

Predicting Outcome After Traumatic Brain Injury in Bouaké, Côte d'Ivoire: Performance of IMPACT, CRASH-Basic and a Simplified Glasgow-Pupils Score in a Hospital Series of 772 Patients

Dr. Landry Teti¹; Dr. Yao Bernard Fionko²; Dr. Konan Serge Yao³;
Dr. Adama Kouadio⁴; Dr. Adama Simaga⁵; Adéréhime Haidara⁶

¹MD, Neurochirurgien, Department of Neurosurgery, Bouaké University Hospital, Bouaké, Côte d'Ivoire
^{2,3,4,5,6}MD, Department of Neurosurgery, Bouaké University Hospital, Bouaké, Côte d'Ivoire

Publication Date: 2026/10/30

Abstract

➤ Background:

Prognostic scores for traumatic brain injury (TBI) were built in wealthy countries. Few have been tested in sub-Saharan Africa.

➤ Objective:

To test, in Bouaké, IMPACT (computed from available variables), CRASH-Basic and a simplified score, the GCS-Pupils adapted to anisocoria (adapted GCS-P), and to explore local recalibration.

➤ Methods:

Retrospective analysis of the consecutive-admissions register of 772 patients (Bouaké University Hospital, 2025–2026), following TRIPOD. Primary endpoint: in-hospital death (36 deaths, all by day 14). Secondary endpoint: unfavourable outcome (GOS 1–3, median follow-up 3.2 months). Discrimination: AUC, DeLong test. Calibration: intercept, slope, expected/observed ratio, Brier score compared with a null model. Analyses: whole cohort, moderate-to-severe TBI (IMPACT's domain), adults; post hoc additions.

➤ Results:

In-hospital mortality was 4.7%. IMPACT overestimated risk (expected/observed ratio 3.4; 2.0 in moderate-to-severe TBI), discriminated poorly (AUC 0.64; 0.66), with a Brier score above the null model's. The adapted GCS-P (AUC 0.79), GCS alone (0.78) and CRASH-Basic (0.78; expected/observed 1.46) discriminated better. Anisocoria added almost nothing to the GCS. In moderate-to-severe TBI the difference with IMPACT was not significant ($p = 0.16$).

➤ Conclusions:

IMPACT clearly overestimates risk: its horizon (six months) is not that of our endpoint, and 56% of patients have mild TBI. The GCS and CRASH-Basic rank patients better, CRASH-Basic being less miscalibrated. Calibration of the adapted GCS-P remains to be evaluated. Multicentre validation is needed.

Keywords: Traumatic Brain Injury; Prognostic Models; Calibration; IMPACT; Sub-Saharan Africa.

Teti, D. L., Fionko, D. Y. B., Yao, D. K. S., Kouadio, D. A., Simaga, D. A., & Haidara, A. (2026). Predicting Outcome After Traumatic Brain Injury in Bouaké, Côte d'Ivoire: Performance of IMPACT, CRASH-Basic and a Simplified Glasgow-Pupils Score in a Hospital Series of 772 Patients: External Validation of Prognostic Models in a Resource-Limited Neurosurgical Setting. *International Journal of Scientific Research and Modern Technology*, 5(10), 61–71. <https://doi.org/10.38124/ijsrmt.v5i10.1715>

I. INTRODUCTION

Traumatic brain injury (TBI) kills and disables mostly in low- and middle-income countries. Sub-Saharan Africa bears a heavy share: a young population, frequent road crashes, and limited access to emergency neurosurgery, neurocritical care and imaging [1]. In this setting, estimating prognosis at admission is harder, and more useful: it guides decisions and what families are told.

Two models are the reference. IMPACT predicts mortality and unfavourable outcome at six months [2]. CRASH predicts mortality at 14 days and unfavourable outcome at six months, with different equations according to the country's income level [3]; the CRASH trial itself recruited many patients in low- and middle-income countries. The GCS-Pupils score (GCS-P) simplifies the approach: the number of unreactive pupils is subtracted from the Glasgow Coma Scale score [4] [5]. It needs only data that are immediately available.

A model is worth only as much as it keeps, away from home, two qualities: ranking patients correctly (discrimination) and predicting accurate risks (calibration). The second often erodes when the setting changes, because recruitment, prehospital care, access delays and critical-care capacity shift baseline risk [6]. A model should therefore be validated, and recalibrated if needed, before local use. The TRIPOD reporting guideline and the PROBAST tool frame this process [7] [8].

These models are rarely validated in sub-Saharan Africa. Our non-systematic literature search (PubMed/MEDLINE, Scopus and Web of Science; queries combining IMPACT, CRASH, traumatic brain injury, validation and calibration; last run on 4 October 2026) identified no published validation in West Africa. In East Africa, a feasibility study has been reported [9]. We therefore tested in Bouaké, in a consecutive hospital series, the IMPACT model computed from the available variables (primary objective), a simplified score derived from the GCS-P and the CRASH-Basic model (secondary objective). We then explored local recalibration and derived models (exploratory objective).

II. METHODS

➤ *Study Design and Reporting*

We retrospectively analysed the institutional TBI database of the neurosurgery department of Bouaké University Hospital, fed by the consecutive-admissions register and the operating register. Only the recent series was kept (January 2025–June 2026; $n = 772$). The study follows TRIPOD 2015 [7]; TRIPOD+AI 2024 [10] is cited only as a complement, since we use classical logistic regression. PROBAST [8] guided the appraisal of bias; the domain-by-domain appraisal is in the supplementary material (Table S4), together with the completed TRIPOD checklist. No protocol was registered.

The models were developed elsewhere, so this is an independent single-centre geographical validation. We use “external validation” in this sense.

Three levels of analysis are distinguished throughout. (i) Evaluation of IMPACT, the pre-specified primary analysis, run in the whole cohort and in moderate-to-severe TBI, which is the model's development domain. (ii) Evaluation of the adapted GCS-P, the pre-specified secondary analysis. (iii) Development and local recalibration of derived models, an exploratory analysis done after seeing the results (post hoc). The validation of CRASH-Basic, the decomposition of discrimination (age, anisocoria and GCS taken separately) and the sensitivity analyses (motor-score tables, pupil review among deaths) were added afterwards and are also post hoc. Our figures describe IMPACT computed from the available variables (Section 2.4), not its original implementation.

The same series feeds several works. A principal manuscript covers epidemiology, mortality and prognostic models. Others address the prehospital pathway and admission delays, the paediatric subgroup, and lesion mapping and surgical performance (unpublished manuscripts). The question and the endpoint of this article are its own. We share with those works the general characteristics of the cohort and, with the principal manuscript, the admission variables (age, GCS, pupils) of the exploratory local models in Section 3.4; their performance is close to that of the principal manuscript (corrected AUC 0.79 with the continuous GCS). We declare this overlap here and in the cover letter, in line with ICMJE recommendations.

➤ *Population*

Over the period, 826 admissions for suspected TBI were recorded. We excluded 54 records: 23 missing or lacking essential data, 14 without usable CT, 9 duplicates or readmissions, 5 with a final non-traumatic diagnosis and 3 patients transferred before complete assessment. This leaves 772 patients (Figure 1).

Patients who died before admission are not included. Deaths after transfer to intensive care are counted in hospital mortality. Selection bias in favour of patients who arrive alive cannot be excluded. Ages range from 1 to 90 years; 86 patients (11.1%) are under 16. As the models were built in adults, we repeated the analysis in patients aged 16 and over. We checked the consistency of admission variables: GCS (5–15, no patient with a GCS of 3 or 4), blood glucose and age were plausible, with no outlier to correct or exclude.

➤ *Endpoints*

The primary endpoint, set before the analyses, is in-hospital death (all causes, during the stay). All 36 deaths occurred before day 15 (median: 1 day; 78% within the first 48 hours). This endpoint therefore equals 14-day mortality, the horizon of CRASH. It is only a substitute for six-month mortality, the horizon of IMPACT [2], which we could not measure for lack of standardised follow-up.

Since six-month mortality cannot be lower than in-hospital mortality, this gap pushes IMPACT towards overestimating risk, in the same direction as our main result (Section 4.2).

The secondary endpoint is unfavourable outcome (Glasgow Outcome Scale [GOS] 1–3) at last follow-up. It was assessed at consultation (386 patients), by telephone (177) or from the file (104), plus the 36 deaths: 703 patients in total. Median follow-up is 3.2 months (IQR 2.1–5.4), well below the six months of the original models;

late sequelae may therefore escape us, which would accentuate overestimation. The 39 unfavourable outcomes comprise the 36 in-hospital deaths and only three survivors (GOS 2–3). This endpoint almost overlaps the primary one and adds little independent information.

➤ *Models Evaluated and Local Adaptation of Variables*

We computed IMPACT probabilities (Core and Extended models) with the published coefficients. Several variables required local adaptation, summarised in Table 1.

Table 1 Local Adaptation of IMPACT Variables.

Model variable	Available data	Adaptation	Expected impact
Age, total GCS	Yes	Used directly	None
GCS motor score	Not recorded	Estimated from total GCS (3→M1; 4–5→M2; 6–7→M3; 8–9→M4; 10–12→M5; 13–15→M6; M5 and M6 are pooled in IMPACT)	Loss of specific information above GCS 10; lower discrimination
Pupillary reactivity	Anisocoria (proxy)	No anisocoria → 2 reactive; anisocoria → 1 reactive	Bilateral unreactive mydriasis classed as “2 reactive”; most severe forms underestimated
Hypoxia / hypotension	SpO ₂ < 90% / SBP < 90 mmHg	Used directly	None
Traumatic SAH, EDH	Yes	Used directly	None
Marshall classification	Derived	Reconstructed from CT items	Possible misclassification

GCS: Glasgow Coma Scale; SBP: Systolic Blood Pressure; SAH: Subarachnoid Haemorrhage; EDH: Extradural Haematoma.

The GCS motor score was not recorded. We estimated it from the total GCS with a correspondence table defined in advance and never validated in West Africa. This table makes the motor score a step function of the total GCS: any GCS ≥ 10, that is all mild TBI (56% of patients) and part of moderate TBI, falls into the same IMPACT category M5–6. For most patients the motor score therefore carries no information of its own. Pupillary reactivity was approximated by anisocoria. Without anisocoria, the patient is coded “two reactive pupils”: bilateral unreactive mydriasis, which carries the worst prognosis, falls into the most favourable category. This limitation affects IMPACT and the adapted GCS-P alike. In the 111 intubated patients, we used the last GCS recorded before intubation or sedation. IMPACT-Lab could not be evaluated: admission haemoglobin is not routinely measured.

The “adapted GCS-P” is the total GCS minus the number of unreactive pupils estimated from anisocoria (0 or 1). It differs from the original definition by Brennan et al. [4]. We evaluate only its discrimination: its probabilities were not compared with those published for the GCS-P [5]. Our results therefore concern a pragmatic adaptation of IMPACT, not its exact reproduction.

We also computed the original CRASH-Basic model [3] with the 14-day mortality equation intended for low- and middle-income countries. It uses age, total GCS, pupillary reactivity (approximated by anisocoria) and major extracranial injury. Its horizon matches that of our primary endpoint. Post hoc analyses finally decomposed

discrimination (AUC of age, anisocoria and total GCS taken alone), tested two other motor-score tables for IMPACT-Core and recalculated the GCS-P after reviewing the files of the deaths for bilateral unreactive mydriasis. The conservative table is: GCS 3–4→M1; 5–6→M2; 7–8→M3; 9–10→M4; 11–13→M5; 14–15→M6. The liberal table is: 3→M1; 4→M2; 5–6→M3; 7–8→M4; 9–11→M5; 12–15→M6 (Table S3).

➤ *Missing Data*

The primary endpoint and the admission predictors are complete. Only functional outcome (GOS) is missing, for 69 patients (9%), all in-hospital survivors. Their age and severity are close to those of followed patients, and we found no clinically relevant difference: data missing at random (MAR) is plausible. The secondary endpoint is analysed on complete cases. As all patients without GOS are survivors, this approach slightly overestimates the share of unfavourable outcomes, which works against our conclusion on risk overestimation.

➤ *Statistical Analysis*

• *Discrimination:*

We report the AUC with its 95% confidence interval (95% CI), obtained by bootstrap (2000 draws with replacement, of the cohort size). The DeLong test compares the AUCs of models evaluated in the same patients.

- *Calibration:*

We report: the calibration intercept (calibration-in-the-large, slope fixed at 1); the calibration slope (below 1: predictions too extreme); the expected/observed ratio (above 1: risk overestimated, below 1: underestimated); the mean calibration error (ECE: weighted mean absolute difference between observed frequencies and predicted probabilities in 10 decile groups; close to 0, agreement is good); the Brier score; and calibration curves. We compare the Brier score with that of a null model that would predict the observed frequency for everyone, that is $p(1 - p)$: a model with a higher Brier score is less informative than this single rate. The Hosmer-Lemeshow test depends on sample size and grouping; we give it descriptively, alongside the curves, intercept, slope and Brier score. The 95% CIs of calibration metrics come from the bootstrap.

- *Local recalibration:*

Recalibration-in-the-large updates the intercept only. By construction it brings the expected/observed ratio to 1: it therefore does not, by itself, prove good calibration. Logistic recalibration updates the intercept and the overall slope; a slope different from 1 signals a change in the strength of predictor effects. These are the first two steps of model updating [11]. Carried out exploratorily on the whole cohort, with optimism correction by bootstrap (1000 draws), it demonstrates feasibility. It does not yield a ready-to-use model without independent validation.

- *Clinical utility:*

A decision curve [12] compares the net benefit of the models with the “treat all” and “treat none” strategies. It relies on the same data as the recalibration, and the clinical decision that the thresholds would inform is not defined: it remains illustrative.

A pre-specified subgroup (moderate-to-severe TBI, IMPACT’s domain) and a sensitivity analysis in adults (≥ 16 years) complete the analysis. Logistic regressions were fitted by iteratively reweighted least squares. Analyses were run in Python (numpy, scipy, scikit-learn, pandas); versions are given in the supplementary material (Table S5) and scripts are available on request from the corresponding author.

- *Sample Size*

With 36 deaths and 39 unfavourable outcomes, we remain below the threshold of 100 events usually recommended for external validation [13]. Calibration estimates, especially the slope, remain exploratory and their intervals are wide; comparisons between models call for caution. This first single-centre geographical evaluation will need confirmation by independent multicentre validation.

- *Ethical Considerations*

The study complies with the Declaration of Helsinki. The Institutional Ethics Committee of Bouaké University Hospital approved it (approval No. CEI/CHU-BKE/141/2026) and granted a waiver of individual informed consent: the study is retrospective and uses only anonymised routine-care records, from which no patient can be identified.

III. RESULTS

- *Cohort Characteristics*

Our cohort includes 772 patients (Table 2, Figure 1). It is young (median age: 30 years; range 1–90), predominantly male (sex ratio 5.2) and mostly injured in road-traffic crashes (89.4%). Mild TBI dominates; in-hospital mortality is 4.7%. The setting is one of limited resources: median delay of 2 days between injury and admission, transport most often non-medicalised (60.1%), one patient in two referred from another facility (50.4%). Delay is counted in whole days, from injury to admission to neurosurgery, transfers included.

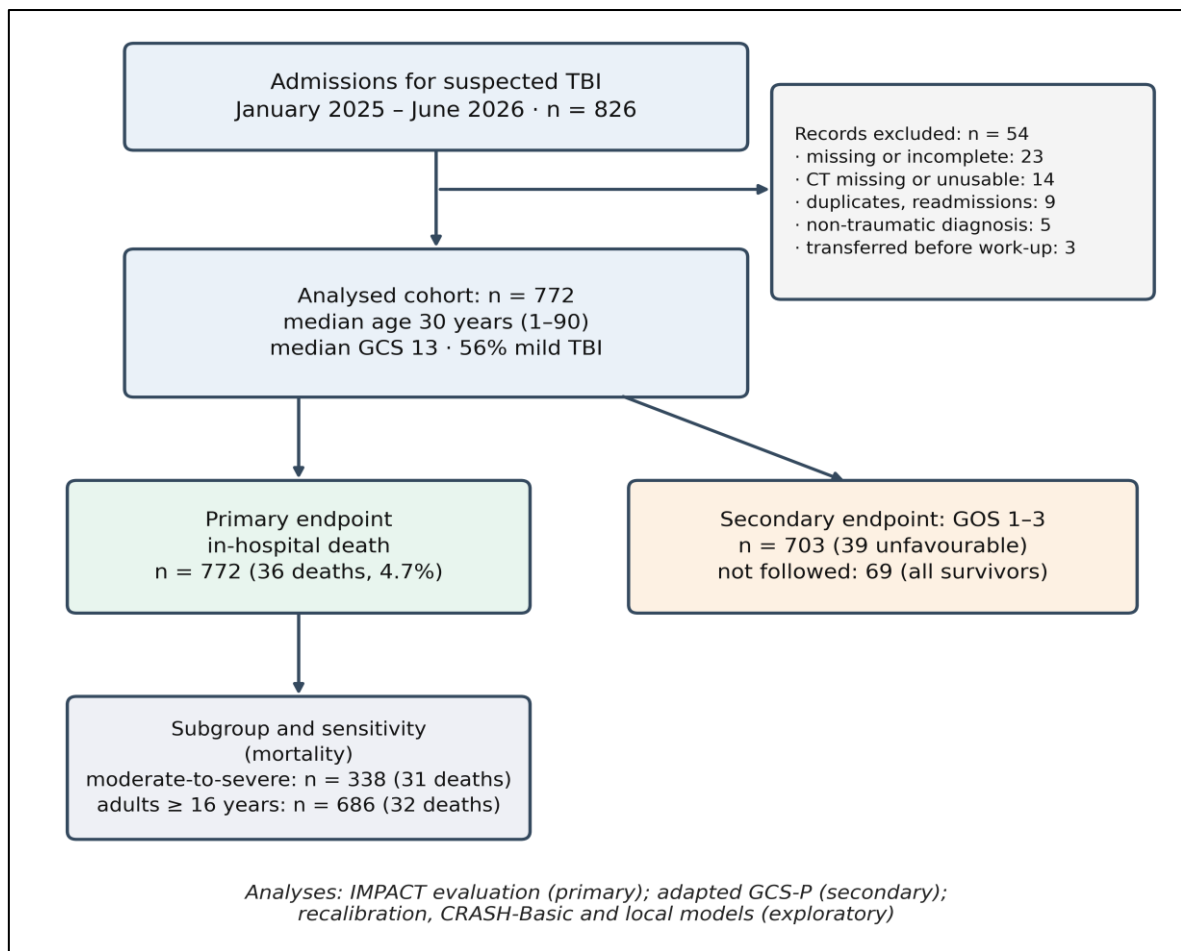


Fig 1 Flow Diagram: Admissions, Exclusions and Numbers Analysed by Endpoint and Subgroup.

Table 2 Characteristics of the Cohort (n = 772).

Variable	Value
Age (years), median [IQR] (range)	30 [20–44] (1–90)
Patients < 16 years	86 (11.1%)
Male sex (sex ratio)	647 (83.8%) (5.2)
Mechanism: road-traffic crash	690 (89.4%)
of which two- or three-wheeled motorised vehicles	532/690 (77.1%)
Injury-to-admission delay (days), median [IQR]	2 [1–3]
Delay ≥ 3 days	285 (36.9%)
Mode of admission: direct / referred	383 (49.6%) / 389 (50.4%)
Transport: non-medicalised / ambulance / emergency service	464 (60.1%) / 189 (24.5%) / 119 (15.4%)
Chronic alcoholism	156 (20.2%)
GCS, median [IQR]	13 [11–14]
Mild TBI (GCS 13–15)	434 (56.2%)
Moderate TBI (GCS 9–12)	277 (35.9%)
Severe TBI (GCS ≤ 8)	61 (7.9%)
Anisocoria	81 (10.5%)
Hypotension (SBP < 90) / hypoxia (SpO ₂ < 90%)	8 (1.0%) / 9 (1.2%)
Intubation at admission	111 (14.4%)
SAH / EDH / acute SDH	116 (15.0%) / 100 (13.0%) / 110 (14.2%)
Contusion	336 (43.5%)
Mass effect	230 (29.8%)
Surgery documented in the operating register	369 (47.8%)
In-hospital mortality	36 (4.7%)
Unfavourable outcome (GOS 1–3), among the 703 patients followed	39 (5.5%)
Follow-up (months), median [IQR]	3.2 [2.1–5.4]

IQR: Interquartile Range; GCS: Glasgow Coma Scale; SBP: Systolic Blood Pressure; SpO₂: Pulse Oxygen Saturation; SAH: Subarachnoid Haemorrhage; EDH: Extradural Haematoma; SDH: Subdural Haematoma; GOS: Glasgow Outcome Scale. Percentages are Relative to the 772 Patients, Except for Two- or Three-Wheeled Vehicles (Relative to the 690 Road-Traffic Crashes) and Unfavourable Outcome (Relative to the 703 Patients Followed).

The 69 patients without GOS are all survivors. They resemble the 703 followed patients in age (median 27 versus 30 years) and severity (7.2% versus 8.0% of severe TBI).

➤ *Primary Analysis (Pre-Specified): Performance of IMPACT*

IMPACT-Core discriminates poorly and overestimates risk. Its AUC is 0.64 in the whole cohort and 0.66 in moderate-to-severe TBI (Table 3). For mortality in the whole cohort, the calibration intercept is -1.44 (95% CI -1.80 to -1.13) and the expected/observed ratio 3.39 (2.56 to 4.75): IMPACT announces about 3.4 times more deaths than occur. In moderate-to-severe TBI, the model's development domain, overestimation decreases without

disappearing (ratio 2.03; intercept -0.90). The calibration slope is below 1 but very uncertain (0.81; 95% CI 0.34–1.28 in the whole cohort), for lack of events.

In all eight situations of Table 3, the Brier score of the models exceeds that of the null model (0.044 to 0.097): the raw IMPACT probabilities are less informative than a single rate equal to the observed frequency. The Extended model does no better than the Core model, probably because the reconstructed CT variables are imperfect. The Hosmer-Lemeshow test is significant for all models ($p < 0.001$; $p = 0.001$ in moderate-to-severe TBI). The calibration curve (Figure 2A) shows overestimation across the whole risk spectrum, not only in high-risk patients.

Table 3 Performance of the IMPACT models (Core and Extended).

Population	Model	Endpoint	n / events	AUC (95% CI)	Intercept	Slope	E:O	Brier	Null Brier	ECE
Whole	Core	Mortality	772 / 36	0.64 (0.53–0.74)	-1.44	0.81	3.39	0.061	0.044	0.11
Whole	Core	Unfavourable	703 / 39	0.66 (0.55–0.76)	-1.99	0.79	4.67	0.105	0.052	0.20
Whole	Extended	Mortality	772 / 36	0.59 (0.47–0.70)	-1.05	0.54	2.44	0.052	0.044	0.08
Whole	Extended	Unfavourable	703 / 39	0.60 (0.49–0.72)	-1.64	0.60	3.58	0.083	0.052	0.15
Mod.-severe	Core	Mortality	338 / 31	0.66 (0.56–0.76)	-0.90	0.71	2.03	0.094	0.083	0.10
Mod.-severe	Core	Unfavourable	312 / 34	0.67 (0.56–0.77)	-1.49	0.68	2.80	0.144	0.097	0.20
Mod.-severe	Extended	Mortality	338 / 31	0.62 (0.49–0.73)	-0.55	0.51	1.53	0.087	0.083	0.08
Mod.-severe	Extended	Unfavourable	312 / 34	0.64 (0.52–0.75)	-1.16	0.54	2.23	0.124	0.097	0.15

Whole: Whole Cohort; Mod.-severe: Moderate-to-Severe TBI. AUC: Area Under the ROC Curve; Intercept: Calibration Intercept; E:O: Expected/Observed Ratio; Null Brier: $p(1 - p)$, the Score of a Model Predicting the Observed Frequency for Everyone; ECE: Mean Calibration Error; CI: Confidence Interval. Ideal Values: High AUC, Intercept 0, Slope 1, E:O 1, Brier Below Null Brier, ECE Close to 0. The 95% CIs of All Calibration Metrics are in Table S1 (Supplementary Material).

Recalibration-in-the-large (updated intercept: -1.44 for mortality, -1.99 for unfavourable outcome) removes the overall gap by construction (expected/observed ratio: 1.0). It therefore does not, alone, prove good calibration.

Logistic recalibration gives a slope of 0.68 to 0.81 for the Core model and 0.51 to 0.60 for the Extended model, with wide intervals. Discrimination does not change: recalibration is a monotone transformation (Figure 2).

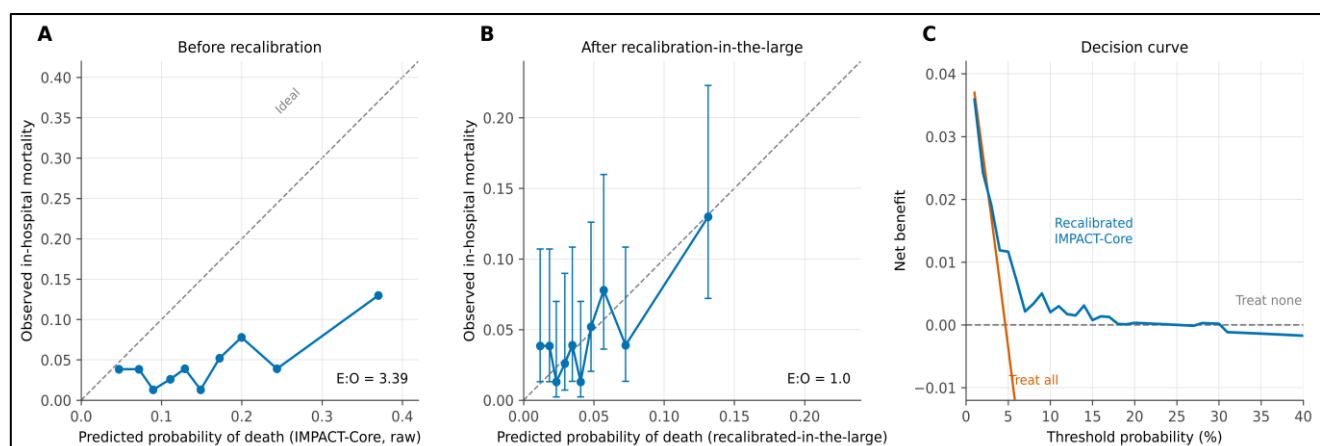


Fig 2 Calibration of IMPACT-Core Before (A) and After Recalibration-in-the-Large (B, with 95% CI), and Decision Curve (C). n = 772; 36 Deaths.

On the decision curve (Figure 2C), the recalibrated model brings a net benefit above the “treat all” and “treat none” strategies for thresholds of about 5% to 30%. This benefit peaks around 5–10% (about 0.01), stays very small beyond 15% (0.003 at most) and turns slightly negative around 31–40%. The raw model brings no net benefit, because of its overestimation. Computed on the recalibration data and without a defined clinical decision, this curve remains illustrative.

➤ Secondary Analysis (Pre-Specified): Adapted GCS-P, Decomposition and CRASH-Basic

The adapted GCS-P discriminates better than IMPACT-Core (Table 4, Figure 3). In the whole cohort, its AUC for mortality is 0.79 (95% CI 0.71–0.88), against 0.64; the difference is significant (DeLong test, $p < 0.001$), for mortality as for unfavourable outcome. In moderate-to-severe TBI it is not ($p = 0.16$ for mortality, $p = 0.41$ for unfavourable outcome). We evaluated only the discrimination of this integer score, not its calibration.

Table 4 Discrimination of the Adapted GCS-P Versus IMPACT-Core (DeLong Test).

Population	Endpoint	AUC adapted GCS-P (95% CI)	AUC IMPACT-Core	Δ AUC	p (DeLong)
Whole	Mortality	0.79 (0.71–0.88)	0.64	+0.15	< 0.001
Whole	Unfavourable	0.79 (0.71–0.87)	0.66	+0.14	< 0.001
Mod.-severe	Mortality	0.73 (0.61–0.83)	0.66	+0.06	0.16
Mod.-severe	Unfavourable	0.70 (0.58–0.80)	0.67	+0.03	0.41

AUC: Area Under the ROC Curve; CI: Confidence Interval; Δ AUC: Difference in AUC. The Higher the AUC, the Better the Discrimination.

Table 5 presents the post hoc decomposition. Total GCS alone reaches 0.78, almost as much as the adapted GCS-P (0.79). Anisocoria alone discriminates modestly (0.65) and adds almost nothing to the GCS; age alone discriminates poorly (0.56). We reviewed the files of the 36 deaths: five patients had bilateral unreactive mydriasis and were moved to “no reactive pupil”. The AUC of the GCS-P so corrected reaches 0.80 (95% CI 0.72–0.89). As only the files of the deaths were reviewed, this reclassification is asymmetric and tends to inflate the AUC; it is a sensitivity analysis and not an estimate. With

the two alternative motor-score tables, IMPACT-Core gives 0.63 and 0.65 (0.64 with the main table): conclusions do not change.

The original CRASH-Basic reaches an AUC of 0.78 (95% CI 0.69–0.86) for mortality. Its calibration is clearly better than that of IMPACT, with moderate overestimation of risk: intercept -0.38 (-0.72 to -0.08), expected/observed ratio 1.46 (1.12–1.92), slope 0.89 (0.50–1.29), a wide interval.

Table 5 Decomposition of Discrimination for in-Hospital Mortality (Whole Cohort, $n = 772$; 36 Deaths).

Score or Model	AUC (95% CI)	Remark
Age alone	0.56 (0.46–0.66)	—
Anisocoria alone	0.65 (0.56–0.74)	—
Total GCS alone	0.78 (0.69–0.86)	—
Adapted GCS-P (GCS and anisocoria)	0.79 (0.71–0.88)	—
Adapted GCS-P after pupil review among deaths	0.80 (0.72–0.89)	Sensitivity; 5 deaths reclassified
CRASH-Basic (original equation, low- and middle-income countries)	0.78 (0.69–0.86)	Expected/observed ratio 1.46
IMPACT-Core	0.64 (0.53–0.74)	0.63–0.65 depending on the motor-score table

AUC: Area Under the ROC Curve; GCS: Glasgow Coma Scale. Values Rounded to Two Decimals; Post hoc Analyses.

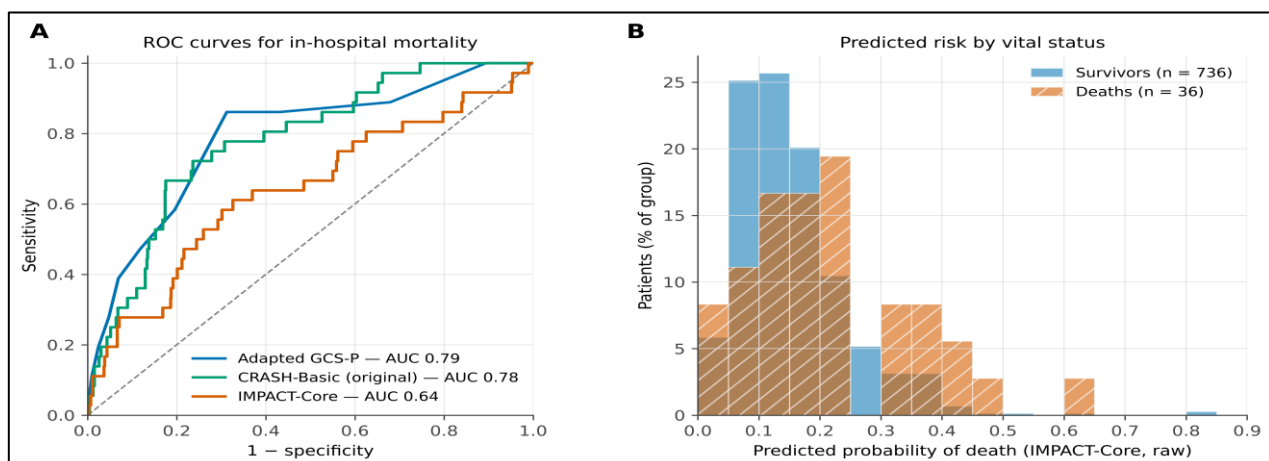


Fig 3 Discrimination of the models compared (A) and distribution of the risk predicted by IMPACT-Core by vital status (B). $n = 772$; 36 Deaths. AUC: Adapted GCS-P 0.79; Original CRASH-Basic 0.78; IMPACT-Core 0.64.

➤ *Exploratory Analyses: Local Models*

A local “GCS-PA” model (adapted GCS-P, age, CT category) reaches an optimism-corrected AUC of 0.79 for mortality, with a slope close to 1 (0.97). Neither age nor CT category brings independent information, but 36 deaths cannot exclude a modest effect. A model reusing the CRASH-Basic variables (age, total GCS, pupillary reactivity, extracranial injury), fitted locally, reaches a corrected AUC of 0.78, comparable to that of the adapted GCS-P, with no gain from CT variables. This model is local development; it must not be confused with the validation of the original CRASH-Basic (Section 3.3). Fitted and validated internally on the same data, these models generate hypotheses. They require independent external validation before any clinical use. Their performance proves no superiority over existing models: it only shows that local adaptation is feasible.

➤ *Sensitivity Analysis in Adults (≥ 16 Years)*

In the 686 adults (32 deaths, 35 unfavourable outcomes), the results hold (details in Table S2). IMPACT still overestimates risk (expected/observed ratio 3.7 for mortality and 5.0 for unfavourable outcome; intercept -1.52 and -2.07) and discriminates at 0.68. The adapted GCS-P remains superior (AUC 0.83; $p < 0.001$ for both endpoints). The estimates are close to those of the main analysis and conclusions do not change.

IV. DISCUSSION

➤ *Main Finding*

We wanted to know what IMPACT is worth in an independent cohort, then what local recalibration adds. The central result concerns calibration: IMPACT-Core strongly overestimates risk, by a factor of 2 to nearly 5 depending on population and endpoint, and its Brier score remains above that of a null model. Recalibrating the intercept corrects the overall gap, but by construction, and does not create the discrimination that is lacking. Ranking patients well does not guarantee predicting an accurate individual risk: calibration decides clinical use.

➤ *Why Does IMPACT Overestimate Risk Here?*

Several mechanisms add up. The first is the horizon. IMPACT predicts six-month mortality; we measure in-hospital mortality, and all our deaths occur before day 15. Part of the predicted risk therefore corresponds to late deaths that our endpoint does not count. Without six-month follow-up we cannot quantify this effect, but it points in the direction of the observed gap; other validations also used in-hospital mortality [14]. CRASH-Basic, whose 14-day horizon matches our endpoint, overestimates far less (expected/observed ratio 1.46 versus 3.4). This points the same way, without separating the effect of horizon from that of the composition of the two models.

The second is the domain. IMPACT was developed in moderate-to-severe TBI; 56% of our series is mild TBI. Overestimation is stronger in the whole cohort (ratio 3.4) than in moderate-to-severe TBI (2.0). That subgroup is

therefore the true validation population. The domain does not explain everything: IMPACT-Core predicted in-hospital mortality well beyond the GCS 12 cut-off in a US national database [14]. The third is baseline risk. With equal predictors, an observed mortality far below that of the development cohorts shifts the intercept towards negative values. Care may explain it, and so may selection: patients who died before admission are absent, and extremely deep comas are probably under-represented. A younger and milder population mostly changes the spread of predictors, hence discrimination, rather than the level of conditional risk. Invoking “case-mix” alone does not explain an intercept defect.

The fourth is implementation. The motor score rebuilt from the total GCS and anisocoria used in place of pupillary reactivity move our model away from the original (Section 2.4). This explanation remains partial: IMPACT-Core never exceeds 0.65, whatever the motor-score table, while GCS alone reaches 0.78. Age, poorly discriminating in this young cohort (AUC 0.56), and the loss of information in the rebuilt motor score are the most plausible explanations; our analyses cannot tell them apart. These leads illustrate the limits of transportability of prognostic models to populations whose case-mix, care pathways and practices differ. External validation is often only the first step of local adaptation.

➤ *Why Does the Simplified Score Discriminate Better than IMPACT?*

The adapted GCS-P ranks patients better than IMPACT-Core, but the comparison calls for caution. Total GCS alone already reaches 0.78, against 0.79 for the adapted GCS-P: the advantage over IMPACT comes almost entirely from the GCS. Anisocoria discriminates modestly (0.65), does not detect bilateral mydriasis and adds almost nothing. After reclassification of five deaths, the AUC reached 0.80, a value potentially inflated by the asymmetric nature of the review. IMPACT-Core, computed with a motor score rebuilt in a few categories and an age that is poorly informative in a young cohort (AUC 0.56), stays at 0.63–0.65 whatever the motor-score table. The 0.15 gap therefore reflects the implementation and composition of the model, not necessarily its worth in its own domain; it was moreover not significant in moderate-to-severe TBI.

Total GCS keeps a practical advantage in low-resource registers: it is recorded routinely, whereas the motor score requires a standardised assessment that is often undocumented. In an intubated or sedated patient, the GCS must be taken before intubation or sedation, as we did; this requires recording from arrival. The local models, CRASH-Basic and the two scores converge: most of the discriminating power comes from the total GCS, pupillary reactivity, as we could record it, adding little. These results do not show that the GCS-P is better than IMPACT in itself. They show that a score built on routinely available variables can rank patients better when a complex model cannot be applied under its original conditions.

➤ *African Specificities*

The median delay of two days, mostly non-medicalised transport and one patient in two referred from another facility draw a long and uneven prehospital pathway. It may select patients who arrive alive at the referral centre. Late access to CT, limited critical care and different neurosurgical practices also change baseline risk relative to development cohorts. This selection may contribute to the low mortality observed, hence to overestimation by imported models.

➤ *Clinical Applicability*

IMPACT should not be used in its raw form: it strongly exaggerates the risk of death. A locally recalibrated version will be interpretable only after independent validation, because our recalibration is in-sample. The adapted GCS-P, computed from total GCS and pupillary reactivity, can help order patients at admission (triage, prioritisation). It does not, as it stands, give an individual risk, because we did not evaluate its calibration. No model replaces clinical judgement, and none suffices alone to justify limiting treatment. To choose a tool, one must look at its calibration in the population concerned, not only its discrimination. Impact studies will show whether a recalibrated model or a simplified score changes decisions or patient outcomes.

➤ *Comparison with the Literature*

Our results confirm a repeated finding: the discrimination of prognostic models varies little from one cohort to another, whereas their calibration depends on context and recruitment [6]. In moderate-to-severe TBI cohorts from high-income countries, IMPACT and CRASH discriminate well (AUC about 0.77 to 0.88 depending on the cohort) [15] [16], with variable calibration gaps, including overestimation of mortality in a recent North American cohort [17] and calibration of moderate quality in a European one [16]. Validations in low- and middle-income countries also show variable calibration [18] [19], and a systematic review of moderate-to-severe TBI models stresses the scarcity of validations in low-resource settings [20].

IMPACT was designed for moderate-to-severe TBI; applying it to mild forms, the majority here, may disadvantage it, but not necessarily: in a US national database IMPACT-Core still predicted in-hospital mortality well beyond GCS 12 [14]. CRASH-Basic, whose equations are published and which exists in a version for low- and middle-income countries, obtains here an AUC of 0.78, comparable to that of GCS alone, with a much smaller overestimation of risk than IMPACT (expected/observed ratio 1.46). In this series it is the published model that fits best. Our non-systematic search identified no published validation in West Africa. This work illustrates the value of a simplified score when resources are limited.

➤ *Generalisability and Perspectives*

The Bouaké cohort resembles other West African settings (Mali, Burkina Faso, Ghana, Benin): a young population, road crashes, a constrained prehospital

pathway. Our results may transpose there; this remains to be checked locally. The most direct next step is to confirm CRASH-Basic's performance in an independent cohort: its 14-day horizon matches the available endpoint, and its calibration, better than IMPACT's, remains to be confirmed. Multicentre validation, first in Côte d'Ivoire then regionally, should follow and, eventually, a model built on the variables truly available in the region, along the lines of recent local-development work [21].

➤ *Strengths*

Our study rests on consecutive recruitment from an admissions register. It evaluates calibration in detail (intercept, slope, expected/observed ratio, mean calibration error, Brier score against a null model, curves), beyond discrimination alone. It follows TRIPOD and compares IMPACT with a simplified score that can be applied immediately. To our knowledge, it is one of the first geographical evaluations of IMPACT in West Africa.

➤ *Limitations*

• *Method:*

The GCS motor score was not recorded; we estimated it from the total GCS with a table not validated locally. Pupillary reactivity was approximated by anisocoria, which classes bilateral unreactive mydriasis among reactive pupils. The Marshall classification was reconstructed. In intubated patients, we used the last GCS recorded before intubation or sedation. No patient has a GCS below 5; this must not be read as an absence of severe TBI (61 patients, 7.9%, have a $GCS \leq 8$), but extremely deep comas are probably under-represented. IMPACT-Lab could not be evaluated for lack of haemoglobin. The validation of CRASH-Basic, the decomposition of discrimination and the sensitivity analyses are post hoc. The pupil review covered only the deaths, which is likely to inflate the AUC of the corrected GCS-P. The recalibrated model, the adapted GCS-P and the local models were not independently validated; the calibration of the adapted GCS-P was not evaluated. Recalibration and the decision curve are evaluated in-sample.

• *Context:*

With 36 deaths, estimates remain imprecise, especially the calibration slope, and the power of comparisons is limited; modest differences between models call for caution. The study is single-centre, retrospective and register-based. It includes children (11.1% under 16) although the models target adults; the analysis in adults nevertheless confirms the results. In-hospital mortality equals 14-day mortality here, but only imperfectly replaces IMPACT's six-month mortality. Functional outcome was assessed with a median follow-up of 3.2 months instead of six, and 69 patients have no GOS. Only three survivors are classed GOS 2–3: the secondary endpoint almost merges with mortality, and disability may have gone unnoticed in telephone calls or file reviews. Selection bias is likely, since patients who died before admission are not included. Acute intoxication was not coded in the register. Finally, our study is not a temporal validation.

V. CONCLUSIONS

In this young, mostly mild West African cohort, IMPACT, computed with the available variables, strongly overestimates risk. Its six-month horizon is not that of our endpoint, and most patients fall outside its domain. It must be recalibrated locally, then independently validated, before any clinical use. Total GCS, alone or combined with anisocoria (adapted GCS-P), and CRASH-Basic obtain similar AUCs (0.78 to 0.79), above that of IMPACT; CRASH-Basic is moreover far less miscalibrated (expected/observed ratio 1.46). The adapted GCS-P is the simplest tool to apply with our data. This does not prove that it outperforms IMPACT applied under its original conditions, and its calibration remains to be evaluated. Multicentre validation in West Africa is needed. More broadly, the calibration of any prognostic model must be checked before it is used in a setting different from that of its development.

➤ *Authors' Contributions*

All authors made a substantial contribution to the conception of the study, data collection, analysis and interpretation of results, or drafting and critical revision of the manuscript. All read and approved the submitted version and collectively take responsibility for it.

➤ *Availability of Data and Materials*

The anonymised data and the analysis code are available from the corresponding author on reasonable request, in compliance with applicable ethics regulations.

➤ *Consent to Publish*

Not applicable: the manuscript contains no individual data allowing a patient to be identified.

➤ *Funding*

This study received no specific funding.

➤ *Conflicts of Interest*

The authors declare no conflicts of interest regarding the publication of this article.

➤ *Acknowledgements*

The authors thank the staff of the neurosurgery department of Bouaké University Hospital for their help with data collection.

➤ *List of Abbreviations*

AUC, area under the ROC curve; CRASH, Corticosteroid Randomisation After Significant Head Injury; ECE, mean calibration error; E:O, expected/observed ratio; EDH, extradural haematoma; GCS, Glasgow Coma Scale; GCS-P, Glasgow Coma Scale-Pupils score; GOS, Glasgow Outcome Scale; IMPACT, International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury; IQR, interquartile range; PROCAST, Prediction model Risk Of Bias ASsessment Tool; SAH, subarachnoid haemorrhage; SBP, systolic blood pressure; SDH, subdural haematoma; SpO₂, pulse oxygen saturation; TBI, traumatic brain injury; TRIPOD, Transparent Reporting of a multivariable

prediction model for Individual Prognosis Or Diagnosis; 95% CI, 95% confidence interval.

REFERENCES

- [1]. Eaton, J., Hanif, A.B., Grudziak, J. and Charles, A. (2017) Epidemiology, Management, and Functional Outcomes of Traumatic Brain Injury in Sub-Saharan Africa. *World Neurosurgery*, 108, 650-655.
- [2]. Steyerberg, E.W., Mushkudiani, N., Perel, P., et al. (2008) Predicting Outcome after Traumatic Brain Injury: Development and International Validation of Prognostic Scores Based on Admission Characteristics. *PLoS Medicine*, 5, e165. <https://doi.org/10.1371/journal.pmed.0050165>
- [3]. 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. <https://doi.org/10.1136/bmj.39461.643438.25>
- [4]. Brennan, P.M., Murray, G.D. and Teasdale, G.M. (2018) Simplifying the Use of Prognostic Information in Traumatic Brain Injury. Part 1: The GCS-Pupils Score. *Journal of Neurosurgery*, 128, 1612-1620.
- [5]. Murray, G.D., Brennan, P.M. and Teasdale, G.M. (2018) Simplifying the Use of Prognostic Information in Traumatic Brain Injury. Part 2: Graphical Presentation of Probabilities. *Journal of Neurosurgery*, 128, 1621-1634.
- [6]. Van Calster, B., McLernon, D.J., van Smeden, M., et al. (2019) Calibration: The Achilles Heel of Predictive Analytics. *BMC Medicine*, 17, Article No. 230.
- [7]. Collins, G.S., Reitsma, J.B., Altman, D.G. and Moons, K.G.M. (2015) Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Annals of Internal Medicine*, 162, 55-63.
- [8]. Wolff, R.F., Moons, K.G.M., Riley, R.D., et al. (2019) PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, 170, 51-58.
- [9]. Vaca, S.D., Kuo, B.J., Nickenig Vissoci, J.R., et al. (2023) A Feasibility Assessment of a Traumatic Brain Injury Predictive Modelling Tool at Kilimanjaro Christian Medical Center and Duke University Hospital. *PLOS Global Public Health*, 3, e0002154.
- [10]. Collins, G.S., Moons, K.G.M., Dhiman, P., et al. (2024) TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods. *BMJ*, 385, e078378.
- [11]. Steyerberg, E.W. (2019) *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2nd Edition, Springer.
- [12]. Vickers, A.J. and Elkin, E.B. (2006) Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Medical Decision Making*, 26, 565-574.

- [13]. Riley, R.D., Debray, T.P.A., Collins, G.S., et al. (2021) Minimum Sample Size for External Validation of a Clinical Prediction Model with a Binary Outcome. *Statistics in Medicine*, 40, 4230-4251.
- [14]. Camarano, J.G., Ratliff, H.T., Korst, G.S., Hrushka, J.M. and Jupiter, D.C. (2021) Predicting In-Hospital Mortality after Traumatic Brain Injury: External Validation of CRASH-Basic and IMPACT-Core in the National Trauma Data Bank. *Injury*, 52, 147-153.
- [15]. Roozenbeek, B., Lingsma, H.F., Lecky, F.E., et al. (2012) Prediction of Outcome after Moderate and Severe Traumatic Brain Injury: External Validation of the IMPACT and CRASH Prognostic Models. *Critical Care Medicine*, 40, 1609-1617. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3335746>
- [16]. Dijkland, S.A., Helmrich, I.R.A.R., Nieboer, D., et al. (2021) Outcome Prediction after Moderate and Severe Traumatic Brain Injury: External Validation of Two Established Prognostic Models in 1742 European Patients. *Journal of Neurotrauma*, 38, 1377-1388. <https://doi.org/10.1089/neu.2020.7300>
- [17]. Yue, J.K., Lee, Y.M., Sun, X., et al. (2024) Performance of the IMPACT and CRASH Prognostic Models for Traumatic Brain Injury in a Contemporary Multicenter Cohort: A TRACK-TBI Study. *Journal of Neurosurgery*, 141, 417-429. <https://doi.org/10.3171/2023.11.JNS231425>
- [18]. Wongchareon, K., Thompson, H.J., Mitchell, P.H., Barber, J. and Temkin, N. (2020) IMPACT and CRASH Prognostic Models for Traumatic Brain Injury: External Validation in a South-American Cohort. *Injury Prevention*, 26, 546-554. <https://doi.org/10.1136/injuryprev-2019-043466>
- [19]. Hashemi, B., Amanat, M., Baratloo, A., et al. (2016) Validation of CRASH Model in Prediction of 14-Day Mortality and 6-Month Unfavorable Outcome of Head Trauma Patients. *Emergency (Tehran)*, 4, 196-201.
- [20]. Dijkland, S.A., Foks, K.A., Polinder, S., Dippel, D.W.J., Maas, A.I.R., Lingsma, H.F. and Steyerberg, E.W. (2020) Prognosis in Moderate and Severe Traumatic Brain Injury: A Systematic Review of Contemporary Models and Validation Studies. *Journal of Neurotrauma*, 37, 1-13.
- [21]. Karabacak, M., Jagtiani, P., Dams-O'Connor, K., Legome, E., Hickman, Z.L. and Margetis, K. (2024) The MOST (Mortality Score for TBI): A Novel Prediction Model beyond CRASH-Basic and IMPACT-Core for Isolated Traumatic Brain Injury. *Injury*, 55, Article ID: 111956. <https://doi.org/10.1016/j.injury.2024.111956>