Prediction of Outcome after Moderate and Severe Traumatic Brain Injury: External Validation of the IMPACT and CRASH Prognostic Models

Bob Roozenbeek, MD[1,2], **Hester F. Lingsma, PhD**[1], **Fiona E. Lecky, MD, PhD**[3,4], **Juan Lu, MD**[5], **James Weir, MSc**[6], **Isabella Butcher, PhD**[6], **Gillian S. McHugh, MSc**[6], **Gordon D. Murray, PhD**[6], **Pablo Perel, PhD**[7], **Andrew I.R. Maas, MD, PhD**[2], and **Ewout W. Steyerberg, PhD**[1] on behalf of the International Mission on Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury (IMPACT) Study Group, the Corticosteroid Randomisation After Significant Head Injury (CRASH) Trial Collaborators, and the Trauma Audit and Research Network (TARN)

[1]Department of Public Health, Center for Medical Decision Making, Erasmus MC, Rotterdam, The Netherlands [2]Department of Neurosurgery, Antwerp University Hospital and University of Antwerp, Edegem, Belgium [3]Trauma Audit and Research Network, Manchester, United Kingdom [4]University of Manchester / Salford Royal Hospitals NHS Foundation Trust, Manchester Medical Academic Health Sciences Centre, Manchester, United Kingdom [5]Department of Neurosurgery, Virginia Commonwealth University, Richmond, Virginia, USA [6]Centre for Population Health Sciences, University of Edinburgh, Scotland [7]Nutrition and Public Health Intervention Department, London School of Hygiene and Tropical Medicine, London, United Kingdom

Abstract

Objective—The International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) prognostic models predict outcome after traumatic brain injury (TBI) but have not been compared in large datasets. The objective of this is study is to validate externally and compare the IMPACT and CRASH prognostic models for prediction of outcome after moderate or severe TBI.

Design—External validation study.

Patients—We considered 5 new datasets with a total of 9036 patients, comprising three randomized trials and two observational series, containing prospectively collected individual TBI patient data.

Measurements—Outcomes were mortality and unfavourable outcome, based on the Glasgow Outcome Score (GOS) at six months after injury. To assess performance, we studied the discrimination of the models (by AUCs), and calibration (by comparison of the mean observed to predicted outcomes and calibration slopes).

Main Results—The highest discrimination was found in the TARN trauma registry (AUCs between 0.83 and 0.87), and the lowest discrimination in the Pharmos trial (AUCs between 0.65 and 0.71). Although differences in predictor effects between development and validation populations were found (calibration slopes varying between 0.58 and 1.53), the differences in

Correspondence: Bob Roozenbeek, MD, Department of Neurosurgery, Antwerp University Hospital, Wilrijkstraat 10, 2650 Edegem, Belgium, Phone: +32-3-8214632, b.roozenbeek@erasmusmc.nl.

discrimination were largely explained by differences in case-mix in the validation studies. Calibration was good, the fraction of observed outcomes generally agreed well with the mean predicted outcome. No meaningful differences were noted in performance between the IMPACT and CRASH models. More complex models discriminated slightly better than simpler variants.

Conclusions—Since both the IMPACT and the CRASH prognostic models show good generalizability to more recent data, they are valid instruments to quantify prognosis in TBI.

Keywords

external validation; outcome; prediction models; traumatic brain injury

INTRODUCTION

Traumatic brain injury (TBI) remains the main cause of death and disability in young adults worldwide (1) (2). It is a heterogeneous disease with respect to cause, pathology, severity and prognosis. This causes considerable uncertainty in the expected outcome of individual patients. Prognostic models can be used to combine different characteristics of an individual patient to predict outcome. Reliable outcome predictions can be used to provide realistic information to relatives, for more efficient design and analysis of clinical trials in TBI and can provide a reference for assessing the quality of health care delivery (3).

Many prognostic models for TBI have been developed in the past decades, with varying methodological quality (4) (5). Only two sets of prognostic models were developed on large datasets and according to the latest methodological insights. These are the models developed on the International Mission on Prognosis and Analysis of Clinical trials in Traumatic brain injury database (IMPACT models) and the Corticosteroid Randomisation After Significant Head Injury trial data (CRASH models) (6) (7).

Prognostic models are mathematical models developed from specific populations. External validation is necessary to determine the ability of a model to reliably predict outcome in populations, different from the development setting. Both the IMPACT models and the CRASH models have been externally validated previously, using each other's development datasets, indicating satisfactory generalizability. Nevertheless, a broader validation and comparison of these models has not yet been performed (8).

In this study we aimed to validate externally the IMPACT and CRASH prognostic models in five relatively recent datasets and to compare these different models to predict mortality and unfavourable outcome.

MATERIALS AND METHODS

Models

Details of the development of the prognostic models were reported previously (6) (7). In summary, the IMPACT models were developed on a dataset with 8509 patients from eleven studies (i.e., three surveys and eight randomized controlled trials) on moderate and severe TBI (9). Three different models were developed for the prediction of mortality and unfavourable outcome (Death, Vegetative State, Severe Disability), based on the dichotomized Glasgow Outcome Score (GOS) at six months after injury. These included the Core Model, the Extended Model and the Lab Model (Table 1). External validation was performed in patients with moderate to severe TBI enrolled in the CRASH trial dataset (n=6681).

The CRASH models were developed on the CRASH trial dataset (n=10008) with patients who sustained mild, moderate and severe TBI, mostly from low and middle-income countries (n=7526, 75%) (10) (11). Two different models were developed for both the prediction of mortality at two weeks after injury and unfavourable outcome at six months, using the dichotomized GOS. The models included the Basic Model and the more extensive CT Model (Table 1). The IMPACT dataset (n=8509) was used for external validation.

Because the original CRASH models were developed for prediction of outcome in all TBI severity groups (including mild: GCS scores 13–15) and not all variables of the CRASH models were available in the validation datasets (see below), we refitted an adapted version of the CRASH models in a selection of 6681 patients with moderate to severe TBI (GCS 12), 14 years of age and with available outcome data. In the refitted models, the total GCS score was replaced by the GCS motor score. Further, the Marshall CT classification and the presence of traumatic subarachnoid hemorrhage were included in the refitted CRASH CT model, instead of the original CT variables (Table 1). New regression coefficients and intercepts were obtained for both the CRASH Basic and the CRASH CT models for prediction of mortality at two weeks and unfavourable outcome at six months (Supplemental Digital Content, Appendix 2).

Datasets for external validation

NABIS Hypothermia—The NABIS Hypothermia study (1994–1998) was a multicenter randomized trial performed in the United States investigating the effect of hypothermia in severe TBI (12). The primary outcome was the GOS at six months after injury. The trial was terminated after inclusion of 392 of the planned 500 patients, because an interim analysis showed that the probability of detecting a treatment effect after expansion to 500 included patients was smaller than 1%. For the validation, we excluded patients younger than 14 years of age or with missing outcome (n=3), resulting in a dataset of 385 patients.

Cerestat—Cerestat is an unpublished multicenter randomized controlled trial in a combined group of North American and European centers investigating the effect of Aptiganel HCl, a non-competitive NMDA (*N*-methyl-*D*-aspartate) receptor antagonist, in severe TBI (GCS 8). In 1996 and 1997, 547 patients were enrolled in the trial. An interim analysis performed at the end of 1997 resulted in the decision that continuation of the trial was unjustified (13). No benefit of the study drug was found in TBI patients. As outcome was missing in 30 patients, a total of 517 were available for validation.

APOE—This single-center observational cohort study (1996–1999, Glasgow, Scotland) investigated the hypothesis that the possession of the APOE e4 allele is associated with a poor outcome after TBI (14). In total, 1094 patients were enrolled, of whom 513 (54%) had a mild TBI (GCS 13–15). The outcome was the GOS at six months after injury. No overall association between the APOE genotype and outcome was found. To be consistent, we excluded patients with GCS > 12 and age < 14, resulting in a dataset with 404 patients.

Pharmos—The Pharmos Dexanabinol study (2001–2004) was a randomized multicenter, placebo-controlled, phase III trial investigating the safety and efficacy of dexanabinol in patients with severe TBI carried out in Europe, Israel, Australia and the United States (15). The study enrolled 861 subjects. The study publication reported dexanabinol to be safe but not efficacious in the treatment of TBI. Because of missing 6-month outcome, five patients were excluded from our analyses, resulting in a dataset of 856 patients.

TARN—The Trauma Audit and Registry Network (TARN) is a hospital-based trauma registry in England and Wales including all patients with trauma resulting in immediate

admission to hospital for three days or longer or death. The outcome measure is in-hospital mortality. For the validation, we selected patients 14 years of age enrolled between 2001 and 2009 with moderate or severe TBI (GCS 12), defined as having an Abbreviated Injury Scale Head of 3 or higher, which was not resulting from scalp laceration, avulsion or penetrating brain injury, with outcome data available, resulting in a dataset of 6874 patients.

Data analysis

The IMPACT and CRASH models (Table 1) were validated in these five validation datasets, all originated from high income countries (Europe, North-America and Australia). Some models could not be applied to all datasets, because not all required predictor or outcome variables were available. Coding of the different predictors in the validation datasets was according to the IMPACT predictor coding as described previously (9). Data on age and GCS motor score, pupillary reactivity, hypoxia and hypotension were available for all validation datasets. However, in TARN data on pupils were only collected in the latter years of inclusion, resulting in a large number of missing values with only 693 pupils available for 6874 patients, 10%. Missing values occurred for several other predictors, mainly because some predictors were not collected in all studies. Table 2 includes information on the total number of patients included in the dataset and on the availability of the parameters of interest in these datasets. In TARN, some CT data were available (EDH and tSAH), but the Marshall CT Classification was not scored directly, and therefore derived with an algorithm using the AIS as a basis (16). Missing values were assumed to be missing at random and were statistically imputed. For the imputation procedure, the *aregImpute* function in R software was used (17).

To quantify differences in the distribution of patient characteristics between the different studies, we calculated a summary measure of heterogeneity. This measure is defined as the standard deviation (SD) of the linear predictor (LP) values of a logistic regression model, with mortality as the outcome variable and age, GCS motor score and pupillary reactivity as predictor variables. The LP value is directly related to the predicted probability of mortality. In a relatively homogeneous population the predicted probabilities - and thus LP values - for all patients will be relatively close together, with a small SD. In a more heterogeneous population, the LPs will be more dispersed and the SD is larger.

Model performance

We assessed the performance of the models in the new datasets in terms of discrimination and calibration. Discrimination describes how well a model distinguishes between patients with and without the outcome of interest. Calibration indicates how closely predicted outcomes match observed outcomes.

To assess discrimination, we calculated the area under the receiver operating characteristic curve (AUC). An AUC of 1 implies perfect discrimination, whereas an AUC of 0.5 implies that a model's discrimination is no better than chance. Discrimination in an external validation setting can be influenced in several ways. The predictor effects in the model can be different, resulting in an invalid model. But differences in the distribution of patient characteristics ("case-mix") between the development and validation data may also influence discriminative power of the model. In a population with a prognostically homogeneous case-mix, it will be more difficult to distinguish between patients with a good or poor outcome with, than in a heterogeneous population. For example, in an observational study with broad inclusion criteria, discrimination of a model will be better than in a strictly selected trial population.

To take these effects into account, we calculated two recently proposed benchmark values for the AUC: the case-mix-corrected and the refitted AUC (18). The case-mix-corrected AUC indicates the discriminative power of a model, under the assumption that the predictor effects are fully correct for the validation population. It was calculated by simulating new outcome values for all patients in the validation dataset, based on the predicted risks for each patient calculated by the prognostic model. We performed 1000 repetitions to obtain stable estimates. The refitted AUC indicates the discriminative power that can be obtained by refitting the model in the validation dataset, resulting in new, optimal predictor effects. This resulted in an upper bound for the discriminative power, which would be found if the predictor effects in the validation dataset were exactly the same as in the development population.

To assess calibration, we plotted observed versus predicted outcome. We assessed calibration-in-the-large by fitting a logistic regression model with the model predictions as an offset variable. The intercept indicates whether predictions are systematically too low or too high, and should ideally be zero. The calibration slope reflects the average effects of the predictors in the model and was estimated in a logistic regression model with the logit of the model predictions as the only predictor. For a perfect model, the slope is equal to 1.

The current study is part of the IMPACT project that was exempt from institutional review board approval. The analyses were performed using SPSS Statistics (Version 17.0.2, IBM Corporation, Somers, NY, USA) and the R software environment (Version 2.7.1, The R Foundation for Statistical Computing, Vienna, Austria) with the *Design* and *Hmisc* packages.

RESULTS

Datasets

The observational series included a more heterogeneous case-mix than the trial populations (Table 2). This was illustrated by the summary measure for heterogeneity: the strictly selected trial populations had a lower value (e.g., Pharmos: 0.58) than the observational series (e.g., TARN: 1.75).

The distributions of the Glasgow Outcome Scale score for both the IMPACT and CRASH development datasets were U-shaped (Figure 1). In NABIS, APOE and Pharmos, the U-shaped outcome distribution had largely disappeared because a higher proportion of patients had severe disability: 26% in both NABIS and APOE and 29% in Pharmos, compared to 16% in IMPACT and 15% in CRASH. This resulted in a greater proportion of patients with unfavourable outcome. In IMPACT and CRASH the proportions of unfavourable outcome were 48% and 47% respectively, while in NABIS this was 56%, in APOE 53% and in Pharmos 51%.

Model performance

Discrimination—Discrimination of the different models varied between the datasets (Table 3). For prediction of mortality, AUCs varied between 0.65 and 0.83 for the IMPACT Core model and between 0.66 and 0.85 for the CRASH Basic model. For prediction of unfavourable outcome, AUCs between 0.66 and 0.76 were found for the IMPACT Core model and between 0.68 and 0.76 for the CRASH Basic model. Discrimination improved for both the IMPACT and CRASH models with increasing complexity of the models. Discrimination was better in the observational datasets (TARN and APOE) than in the trials, which was largely explained by a more heterogeneous case-mix, compared to the development population (small difference between development AUC and case-mix-corrected AUC, Table 3). When the model was refitted in the validation data, the

performance was similar to the performance of the externally validated original model (small difference between external validation AUC and refitted AUC).

Calibration—The calibration slope and intercept values of the validation datasets and models are summarized in Table 3.

In Pharmos, the calibration plots indicated some degree of miscalibration (Figure 2a and 2b). This was also expressed in the calibration slopes, indicating differences in predictor effects between the development and validation populations. For the IMPACT Core model the slopes were 0.733 for mortality and 0.634 for unfavourable outcome. The CRASH Basic model showed calibration slopes of 0.641 (mortality) and 0.645 (unfavourable outcome). In TARN, miscalibration was less apparent in the calibration plots. Calibration slopes were 1.318 (IMPACT Core model) and 1.375 (CRASH Basic model).

With regard to calibration-in-the-large, in Pharmos, an intercept of −0.774 was found for mortality prediction with the IMPACT Core model, indicating that the model's predictions were moderately higher than the observed mortality. However, prediction of unfavourable outcome with the IMPACT Core model showed an intercept of 0.286, indicating mild underestimation of the outcome by the model. A similar pattern was observed for the CRASH Basic model. In TARN, the IMPACT Core model gave an intercept of 0.124 and the CRASH Basic model of 0.110 for prediction of mortality.

Sensitivity analysis—Because of the high proportion of patients with missing data on pupillary reactivity in the TARN data, we also performed the validation in TARN patients with available data on pupillary reactivity (n=693). Compared to the validation with the imputed dataset, the results were very similar for both the IMPACT Models (Core: AUC=0.836, Slope=1.398, Intercept=1081; Extended: AUC=0.864, Slope=1.458, Intercept=0.621) and the CRASH Models (Basic: AUC=0.850, Slope=1.333, Intercept=−0.006; CT: AUC=0.863, Slope=1.330, Intercept=−0.031).

DISCUSSION

We confirmed the external validity of the IMPACT and CRASH prognostic models for prediction of outcome after moderate and severe TBI in five new datasets. Importantly, these datasets included both observational studies and clinical trials, thus permitting conclusions on generalizability. The predictive performance of the IMPACT and CRASH models was largely similar. Discrimination of the models varied between the different validation populations (clinical trials vs. observational studies). Overall, the more complex models performed better than the more simple models.

On average, performance measures did not substantially differ from those obtained in the development populations. Although developed on 6-month mortality or unfavourable outcome, the models also demonstrated good performance when predicting in-hospital mortality in the TARN registry.

These conclusions support the validity of these instruments to quantify prognostic risk in moderate and severe TBI.

Performance and generalizability of models

The CRASH and IMPACT prognostic models have been developed using state of the art approaches on large datasets, and were externally validated reciprocally. They therefore meet methodological quality standards (17). The IMPACT models have however been criticized for having in part been developed on older datasets, and the CRASH models for

having been developed exclusively on data from a clinical trial – albeit one with broad enrollment criteria. Establishing their validity for current practice and their generalizability to other settings is therefore important. Panczykowski et al reported good performance of the IMPACT models (AUC: 0.76–0.83) at validation on an unselected series of 587 patients with severe TBI admitted to a single level I trauma center (19). Honeybul et al applied the CRASH model to a selection 147 of 1786 TBI patients who underwent a decompressive craniectomy, to predict outcome at 18 months after injury (20). The authors noted an overestimation of the predicted risk of unfavourable outcome, predominantly when the predicted risk was < 80%. To our knowledge, no other validation studies on the IMPACT and CRASH models have been performed.

Assessing performance – and its interpretation – is a complex procedure, involving multiple levels. Many factors other than those directly related to the model itself may influence the performance of a model. These include changing epidemiology, trauma organization, different treatment policies, but also changes in approaches to outcome assessment. Such external factors could influence the validity of the predictor effects (regression coefficients) and the distribution of the predictor values in the new population (case-mix).

We observed clear effects of the case-mix on discrimination: higher AUCs were found in the observational series (TARN and APOE) compared to the RCTs. These higher AUCs reflect the less restrictive enrollment criteria: the greater the heterogeneity, the better the models can distinguish patients with very low or very high predictions. We therefore strongly discourage the indiscriminate comparisons of reported AUCs between different studies without taking the case-mix of the development and validation population into consideration. Use of the case-mix corrected AUC as proposed by Vergouwe et al would appear more appropriate, as illustrated by the decrease in AUC observed in the observational series versus the increase observed in the RCTs when calculating the case-mix-corrected AUC (Table 3) (18).

With regard to calibration, we did observe some indication for differences in predictor effects between the development and validation datasets. Particularly noteworthy is the discrepancy in the calibration slopes for the Pharmos and NABIS trials: the observed mortality is lower than predicted, whilst the number of patients with an unfavorable outcome is substantially higher than predicted. This discrepancy could be explained by a greater number of patients with Severe Disability than expected. Although some difference in predictor effects may be present, we consider the different outcome distribution (Figure 1) a more likely explanation. The main question then becomes what may have caused this change. It may be that approaches to outcome assessment, and in particular to the assignment of outcome categories, have changed. Pharmos and NABIS used the structured interview as a means to better standardize outcome assignment and in the Pharmos trial a central review of outcome assignments was performed (21). Wilson et al reported that central assignments were generally worse, causing a shift towards more severe disability (22). Our findings indicate that changes in the approach to assessment of the GOS, may influence both outcome distribution and predictor effects. Clarification of this issue is considered essential and requires a dedicated prospective study. The need for such a study is becoming more pertinent following the recent proposals by Lu et al for yet another new approach to assessing the extended GOS (23).

IMPACT or CRASH models?

We have often been asked the question which model should be preferred: CRASH or IMPACT? The current validation studies do not indicate any clear preference. Both models demonstrated comparable performance on external validation. A difference between the IMPACT and CRASH models is the inclusion of major extracranial injury (MEI) as a

predictor in the CRASH models. We, however, found that the added prognostic effect of MEI in our validation series was negligible (results not shown), which may have resulted from the case-mix of the validation datasets. In a recent meta-analysis of IMPACT, CRASH and TARN data, we found an interaction between MEI and the clinical severity: only small effects of MEI were noted in patients with severe TBI, whilst in mild TBI, the presence of extracranial injury had a strong prognostic effect (24).

Perhaps the most appropriate answer to the question which model should be preferred is to consider the setting and case-mix of the population under study. The IMPACT models were developed on patients with moderate and severe TBI, mostly from high-income countries. The development population for the CRASH models included also patients with mild TBI and a substantial number of patients from low and middle-income countries. Consequently, a preference may be expressed for the CRASH models when the population under study includes patients with mild TBI or patients from low and middle-income countries.

Limitations

Our study has some limitations. First, due to differences in recorded predictor and outcome variables, it was not possible to validate the original CRASH models. As described above, we refitted the CRASH models in a selection of the CRASH trial data (GCS 12) with a different definition of the CT variables and the GCS motor score instead of the full GCS score. This resulted in an adapted version of the CRASH models, making them more similar to the IMPACT models. Nevertheless, some aspects of the adapted CRASH models were still substantially different from the IMPACT models. These differences included the inclusion of the presence of major extracranial injury as a predictor in the CRASH models and the inclusion of second insults (hypoxia and hypotension) as predictors to the IMPACT Extended and Lab models. Further, the adapted CRASH models were developed to predict early (14-day) mortality and the IMPACT models late (6-month) mortality.

It was not possible to validate all models to all new datasets, since not all required predictor and outcome variables were recorded for each dataset. In TARN, the 6-month GOS was not recorded as an outcome measure and only the models predicting mortality were validated on these data. Although the outcome measure was in-hospital mortality, instead of 14-day or 6-month mortality, both IMPACT and CRASH models performed well.

In TARN, actual CT data were not available, but derived from extensive coding of the Abbreviated Injury Score (AIS) according to an algorithm proposed by Lesko et al. (16). The AIS codes were, however, not only based on data of imaging studies on admission, but also upon information of later imaging studies (CT and MRI), as well as findings during surgery. Servadei et al have previously demonstrated that the "worst" CT may demonstrate more abnormalities than the admission CT (25).

A further limitation was the high proportion of missing values on pupillary reactivity in the TARN dataset (90%). Since it would be a waste of information to exclude all patients with missing pupils (6181 patients with a total of 49496 data points), we decided to include these patients for the validation and to statistically impute the missing values with an imputation model. Both theoretical 15 and empirical support is growing for the use of such imputation methods, instead of traditional complete case analyses (17). As a sensitivity analysis, we also performed the validation in a complete case dataset of TARN, which showed very similar results, supporting the validity of the approaches taken. The lack of availability of some important predictors in some datasets and missing data in others illustrates the paramount importance of standardization of data collection and coding of variables as proposed in the common data elements initiative (26).

CONCLUSIONS

This external validation study confirms the validity of both the IMPACT and CRASH models in more recent datasets, with best performance in more heterogeneous observational series. Model performance improved slightly with increasing complexity of the models, but most information is captured in three key predictors: age, GCS motor score, and pupillary reactivity. No relevant difference in performance was found between the IMPACT and the CRASH models. These findings confirm generalizability of the models in new data and support the use of the models for a wide a range of applications. These include the classification of future TBI patient populations (especially exploiting the discriminative ability of the models) and calculation of the baseline prognostic risk of individual patients (requiring good calibration of the models).

It may be expected that the effects of treatment and health care organization on outcome of individual patients change over time. This underlines the necessity of re-validation of these prognostic models in the future, to re-confirm generalizability or to update the models on more recent patient populations. This should become a continuing process, and highlights the need for a prospective high quality observational study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Maas AI, Stocchetti N, Bullock R. Moderate and severe traumatic brain injury in adults. Lancet Neurol. 2008; 7:728–41. [PubMed: 18635021]

2. Ghajar J. Traumatic brain injury. Lancet. 2000; 356:923–9. [PubMed: 11036909]

3. Lingsma HF, Roozenbeek B, Steyerberg EW, et al. Early prognosis in traumatic brain injury: From prophecies to predictions. Lancet Neurol. 2010; 9:543–554. [PubMed: 20398861]

4. Perel P, Edwards P, Wentz R, et al. Systematic review of prognostic models in traumatic brain injury. BMC Med Inform Decis Mak. 2006; 6:38. [PubMed: 17105661]

5. Mushkudiani NA, Hukkelhoven CW, Hernandez AV, et al. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. J Clin Epidemiol. 2008; 61:331–343. [PubMed: 18313557]

6. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: Development and international validation of prognostic scores based on admission characteristics. PLoS Med. 2008; 5:e165. discussion e165. [PubMed: 18684008]

7. Perel P, Arango M, et al. MRC CRASH Trial Collaborators. Predicting outcome after traumatic brain injury: Practical prognostic models based on large cohort of international patients. BMJ. 2008; 336:425–429. [PubMed: 18270239]

8. Menon DK, Zahed C. Prediction of outcome in severe traumatic brain injury. Curr Opin Crit Care. 2009; 15:437–441. [PubMed: 19713837]

9. Marmarou A, Lu J, Butcher I, et al. IMPACT database of traumatic brain injury: Design and description. J Neurotrauma. 2007; 24:239–250. [PubMed: 17375988]

10. Roberts I, Yates D, Sandercock P, et al. Effect of intravenous corticosteroids on death within 14 days in 10008 adults with clinically significant head injury (MRC CRASH trial): Randomised placebo-controlled trial. Lancet. 2004; 364:1321–1328. [PubMed: 15474134]

11. Edwards P, Arango M, Balica L, et al. Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months. Lancet. 2005; 365:1957–1959. [PubMed: 15936423]

12. Clifton GL, Miller ER, Choi SC, et al. Lack of effect of induction of hypothermia after acute brain injury. N Engl J Med. 2001; 344:556–563. [PubMed: 11207351]

13. Miller, LP.; Hayes, R.; Newcomb, J., editors. Head Trauma: Basic, Pre-clinical and Clinical Directions. Hoboken, NJ: Wiley-Liss Inc; 2001.

14. Teasdale GM, Murray GD, Nicoll JA. The association between APOE epsilon4, age and outcome after head injury: A prospective cohort study. Brain. 2005; 128:2556–2561. [PubMed: 16033781]

15. Maas AI, Murray G, Henney H 3rd, et al. Efficacy and safety of dexanabinol in severe traumatic brain injury: Results of a phase III randomised, placebo-controlled, clinical trial. Lancet Neurol. 2006; 5:38–45. [PubMed: 16361021]

16. Lesko MM, Woodford M, White L, et al. Using abbreviated injury scale (AIS) codes to classify computed tomography (CT) features in the marshall system. BMC Med Res Methodol. 2010; 10:72. [PubMed: 20691038]

17. Steyerberg, EW. Clinical prediction models. Springer Science+Business Media; 2009.

18. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol. 2010; 172:971–980. [PubMed: 20807737]

19. Panczykowski D, Puccio A, Scruggs BJ, et al. Prospective Independent Validation of IMPACT Modeling as a Prognostic Tool in Severe Traumatic Brain Injury. J Neurotrauma. 2011 Epub ahead of print.

20. Honeybul S, Ho KM, Lind CR, et al. Observed versus predicted outcome for decompressive craniectomy: A population-based study. J Neurotrauma. 2010; 27:1225–1232. [PubMed: 20412014]

21. Wilson JT, Pettigrew LE, Teasdale GM. Structured interviews for the glasgow outcome scale and the extended glasgow outcome scale: Guidelines for their use. J Neurotrauma. 1998; 15:573–585. [PubMed: 9726257]

22. Wilson JT, Slieker FJ, Legrand V, et al. Observer variation in the assessment of outcome in traumatic brain injury: Experience from a multicenter, international randomized clinical trial. Neurosurgery. 2007; 61:123–8. discussion 128–9. [PubMed: 17621027]

23. Lu J, Marmarou A, Lapane K, et al. A method for reducing misclassification in the extended glasgow outcome score. J Neurotrauma. 2010; 27:843–852. [PubMed: 20334503]

24. van Leeuwen N, Lingsma HF, Perel P, et al. Prognostic Value of Major Extracranial Injury in Traumatic Brain Injury: An Individual Patient Data Meta-analysis in 39,274 Patients. Neurosurgery. 2011 Epub ahead of print.

25. Servadei F, Murray GD, Penny K, et al. The value of the "worst" computed tomographic scan in clinical studies of moderate and severe head injury. European Brain Injury Consortium. Neurosurgery. 2000; 46:70–5. discussion 75–7. [PubMed: 10626937]

26. Maas AI, Harrison-Felix CL, Menon D, et al. Standardizing data collection in traumatic brain injury. J Neurotrauma. 2011; 28:177–187. [PubMed: 21162610]
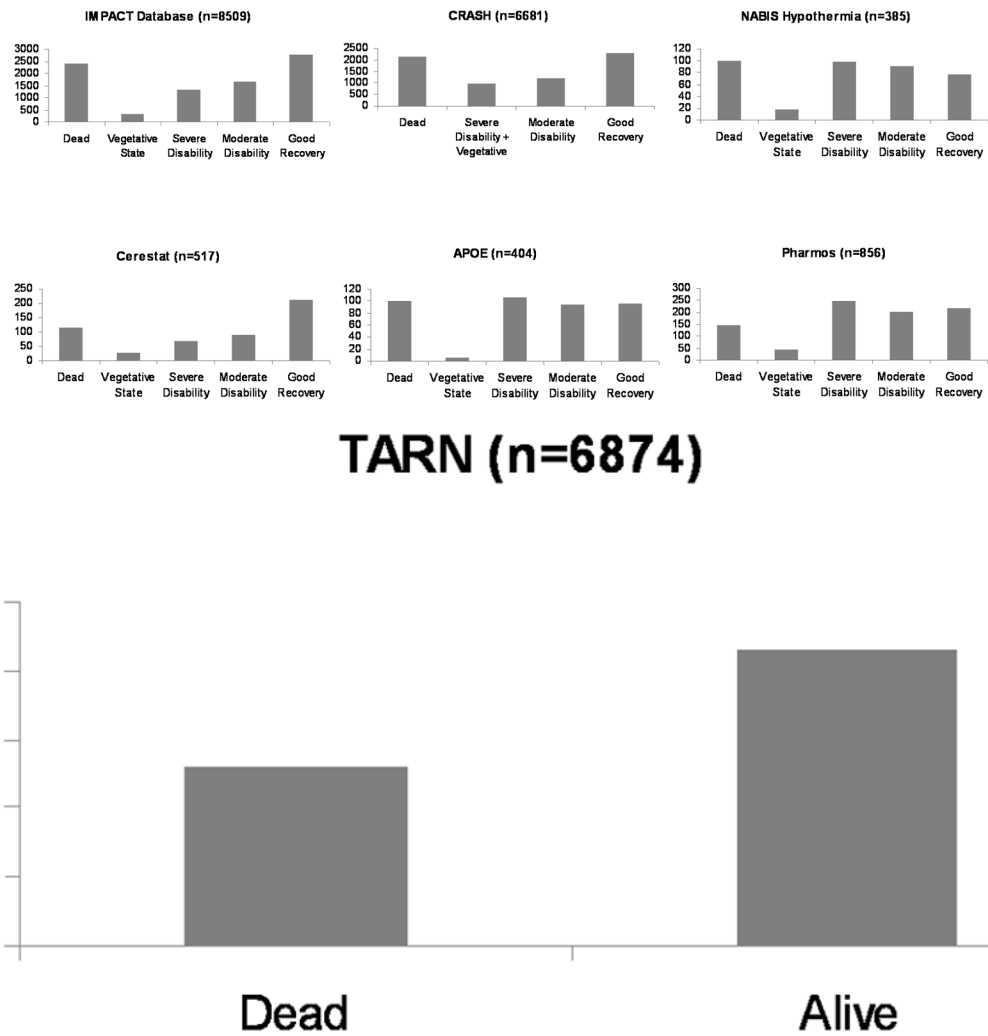
Figure 1.
Figure 1a. Six-month Glasgow Outcome Scale in development (IMPACT and CRASH) and validation datasets (NABIS Hypothermia, Cerestat, APOE and Pharmos).
Figure 1b. In-hospital mortality in the TARN-TBI registry.
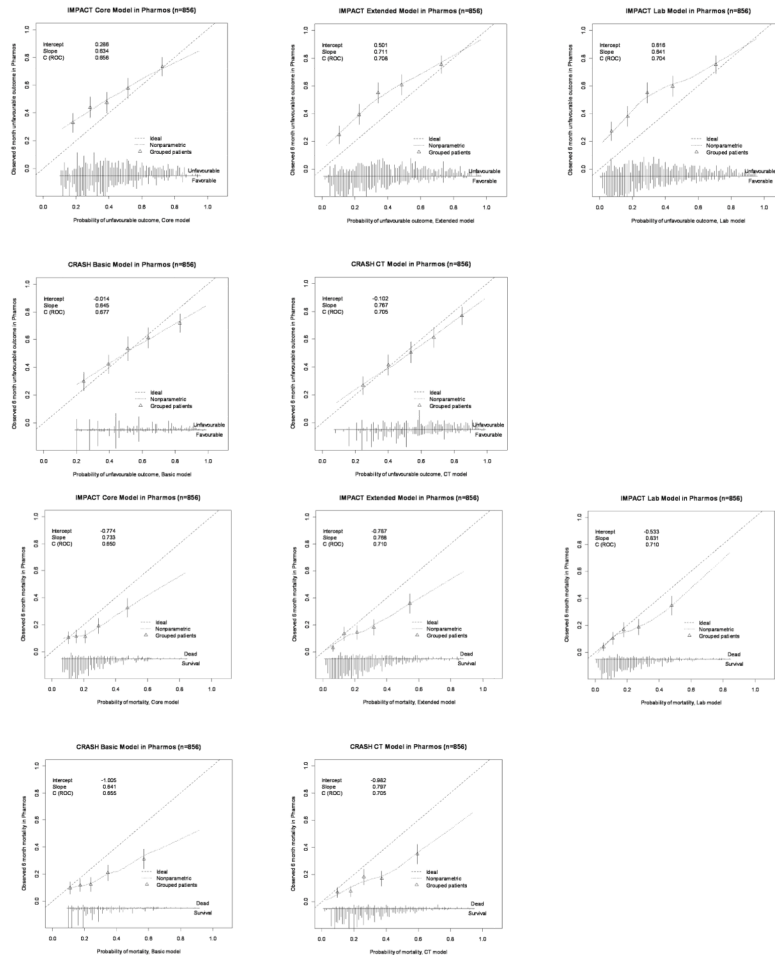
Figure 2.
Figure 2a. Calibration plots for external validation of the IMPACT and CRASH models for prediction of 6-month unfavourable outcome in the Pharmos trial data. Predicted probabilities are on the x-axis and observed outcomes on the y-axis. The triangles indicate the observed frequencies by quantiles of predicted probability with 95% confidence intervals (vertical lines). The distribution of the predicted probabilities is shown at the bottom of the graphs, separate for those with and without the outcome of interest.
Figure 2b. Calibration plots for external validation of the IMPACT and CRASH models for prediction of mortality in the Pharmos trial data.

IMPACT Core Model in TARN TBI (n=6874)

IMPACT Extended Model in TARN TBI (n=6874)

CRASH Basic Model in TARN TBI (n=6874)

CRASH CT Model in TARN TBI (n=6874)
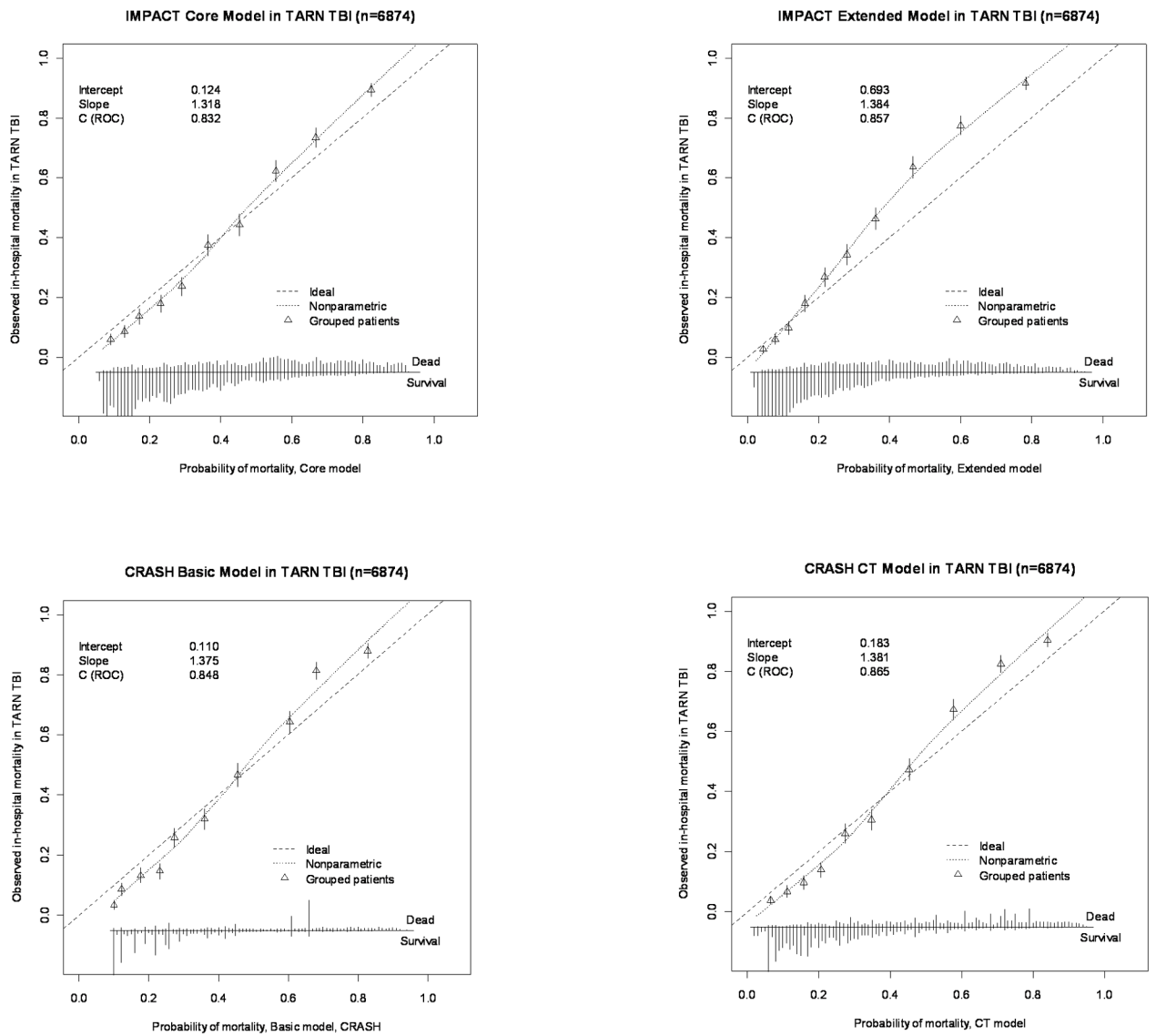
Figure 3.
Calibration plots for external validation of the IMPACT and CRASH models for prediction of mortality in the TARN TBI dataset.

Table 1

Variables included in the IMPACT models and adapted CRASH models.

IMPACT Models: predicting unfavourable outcome and mortality at six months		
Core Model	*Extended Model*	*Lab Model*
Age	Core Model variables +	Extended Model variables +
GCS motor score	CT Classification	Glucose
Pupillary reactivity	EDH	Hemoglobin
	tSAH	
	Hypoxia	
	Hypotension	

CRASH Models: predicting unfavourable outcome at six months and mortality at 14 days	
Basic Model	*CT Model*
Age	Basic Model variables+
GCS motor score	CT Classification
Pupillary reactivity	tSAH
Major extracranial injury	

IMPACT= International Mission on Prognosis and Analysis of Clinical Trials in TBI; GCS= Glasgow Coma Scale; CT= computed tomography; EDH= epidural hematoma; tSAH= traumatic subarachnoid hemorrhage; CRASH= Corticosteroid Randomisation After Significant Head Injury

Table 2

Patient characteristics of the IMPACT Database (used to develop the IMPACT models), the CRASH trial dataset (used to develop the CRASH models) and the NABIS Hypothermia, Cerestat, APOE, Pharmos, and TARN TBI datasets (used for external validation). All datasets were selected for age 14, GCS 12 and non-missing outcome data.

Characteristics	Measure or Category	IMPACT Database (n = 8509)	CRASH Trial (n = 6681)	NABIS (n = 385)	Cerestat (n = 517)	APOE (n = 404)	Pharmos (n = 856)	TARN TBI (n = 6874)
	Number of patients included in study							
Origin	Study design	RCT + OBS	RCT	RCT	RCT	OBS	RCT	OBS
Inclusion period		1984–1997	1999–2004	1994–1998	1996–1997	1996–1999	2001–2004	2001–2009
Age, years	Median (25–75 percentile)	30 (21–45)	32 (23–47)	30 (21–40)	28 (20–41)	39 (26–58)	33 (23–46)	39 (24–58)
Motor score of GCS	Total	8509 (100%)	6681 (100%)	385 (100%)	517 (100%)	404 (100%)	856 (100%)	6874 (100%)
	None (1)	1395 (16%)	785 (12%)	82 (21%)	40 (8%)	5 (1%)	43 (5%)	2047 (30%)
	Extension (2)	1042 (12%)	515 (8%)	62 (16%)	88 (17%)	2 (<1%)	91 (11%)	366 (5%)
	Abnormal flexion (3)	1085 (13%)	658 (10%)	55 (14%)	86 (16%)	18 (5%)	136 (16%)	395 (6%)
	Normal flexion (4)	1940 (23%)	1156 (17%)	76 (20%)	180 (35%)	32 (8%)	225 (26%)	668 (10%)
	Localizes/obeys (5/6)	2591 (30%)	3567 (53%)	100 (26%)	123 (24%)	16 (4%)	235 (27%)	2117 (25%)
	Untestable or missing (9)	456 (5%)	0	10 (3%)	0	331 (82%)	126 (15%)	1670 (24%)
Pupillary reactivity	Total	7126 (84%)	6272 (94%)	370 (96%)	426 (82%)	402 (>99%)	822 (96%)	693 (10%)
	Both pupils reactive	4486 (63%)	4956 (74%)	229 (62%)	302 (71%)	299 (74%)	642 (78%)	448 (65%)
	One pupil reactive	886 (12%)	530 (8%)	48 (13%)	72 (17%)	27 (7%)	147 (18%)	63(9%)
	No pupil reactive	1754 (25%)	786 (12%)	93 (25%)	52 (12%)	76 (19%)	33 (4%)	182(26%)
Hypoxia	Total	5452 (64%)	0	371 (96%)	478 (93%)	381 (94%)	856 (100%)	4772 (70%)
	Yes or expected	1116 (20%)	NA	124 (33%)	62 (13%)	162 (40%)	212 (25%)	321 (7%)
Hypotension	Total	6420(75%)	0	371 (96%)	506 (98%)	382 (95%)	855 (>99%)	5700 (83%)
	Yes or expected	1171 (18%)	NA	56 (15%)	85 (17%)	69 (17%)	132 (15%)	419 (6%)
CT Classification	Total	5192 (61%)	5654 (85%)	349 (91%)	0	0	849 (99%)	6874 (100%)
	No abnormalities (I)	360 (7%)	954 (17%)	4 (1%)	NA	NA	15 (2%)	482 (7%)
	Diffuse injury (II)	1838 (35%)	1517 (27%)	31 (9%)	NA	NA	412 (49%)	3237 (47%)

Characteristics	Measure or Category	IMPACT Database (n = 8509)	CRASH Trial (n = 6681)	NABIS (n = 385)	Cerestat (n = 517)	APOE (n = 404)	Pharmos (n = 856)	TARN TBI (n = 6874)
	Number of patients included in study							
	Diffuse injury plus obliteration basal cisterns (III)	863 (17%)	604 (11%)	193 (55%)	NA	NA	203 (24%)	1462 (21%) [a]
	Shift, no mass lesion (IV)	187 (4%)	133 (2%)	5 (2%)	NA	NA	56 (6%)	-
	Any mass lesion (V + VI)	1944 (38%)	2446 (43%)	116 (33%)	NA	NA	163 (19%)	1693 (25%)
Traumatic subarachnoid hemorrhage	Total	7393 (87%)	5653 (85%)	0	0	398 (99%)	856 (100%)	6873 (100%)
	Yes	3313 (45%)	2045 (36%)	NA	NA	141 (35%)	511 (60%)	2468 (36%)
Epidural hematoma	Total	7409 (87%)	0	0	511 (99%)	404 (100%)	856 (100%)	6874 (100%)
	Yes	999 (13%)	NA	NA	60 (12%)	38 (9%)	166 (19%)	1113 (16%)
Glucose (mmol/L)	Total	4830 (57%)	0	217 (56%)	0	0	838 (98%)	0
	Median (25–75 percentile)	8.2 (6.7–10.4)	NA	8.7 (7.4–11.4)	NA	NA	7.4 (6.2–9.1)	NA
Hemoglobin (g/dL)	Total	4376 (51%)	0	236 (61%)	0	0	846 (99%)	0
	Median (25–75 percentile)	12.7 (10.8–14.3)	NA	13.0 (11.6–14.3)	NA	NA	12.7 (11.1–14.1)	NA
Major extracranial injury	Total	0	6524 (98%)	385 (100%)	517 (100%)	0	856 (100%)	6874 (100%)
	Yes	NA	1735 (27%)	128 (33%)	275 (53%)	NA	440 (51%)	2716 (40%)
Summary measure for heterogeneity	Standard deviation of the LP [b]	1.06	1.09 [c]	0.88	1.04	1.36	0.58	1.75 [c]

[a] CT Classes III and IV were combined and scored as III in TARN TBI.

[b] Standard deviation of the linear predictors for the following logistic regression model: p(dead at 6-mo) = a + b_1· age + b_2· GCS motor + b_3· pupillary reactivity

[c] For this dataset 14-d mortality was used instead of 6-mo mortality as outcome for the model

CT, Computed Tomography; GCS, Glasgow Coma Scale; NA, not available.

Table 3

Performance of the IMPACT Core and CRASH Basic models for prediction of mortality and unfavourable outcome after traumatic brain injury in five new datasets.

Studies		Mortality			Unfavourable outcome	
		IMPACT Core	CRASH Basic		IMPACT Core	CRASH Basic
Development	AUC	0.766	0.776		0.779	0.784
NABIS (n=385)	AUC (external validation)	0.702	0.693		0.728	0.736
	Slope	0.838	0.670		0.890	0.804
	Intercept	−0.403	−0.652		0.424	0.040
	AUC (case-mix-corrected)	0.740	0.771		0.754	0.781
	AUC (refitted)	0.724	0.709		0.751	0.749
Cerestat (n=517)	AUC (external validation)	0.746	0.736		0.712	0.755
	Slope	1.047	0.889		0.815	0.947
	Intercept	−0.222	−0.651		−0.357	−0.725
	AUC (case-mix-corrected)	0.723	0.738		0.751	0.758
	AUC (refitted)	0.776	0.768		0.773	0.769
APOE (n=404)	AUC (external validation)	0.806	NA		0.758	NA
	Slope	1.534	NA		1.154	NA
	Intercept	−0.006	NA		0.144	NA
	AUC (case-mix-corrected)	0.728	NA		0.736	NA
	AUC (refitted)	0.810	NA		0.769	NA
Pharmos (n=856)	AUC (external validation)	0.650	0.655		0.656	0.677
	Slope	0.733	0.641		0.634	0.645
	Intercept	−0.774	−1.005		0.286	−0.014
	AUC (case-mix-corrected)	0.702	0.721		0.728	0.741
	AUC (refitted)	0.658	0.664		0.663	0.682
TARN TBI (n=6874)	AUC (external validation)	0.832	0.848		NA	NA

Studies		Mortality		Unfavourable outcome	
		IMPACT Core	CRASH Basic	IMPACT Core	CRASH Basic
	Slope	1.318	1.375	NA	NA
	Intercept	0.124	0.110	NA	NA
	AUC (case-mix-corrected)	0.781	0.784	NA	NA
	AUC (refitted)	0.842	0.857	NA	NA