

**Title:** Predicting In-Hospital Mortality After Traumatic Brain Injury: External Validation of CRASH and IMPACT in the National Trauma Data Bank

**Key Words:** traumatic brain injury; prognosis; mortality; prediction models; external validation

**Pre-print manuscript**

Copyright: CC BY-NC-ND 3.0 US

Elsevier sharing policy: <https://www.elsevier.com/about/policies/sharing>

DOI: <https://doi.org/10.1016/j.injury.2020.10.051>

Pre-print manuscript

## Abstract

**Background:** Traumatic brain injury (TBI) prognostic prediction models offer value to individualized treatment planning, systematic outcome assessments and clinical research design but require continuous external validation to ensure generalizability to different settings. The Corticosteroid Randomization After Significant Head Injury (CRASH) and International Mission on Prognosis and Analysis on Clinical Trials in TBI (IMPACT) models are widely available but lack robust assessments of performance in a current national sample of patients. The purpose of this study is to assess the performance of the CRASH-Basic and IMPACT-Core models in predicting in-hospital mortality using a nationwide retrospective cohort from the National Trauma Data Bank (NTDB).

**Methods:** The 2016 NTDB was used to analyze an adult cohort with moderate-severe TBI (Glasgow Coma Scale [GCS]  $\leq 12$ , head Abbreviated Injury Scale of 2-6). Observed in-hospital mortality or discharge to hospice was compared to the CRASH-Basic and IMPACT-Core models' predicted probability of 14-day or 6-month mortality, respectively. Performance measures included discrimination (area under the receiver operating characteristic curve [AUC]) and calibration (calibration plots and Brier scores). Further sensitivity analysis included patients with GCS  $\leq 14$  and considered patients discharged to hospice to be alive at 14-days.

**Results:** A total of 26,228 patients were included in this study. Both models demonstrated good ability in differentiating between patients who died and those who survived, with IMPACT demonstrating a marginally greater AUC (0.863; 95% CI: 0.858 – 0.867) than CRASH (0.858; 0.854 – 0.863);  $p < 0.001$ . On calibration, IMPACT overpredicted at lower scores and underpredicted at higher scores but had good calibration-in-the-large (indicating no systemic over/underprediction), while CRASH consistently underpredicted mortality. Brier scores were

similar (0.152 for IMPACT, 0.162 for CRASH). Both models showed slight improvement in performance when including patients with  $GCS \leq 14$ .

**Conclusion:** Both CRASH-Basic and IMPACT-Core models accurately predict in-hospital mortality following moderate-severe TBI, and IMPACT-Core performs well beyond its original GCS cut-off of 12. By demonstrating validity in the NTDB, these models appear generalizable to new data and offer value to current practice in diverse settings as well as to large-scale research design.

Pre-print manuscript

## Introduction

Traumatic brain injury (TBI) affects an estimated sixty-nine million people per year worldwide and contributes to global death and disability more than any other traumatic injury [1]. In the United States, there were approximately 2.5 million emergency department (ED) visits and 56,000 deaths related to TBI in 2013 [2]. The economic impact of TBI is troubling as well. In 2010, the lifetime direct and indirect costs of TBI were approximately \$76.5 billion, with 90% of total cost resulting from TBIs that are fatal or require hospitalization [3,4].

Ultimately, effective and efficient management is imperative to minimize disability and mortality.[5] Recently, much focus has been applied to generating standardized management guidelines for acute care of TBI patients [6,7]. Still, there are a myriad of ways to categorize patients with respect to mechanism of insult, clinical severity and pathophysiology, each of which may influence management and prognosis. The complex and heterogeneous nature of TBI often leads to uncertainty in expected patient outcomes and has significantly limited research design and clinical trial results [8–10]. As a result, a lack of robust randomized controlled trials (RCTs) has weakened evidence that supports the benefit of many treatment concepts [11–13]. Predictive prognostic models for TBI have previously been developed with the intent of producing an effective method for facilitating early clinical decisions, and validated prognostic models offer value to clinical care, research design and policy making [14].

When treating patients, these models may be used alongside clinical assessment in order to support decisions on treatment and counsel family members on expected outcomes and risks [15–17]. Utilizing prognostic tools in goals-of-care decisions helps to limit subjective variability

among physicians, and exposure to CRASH risk score data has been shown to reduce overoptimistic prognostication [18–20]. Additionally, decision making surrogates prefer numeric estimates to reduce uncertainty surrounding a loved one’s prognosis [21]. In thinking about the health care system as a whole, prognostic models are critical in the optimization of resource allocation and to classify severity based on prognostic risk and establish a baseline for clinical audits.[9,16] In research, they provide tools for comparative analysis research and development of stratified randomization protocols during enrollment into clinical trials [9,22].

In order to meet methodological standards, prediction models must include adequately large cohort sizes and undergo internal and external validation [23]. Currently, the best-established prognostic models are the International Mission on Prognosis and Analysis of randomized Controlled Trials in TBI (IMPACT) [14] and the Corticosteroid Randomization After Significant Head Injury (CRASH) models [24]. IMPACT predicts six-month mortality after moderate and severe TBI, while the CRASH model predicts two week and six-month mortality in these injuries. Both models were developed using large datasets and were externally validated against each other’s patient dataset [14,24]. Still, these models must be routinely evaluated to assure their generalizability to other settings and their validity in current practice.

The CRASH and IMPACT models have only been validated in cohorts much smaller than the original developmental datasets. Current validation studies have mainly utilized single-institution, multi-center or regional patient populations, further limiting generalizability to a nationwide population. Additionally, no study has evaluated the performance of IMPACT

alongside CRASH when predicting short-term outcomes in a heterogeneous, nationwide database, nor has any study evaluated IMPACT's performance in mild-moderate TBI patients. The purpose of this study is to assess the ability of these models for prediction of in-hospital mortality using a large nationwide retrospective cohort derived from the National Trauma Data Bank (NTDB). Additionally, we aim to assess the performance of IMPACT using an expanded GCS cut-off of 14 and below (as was done in the CRASH study) in order to determine its predictive value with inclusion of mild TBI patients.

## Materials and Methods

### Predictive models

The development of the IMPACT [14] and CRASH [24] models have been previously described and their calculators are available online ([IMPACT](#), [CRASH](#)). Each model includes base versions as well as more elaborate iterations with additional variables. Previous external validation efforts have demonstrated good prognostic accuracy for the base CRASH and IMPACT predictive models with marginal improvement when using the more elaborate iterations [14,25]. For the present investigation, we used the CRASH-Basic and IMPACT-Core models to predict in-hospital mortality.

### *CRASH*

The CRASH model was derived from a cohort of 10,008 adults in the CRASH trial (1999 to 2005). Participants in this trial had a GCS  $\leq$  14 and came from 49 countries (with 75% coming from low and middle-income countries). The CRASH model predicts mortality at 14 days and unfavorable outcome at six months. The base model includes age, total GCS, pupillary reactivity, and major extracranial injury.

## *IMPACT*

The IMPACT model was derived from 8,509 participants across 8 randomized trials and 3 observational studies (from 1984 to 1997). Participants in the IMPACT database had a GCS  $\leq 12$  and were aged  $\geq 14$  years old. The IMPACT model predicts mortality and unfavorable outcome at six months, and the base model includes age, motor GCS, and pupillary reactivity. Relevant differences in the variables included and predicted outcomes from the two models are summarized in **TABLE 1**, and an overview of the two study populations the models were trained on is summarized in **TABLE 2**.

### **Validation data**

To validate the CRASH and IMPACT scoring systems, we used data from the 2016 National Trauma Data Bank (NTDB). In our primary analysis, we considered patients with TBI as those having a head Abbreviated Injury Scale (AIS) between 2 to 6, total GCS  $\leq 12$ , and who were aged 14 years or older. Our primary endpoint of mortality was defined as a composite of in-hospital mortality (including death in the emergency department) or discharge to hospice. We compared the observed mortality in the NTDB to the probability of 14-day mortality predicted by the CRASH-Basic model, and the probability of 6-month mortality predicted by the IMPACT-Core model.

We defined major extracranial injury (used in the CRASH model) as an AIS greater than 2 in a region that is not the head, as has been previously described [26]. All other predictors (age, total

GCS, motor GCS, pupillary response) were used as reported in the NTDB. Patients with missing data for age, pupillary response, total GCS, or hospital disposition were excluded from the analysis, as were those who were dead on arrival to the emergency department (ED) or transferred directly from the ED to another facility.

### *Sensitivity analysis*

In a sensitivity analysis, we examined both models using a GCS cutoff  $\leq 14$ : this is the cutoff used in the CRASH model, whereas the cutoff of total GCS  $\leq 12$ , as used in the primary analysis, is the GCS cutoff from the IMPACT model. In a separate analysis of the CRASH model, we used a more conservative estimate of mortality by considering patients discharged to hospice to be alive at 14-days.

### **Statistical Methods**

For both predictive models, we assessed discrimination (the model's ability to discriminate between patients who died and patients who lived) using the area under the receiver operating characteristic curve (ROC). The area under the ROC curve (AUC) is equivalent to the C index [27]. ROC curves for both models were compared for statistical significance using the nonparametric methodology outlined by DeLong [28]. For each ROC curve, we determined optimal cutoff points based on Youden's Index (which maximizes the sum of the sensitivity and specificity) and calculated the sensitivity and specificity at that cutoff threshold (where patients with predicted scores higher than the cutoff are considered to have died). Values at additional cutoffs are available in the **Supplemental Table 1**.

Calibration was assessed graphically using calibration plots, and formally using Brier's score. Calibration refers to the agreement between predicted risk and observed outcomes, and when visualized on calibration plots the 45-degree line represents perfect calibration. Brier's score is the mean standard error between the predicted and observed outcomes. All analysis was conducted using the R software environment [29] (version 3.6.3) with the *OptimalCutpoints* [30] and *rms* [31] packages.

## Results

We included 26,228 patients in our primary analysis, with a mean (95% CI) age of 46.5 (46.2 – 46.8) years old. The majority (n=18,075; 69%) had a total GCS of 8 or below, and the most common pupillary response was both pupils reactive (n=15,521; 59%). Motor GCS followed a bimodal distribution, with nearly identical proportions of patients having a motor GCS of 1 (n=10,185; 39%) or score of 5 or 6 (n=10,336; 39%). 11,179 (43%) had major extracranial injury, and 10,919 (42%) died in the hospital or were discharged to hospice. Among the 10,919 patients who developed this composite outcome, 6.9% (n=752) were discharged to hospice (**TABLE 2**). Further details on the number of patients excluded during selection for the primary analysis are provided in **Supplemental Figure 1**.

### External validation of predictive models

When discrimination was assessed with ROC curves, the IMPACT model had a marginally greater AUC (0.863; 95% CI: 0.858 – 0.867) than the CRASH model (0.858; 95% CI: 0.858 – 0.863);  $p < 0.001$  (DeLong's test). Calibration of the IMPACT model showed overprediction at

lower scores but underprediction at higher scores, while the CRASH model showed consistent underprediction (**FIGURE 1**). The Brier score for the IMPACT and CRASH models were similar at 0.152 and 0.162, respectively (**TABLE 3**).

When optimal cutoff thresholds were determined using Youden's Index, the cutoff value for the CRASH model (33.1%) was lower than the cutoff for the IMPACT model (42.8%). At these respective cutoff points, both models showed similar sensitivities (CRASH: 78.2%, IMPACT: 80.1%) and specificities (CRASH: 80.3%, IMPACT: 77.9%). These metrics are available for various other cutoffs in the **Supplemental Table 1**.

### **Sensitivity analysis**

In our sensitivity analysis, we raised the GCS threshold for inclusion from  $GCS \leq 12$  to  $GCS \leq 14$ . As expected, the AUC in the CRASH model improved from 0.858 in the primary analysis to 0.872 (95% CI: 0.869 – 0.876) in the sensitivity analysis. The AUC for the IMPACT model also increased slightly from 0.863 to 0.865 (95% CI: 0.861 – 0.869). The calibration of both models improved in the sensitivity analysis, with the CRASH model having a Brier score of 0.133 and the IMPACT model having a Brier score of 0.139 (**Supplemental Figure 2**).

When we tested the CRASH model with a more conservative estimate of mortality by considering patients discharged to hospice to be alive at 14-days, the AUC significantly worsened from 0.858 in the primary analysis to 0.847 (95% CI: 0.842 – 0.851) in the sensitivity

analysis ( $p < 0.001$ ). However, the Brier score marginally improved to 0.161 (**Supplemental Figure 3**).

## Discussion

We found that both the IMPACT and CRASH models demonstrated good discrimination and calibration when externally validated using the NTDB. With a sample size that surpasses the number of patients in the IMPACT and CRASH datasets combined, this study represents the largest external validation of these models to date. Given the lack of universal decision-making guidelines, heterogeneity in patient populations, and varying treatment preferences among institutions, the NTDB provides a valuable tool for validation of generalizable prognostic models for TBI patients.

The CRASH model was developed almost exclusively using data from an international clinical trial, and the IMPACT model was in part developed from data collected over 20 years ago [14,24]. Roozenbeek et al. externally validated the IMPACT and a modified CRASH model for predicting 6-month outcome using five contemporary datasets, as well as for 14-day mortality in 2513 patients from a New York Brain Trauma Foundation database [32,33]. Sun et al. demonstrated validity of the IMPACT model using 1124 patients derived from the SyNAPSe trial; however, they found calibration to be poorer than previously reported, with overestimations of mortality and underestimations of unfavorable outcome [34]. It is unclear whether this finding was primarily driven by effects of case-mix on model performance or by changes in current standards of care [34]. Other external validation studies have remained somewhat limited by small sample sizes or datasets derived from a single institution [5,25,35–37].

Our findings demonstrate that both the IMPACT-Core and CRASH-Basic show good discrimination when predicting in-hospital mortality in patients with moderate to severe TBI (GCS  $\leq$  12 and head AIS of 2-6). These measures show slightly better performance of both basic models than previously described for 14-day and 6-month mortality, indicating good generalizability of these models for early mortality prediction in TBI patients. The calibration of the IMPACT model showed that estimates of risk were too moderate (i.e., estimates were too high for low risk patients, and too low for high risk patients) but had good calibration-in-the-large (meaning that there was no systematic over/under-prediction). The CRASH model showed consistent underestimation across patients of all levels of risk (**FIGURE 1**). As the IMPACT database was collected between 1984 and 1997, it is possible that improvements in standards of care have resulted in better early management of lower risk patients. Still, it is possible that validation is also influenced by differences in distribution of variables between datasets. Similar to Han et al., who also noted underprediction of 14-day mortality using the CRASH-Basic model, our dataset had a higher average age, prevalence of patients with GCS less than 9 and presence of bilateral pupillary defects than the CRASH dataset [25].

Furthermore, although the IMPACT model was originally designed for use in moderate to severe TBI (GCS of 12 and below), sensitivity analysis revealed similar discrimination and an unexpected improvement in calibration when including patients with GCS scores of 13-14. Mild TBI (mTBI) has its own unique range of severity and potential neuropsychiatric outcomes, and therefore necessitates its own clinical guidelines for recognition and management [38]. Although reported short-term mortality following mTBI is relatively low, previous studies have

demonstrated that a higher GCS score does not preclude poor outcomes, and the definition of mild TBI based on GCS cutoff (i.e., 13 versus 14) impacts the performance of mortality prediction models [39–42]. As such, models that can accurately distinguish mortality risk when including mTBI patients are beneficial in identifying who may benefit from initiating a unique treatment protocol. Further, they could be utilized in future efforts to determine appropriate severity classifications, unique risk factors, and best clinical practices [43].

Of note, previous studies have described only minor differences in performance between the two models. When applied to our large dataset, the performance of the two models was nearly identical. Although the criteria variables included in these models are similar, the application of one versus the other may be limited by what patient criteria can be reasonably and accurately obtained. Given their near equivalent performances, it may be reasonable to assume that choice of which model to use can be tailored to fit what patient information is available.

### **Strengths & Limitations**

This study has some limitations, primarily that our primary outcome (mortality or discharge to hospice) was not the same endpoint used by the CRASH or IMPACT models in their original development. This was less of a concern for the CRASH model (which predicts 14-day mortality), as the average length of stay in our population was 14.3 days (95% CI: 14.1 – 14.5; data not shown). However, IMPACT predicts 6-month mortality, and thus using in-hospital mortality could potentially misclassify patients who lived to hospital discharge but died in the following months. While this is certainly a limitation, Roozenbeek et al. previously demonstrated

validity of the IMPACT model in predicting 14-day mortality [33], potentially due to the disproportionate number of TBI-related deaths occurring early in hospital stay [44].

Additionally, we were unable to evaluate the other versions of the CRASH and IMPACT models that include additional variables (e.g., CRASH-CT [24], IMPACT-Extended [45]) that are unavailable in the NTDB. Although previous studies have shown these more complex models had higher predictive value than the base model, the differences are marginal. For example, in an external validation of the IMPACT model, the AUC for mortality only increased by 0.025 when the extended model was used over the core model [14]. Several other studies externally validating the core and extended models further demonstrate the marginality of difference [25,33,36].

Our results demonstrate good performance in the models which utilize the least number of parameters. In both clinical practice and research design, prognostic tools should be both accurate and easy to calculate and apply [15]. However, further studies should continue to evaluate performance of extended iterations of these models and their efficacy in predicting long-term and functional outcomes using similarly large sample sizes.

## **Conclusions**

Models which accurately predict in-hospital mortality may be used alongside clinical data in order to identify individuals likely to benefit from more aggressive early intervention and to avoid overoptimistic prognostication. In addition, they may be used in comparative analysis research to identify best practices or for selecting appropriate exclusion criteria in RCTs, both of

which have been especially challenging given the inherent heterogeneity of TBI populations. In order to be of significant value, physicians and researchers must be confident that these models are applicable to any given patient population.

Using a nationwide cohort of 26,228 adults, this study demonstrates that both the CRASH-Basic and IMPACT-Core models, although originally designed to predict 14 day and/or 6-month outcomes, can accurately predict in-hospital mortality following moderate-severe TBI. Although we did not include patients with a GCS of 15, these results also highlight their potential value in predicting mild TBI outcomes as well. In addition, we have shown that the National Trauma Data Bank may be a valuable tool for validation and optimization of these prognostic models. As this dataset represents a large, generalizable cohort its use can serve to minimize the effects of variation between institutions and patient populations.

**Acknowledgments:**

Committee on Trauma, American College of Surgeons. NTDB Version 2016. Chicago, IL, 2018.

The content reproduced from the NTDB remains the full and exclusive copyrighted property of the American College of Surgeons. The American College of Surgeons is not responsible for any claims arising from works based on the original data, text, tables, or figures.

Pre-print manuscript

## References

- [1] Dewan MC, Rattani A, Gupta S, Baticulon RE, Hung Y-C, Punchak M, et al. Estimating the global incidence of traumatic brain injury. *J Neurosurg* 2019;130:1080–97. <https://doi.org/10.3171/2017.10.JNS17352>.
- [2] Taylor CA, Bell JM, Breiding MJ, Xu L. Traumatic brain injury-related emergency department visits, hospitalizations, and deaths - United States, 2007 and 2013. *MMWR Surveill Summ* 2017;66:1–16. <https://doi.org/10.15585/mmwr.ss6609a1>.
- [3] Finkelstein EA, Corso PS, Miller TR. The Incidence and Economic Burden of Injuries in the United States. 2009. <https://doi.org/10.1093/acprof:oso/9780195179484.001.0001>.
- [4] Faul M, Coronado V. Epidemiology of traumatic brain injury. *Handb. Clin. Neurol.*, vol. 127, Elsevier B.V.; 2015, p. 3–13. <https://doi.org/10.1016/B978-0-444-52892-6.00001-5>.
- [5] Maeda Y, Ichikawa R, Misawa J, Shibuya A, Hishiki T, Maeda T, et al. External validation of the TRISS, CRASH, and IMPACT prognostic models in severe traumatic brain injury in Japan. *PLoS One* 2019;14:e0221791. <https://doi.org/10.1371/journal.pone.0221791>.
- [6] Carney N, Totten AM, O'Reilly C, Ullman JS, Hawryluk GWJ, Bell MJ, et al. Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. *Neurosurgery* 2017;80:6–15. <https://doi.org/10.1227/NEU.0000000000001432>.
- [7] Carney NA, Chesnut R, Kochanek PM. Guidelines for the acute medical management of severe traumatic brain injury in infants, children, and adolescents. *J Trauma - Inj Infect Crit Care* 2003. <https://doi.org/10.1097/00005373-200306001-00002>.
- [8] Mushkudiani NA, Hukkelhoven CWPM, Hernández A V., Murray GD, Choi SC, Maas AIR, et al. A systematic review finds methodological improvements necessary for

- prognostic models in determining traumatic brain injury outcomes. *J Clin Epidemiol* 2008;61:331–43. <https://doi.org/10.1016/j.jclinepi.2007.06.011>.
- [9] Maas AI, Stocchetti N, Bullock R. Moderate and severe traumatic brain injury in adults. *Lancet Neurol* 2008;7:728–41. [https://doi.org/10.1016/S1474-4422\(08\)70164-9](https://doi.org/10.1016/S1474-4422(08)70164-9).
- [10] Stein DG, Howard RB, Sayeed I. Why Did the Phase III Clinical Trials for Progesterone in TBI Fail? An Analysis of Three Potentially Critical Factors. *New Ther. Trauma. Brain Inj. Prev. Second. Brain Damage Enhanc. Repair Regen.*, 2017. <https://doi.org/10.1016/B978-0-12-802686-1.00001-8>.
- [11] Volovici V, Steyerberg EW, Cnossen MC, Haitsma IK, Dirven CMF, Maas AIR, et al. Evolution of Evidence and Guideline Recommendations for the Medical Management of Severe Traumatic Brain Injury. *J Neurotrauma* 2019;36:3183–9. <https://doi.org/10.1089/neu.2019.6474>.
- [12] Carroll LJ, Cassidy JD, Holm L, Kraus J, Coronado VG. Methodological issues and research recommendations for mild traumatic brain injury: The WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *J Rehabil Med Suppl* 2004:113–25. <https://doi.org/10.1080/16501960410023877>.
- [13] Menon DK, Schwab K, Wright DW, Maas AI. Position statement: Definition of traumatic brain injury. *Arch Phys Med Rehabil* 2010;91:1637–40. <https://doi.org/10.1016/j.apmr.2010.05.017>.
- [14] Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: Development and international validation of prognostic scores based on admission characteristics. *PLoS Med* 2008;5:1251–61. <https://doi.org/10.1371/journal.pmed.0050165>.

- [15] Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak* 2006;6:38.  
<https://doi.org/10.1186/1472-6947-6-38>.
- [16] Lingsma HF, Steyerberg EW, Maas AIR, Lingsma HF, Roozenbeek B, Steyerberg W, et al. Early prognosis in traumatic brain injury: from prophecies to predictions. vol. 9. 2010.  
[https://doi.org/10.1016/S1474-4422\(10\)70065-X](https://doi.org/10.1016/S1474-4422(10)70065-X).
- [17] Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A. Developing a quality criteria framework for patient decision aids: Online international Delphi consensus process. *Br Med J* 2006;333:417–9. <https://doi.org/10.1136/bmj.38926.629329.AE>.
- [18] Sacks GD, Dawes AJ, Ettner SL, Brook RH, Fox CR, Russell MM, et al. Impact of a risk calculator on risk perception and surgical decision making: A randomized trial. *Ann Surg* 2016. <https://doi.org/10.1097/SLA.0000000000001750>.
- [19] Elahi C, Williamson T, Spears CA, Williams S, Nambi Najjuma J, Staton CA, et al. Estimating prognosis for traumatic brain injury patients in a low-resource setting: how do providers compare to the CRASH risk calculator? *J Neurosurg* 2020.  
<https://doi.org/10.3171/2020.2.jns192512>.
- [20] Moskowitz J, Quinn T, Khan MW, Shutter L, Goldberg R, Col N, et al. Should We Use the IMPACT-Model for the Outcome Prognostication of TBI Patients? A Qualitative Study Assessing Physicians' Perceptions. *MDM Policy Pract* 2018;3:238146831875798.  
<https://doi.org/10.1177/2381468318757987>.
- [21] Quinn T, Moskowitz J, Khan MW, Shutter L, Goldberg R, Col N, et al. What Families Need and Physicians Deliver: Contrasting Communication Preferences Between Surrogate Decision-Makers and Physicians During Outcome Prognostication in Critically Ill TBI

- Patients. *Neurocrit Care* 2017;27:154–62. <https://doi.org/10.1007/s12028-017-0427-2>.
- [22] Young FB, Lees KR, Weir CJ. Improving trial power through use of prognosis-adjusted end points. *Stroke* 2005;36:597–601.  
<https://doi.org/10.1161/01.STR.0000154856.42135.85>.
- [23] Cowley LE, Farewell DM, Maguire S, Kemp AM. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. *Diagnostic Progn Res* 2019;3. <https://doi.org/10.1186/s41512-019-0060-y>.
- [24] Perel PA, Ollidashi F, Muzha I, Filipi N, Lede R, Copertari P, et al. Predicting outcome after traumatic brain injury: Practical prognostic models based on large cohort of international patients. *Bmj* 2008;336:425–9.  
<https://doi.org/10.1136/bmj.39461.643438.25>.
- [25] Han J, King NKK, Neilson SJ, Gandhi MP, Ng I. External validation of the CRASH and IMPACT prognostic models in severe traumatic brain injury. *J Neurotrauma* 2014;31:1146–52. <https://doi.org/10.1089/neu.2013.3003>.
- [26] Van Leeuwen N, Lingsma HF, Perel P, Lecky F, Roozenbeek B, Lu J, et al. Prognostic value of major extracranial injury in traumatic brain injury: An individual patient data meta-analysis in 39 274 patients. *Neurosurgery* 2012;70:811–8.  
<https://doi.org/10.1227/NEU.0b013e318235d640>.
- [27] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.  
<https://doi.org/10.1148/radiology.143.1.7063747>.
- [28] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach.

- Biometrics 1988;44:837. <https://doi.org/10.2307/2531595>.
- [29] R Development Core Team 3.0.1. A Language and Environment for Statistical Computing. R Found Stat Comput 2013;2:<https://www.R-project.org>.
- [30] López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, Gude-Sampedro F. Optimalcutpoints: An R package for selecting optimal cutpoints in diagnostic tests. J Stat Softw 2014;61:1–36. <https://doi.org/10.18637/jss.v061.i08>.
- [31] Harrell Jr FE. Regression Modeling Strategies 2019:1–246.
- [32] Roozenbeek B, Lingsma HF, Lecky FE, Lu J, Weir J, Butcher I, et al. Prediction of outcome after moderate and severe traumatic brain injury: External validation of the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation after Significant Head injury (CRASH) prognostic mod. Crit Care Med 2012;40:1609–17. <https://doi.org/10.1097/CCM.0b013e31824519ce>.
- [33] Roozenbeek B, Chiu YL, Lingsma HF, Gerber LM, Steyerberg EW, Ghajar J, et al. Predicting 14-day mortality after severe traumatic brain injury: Application of the IMPACT models in the brain trauma foundation TBI-trac® New York state database. J Neurotrauma 2012;29:1306–12. <https://doi.org/10.1089/neu.2011.1988>.
- [34] Sun H, Lingsma HF, Steyerberg EW, Maas AIR. External Validation of the International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury: Prognostic Models for Traumatic Brain Injury on the Study of the Neuroprotective Activity of Progesterone in Severe Traumatic Brain I. J Neurotrauma 2016;33:1535–43. <https://doi.org/10.1089/neu.2015.4164>.
- [35] Lingsma H, Andriessen TMJC, Haitsema I, Horn J, Van Der Naalt J, Franschman G, et al. Prognosis in moderate and severe traumatic brain injury: External validation of the

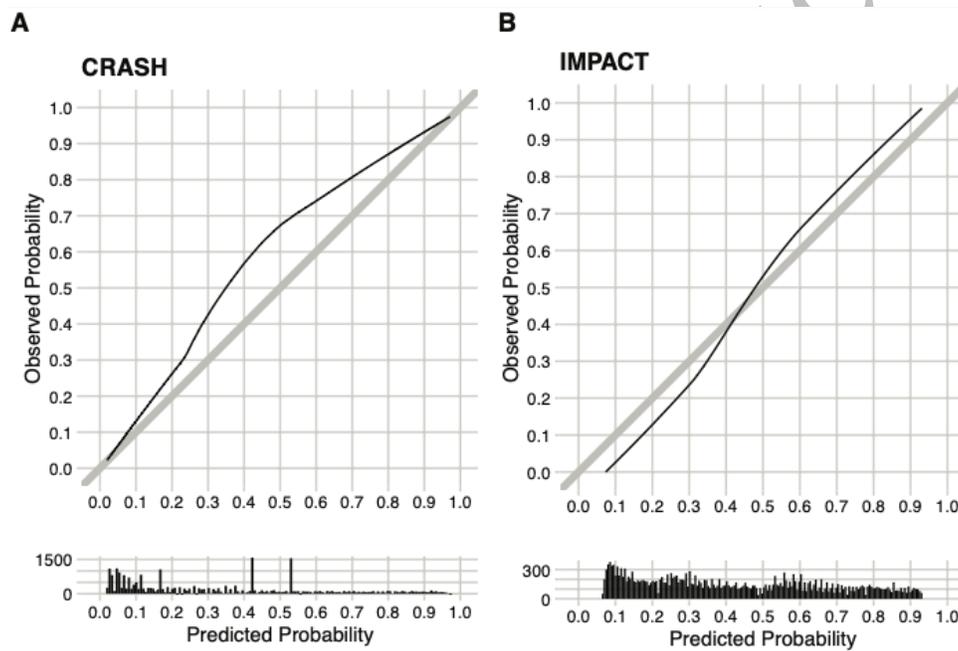
- IMPACT models and the role of extracranial injuries. *J Trauma Acute Care Surg* 2013;74:639–46. <https://doi.org/10.1097/TA.0b013e31827d602e>.
- [36] Panczykowski DM, Puccio AM, Scruggs BJ, Bauer JS, Hricik AJ, Beers SR, et al. Prospective independent validation of IMPACT modeling as a prognostic tool in severe traumatic brain injury. *J Neurotrauma* 2012;29:47–52. <https://doi.org/10.1089/neu.2010.1482>.
- [37] Majdan M, Lingsma HF, Nieboer D, Mauritz W, Rusnak M, Steyerberg EW. Performance of IMPACT, CRASH and Nijmegen models in predicting six month outcome of patients with severe or moderate TBI: An external validation study. *Scand J Trauma Resusc Emerg Med* 2014;22:68. <https://doi.org/10.1186/s13049-014-0068-9>.
- [38] Debaquey Veterans ME, Levin HS, Levin HS, Diaz-Arrastia RR. Diagnosis, prognosis, and clinical management of mild traumatic brain injury. *Lancet Neurol* 2015;14:506–23. [https://doi.org/10.1016/S1474-4422\(15\)00002-2](https://doi.org/10.1016/S1474-4422(15)00002-2).
- [39] Carroll LJ, Cassidy JD, Cancelliere C, Côté P, Hincapié CA, Kristman VL, et al. Systematic review of the prognosis after mild traumatic brain injury in adults: Cognitive, psychiatric, and mortality outcomes: Results of the international collaboration on mild traumatic brain injury prognosis. *Arch Phys Med Rehabil* 2014;95:S152–73. <https://doi.org/10.1016/j.apmr.2013.08.300>.
- [40] Humberto Mena J, Ignacio Sanchez A, Rubiano AM, Peitzman AB, Sperry JL, Isabel Gutierrez M, et al. Effect of the Modified Glasgow Coma Scale Score Criteria for Mild Traumatic Brain Injury on Mortality Prediction: Comparing Classic and Modified Glasgow Coma Scale Score Model Scores of 13. *J Trauma* 2011;71:1185–93. <https://doi.org/10.1097/TA.0b013e31823321f8>.

- [41] Binder LM. A review of mild head trauma. Part II: Clinical implications. *J Clin Exp Neuropsychol* 1997;19:432–57. <https://doi.org/10.1080/01688639708403871>.
- [42] Tucker B, Aston J, Dines M, Caraman E, Yacyshyn M, McCarthy M, et al. Early Brain Edema is a Predictor of In-Hospital Mortality in Traumatic Brain Injury. *J Emerg Med* 2017;53:18–29. <https://doi.org/10.1016/j.jemermed.2017.02.010>.
- [43] Lingsma HF, Yue JK, Maas AIR, Steyerberg EW, Manley GT, Cooper SR, et al. Outcome Prediction after Mild and Complicated Mild Traumatic Brain Injury: External Validation of Existing Models and Identification of New Predictors Using the TRACK-TBI Pilot Study. *J Neurotrauma* 2015;32:83–94. <https://doi.org/10.1089/neu.2014.3384>.
- [44] Ollidashi F, Muzha I, Filipi N, Lede R, Copertari P, Traverso C, 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;364:1321–8. [https://doi.org/10.1016/S0140-6736\(04\)17188-2](https://doi.org/10.1016/S0140-6736(04)17188-2).
- [45] Van Beek JGM, Mushkudiani NA, Steyerberg EW, Butcher I, McHugh GS, Lu J, et al. Prognostic value of admission laboratory parameters in traumatic brain injury: Results from the IMPACT study. *J Neurotrauma* 2007;24:315–28. <https://doi.org/10.1089/neu.2006.0034>.

## Figure Legends:

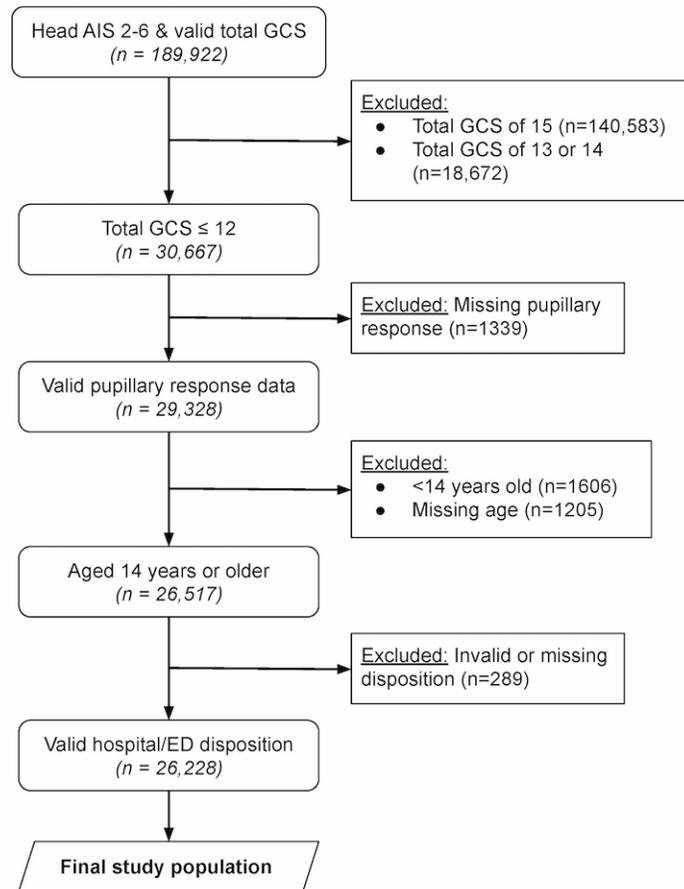
**Figure 1 text:** Calibration plots demonstrating the observed versus predicted mortality for the CRASH (panel A) and IMPACT (panel B) models. Points on the curve that are above the grey 45-degree line indicate the model underestimates mortality and points below indicate overestimation. The histogram below the plot demonstrate the distribution of predicted probabilities for each model.

Filename: Figure1.eps



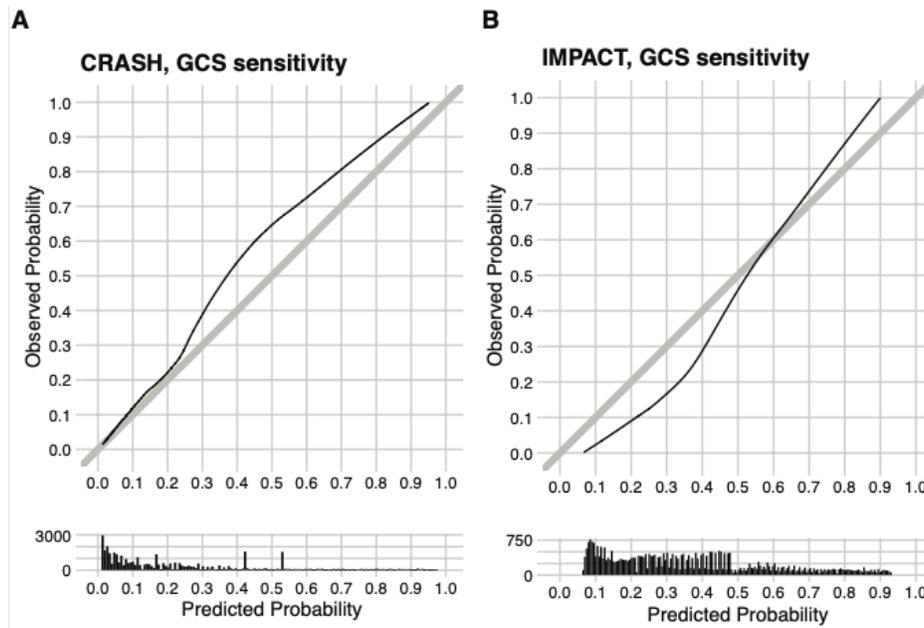
**Supplemental Figure 1 text:** Study flow diagram.

Filename: SDC\_Figure1.tiff



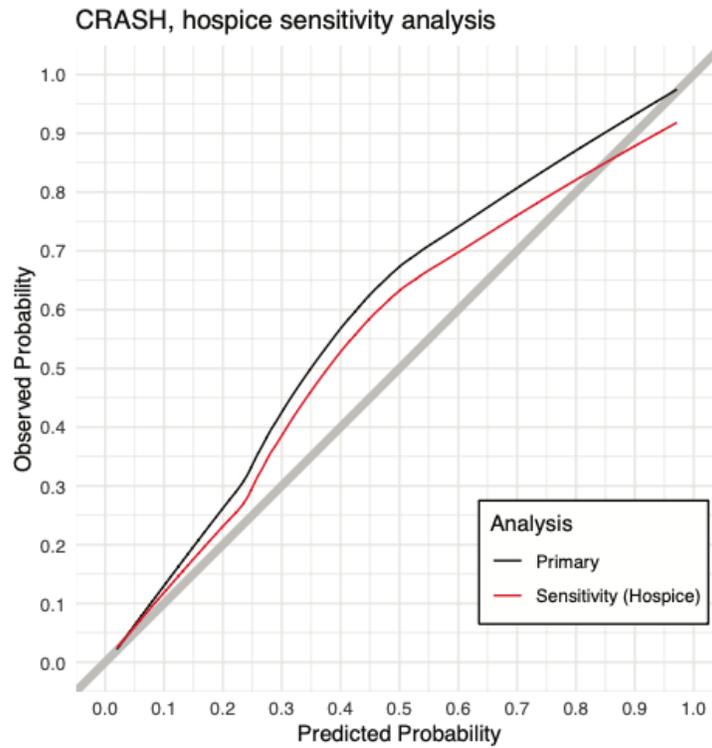
**Supplemental Figure 2 text:** Calibration plots for the sensitivity analysis using the GCS threshold for inclusion as  $GCS \leq 14$  for both the CRASH (panel A) and IMPACT (panel B) models.

Filename: SDC\_Figure2.eps



**Supplemental Figure 3 text:** Calibration plot comparing CRASH model from primary analysis (black line, same as Figure 1A in the main text) to the sensitivity analysis where patients discharged to hospice were considered to have lived (red line).

Filename: SDC\_Figure3.eps



**TABLE 1: Variables and outcomes of predictive models**

	<b>CRASH-Basic</b>	<b>IMPACT-Core</b>
<b>Variables included</b>	Age* Total GCS Pupillary reactivity Major extra-cranial injury	Age Motor GCS Pupillary reactivity
<b>Outcomes predicted†</b>	14-day mortality 6-month unfavorable outcome	6-month mortality 6-month unfavorable outcome

\* In the CRASH model, age is calculated as number of years over 40 years old. Patients 40 years or younger are given an age score of zero

† Unfavorable outcome is defined as death, vegetative state, or severe disability

**TABLE 2:** Descriptive statistics

	<b>NTDB</b> <i>n</i> = 26,228	<b>CRASH</b> <i>n</i> = 10,008	<b>IMPACT</b> <i>n</i> = 8,509
<b>Age, years</b>			
Mean (95% CI)	46.50 (46.24 – 46.76)	37.0 (36.83 – 37.17)	NR
Median (Q25, Q75)	45 (27, 63)	32 (23, 47)	30 (21, 45)
<b>Total GCS</b>			
Mean (95% CI)	6.40 (6.36 – 6.44)	-	-
13 or higher (%)	0 (0%)	30.2%	0%
12 to 9 (%)	8,153 (31%)	30.4%	18%
8 or lower (%)	18,075 (69%)	39.5%	82%
<b>Motor GCS</b>			
5 or 6	10,336 (39%)	NR	30%
4	3,585 (14%)	NR	23%
3	1,080 (4%)	NR	13%
2	854 (3%)	NR	12%
1	10,185 (39%)	NR	16%
Untestable or missing	188 (1%)	0%	5%
<b>Pupils*</b>			
Both reactive	15,521 (59%)	82.8%	63%
One reactive	1,880 (7%)	6.3%	12%
Neither reactive	8,827 (34%)	8.2%	25%
<b>Major extracranial injury</b>			
Yes	11,179 (43%)	77.3%	NR
No	15,049 (57%)	22.7%	NR
<b>In hospital mortality or discharge to hospice</b>			
Died	10,167 (39%)	NR	NR
Hospice	752 (3%)	NR	NR
Lived	15,309 (58%)	NR	NR

\* The total percentages for CRASH do not sum to 100%, as 2.7% had untestable pupillary responses

Abbreviations: *NR* = Not reported; *CI* = Confidence interval; *Q25* = 25<sup>th</sup> percentile; *Q75* = 75% percentile

**TABLE 3:** Discrimination, calibration, and optimal cutoffs for the CRASH and IMPACT scores

	<b>CRASH</b>	<b>IMPACT</b>
<b>Discrimination: AUC (95% CI)</b>		
Primary Analysis	0.858 (0.854 – 0.863)	0.863 (0.858 – 0.867)
GCS $\leq$ 14*	0.872 (0.869 – 0.876)	0.865 (0.861 – 0.869)
Only hospital mortality†	0.847 (0.842 – 0.851)	Not tested
<b>Calibration: Brier score (95% CI)</b>		
Primary Analysis	0.162 (0.160 – 0.165)	0.152 (0.150 – 0.154)
GCS $\leq$ 14*	0.133 (0.131 – 0.135)	0.139 (0.138 – 0.141)
Only hospital mortality†	0.161 (0.158 – 0.164)	Not tested
<b>Optimal cutoff value: Based on Youden's Index</b>		
Cutoff value	0.331	0.428
Sensitivity	78.2%	80.1%
Specificity	80.3%	77.9%
PPV	73.9%	72.1%
NPV	83.8%	84.6%

\* Sensitivity analysis using GCS threshold for inclusion of  $\leq$  14

† Sensitivity analysis in which patients discharged to hospice were no longer considered to have died

Abbreviations: AUC = *Area under curve*; CI = *Confidence interval*; PPV = *Positive predicative value*; NPV = *Negative predictive value*

**Supplemental Table 1:** Metrics for CRASH and IMPACT models at various cutoff values

Cutoff	CRASH				IMPACT			
	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
0.05	99.0%	21.6%	47.4%	96.8%	100.0%	0.0%	41.6%	NaN
0.10	95.7%	44.6%	55.2%	93.5%	99.5%	12.7%	44.8%	97.3%
0.15	91.8%	56.7%	60.2%	90.7%	98.2%	28.3%	49.4%	95.6%
0.20	87.1%	67.2%	65.4%	88.0%	96.6%	39.0%	53.0%	94.2%
0.25	84.2%	73.0%	69.0%	86.6%	94.6%	49.4%	57.2%	92.8%
0.30	80.8%	77.3%	71.7%	85.0%	91.1%	60.2%	62.0%	90.4%
0.35	76.5%	81.7%	74.9%	83.0%	87.2%	68.8%	66.6%	88.3%
0.40	72.0%	85.3%	77.8%	81.1%	82.7%	75.0%	70.2%	85.9%
0.45	57.3%	89.1%	79.0%	74.5%	77.8%	80.0%	73.5%	83.5%
0.50	53.4%	90.9%	80.8%	73.3%	73.0%	83.6%	76.0%	81.3%
0.55	39.1%	94.2%	82.7%	68.4%	65.4%	87.5%	78.8%	78.0%
0.60	35.1%	95.5%	84.6%	67.3%	54.8%	91.1%	81.4%	73.8%
0.65	30.6%	96.5%	86.0%	66.1%	45.5%	93.8%	83.9%	70.7%
0.70	26.2%	97.4%	87.7%	64.9%	37.3%	95.8%	86.3%	68.2%
0.75	21.8%	98.2%	89.4%	63.8%	30.0%	97.2%	88.3%	66.0%
0.80	17.4%	98.8%	91.0%	62.6%	21.1%	98.4%	90.6%	63.6%
0.85	12.5%	99.4%	93.3%	61.4%	12.9%	99.2%	92.3%	61.5%
0.90	7.3%	99.7%	94.9%	60.1%	4.8%	99.8%	95.8%	59.5%
0.95	1.3%	100.0%	97.9%	58.7%	0.0%	100.0%	NaN	58.4%
1.00	0.0%	100.0%	NaN	58.4%	0.0%	100.0%	NaN	58.4%

Abbreviations: *PPV = Positive predicative value; NPV = Negative predictive value; NaN = Not a number (occurs when denominator is zero)*