, 25%) and unfavorable outcome (456/1,118, 41%). Predictions can be calculated from an Excel spreadsheet and from a Web page (Text S1 is also available at http://www.tbi-impact.org/).

Results

The characteristics of IMPACT and CRASH patients with GCS $\leq$ 12 were fairly comparable (Table 1). CRASH trial patients were marginally older than in IMPACT, and admission motor scores were somewhat higher. Six-month mortality was 28% in IMPACT and 32% in CRASH, and unfavorable outcomes occurred in nearly half of the patients (48% in IMPACT, 47% in CRASH). Mortality was slightly lower in the placebo group of the selected CRASH patients (mortality 988/3,287, 30%; unfavorable outcome 1,524/3,287, 46%), and in the patients from high-income countries (mortality 405/1,588, 26%; unfavorable outcome 747/1,588, 47%).

All predictors had statistically significant associations with 6 mo GOS in univariate and multivariable analyses (Table 2). An increase in age equal to the interquartile range (24 y) was associated with approximately a doubling of the risk of poor outcome. A poor outcome occurred especially for those with motor scores 1 (none) or 2 (extension). Pupillary reactivity, hypoxia, and hypotension also had strong prognostic effects. CT classifications showing mass lesions or signs of raised Intracranial Pressure (CT class III to VI) had similar increases in risk as the presence of tSAH (OR around 2). An EDH was a relatively favorable sign on a CT (compared to not having an EDH on CT). Higher glucose levels and lower Hb levels were associated with a poor outcomes, but effects were more moderate than, for example, for age.

A simple score chart for the sequential application of the models is presented in Figure 1, which can be used in combination with Figure 2 to obtain approximate predictions for individual patients. For example, a 35-y-old patient, with a motor score of 3 (abnormal flexion), and both pupils reacting, has a core model score of $1 + 4 + 0 = 5$ points. According to Figure 2, this score corresponds to risks of mortality and unfavorable outcome of approximately 20% and 50%, respectively. If this patient had suffered from hypoxia but not hypotension before admission, and the CT showed a mass lesion and tSAH, the extended model score becomes 5 for the core model + $1 + 2 + 2 + 0 = 10$ points. The corresponding risks are approximately 40% for mortality and 70% for unfavorable outcome. When glucose is 10 mmol/l and Hb 11 g/dl, the lab model score increases by $2 + 2$ to 14 points, which corresponds to slightly higher predictions of mortality and unfavorable outcome than those estimated with the extended model (Figure 3).

Cross-Validation and External Validation

The discriminatory ability of the models increased with increasing complexity (Table 3). Within the IMPACT data, the

Table 1. Patient Characteristics of 11 Studies in the IMPACT Database and the CRASH Trial

Characteristics	Measure or Category	IMPACT Database (n = 8,509)	CRASH Trial[a] (n = 6,681)
Age, years	Median (25−75 percentile)	30 (21–45)	32 (23–47)
Motor score	None (1)	1,395 (16%)	785 (12%)
	Extension (2)	1,042 (12%)	515 (8%)
	Abnormal flexion (3)	1,085 (13%)	658 (10%)
	Normal flexion (4)	1,940 (23%)	1,156 (17%)
	Localizes/obeys (5/6)	2,591 (30%)	3,567 (53%)
	Untestable/missing (9)	456 (5%)	0
Pupillary reactivity	Total	7,126	6,272
	Both pupils reacted	4,486 (63%)	4,956 (74%)
	One pupil reacted	886 (12%)	530 (8%)
	No pupil reacted	1,754 (25%)	786 (12%)
Hypoxia	Total	5,452	NA
	Yes or suspected	1,116 (20%)	NA
Hypotension	Total	6,420	NA
	Yes or suspected	1,171 (18%)	NA
CT classification[b]	Total	5,192	5,654
	I	360 (7%)	954 (17%)
	II	1,838 (35%)	1,517 (27%)
	III	863 (17%)	604 (11%)
	IV	187 (4%)	133 (2%)
	V	1,435 (28%)	815 (14%)
	VI	509 (10%)	1,631 (29%)
Traumatic subarachnoid hemorrhage	Total	7,393	5,653
	Yes	3,313 (45%)	2,045 (36%)
Epidural hematoma	Total	7,409	NA
	Yes	999 (13%)	NA
Glucose (mmol/l)	Total	4,830	NA
	Median (25–75 percentile)	8.2 (6.7–10.4)	NA
Hb (g/dl)	Total	4,376	NA
	Median (25–75 percentile)	12.7 (10.8–14.3)	NA
Six-month outcome	Dead	2,396 (28%)	2,146 (32%)
	Vegetative	351 (4%)	993 (15%)[c]
	Severe disability	1,335 (16%)	—
	Moderate disability	1,666 (20%)	1,224 (18%)
	Good recovery	2,761 (32%)	2,318 (35%)

All values are given as number (percentage of listed total) unless indicated otherwise.
[a]Selected for GCS ≤ 12
[b]CT classification: I, no visible intracranial pathology on CT scan; II, midline shift 0–5 mm; III, cisterns compressed or absent with midline shift 0–5 mm; IV, midline shift > 5 mm; V, any lesion surgically evacuated; VI, high- or mixed-density lesion > 25 mm, not surgically evacuated.
[c]Vegetative and severe disability were not separately recorded in CRASH.
NA, not available.
doi:10.1371/journal.pmed.0050165.t001

best cross-validated performance was seen for the three observational studies, with AUCs over 0.80. Evaluation in the RCTs showed lower AUCs. External validation confirmed the discriminatory ability of the core model in the CRASH trial (AUC 0.776 and 0.780 for mortality and unfavorable outcome, respectively, Figures 4 and 5). When CT classification and tSAH were considered as well, the performance increased to 0.801 and 0.796 for mortality and unfavorable outcome, respectively, for 5,309 patients in CRASH. Outcomes in CRASH were systematically poorer than those predicted for both the core and core + CT models, for both mortality and unfavorable outcome (Hosmer-Lemeshow tests, $p < 0.001$, Figures 4 and 5). This miscalibration was slightly less but did not disappear when only the placebo patients were considered. Calibration was better for the patients from high-income countries, with near perfect calibration (Hosmer-Lemeshow tests, $p > 0.1$) for the extended model predicting mortality ($n = 1,351$, Figure 4) and the core model predicting unfavorable outcomes ($n = 1,466$, Figure 5).

Discussion

In this paper we describe the development of a series of prognostic models of increasing complexity, based on admission characteristics, to predict the risk of 6-mo mortality and unfavorable outcomes in individual patients after moderate or severe TBI. The models discriminated adequately between patients with poor and good outcomes, especially in the relatively unselected observational studies. Patients in the randomized trials were selected according to various enrollment criteria, which led to more homogeneous samples, as reflected in a lower discriminative ability of the models. We found a small but systematic difference between predicted and observed outcome in a large, relatively recent, external validation set [13] with recently treated patients from both high- and low/middle-income countries. This miscalibration largely disappeared when we considered only patients from high-income countries.

The largest amount of prognostic information was contained in a core set of three predictors: age, motor score, and

Table 2. Associations between Predictors and 6-Month Outcome in the IMPACT Data ($n = 8,509$)

Characteristics	Coding	6 Month Outcome Number (%)		Odds Ratios (95% CI)			
		Dead ($n = 2,396$)	Unfavorable ($n = 4,082$)	Univariate	Core Model ($n = 8,509$)	Extended Model[a] ($n = 6,999$)	Lab Model[b] ($n = 3,554$)
Age, years	45 versus 21 years	—	—	2.2 (2.0–2.3)	2.4 (2.2–2.5)	2.2 (2.0–2.3)	1.9 (1.7–2.1)
Motor score	None (1)	625 (45%)	894 (64%)	4.9 (4.3–5.5)	3.9 (3.4–4.5)	3.4 (2.9–4.0)	2.8 (2.1–3.7)
	Extension (2)	496 (48%)	807 (77%)	7.2 (6.3–8.3)	5.7 (4.9–6.6)	4.6 (3.9–5.4)	4.3 (3.5–5.4)
	Abnormal flexion (3)	326 (30%)	619 (57%)	3.5 (3.1–4)	3.0 (2.6–3.5)	2.8 (2.4–3.2)	2.7 (2.2–3.3)
	Normal flexion (4)	411 (21%)	800 (41%)	1.8 (1.6–2)	1.7 (1.5–1.9)	1.6 (1.4–1.8)	1.5 (1.3–1.8)
	Localizes/obeys (5/6)	383 (15%)	699 (27%)	1.0 (ref)	1.0 (ref)	1.0 (ref)	1.0 (ref)
	Untestable/missing (9)	155 (34%)	263 (58%)	2.2 (1.8–2.7)	2.1 (1.7–2.6)	2.0 (1.7–2.5)	1.3 (0.6–2.6)
Pupillary reactivity	Both pupils reacted	790 (18%)	1,578 (35%)	1.0 (ref)	1.0 (ref)	1.0 (ref)	1.0 (ref)
	One pupil reacted	295 (33%)	521 (59%)	2.7 (2.4–3.1)	1.8 (1.6–2.1)	1.6 (1.4–1.8)	1.4 (1.1–1.7)
	No pupil reacted	946 (54%)	1,351 (77%)	5.9 (5.3–6.6)	3.3 (3.0–3.7)	2.7 (2.4–3.1)	2.1 (1.6–2.6)
Hypoxia	Yes or suspected	481 (43%)	713 (64%)	2.1 (1.9–2.4)	—	1.3 (1.1–1.5)	1.4 (1.2–1.7)
	No	1,158 (27%)	1,928 (44%)	1.0 (ref)	—	1.0 (ref)	1.0 (ref)
Hypotension	Yes or suspected	578 (49%)	794 (68%)	2.7 (2.4–3.1)	—	1.8 (1.6–2.1)	1.5 (1.2–1.8)
	No	1,315 (25%)	2,263 (43%)	1.0 (ref)	—	1.0 (ref)	1.0 (ref)
CT classification[c]	I	24 (7%)	50 (14%)	0.41 (0.33–0.52)	—	0.64 (0.51–0.82)	0.65 (0.47–0.89)
	II	256 (14%)	582 (32%)	1.0 (ref)	—	1.0 (ref)	1.0 (ref)
	III	287 (33%)	456 (53%)	2.6 (2.3–3)	—	1.7 (1.5–2.0)	1.7 (1.4–2.0)
	IV	86 (46%)	107 (57%)	—	—		
	V	422 (29%)	709 (49%)	2.3 (2–2.6)	—	1.6 (1.4–1.9)	1.8 (1.5–2.2)
	VI	217 (43%)	293 (58%)	—	—		
Traumatic subarachnoid	Yes	1,193 (36%)	1,925 (58%)	2.6 (2.4–2.9)	—	1.7 (1.5–1.8)	1.8 (1.6–2.1)
Hemorrhage	No	724 (18%)	1,462 (36%)	1.0 (ref)	—	1.0 (ref)	1.0 (ref)
Epidural hematoma	Yes	207 (21%)	358 (36%)	0.64 (0.56–0.72)	—	0.61 (0.53–0.70)	0.56 (0.46–0.69)
	No	1,794 (28%)	3,101 (48%)	1.0 (ref)	—	1.0 (ref)	1.0 (ref)
Glucose	10.4 versus 6.7.mmol/l	—	—	1.7 (1.6–1.8)	—	—	1.3 (1.2–1.4)
Hb	14.3 versus 10.8 g/dl	—	—	0.66 (0.61–0.72)	—	—	0.78 (0.70 – 0.87)

ORs were calculated with logistic regression for the 6 mo GOS with five categories (proportional odds analysis, see text).
[a]Core plus hypoxia, hypotension, and CT characteristics.
[b]Extended plus glucose and Hb.
[c]CT classification: I, no visible intracranial pathology on CT scan; II, midline shift 0–5 mm; III, cisterns compressed or absent with midline shift 0–5 mm; IV, midline shift > 5 mm; V, any lesion surgically evacuated; VI, high- or mixed-density lesion > 25 mm, not surgically evacuated. In the analysis, category III and IV were combined, as well as V and VI.
doi:10.1371/journal.pmed.0050165.t002

pupillary reactivity at admission. These characteristics were already considered in the first well-known model for TBI [3] and in many subsequent prognostic models [7,8]. Information from the CT scan provided additional prognostic information, although we did not exploit all the prognostic information contained in a CT scan. The Marshall CT classification combines several characteristics, and we previously proposed a more detailed scoring for prognostic purposes [28]. Further validation of this score is necessary, but the required data were not sufficiently available in most studies from IMPACT. The presence of EDH was associated with a better outcome after trauma, which may be explained by the possibility of emergent surgical evacuation of such hematomas. An EDH often disturbs brain function because of compression, although there is generally little intrinsic brain damage. If compression is relieved in time, full recovery will more likely occur. Laboratory parameters have not yet been widely used for prognosis after TBI [29]. Glucose and hemoglobin were shown to contribute to outcome prediction, although their effects are smaller than other predictors, e.g., age. Coagulation parameters may also be very relevant for outcome prediction [29], but these parameters were not sufficiently available in our studies. These biochemical parameters warrant further exploration, especially since they are amenable to intervention. For example, in critical care,

intensive hyperglycemia management has been shown to reduce mortality [30]. We could not include effects of extracranial injuries, since measures such as the ISS (injury severity score) were not consistently recorded in the IMPACT studies. Major extracranial injury was included as a predictor in recently developed prognostic models from the CRASH trial [14]. It is likely that the AUC of our models would have been even better if this variable had been available [31,32]

Relationship of Our Model to Previously Published Models

Several models have been derived to estimate the probability of hospital mortality of adult intensive care unit patients with physiological characteristics collected during the first day(s), including APACHE (Acute Physiology and Chronic Health Evaluation), SAPS (Simplified Acute Physiology Score), and MPM (Mortality Prediction Model) [33–35]. Our models differ in several aspects, since we predicted long term outcome, specifically for TBI patients, and used only baseline characteristics. Recently, prognostic models for 14 d mortality and 6 mo outcome were published by the MRC CRASH trial collaborators. CRASH was a mega-trial, including mild TBI (30% of $n = 10,008$), with a relatively simple data collection in mostly patients from low-income countries (75% of $n = 10,008$) [14]. The IMPACT database involves merged individual patient data from eight clinical trials and three observational series, conducted over approximately 15

Characteristics	Value	Score	Sum
Age (years)	≤ 30	0	
	30 – 39	1	
	40 – 49	2	
	50 – 59	3	
	60 – 69	4	
	70 +	5	
Motor score	None/extension	6	
	Abnormal flexion	4	
	Normal flexion	2	
	Localizes/obeys	0	
	Untestable/missing	3	
Pupillary reactivity	Both pupils reacted	0	
	One pupil reacted	2	
	No pupil reacted	4	
Sum score core model			
Hypoxia	Yes or suspected	1	
	No	0	
Hypotension	Yes or suspected	2	
	No	0	
CT classification	I	−2	
	II	0	
	III/IV	2	
	V/VI	2	
Traumatic subarachnoid	Yes	2	
Hemorrhage	No	0	
Epidural hematoma	Yes	−2	
	No	0	
Subscore CT			
Sum score extended model			
Glucose (mmol/l)	< 6	0	
	6–8.9	1	
	9–11.9	2	
	12–14.9	3	
	15 +	4	
Hb (g/dl)	< 9	3	
	9–11.9	2	
	12–14.9	1	
	15 +	0	
Subscore lab			
Sum score lab model			

Figure 1. Score Chart for 6 Month Outcome after TBI

Sum scores can be calculated for the core model (age, motor score, pupillary reactivity), the extended model (core + hypoxia + hypotension + CT characteristics), and a lab model (core + hypoxia + hypotension + CT + glucose + Hb). The probability of 6 mo outcome is defined as $1 / (1 + e^{-LP})$, where LP refers to the linear predictor in a logistic regression model. Six LPs were defined as follows:

$LP_{core,\ mortality} = -2.55 + 0.275 \times$ sum score core

$LP_{core,\ unfavorable\ outcome} = -1.62 + 0.299 \times$ sum score core

$LP_{extended,\ mortality} = -2.98 + 0.256 \times$ (sum score core + subscore CT)

$LP_{extended,\ unfavorable\ outcome} = -2.10 + 0.276 \times$ (sum score core + subscore CT)

$LP_{lab,\ mortality} = -3.42 + 0.216 \times$ (sum score core + subscore CT + subscore lab)

$LP_{lab,\ unfavorable\ outcome} = -2.82 + 0.257 \times$ (sum score core + subscore CT + subscore lab)

The logistic functions are plotted with 95% confidence intervals in Figure 2.

doi:10.1371/journal.pmed.0050165.g001

y, and focused on severe TBI. The IMPACT data are available in greater detail, especially with respect to CT scan characteristics.

We externally validated modified versions of two of our three IMPACT models in selected patients from the CRASH trial with GCS ≤ 12, similar to the external validation of two modified versions of CRASH models for unfavorable outcome at 6 mo in IMPACT [14]. Both studies confirmed the external validity of the presented models. This collaboration with reciprocal validation of CRASH and IMPACT models is important for reliable application of models outside their respective development settings.

Early prediction of outcome permits establishment of a baseline risk profile for individual patients, thus providing a reference for assessing quality of health-care delivery. Prognostic models are particularly relevant for a more efficient design and analysis of RCTs. For example, we can exclude those with a very good or a very poor prognosis [4], perform covariate adjustment of a treatment effect [6,36], and consider other analyses that lead to increases in statistical power [37].
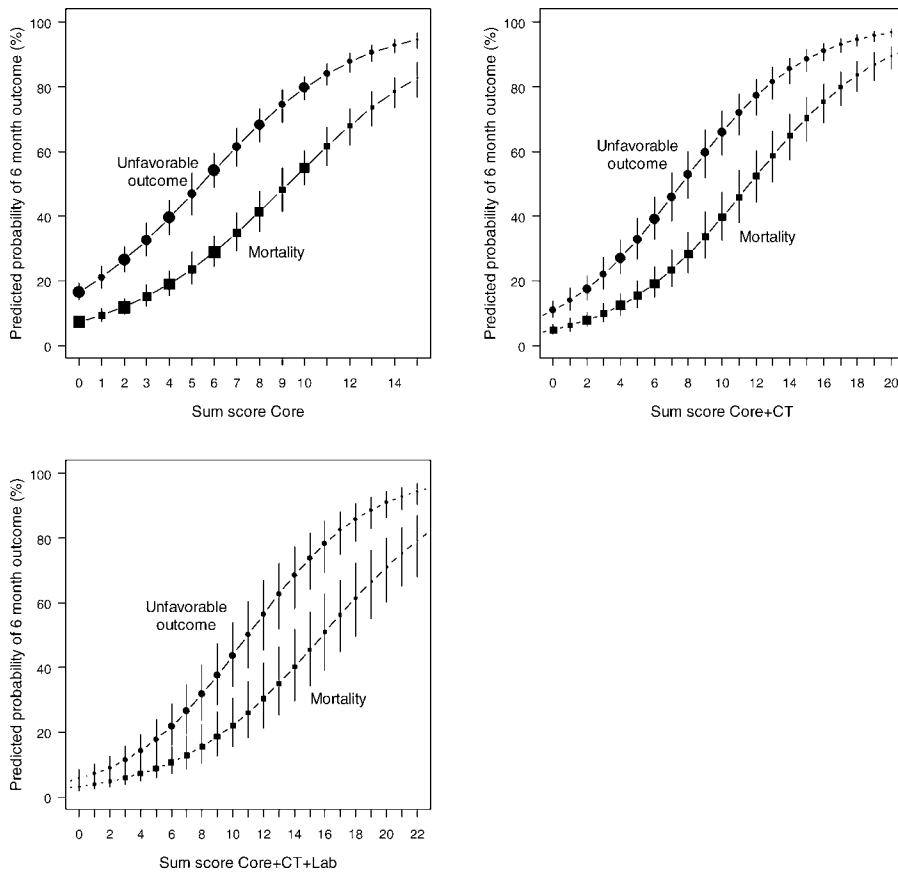
Figure 2. Predicted Probabilities of Mortality and Unfavorable Outcome at 6 Month after TBI in Relation to the Sum Scores from the Core, Extended, and Lab Models

The logistic functions are plotted with 95% confidence intervals. Dot size is proportional to sample size. Sum scores can be obtained from Figure 1.

doi:10.1371/journal.pmed.0050165.g002

The proposed scores may also support clinicians in their initial assessment of the severity and prognosis of a TBI patient. We note, however, that statistical models can only augment, not replace clinical judgment, although it is unlikely that any clinician has the equivalent systematic experience of the outcomes of the thousands of patients underlying our models. Predictions should be regarded with care and not directly be applied for treatment-limiting decisions [38]. The
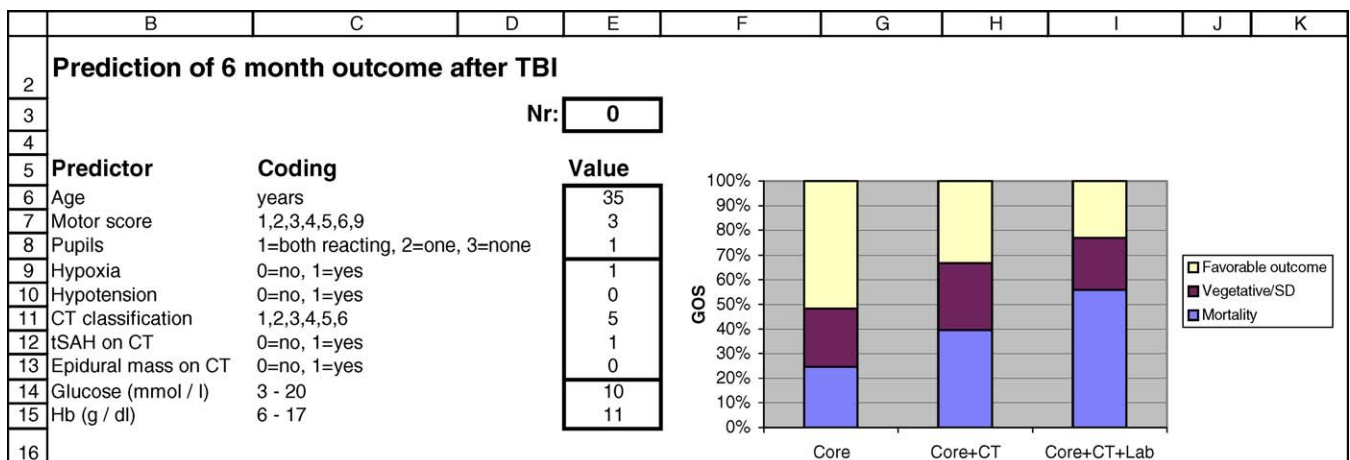

Figure 3. Screenshot of the Spreadsheet with Calculations of Probabilities for the Three Prediction Models

Predictions are calculated for a 35-y-old patient with motor score 3, both pupils reacting, hypoxia before admission, mass lesion and tSAH on admission CT scan, glucose 11 mmol/l, and Hb 10 g/dl. A Web-based calculator is available at http://www.tbi-impact.org/.

doi:10.1371/journal.pmed.0050165.g003

Table 3. Discriminative Ability of the Models at Cross-Validation in IMPACT Patients (Studies with *n* > 500), and External Validation in Patients from the CRASH Trial

Dataset	Selection	Study	*n*	Mortality			Unfavorable Outcome		
				Core	Extended	Lab	Core	Extended	Lab
Cross-validation in IMPACT	RCTs	TirInt	1,118	0.70	0.78	0.80	0.75	0.80	0.80
		TirUS	1,041	0.74	0.77	0.79	0.78	0.80	0.82
		Saphir	919	0.66	0.71	0.72	0.70	0.73	0.75
		Pegsod	1,510	0.76	—	—	0.77	—	—
		HIT-II	819	0.72	0.77	—	0.74	0.78	—
	Observational studies	UK4	791	0.81	0.81	—	0.81	0.80	—
		TCDB	604	0.81	0.83	—	0.82	0.83	—
		EBIC	822	0.84	0.87	—	0.81	0.84	—
External validation in CRASH	All with GCS ≤12	CRASH	6,272	0.78	0.80	—	0.78	0.80	—
	Placebo with GCS ≤12	CRASH	3,075	0.78	0.81	—	0.78	0.79	—
	HIC with GCS ≤12	CRASH	1,466	0.80	0.83	—	0.77	0.80	—

Values are AUCs for models predicting 6 mo mortality (GOS 1) and models predicting 6 mo unfavorable outcome (GOS 1–3).
CRASH, Corticosteroid Randomization after Significant Head Injury trial (1999–2004); EBIC, European Brain Injury Consortium survey (1995); HIC, high-income countries; HIT-II, Nimodipine trial (1989–1991); Pegsod, Pegsod trial (1993–1995); Saphir, Saphir trial (1995–1997); TCDB, Traumatic Coma Data Bank (1984–1987); TirInt, International Tirilazad (enrollment 1992–1994); TirUS, US Tirilazad (1991–1994); UK4, UK 4 Centres study (1986–1988) [15].
doi:10.1371/journal.pmed.0050165.t003

UK 4 Centres Study found that making predictions available as part of a routine clinical service altered deployment of resources [39].

The validity and applicability of the prognostic models is affected by various factors. The local level of care may vary between regions, which may result in differences in outcome. Previously we found unexplained outcome differences between the US and the international part of the Tirilazad
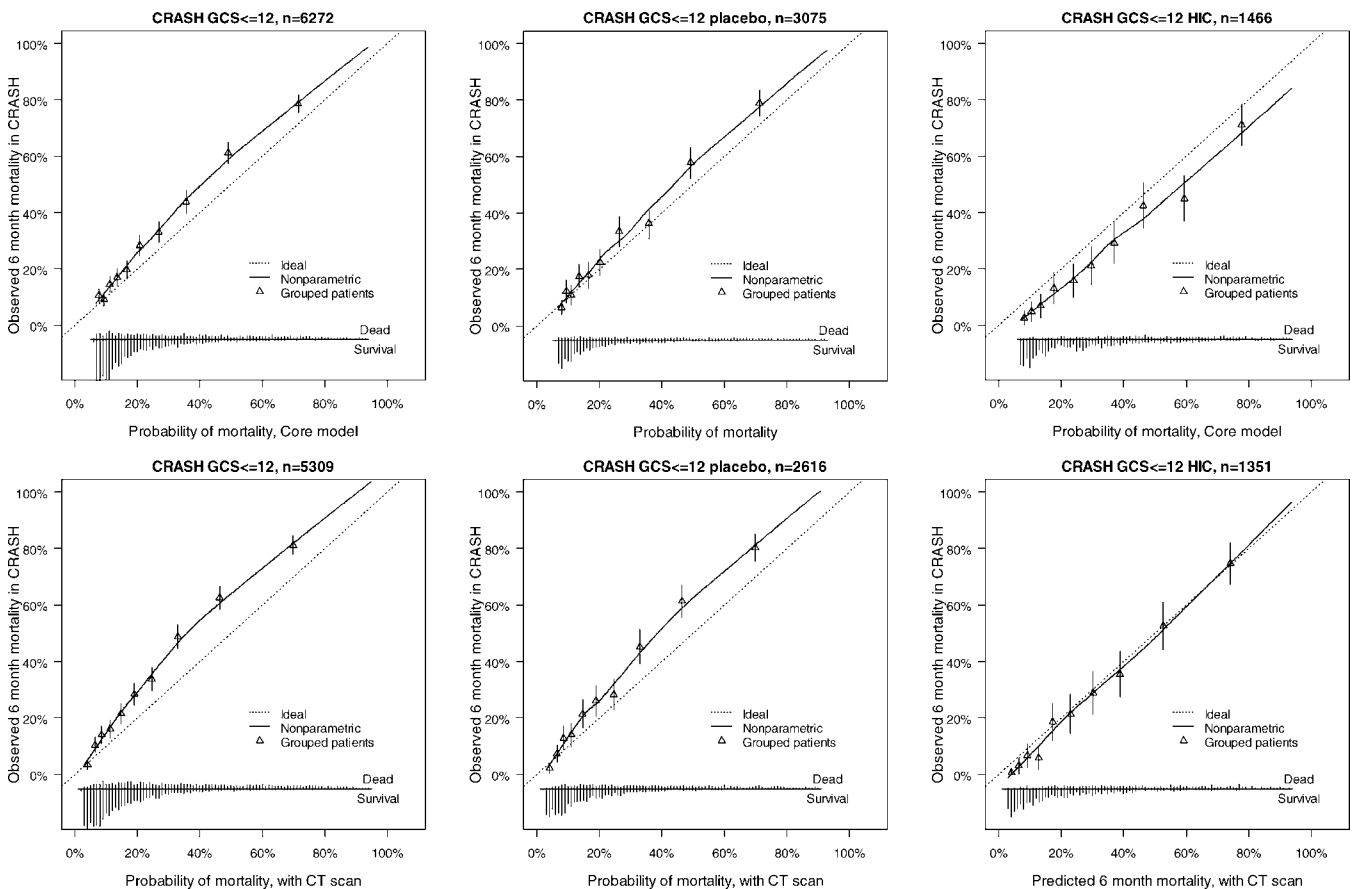

Figure 4. External Validity for the Core and Core + CT Model Characteristics for Prediction of Mortality in the CRASH Trial
The distribution of predicted probabilities is shown at the bottom of the graphs, by 6-mo mortality. The triangles indicate the observed frequencies by decile of predicted probability.
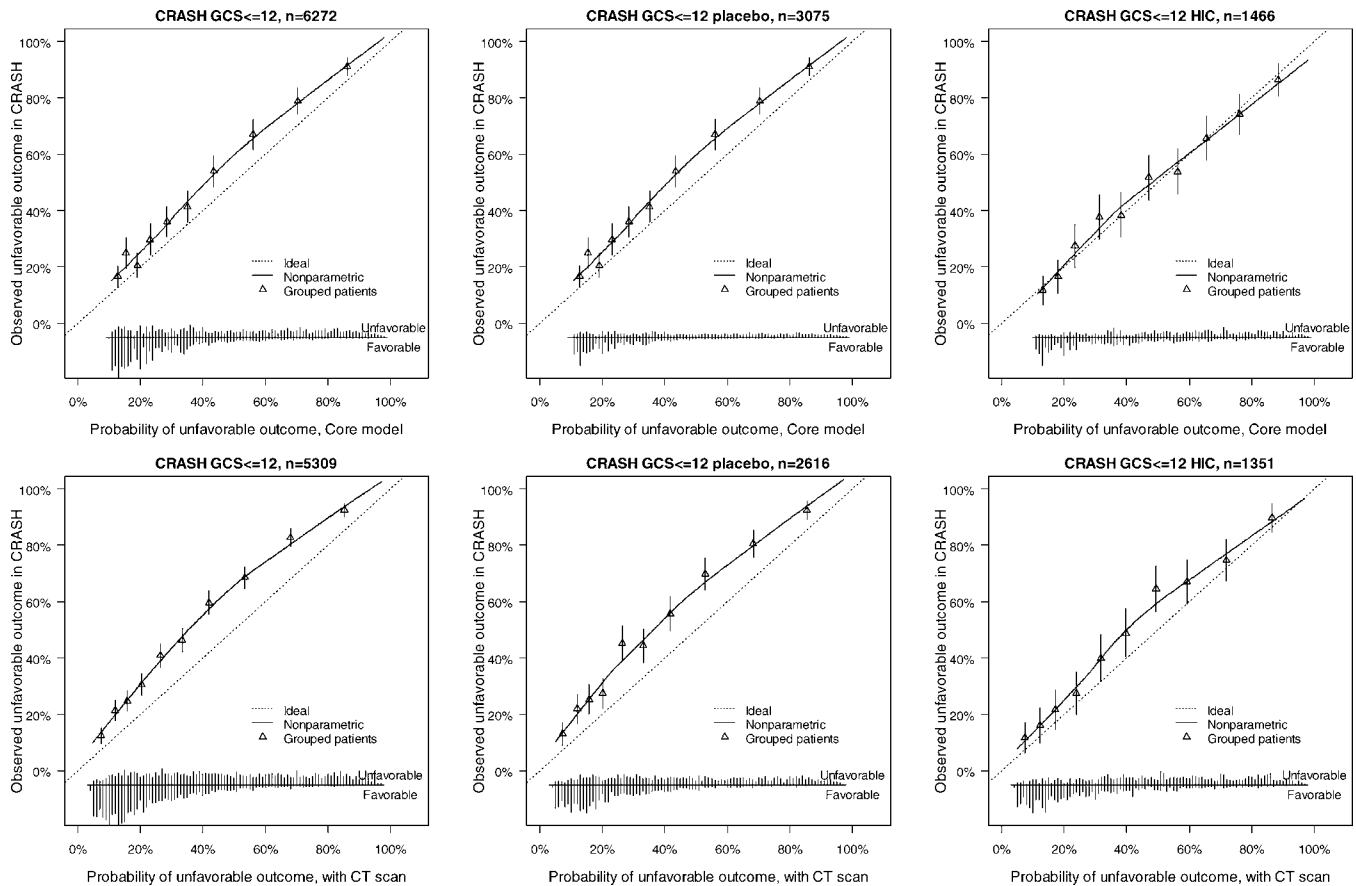doi:10.1371/journal.pmed.0050165.g004

Figure 5. External Validity for the Core and Core + CT Model Characteristics for Prediction of Unfavorable Outcomes in the CRASH Trial
The distribution of predicted probabilities is shown at the bottom of the graphs, by 6-mo outcome. The triangles indicate the observed frequencies by decile of predicted probability.
doi:10.1371/journal.pmed.0050165.g005

trials [40]. In the CRASH trial, outcomes were better for patients from the high-income countries [14]. One explanation is that facilities were more extensive than in the low- and middle-income countries that participated in this trial. Predictions for TBI patients in low- and middle-income countries may best be obtained from the CRASH models that were specifically developed for these countries [14]. Our model predictions may be better than the CRASH predictions for high-income countries, because of the more detailed information in the models and larger patient numbers used in model development. Predictions may, even on average, be too poor, considering that treatment standards have improved over time, including trauma organization, diagnostic facilities such as CT scanning, and critical care management. We did not, however, find a clear trend of better outcomes in more recently treated patients when we applied identical selection criteria to the studies in the IMPACT database. Both the CRASH and IMPACT model predictions may require regular updating according to specific population characteristics, such as calendar year, treatment setting, or local trauma organization [41,42].

Limitations of These Models

Our study has several limitations. Patients in our studies were treated between 1984 and 1997. Even though evaluation

in the more recent CRASH trial data confirmed the validity of the model predictions to more recent times (enrollment between 1999 and 2004), we cannot exclude that better outcomes are obtained nowadays. Also, the motor score is not always available in current clinical practice, or can be unreliable even when it is available, due to the effects of early sedation or paralysis. Furthermore, missing variables and missing values were a problem in the development of the models. Multiple imputation of the relatively few missing values allowed us to use the information from other predictors. Both theoretical and empirical support is growing for the use of such imputation methods instead of traditional complete case analyses [19,43]. However, more complete data would have been preferable. Furthermore, some misclassification may have occurred in classification of unfavorable versus favorable outcome. Mortality at 6 mo has the advantage that it suffers less from such a potential bias.

In conclusion, prognostic models are now available that provide adequate discrimination between patients with good and poor 6-mo outcome. These models may be useful for providing realistic information to relatives on expectations of outcome, for quantifying and classifying the severity of brain injury, for stratification and covariate adjustment in clinical trials, and as a reference for evaluating quality of care.

Supporting Information

Text S1. Excel File That Can Be Used to Calculate Predictions with Increasingly Complex Models

Found at doi:10.1371/journal.pmed.0050165.sd001 (185 KB XLS).

Acknowledgments

References

1. Teasdale G, Jennett B (1974) Assessment of coma and impaired consciousness. A practical scale. Lancet 2: 81–84.
2. Jennett B, Bond M (1975) Assessment of outcome after severe brain damage. Lancet 1: 480–484.
3. Jennett B, Teasdale G, Braakman R, Minderhoud J, Knill-Jones R (1976) Predicting outcome in individual patients after severe head injury. Lancet 1: 1031–1034.
4. Machado SG, Murray GD, Teasdale GM (1999) Evaluation of designs for clinical trials of neuroprotective agents in head injury. European Brain Injury Consortium. J Neurotrauma 16: 1131–1138.
5. Hernandez AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JD, et al. (2005) Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review. Neurosurgery 57: 1244–1253.
6. Hernandez AV, Steyerberg EW, Butcher I, Mushkudiani N, Taylor GS, et al. (2006) Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study. J Neurotrauma 23: 1295–1303.
7. Perel P, Edwards P, Wentz R, Roberts I (2006) Systematic review of prognostic models in traumatic brain injury. BMC Med Inform Decis Mak 6: 38.
8. Mushkudiani NA, Hukkelhoven CW, Hernandez AV, Murray GD, Choi SC, et al. (2008) A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. J Clin Epidemiol 61: 331–343.
9. Justice AC, Covinsky KE, Berlin JA (1999) Assessing the generalizability of prognostic information. Ann Intern Med 130: 515–524.
10. Reilly BM, Evans AT (2006) Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med 144: 201–209.
11. Maas AI, Marmarou A, Murray GD, Teasdale SG, Steyerberg EW (2007) Prognosis and clinical trial design in traumatic brain injury: the IMPACT study. J Neurotrauma 24: 232–238.
12. Murray GD, Butcher I, McHugh GS, Lu J, Mushkudiani NA, et al. (2007) Multivariable prognostic analysis in traumatic brain injury: results from the IMPACT study. J Neurotrauma 24: 329–337.
13. Edwards P, Arango M, Balica L, Cottingham R, El-Sayed H, et al. (2005) Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months. Lancet 365: 1957–1959.
14. MRC CRASH Trial Collaborators, Perel P, Arango M, Clayton T, Edwards P, et al. (2008) Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. BMJ 336: 425–429.
15. Marmarou A, Lu J, Butcher I, McHugh GS, Mushkudiani NA, et al. (2007) IMPACT database of traumatic brain injury: design and description. J Neurotrauma 24: 239–250.
16. Marshall LF, Marshall SB, Klauber MR, Van Berkum Clark M, Eisenberg H, et al. (1992) The diagnosis of head injury requires a classification based on computed axial tomography. J Neurotrauma 9: S287–292.
17. McHugh GS, Butcher I, Steyerberg EW, Lu J, Mushkudiani N, et al. (2007) Statistical approaches to the univariate prognostic analysis of the IMPACT database on traumatic brain injury. J Neurotrauma 24: 251–258.
18. Rubin DB, Schenker N (1991) Multiple imputation in health-care databases: an overview and some applications. Stat Med 10: 585–598.
19. Schafer JL, Graham JW (2002) Missing data: our view of the state of the art. Psychol Methods 7: 147–177.
20. Enders CK (2006) A primer on the use of modern missing-data methods in psychosomatic medicine research. Psychosom Med 68: 427–436.
21. Van Buuren S, Oudshoorn CGM (2006) mice: Multivariate Imputation by Chained Equations. R package version 1.16. Available: http://web.inter.nl.net/users/S.vanBuuren/mi/hmtl/mice.htm. Accessed 12 December 2006.
22. R Development Core Team (2006) R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available: http://www.R-project.org/. Accessed 30 June 2008.
23. McCullagh P (1980) Regression models for ordinal data. J R Stat Soc Ser B 42: 109–142.
24. Harrell FE (2001) Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer. 568 pp.
25. Harrell FE (2006) Design: Design Package. R package version 2.0. http://biostat.mc.vanderbilt.edu/s/Design. Accessed 30 June 2008.
26. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD (2005) Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 58: 475–483.
27. Moons KG, Harrell FE, Steyerberg EW (2002) Should scoring rule be based on odds ratios or regression coefficients? J Clin Epidemiol 55: 1054–1055.
28. Maas AI, Hukkelhoven CW, Marshall LF, Steyerberg EW (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.
29. Van Beek JG, Mushkudiani NA, Steyerberg EW, Butcher I, McHugh GS, et al. (2007) Prognostic value of admission laboratory parameters in traumatic brain injury: results from the IMPACT study. J Neurotrauma 24: 315–328.
30. van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, et al. (2001) Intensive insulin therapy in the critically ill patients. N Engl J Med 345: 1359–1367.
31. Signorini DF, Andrews PJ, Jones PA, Wardlaw JM, Miller JD (1999) Predicting survival using simple clinical variables: a case study in traumatic brain injury. J Neurol Neurosurg Psychiatry 66: 20–25.
32. Schreiber MA, Aoki N, Scott BG, Beck JR (2002) Determinants of mortality in patients with severe blunt head injury. Arch Surg 137: 285–290.
33. Lemeshow S, Le Gall JR (1994) Modeling the severity of illness of ICU patients. A systems update. JAMA 272: 1049–1055.
34. Cho DY, Wang YC (1997) Comparison of the APACHE III, APACHE II and Glasgow coma scale in acute head injury for prediction of mortality and functional outcome. Intensive Care Med 23: 77–84.
35. Hyam JA, Welch CA, Harrison DA, Menon DK (2006) Case mix, outcomes and comparison of risk prediction models for admissions to adult, general and specialist critical care units for head injury: a secondary analysis of the ICNARC Case Mix Programme Database. Crit Care 10: S2.
36. Choi SC (1998) Sample size in clinical trials with dichotomous endpoints: use of covariables. J Biopharm Stat 8: 367–375.
37. Murray GD, Barer D, Choi S, Fernandes H, Gregson B, et al. (2005) Design and analysis of phase III trials with ordered outcome scales: the concept of the sliding dichotomy. J Neurotrauma 22: 511–517.
38. Lemeshow S, Klar J, Teres D (1995) Outcome prediction for individual intensive care patients: useful, misused, or abused? Intensive Care Med 21: 770–776.
39. Murray LS, Teasdale GM, Murray GD, Jennett B, Miller JD, et al. (1993) Does prediction of outcome alter patient management? Lancet 341: 1487–1491.
40. Hukkelhoven CW, Steyerberg EW, Farace E, Habbema JD, Marshall LF, et al. (2002) Regional differences in patient characteristics, case management, and outcomes in traumatic brain injury: experience from the tirilazad trials. J Neurosurg 97: 549–557.
41. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD (2004) Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med 23: 2567–2586.
42. Steyerberg EW (2005) Local applicability of clinical and model-based probability estimates. Med Decis Making 25: 678–680.
43. Moons KG, Donders RA, Stijnen T, Harrell FE Jr. (2006) Using the outcome for imputation of missing predictor values was preferred. J Clin Epidemiol 59: 1092–1101.

Editors' Summary

Background. Traumatic brain injury (TBI) causes a large amount of morbidity and mortality worldwide. According to the Centers for Disease Control, for example, about 1.4 million Americans will sustain a TBI—a head injury—each year. Of these, 1.1 million will be treated and released from an emergency department, 235,000 will be hospitalized, and 50,000 will die. The burden of disease is much higher in the developing world, where the causes of TBI such as traffic accidents occur at higher rates and treatment may be less available.

Why Was This Study Done? Given the resources required to treat TBI, a very useful research tool would be the ability to accurately predict on admission to hospital what the outcome of a given injury might be. Currently, scores such as the Glasgow Coma Scale are useful to predict outcome 24 h after the injury but not before.

Prognostic models are useful for several reasons. Clinically, they help doctors and patients make decisions about treatment. They are also useful in research studies that compare outcomes in different groups of patients and when planning randomized controlled trials. The study presented here is one of a number of analyses done by the IMPACT research group over the past several years using a large database that includes data from eight randomized controlled trials and three observational studies conducted between 1984 and 1997. There are other ongoing studies that also seek to develop new prognostic models; one such recent study was published in *BMJ* by a group involving the lead author of the *PLoS Medicine* paper described here.

What Did the Researchers Do and Find? The authors analyzed data that had been collected prospectively on individual patients from the 11 studies included in the database and derived models to predict mortality and unfavorable outcome at 6 mo after injury for the 8,509 patients with severe or moderate TBI. They found that the strongest predictors of outcome were age, motor score, pupillary reactivity, and characteristics on the CT scan, including the presence of traumatic subarachnoid hemorrhage. A core prognostic model could be derived from the combination of age, motor score, and pupillary reactivity. A better score could be obtained by adding CT characteristics, secondary problems (hypotension and hypoxia), and laboratory measurements of glucose and hemoglobin. The scores were then tested to see how well they predicted outcome in a different group of patients—6,681 patients from the recent Medical Research Council Corticosteroid Randomisation after Significant Head Injury (MRC CRASH) trial.

What Do These Findings Mean? In this paper the authors show that it is possible to produce prognostic models using characteristics collected on admission as part of routine care that can discriminate between patients with good and poor outcomes 6 mo after TBI, especially if the results from CT scans and laboratory findings are added to basic models. This paper has to be considered together with other studies, especially the paper mentioned above, which was recently published in the *BMJ* (MRC CRASH Trial Collaborators [2008] Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *BMJ* 336: 425–429.). The *BMJ* study presented a set of similar, but subtly different models, with specific focus on patients in developing countries; in that case, the patients in the CRASH trial were used to produce the models, and the patients in the IMPACT database were used to verify one variant of the models. Unfortunately this related paper was not disclosed to us during the initial review process; however, during *PLoS Medicine*'s subsequent consideration of this manuscript we learned of it. After discussion with the reviewers, we took the decision that the models described in the *PLoS Medicine* paper are sufficiently different from those reported in the other paper and as such proceeded with publication of the paper. Ideally, however, these two sets of models would have been reviewed and published side by side, so that readers could easily evaluate the respective merits and value of the two different sets of models in the light of each other. The two sets of models are, however, discussed in a Perspective article also published in *PLoS Medicine* (see below).

Additional Information. Please access these Web sites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.0050165.

- This paper and the *BMJ* paper mentioned above are discussed further in a PLoS *Medicine* Perspective article by Andrews and Young
- The TBI Impact site provides a tool to calculate the scores described in this paper
- The CRASH trial, which is used to validate the scores mentioned here, has a Web site explaining the trial and its results
- The R software, which was used for the prognostic analyses, is freely available
- The MedlinePlus encyclopedia has information on head injury
- The WHO site on neurotrauma discusses head injury from a global perspective
- The CDC's National Center for Injury Prevention and Control gives statistics on head injury in the US and advice on prevention