

Predicting therapeutic neurosurgical intervention within 7 days after initial nonoperative management of mild traumatic brain injury with subdural hematoma: development and internal validation of a clinical-radiographic model

Po-Hsin Lin

Linkou Chang Gung Memorial Hospital

Zhuo-Hao Liu

Linkou Chang Gung Memorial Hospital

Ching-Chang Chen

Linkou Chang Gung Memorial Hospital

Yi-Ting Chen

Linkou Chang Gung Memorial Hospital

Chun-Ting Chen

Linkou Chang Gung Memorial Hospital

Ting-Wei Chang

Linkou Chang Gung Memorial Hospital

Yu-Chi Wang

Linkou Chang Gung Memorial Hospital

Shuo-Chi Chien

shuochichien@gmail.com

Linkou Chang Gung Memorial Hospital

Research Article

Keywords: mild traumatic brain injury, subdural hematoma, neurosurgical intervention, prediction model, Firth logistic regression

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11027621/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Purpose

To develop and internally validate an admission model estimating subsequent therapeutic cranial surgery within 7 days among adults with mild traumatic brain injury and traumatic subdural hematoma selected for initial nonoperative management.

Methods

This retrospective cohort included 745 consecutive adults with Glasgow Coma Scale scores of 13–15 and traumatic subdural hematoma, with or without concomitant intracranial lesions, treated at a tertiary center from 2019 to 2021. All were initially managed nonoperatively; patients selected for immediate or emergency surgery were excluded. The endpoint was an adjudicated therapeutic cranial operation for hematoma evacuation or decompression within 7 days. This treatment-decision endpoint did not represent intervention-independent biological progression. A prespecified Firth logistic regression model used admission variables and underwent 1000-resample bootstrap validation.

Results

The 7-day surgical endpoint occurred in 111 patients (14.9%). Maximum subdural hematoma thickness was the dominant predictor (adjusted odds ratio per 5 mm, 3.65; 95% confidence interval, 2.80–4.76). Apparent and optimism-corrected areas under the curve were 0.838 and 0.818; the corrected Brier score was 0.0914 and calibration slope was 0.932. Observed event rates ranged from 3.3% in the low-probability stratum to 62.5% in the very-high-probability stratum.

Conclusion

The model estimated subsequent surgery under observed local practice and may help prioritize surveillance and reassessment after an initial nonoperative decision. It should not be interpreted as predicting biological progression or defining operative indications. External validation at centers with different operative thresholds is required, emphasizing calibration.

Introduction

Mild traumatic brain injury (mTBI) accounts for most traumatic brain injury presentations, and the admission neurological examination is often reassuring. However, patients with a Glasgow Coma Scale (GCS) score of 13–15 and traumatic subdural hematoma (SDH) on computed tomography (CT), with or without concomitant intracranial lesions, constitute a clinically distinct subgroup. When immediate surgery is not indicated, initial nonoperative management with clinical and radiographic observation is

common. Most patients remain stable, but a clinically important minority deteriorate, show radiographic progression, or require subsequent therapeutic neurosurgical intervention. Clinicians must therefore identify which initially observed patients need higher-acuity monitoring, earlier repeat imaging, or intensified neurosurgical surveillance [5, 15, 17–19].

The probability of subsequent therapeutic surgery is not uniform among patients with mTBI and traumatic SDH selected for initial nonoperative management. Large mixed-injury cohorts have shown that intracranial injury pattern is strongly associated with intervention [11, 14, 16]. Studies restricted to traumatic SDH have further demonstrated that quantitative radiographic features, particularly hematoma thickness and mass effect, can support early risk estimation and rule-out strategies [7, 9, 10]. These findings establish the importance of lesion-specific assessment, but isolated-SDH tools do not directly address patients with heterogeneous concomitant EDH, SAH, intracerebral hemorrhage or contusion, and IVH, while categorical management rules do not necessarily provide an individualized probability of subsequent therapeutic surgery.

Important differences also exist in the populations, time horizons, and outcomes used by previous studies. Some models evaluate eventual intervention during hospitalization, whereas others use composite outcomes that include neurological deterioration, mortality, intracranial pressure monitoring, or procedures that are not themselves therapeutic cranial surgery [14, 16–18]. Few studies have developed an internally validated admission model specifically for patients selected for initial nonoperative management, using a prespecified early window and an endpoint restricted to subsequent therapeutic cranial operations. Such a model could quantify the risk of early management escalation, provided that it is used to support surveillance rather than as a substitute for operative judgment.

We therefore developed and internally validated an admission-based clinical-radiographic model to estimate the probability of a subsequent qualifying therapeutic cranial neurosurgical operation within 7 days among patients with mTBI and traumatic SDH, with or without concomitant intracranial lesions, after an initial decision for nonoperative management. The prediction target was the observed treatment event under the study center's clinical pathway, not biological hematoma progression independent of treatment decisions. We assessed discrimination, overall prediction error, calibration, bootstrap-corrected performance, clinically interpretable probability strata, and net benefit across prespecified thresholds, and we examined robustness through targeted sensitivity analyses.

Methods

Study Design and Population

We conducted a retrospective cohort study of consecutive adult patients with mTBI and traumatic SDH identified on the initial CT who presented to Linkou Chang Gung Memorial Hospital between 2019 and 2021. Concomitant EDH, SAH, intracerebral hemorrhage or contusion, and IVH were permitted. Eligible patients had an admission GCS score of 13–15 and were selected for initial nonoperative management

with clinical and radiographic observation after neurosurgical evaluation. Thus, the intended-use population was patients for whom an initial operative indication had not been judged to be present; it did not include the full spectrum of traumatic SDH. Patients selected for immediate or emergency surgery at the initial evaluation (n = 73) were excluded, as were patients with GCS < 13 (n = 449), major extracranial injury defined as a non-head Abbreviated Injury Scale score ≥ 3 (n = 160), or insufficient required clinical, radiographic, or operative data (n = 43).

Time zero was defined as arrival at the emergency department of the treating neurosurgical center. For patients transferred from another hospital, arrival at the study hospital emergency department was used as the reference time because the model was intended to support admission risk stratification after the decision to begin nonoperative observation. All 745 patients in the analytic cohort underwent an initial nonoperative observation period. All predictors in the primary model were available for every patient; therefore, the primary analysis used the complete cohort without exclusion for missing predictor data.

The study protocol was approved by the Institutional Review Board of the Chang Gung Medical Foundation (IRB No. 202600525B0C5001). The requirement for individual informed consent was waived because this was a retrospective analysis of routinely collected, de-identified clinical data. No separate study protocol was prepared, and the study was not prospectively registered.

Outcome Definition

The primary endpoint was a subsequent qualifying therapeutic cranial neurosurgical operation within 7 days after time zero for traumatic intracranial pathology. This was a treatment-decision endpoint: it represented escalation from initial observation under the study center's practice and was not defined as radiographic progression, neurological deterioration, or unavoidable biological need for surgery. A qualifying event required an operation intended to evacuate an intracranial hematoma and/or relieve traumatic mass effect, including craniotomy or craniectomy with hematoma evacuation and decompressive craniectomy. Classification was based on documented therapeutic intent and the operative components actually performed.

Monitoring-only procedures, diagnostic procedures, and drainage-only procedures without qualifying hematoma evacuation or decompression were excluded. All candidate events were adjudicated by Shuo-Chi Chien, MD, a neurosurgeon, through review of the operative note, the procedure actually performed, and the surrounding clinical record; timing or a procedure code alone was insufficient for event classification. A case was classified as an event only when the documented operative components included qualifying hematoma evacuation and/or decompression for traumatic intracranial pathology. When an otherwise excluded procedure accompanied a qualifying therapeutic operation during the same operative episode, classification was based on the qualifying evacuation or decompression. A first qualifying operation at or before 7 days after time zero met the primary endpoint. Patients with no qualifying operation within 7 days, including those with no surgery during hospitalization and those whose first qualifying operation occurred more than 7 days after presentation, were classified as not

meeting the primary endpoint. Outcome adjudication was not blinded to predictor information because endpoint classification required review of operative notes and the surrounding clinical record.

Candidate Predictors and Primary Model Specification

Candidate predictors were selected a priori on the basis of clinical relevance, availability at admission, and expected relevance to subsequent therapeutic surgery after initial nonoperative management. Automated stepwise selection was not used. The prespecified primary model included age, neurological status, vascular comorbidities, antiplatelet exposure, and admission CT lesion characteristics.

Age was modeled per 5-year increase. Neurological status was modeled as GCS 15 versus GCS 13–14. Comorbidities included hypertension, diabetes mellitus, and dyslipidemia. Antiplatelet use before injury was included as a medication-related predictor. CT predictors included maximum subdural hematoma (SDH) thickness, modeled per 5-mm increase; epidural hematoma (EDH); subarachnoid hemorrhage (SAH); intracerebral hemorrhage or contusion (ICH/contusion); and intraventricular hemorrhage (IVH). SAH was defined as the presence of sulcal SAH or basal cistern SAH.

CT predictors were defined from the initial head CT available at admission, before any interval clinical or radiographic evolution. Two physicians (Zhuo-Hao Liu, MD, and Ping-Chien Li, MD) assessed admission CT variables, including maximum SDH thickness and concomitant intracranial lesions. Maximum thickness was the greatest thickness of the traumatic subdural collection on the admission CT; subsequent CT measurements were not used as baseline predictors. Inter-rater and intra-rater reliability were not quantified.

Statistical Analysis

Continuous variables were summarized as mean (standard deviation [SD]) or median (interquartile range [IQR]), as appropriate. Categorical variables were summarized as count (percentage). Baseline characteristics were compared according to 7-day surgical endpoint status using Welch t tests for approximately normally distributed continuous variables, Mann-Whitney U tests for skewed continuous variables, and chi-square or Fisher exact tests for categorical variables, as appropriate. These comparisons were descriptive and were not used for model selection.

The primary multivariable model was estimated using Firth penalized logistic regression. This approach was selected to reduce small-sample and separation-related bias, particularly for uncommon imaging predictors such as IVH. Adjusted odds ratios (aORs), Wald 95% confidence intervals (CIs), and two-sided P values were reported. Statistical significance was defined as a two-sided P value < .05.

Study size was fixed by inclusion of all consecutive eligible patients available during the study period; no a priori formal sample-size calculation was performed. The primary model contained 11 predictor parameters and 111 endpoint events, corresponding to 10.1 events per parameter. This ratio was reported descriptively and was not treated as proof of adequacy. Model complexity was restricted a

priori, and Firth estimation together with bootstrap internal validation was used to reduce small-sample bias and quantify optimism. Because an anticipated Cox-Snell R-squared had not been specified before analysis, the retrospective fixed-cohort assessment was not presented as an a priori Riley sample-size calculation [12].

For reproducibility, the primary model coefficients, predictor coding, and intercept were reported in Supplementary Table S1. Predicted probability of the 7-day surgical endpoint was calculated as:

Predicted risk = $1/[1 + \exp(-\text{linear predictor})]$.

where the linear predictor was the model intercept plus the sum of each regression coefficient multiplied by its coded predictor value.

Model Performance, Calibration, and Internal Validation

Discrimination was assessed using the area under the receiver operating characteristic curve (AUC). Overall prediction accuracy was assessed using the Brier score. Calibration-in-the-large was estimated by logistic regression with the original model linear predictor as an offset, fixing its slope at 1 and estimating only an intercept. Joint calibration intercept and joint calibration slope were estimated in a separate logistic regression of the observed outcome on the original model linear predictor, with both parameters freely estimated. Calibration was also assessed using decile-based observed event rates with Wilson 95% CIs and locally weighted scatterplot smoothing.

Internal validation was performed using 1000 bootstrap resamples. In each resample, the prespecified Firth model was refitted, and model performance was estimated in both the bootstrap sample and original cohort. Optimism was calculated as the difference between bootstrap-sample and original-cohort performance, and optimism-corrected performance was obtained by subtracting mean optimism from apparent performance. Optimism-corrected AUC, Brier score, and calibration slope were reported.

Probability Stratification and Decision-Curve Analysis

Predicted probabilities of the 7-day surgical endpoint were categorized into four prespecified clinical strata: low probability (< 5%), intermediate probability (5% to < 15%), high probability (15% to < 30%), and very high probability ($\geq 30\%$). For each stratum, the number of patients, number of endpoint events, observed event rate, and mean predicted probability were calculated.

Potential clinical utility was evaluated using decision-curve analysis across threshold probabilities from 5% to 30%, corresponding to the prespecified clinically interpretable range. The model was compared with strategies that classified all or no patients as having a high probability of the 7-day surgical endpoint. Results were interpreted as the potential value of risk-directed observation, repeat imaging, and expedited neurosurgical reassessment; they were not interpreted as evidence that the model should independently determine operative indications. The analysis did not compare the model with

contemporaneous treating-surgeon probability estimates and therefore could not establish incremental value over clinical gestalt.

Sensitivity and Exploratory Analyses

Targeted robustness analyses evaluated sensitivity to the 7-day outcome definition, subgroup restriction, and SDH-thickness parameterization; none was used to select or revise the primary model. First, the primary model was repeated after excluding patients who underwent delayed surgery more than 7 days after presentation, thereby restricting the comparison to patients meeting the 7-day surgical endpoint versus those who did not undergo surgery during hospitalization. Second, the model was repeated in the GCS 15 subgroup; the GCS indicator was omitted because it was constant within this subgroup.

Third, SDH thickness was modeled categorically for clinical interpretability using a three-level specification (< 5 mm, 5-9.9 mm, and ≥ 10 mm) and a binary specification (< 10 mm versus ≥ 10 mm). These categories reflected established clinical cut points and were not treated as a formal test of nonlinearity or as an independently derived threshold.

Because male sex differed between outcome groups in unadjusted comparisons but was not part of the prespecified clinical-radiographic primary model, a post hoc sensitivity model additionally including male sex was fitted. Stability of the primary associations was assessed by comparing aORs for shared predictors between the primary and sex-augmented models. Apparent AUC, Brier score, calibration intercept, and calibration slope, together with optimism-corrected AUC, Brier score, and calibration slope, were compared using paired bootstrap resampling. Optimism-corrected calibration intercepts were not reported.

Threshold-Specific Classification and Prediction-Error Analysis

Threshold-specific classification performance was evaluated at predicted-risk thresholds of 5%, 10%, 15%, 20%, 30%, and the Youden-derived threshold. Sensitivity, specificity, positive predictive value, negative predictive value, and corresponding true-positive, false-positive, true-negative, and false-negative counts were calculated.

A post hoc exploratory prediction-error analysis was conducted to characterize clinically relevant model misclassification. False-negative cases were defined as patients with predicted probability $< 10\%$ who nevertheless met the 7-day surgical endpoint. False-positive cases were defined as patients with predicted probability $\geq 30\%$ who did not meet the endpoint. Post-admission deterioration and repeat CT findings were reviewed only to contextualize misclassification and were not entered as baseline predictors.

No formal model-fairness analysis or algorithmic fairness-mitigation procedure was prespecified or performed. The post hoc sex-augmented sensitivity analysis evaluated coefficient stability and

incremental model performance but was not designed as a fairness analysis. All statistical analyses were performed using Python version 3.12 (Python Software Foundation).

Results

Study Cohort and Endpoint Distribution

The final analytic cohort included 745 patients with mTBI, admission GCS scores of 13–15, and traumatic SDH, with or without concomitant traumatic intracranial lesions, all of whom were initially managed nonoperatively after immediate or emergency surgery was excluded (Fig. 1). The 7-day surgical endpoint occurred in 111 patients (14.9%). The remaining 634 patients (85.1%) did not meet the endpoint, including 603 who had no surgery during hospitalization and 31 whose first qualifying operation occurred more than 7 days after presentation.

All predictors included in the primary model were available for the full cohort. Therefore, all 745 patients were included in the complete-case primary Firth logistic regression analysis.

Baseline Characteristics

Baseline characteristics are summarized in Table 1. Mean age was similar between patients who met the 7-day surgical endpoint and those who did not. Male sex was more frequent among patients meeting the endpoint (71.2% vs 54.9%, $P=.001$). Patients meeting the endpoint were less likely to present with GCS 15 and more likely to present with GCS 13–14.

Table 1
Baseline characteristics according to early therapeutic neurosurgical intervention status

Characteristic	Overall (N = 745)	Early intervention ≤ 7 days (n = 111)	No early intervention ≤ 7 days (n = 634)	P value
Age, years, mean (SD)	64.0 (18.7)	63.8 (18.0)	64.0 (18.8)	0.915
Male sex	427 (57.3%)	79 (71.2%)	348 (54.9%)	0.001
Admission GCS category				< 0.001
GCS 15	592 (79.5%)	75 (67.6%)	517 (81.5%)	
GCS 13–14	153 (20.5%)	36 (32.4%)	117 (18.5%)	
Hypertension	318 (42.7%)	43 (38.7%)	275 (43.4%)	0.362
Diabetes mellitus	187 (25.1%)	27 (24.3%)	160 (25.2%)	0.838
Dyslipidemia	15 (2.0%)	2 (1.8%)	13 (2.1%)	1.000
Antiplatelet use	73 (9.8%)	9 (8.1%)	64 (10.1%)	0.516
Maximum SDH thickness, mm, median (IQR)	4.0 (3.0– 5.0)	8.0 (4.3–15.0)	3.6 (2.0–5.0)	< 0.001
EDH	50 (6.7%)	14 (12.6%)	36 (5.7%)	0.007
SAH	355 (47.7%)	40 (36.0%)	315 (49.7%)	0.008
ICH or contusion	162 (21.7%)	32 (28.8%)	130 (20.5%)	0.050
IVH	15 (2.0%)	5 (4.5%)	10 (1.6%)	0.058
Values are mean (SD), median (IQR), or n (%). P values were calculated using Welch t tests, Mann-Whitney U tests, chi-square tests, or Fisher exact tests, as appropriate. Abbreviations: EDH, epidural hematoma; GCS, Glasgow Coma Scale; ICH, intracerebral hemorrhage or contusion; IQR, interquartile range; IVH, intraventricular hemorrhage; SAH, subarachnoid hemorrhage; SD, standard deviation; SDH, subdural hematoma.				

Maximum SDH thickness was greater among patients who met the 7-day surgical endpoint than among those who did not, with median values of 8.0 mm (IQR, 4.3–15.0 mm) and 3.6 mm (IQR, 2.0–5.0 mm), respectively ($P < .001$). EDH was more frequent among patients meeting the endpoint (12.6% vs 5.7%, $P = .007$). ICH or contusion was also more common among patients meeting the endpoint (28.8% vs 20.5%), a difference of borderline statistical significance ($P = .050$). In contrast, SAH was more frequent among patients who did not meet the endpoint in unadjusted comparisons.

Primary Multivariable Firth Logistic Regression Model

Results of the prespecified primary Firth logistic regression model are shown in Table 2 and Fig. 2. Maximum SDH thickness was the strongest predictor of the 7-day surgical endpoint. Each 5-mm increase in SDH thickness was associated with an aOR of 3.65 (95% CI, 2.80–4.76; $P < .001$).

Table 2
Primary multivariable Firth logistic regression model for early therapeutic neurosurgical intervention within 7 days

Predictor	Adjusted OR	95% CI	P value
Age per 5 years	0.98	0.91–1.06	0.671
GCS 15 vs 13–14	0.47	0.28–0.81	0.006
Hypertension	0.52	0.28–0.96	0.037
Diabetes mellitus	1.30	0.69–2.44	0.412
Dyslipidemia	1.80	0.40–8.13	0.446
Antiplatelet use	0.44	0.17–1.15	0.095
SDH thickness per 5 mm	3.65	2.80–4.76	< 0.001
EDH	2.90	1.37–6.13	0.005
SAH	0.77	0.45–1.30	0.322
ICH or contusion	1.99	1.13–3.49	0.017
IVH	4.11	1.12–15.02	0.033
Odds ratios and 95% confidence intervals were estimated using Firth penalized logistic regression. Age was modeled per 5-year increase and maximum SDH thickness per 5-mm increase. SAH included sulcal or basal cistern subarachnoid hemorrhage. The model included 745 patients and 111 early-intervention events.			

Several additional CT features were independently associated with the 7-day surgical endpoint. EDH was associated with an aOR of 2.90 (95% CI, 1.37–6.13; $P = .005$), ICH or contusion with an aOR of 1.99 (95% CI, 1.13–3.49; $P = .017$), and IVH with an aOR of 4.11 (95% CI, 1.12–15.02; $P = .033$). Compared with patients presenting with GCS 13–14, those with GCS 15 had lower odds of meeting the endpoint (aOR, 0.47; 95% CI, 0.28–0.81; $P = .006$).

Hypertension was inversely associated with the 7-day surgical endpoint after multivariable adjustment (aOR, 0.52; 95% CI, 0.28–0.96; $P = .037$). This finding was interpreted as an adjusted statistical association rather than evidence of a protective biological effect. Age, diabetes mellitus, dyslipidemia, antiplatelet use, and SAH were not independently associated with the endpoint in the primary model.

The primary model intercept was -2.5502 . Complete predictor coding, regression coefficients, standard errors, and the prediction equation are provided in Supplementary Table S1.

Model Performance, Calibration, and Internal Validation

Performance of the primary model is summarized in Table 3 and Fig. 3. The model demonstrated an apparent AUC of 0.838 and an apparent Brier score of 0.087. Apparent calibration was favorable, with a joint calibration intercept of 0.000 and a joint calibration slope of 1.03. Calibration-in-the-large was -0.041, indicating slight overall overprediction (mean predicted probability, 15.27%; observed event rate, 14.90%). The calibration plot showed generally close agreement between predicted and observed event rates across the observed risk range, although the majority of patients were concentrated in the lower-risk region.

Table 3

Model performance and bootstrap internal validation of the primary Firth logistic regression model

Performance metric	Apparent performance	Optimism-corrected performance
No. of patients	745	—
No. of early-intervention events	111	—
Area under the ROC curve	0.838	0.818
Brier score	0.087	0.0914
Calibration-in-the-large	-0.041	—
Joint calibration intercept	0.000	—
Joint calibration slope	1.03	0.932
Values are based on the primary Firth penalized logistic regression model for early therapeutic neurosurgical intervention within 7 days after time zero. Apparent performance was evaluated in the development cohort. Optimism-corrected performance was estimated using 1000 bootstrap resamples. Calibration-in-the-large was estimated with the original model linear predictor as an offset and slope fixed at 1. Joint calibration intercept and slope were estimated together in a separate logistic regression. Calibration-in-the-large and joint calibration intercept are apparent estimates; the joint calibration slope is reported before and after optimism correction.		

Bootstrap internal validation with 1000 resamples indicated modest optimism but preserved overall model performance. The optimism-corrected AUC was 0.818, the optimism-corrected Brier score was 0.0914, and the optimism-corrected calibration slope was 0.932. These findings indicate modest overfitting, with preservation of discrimination and acceptable calibration after internal validation.

Probability Stratification

Observed event rates increased across the prespecified predicted-probability strata (Table 4). Among 241 patients classified as low probability (< 5%), 8 met the endpoint, corresponding to an observed event rate of 3.3%. Among 297 intermediate-probability patients (5% to < 15%), 23 met the endpoint, corresponding to an observed event rate of 7.7%.

Table 4
Observed event rate by predicted-risk stratum

Predicted-risk stratum	No. of patients	No. of events	Observed event rate	Mean model-predicted risk
Low risk, < 5%	241	8	3.3%	3.3%
Intermediate risk, 5% to < 15%	297	23	7.7%	8.8%
High risk, 15% to < 30%	111	20	18.0%	20.4%
Very high risk, $\geq$ 30%	96	60	62.5%	59.3%
Predicted-risk strata were based on probabilities estimated by the primary Firth logistic regression model. Observed event rate denotes the proportion of patients undergoing early therapeutic neurosurgical intervention within 7 days after time zero.				

Among 111 high-probability patients (15% to < 30%), 20 met the endpoint, for an observed event rate of 18.0%. In the very-high-probability group ($\geq$ 30%), 60 of 96 patients met the endpoint, corresponding to an observed event rate of 62.5%. Mean predicted probabilities were 3.3%, 8.8%, 20.4%, and 59.3% across the four strata, respectively.

Decision-Curve Analysis

Decision-curve analysis demonstrated that the primary model provided greater net benefit than classify-all and classify-none strategies across the prespecified threshold range of 5% to 30% (Fig. 3). The model's net benefit remained positive throughout this range, whereas the classify-all strategy declined to no net benefit at approximately the observed cohort event rate and became unfavorable at higher threshold probabilities.

These findings suggest potential value for probability-based prioritization of enhanced clinical observation, repeat imaging, or early neurosurgical vigilance. They do not establish superiority to treating-surgeon gestalt because no contemporaneous clinician probability estimate was available, and the decision-curve analysis was not interpreted as support for using the model as an isolated surgical decision rule.

Sensitivity Analyses

After excluding 31 patients who underwent delayed surgery more than 7 days after presentation, the sensitivity cohort included 714 patients: 111 who met the 7-day surgical endpoint and 603 who did not undergo surgery during hospitalization. Discrimination remained strong, with an AUC of 0.849 and a Brier score of 0.0846. SDH thickness remained the dominant predictor, with an aOR of 4.16 per 5-mm increase (95% CI, 3.09–5.60; $P < .001$).

In the GCS 15 subgroup, which included 592 patients and 75 endpoint events, SDH thickness remained strongly associated with the 7-day surgical endpoint (aOR per 5 mm, 3.69; 95% CI, 2.72–5.01; $P < .001$).

The subgroup model had an AUC of 0.836.

Categorical modeling of SDH thickness showed progressively higher odds of the 7-day surgical endpoint. Compared with SDH thickness < 5 mm, thickness of 5-9.9 mm was associated with an aOR of 2.94 (95% CI, 1.71–5.06), whereas thickness $\geq$ 10 mm was associated with an aOR of 30.06 (95% CI, 15.54–58.15). In the binary analysis, thickness $\geq$ 10 mm was associated with an aOR of 19.57 (95% CI, 10.76–35.61; $P<.001$) compared with thickness < 10 mm. Because the categories were defined around established clinical cut points, these estimates did not formally demonstrate nonlinearity or an independently derived 10-mm threshold.

Sensitivity Analysis Including Male Sex

In the sex-augmented sensitivity model, male sex was associated with an aOR of 1.68 (95% CI, 0.99–2.85; $P=.057$) for the 7-day surgical endpoint (Supplementary Table S2). Thus, male sex showed an association of borderline statistical significance after adjustment but did not meet the prespecified threshold for statistical significance.

Importantly, inclusion of sex did not materially alter the associations of the prespecified primary predictors. Across shared predictors, the ratio of aORs comparing the sex-augmented model with the primary model ranged from 0.928 to 1.070. SDH thickness remained the dominant predictor in the sex-augmented model (aOR per 5 mm, 3.62; 95% CI, 2.76–4.73; $P<.001$). EDH, ICH or contusion, IVH, and GCS category retained similar directions and magnitudes of association.

Performance changed minimally after adding sex. The apparent AUC increased from 0.838 to 0.842, while the optimism-corrected AUC increased from 0.818 to 0.821. The apparent Brier score improved from 0.087 to 0.0853, and the optimism-corrected Brier score improved from 0.0914 to 0.0904. The optimism-corrected calibration slope was 0.932 for the primary model and 0.927 for the sex-augmented model. Given the minimal incremental improvement in validated performance and preservation of the prespecified primary associations, the original clinical-radiographic model was retained as the primary model for parsimony and interpretability. Comparative results are presented in Supplementary Table S2.

Threshold-Specific Classification Performance and Prediction-Error Analysis

Threshold-specific classification performance is shown in Supplementary Table S3. At a 5% threshold, the model achieved sensitivity of 92.8% and negative predictive value of 96.7%, with specificity of 36.8%. At a 10% threshold, sensitivity was 82.0%, specificity was 66.2%, positive predictive value was 29.8%, and negative predictive value was 95.5%. At a 30% threshold, specificity increased to 94.3% and positive predictive value increased to 62.5%, whereas sensitivity decreased to 54.1%.

For the exploratory prediction-error analysis, 20 patients had predicted probability < 10% but met the 7-day surgical endpoint. Their mean predicted probability was 6.1%; 95.0% presented with GCS 15, mean SDH thickness was 3.5 mm, no patient had SDH thickness $\geq$ 10 mm, and IVH was absent in all cases.

Clinical deterioration and repeat CT documentation were each identified in 20.0% of these cases. These features occurred after admission and were not available to the baseline model.

Thirty-six patients had predicted probability $\geq 30\%$ but did not meet the 7-day surgical endpoint. Their mean predicted probability was 49.1%, mean SDH thickness was 11.4 mm, 52.8% had SDH thickness ≥ 10 mm, 25.0% had EDH, and 47.2% had ICH or contusion. Eight patients (22.2%) subsequently underwent delayed or other surgery more than 7 days after presentation, whereas 28 patients (77.8%) did not undergo surgery during hospitalization.

Overall, false-negative cases were characterized primarily by relatively mild admission neurological and radiographic features followed by possible subsequent clinical or radiographic progression. False-positive cases tended to have more extensive admission radiographic injury but did not meet the prespecified 7-day surgical endpoint.

Discussion

In this retrospective cohort of 745 patients with mTBI and traumatic SDH selected for initial nonoperative management, 111 patients (14.9%) subsequently underwent a qualifying therapeutic cranial neurosurgical operation within 7 days. The admission-based clinical-radiographic model showed good apparent discrimination (AUC, 0.838), with limited optimism after bootstrap internal validation (optimism-corrected AUC, 0.818; calibration slope, 0.932). Maximum SDH thickness was the dominant predictor, while EDH, intracerebral hemorrhage or contusion, IVH, and admission GCS provided additional predictive information for the observed treatment endpoint. Observed event rates increased from 3.3% in the low-probability stratum to 62.5% in the very-high-probability stratum, and decision-curve analysis suggested potential net benefit across threshold probabilities of 5%-30% when used to target surveillance and reassessment rather than to determine surgery.

The strong association between SDH thickness and the 7-day surgical endpoint is consistent with prior isolated-SDH studies in which increasing hematoma thickness and measures of mass effect were among the most informative radiographic predictors of neurosurgical intervention [7, 9]. The marked increase in the ≥ 10 -mm category is directionally concordant with the classic operative criterion for acute SDH thickness greater than 10 mm [3]. However, because SDH thickness itself informs operative judgment, this concordance is better interpreted as face validity and alignment with prevailing decision policy than as external validation of an intervention-independent biological threshold. Moreover, the category-based sensitivity analysis did not formally establish nonlinearity; large category-specific odds ratios can arise from accumulation of the primary model's steep linear association across multiple 5-mm increments. The linear specification therefore remains a parsimonious working form, but possible miscalibration at the extremes of SDH thickness requires formal spline assessment in an adequately sized dataset.

This study extends, rather than replaces, existing mTBI stratification research. In a National Trauma Data Bank cohort of more than 50,000 patients with GCS 14–15, Sweeney and colleagues demonstrated

substantial variation in intervention rates according to injury type and reported discrimination similar to that of the present model [14]. Subsequent complicated-mTBI studies and the updated Brain Injury Guidelines and Quebec Brain Injury Categories have refined risk classification for disposition, transfer, deterioration, or intervention [16–18]. More recent isolated-SDH models have achieved strong rule-out performance using detailed radiographic variables [9, 10]. The present model addresses a complementary question by estimating the individual 7-day probability of escalation to therapeutic cranial surgery after an initial nonoperative decision among patients with traumatic SDH, including those with concomitant lesion patterns, and by reporting calibration, optimism correction, risk strata, and decision-curve analysis in addition to discrimination.

The lesion-specific associations were clinically coherent but require cautious interpretation. EDH remained independently associated with the 7-day surgical endpoint, and intracerebral hemorrhage or contusion also contributed independently. Dyslipidemia and IVH were each present in only 15 patients. The IVH point estimate was large but imprecise, whereas dyslipidemia provided little apparent signal; neither variable was removed on the basis of its P value because the candidate set was prespecified. Their sparse frequencies nevertheless increase coefficient uncertainty and strengthen the rationale for comparing the full model with a clinically defined parsimonious model in future validation. SAH was not independently associated with the endpoint after adjustment. This is consistent with reports that isolated traumatic SAH in otherwise mild TBI rarely requires neurosurgical intervention [6], but it does not imply that SAH is protective or clinically unimportant in patients with additional lesions.

Admission GCS remained informative despite restriction to the narrow range of 13–15. Patients with GCS 15 had lower adjusted odds of meeting the 7-day surgical endpoint than those with GCS 13–14, yet 75 patients with GCS 15 still met the endpoint and the subgroup model retained good discrimination. Prior work likewise shows that a normal or near-normal examination does not eliminate the risk of deterioration or intervention when traumatic lesions are present [5, 16, 18]. This finding underscores the complementary roles of neurological examination and CT burden in estimating the probability of subsequent intervention under observed practice.

The model was intentionally developed to estimate a policy-conditional treatment outcome after an initial nonoperative decision, not biological progression or whether surgery is indicated. Subsequent therapeutic intervention is influenced by anatomy, clinical evolution, serial imaging, surgeon judgment, institutional practice, treatment limitations, and patient or family preferences. Because several baseline imaging predictors also inform operative decisions, part of the observed discrimination likely reflects reconstruction of the study center's decision process. Interhospital variation in neurosurgical intervention for CT-positive mTBI confirms that the same admission profile may not lead to identical treatment across settings [8]. Accordingly, decision-curve results should be understood as exploratory support for surveillance prioritization. They neither establish that a predicted-probability threshold should trigger surgery nor demonstrate incremental value over treating-surgeon gestalt [21].

The prediction-error analysis reinforces the need for longitudinal reassessment. False-negative cases generally had GCS 15, thin SDH, no IVH, and low estimated admission risk, but a subset later developed clinical or radiographic changes. Conversely, false-positive cases often had substantial CT abnormalities but remained stable, continued nonoperative management, or underwent surgery outside the 7-day window. Delayed deterioration and later surgery after initially nonoperative traumatic SDH are recognized phenomena [1, 5], while contemporary evidence favors selective rather than routine repeat CT in clinically stable mTBI [15]. Future dynamic models should evaluate whether neurological change, interval CT progression, midline shift, basal cistern compression, hematoma volume, and early physiological trajectories add value beyond the admission model.

The sex-augmented sensitivity analysis and the inverse coefficient for hypertension should not be overinterpreted. Male sex differed between outcome groups, but its adjusted association was borderline and its addition produced minimal improvement in optimism-corrected performance while leaving the primary coefficients essentially unchanged. Retaining the prespecified model therefore favored parsimony over post hoc expansion. Similarly, the inverse association with hypertension is unlikely to represent a protective biological effect and may reflect residual confounding, treatment selection, unmeasured frailty or goals of care, or chance. Both findings should be viewed as cohort-specific adjusted associations rather than mechanistic determinants of operative risk.

Several design and reporting features strengthen this study. The cohort was explicitly restricted to patients selected for initial nonoperative management, predictors were chosen a priori from variables available at presentation, and the endpoint was restricted to adjudicated therapeutic cranial surgery within a clinically actionable 7-day window. Model evaluation included discrimination, Brier score, calibration, observed probability strata, decision-curve analysis, and bootstrap optimism correction. This multidimensional assessment is preferable to reliance on AUC alone because discrimination, calibration, and potential clinical utility answer different questions [4, 20, 21]. Reporting the full model equation and coefficients also permits independent external validation and recalibration.

Several limitations remain. First, the retrospective single-center design limits transportability, because operative selection, observation, transfer, repeat-imaging, and operative practices vary across hospitals [8]. Second, excluding patients requiring immediate or emergency surgery intentionally created a narrower, post-selection spectrum; the model applies only after an initial nonoperative decision, not to all patients with traumatic SDH. Third, the endpoint was a decision-mediated treatment event rather than biological progression, and baseline imaging features may influence both prediction and operative judgment. Fourth, decision-curve analysis did not compare the model with clinician gestalt. Fifth, SDH-thickness measurement reliability was not quantified. Sixth, the primary model assumed a linear association between SDH thickness and log odds; the categorical sensitivity analysis did not constitute a formal test of nonlinearity. Seventh, no a priori Riley sample-size calculation was performed, and Firth estimation and bootstrap correction mitigate small-sample bias and quantify optimism but do not establish sample-size adequacy [12]. Eighth, sparse dyslipidemia and IVH frequencies produced uncertain coefficient estimates. Ninth, potentially important mass-effect variables, including midline

shift, basal cistern compression, hematoma volume, lesion laterality, and acute-on-chronic SDH features, were unavailable in structured form [9]. Finally, the model did not incorporate pretransfer evolution, subsequent neurological deterioration, serial CT progression, or dynamic treatment decisions.

The study could not assess potential health inequalities in model performance or predicted intervention risk because detailed socioeconomic and ethnicity-related variables were unavailable. Because the endpoint was treatment-mediated, the model may also reflect local differences in access to neurosurgical observation, follow-up imaging, and operative treatment. Its fairness and transportability should therefore be evaluated before use in other populations or health systems.

The next step is external validation in geographically and clinically distinct centers that use different operative thresholds and observation pathways. Evaluation should emphasize center-specific calibration-in-the-large, calibration slope, and net benefit in addition to discrimination [13, 20]. Systematic differences in calibration may represent transport failure of the prediction model, genuine differences in case mix, differences in decision policy, or a combination of these mechanisms. Future studies should also formally examine nonlinear SDH-thickness effects, quantify measurement reliability, compare the full model with a clinically defined parsimonious model and treating-surgeon estimates, and test whether standardized mass-effect variables and early longitudinal changes add incremental value. Ultimately, a prospective impact study is required before calculator or electronic decision-support implementation [4].

Conclusion

In conclusion, this admission-based clinical-radiographic model showed good discrimination, acceptable calibration, and limited optimism after bootstrap internal validation for estimating subsequent therapeutic cranial surgery within 7 days among patients with mTBI and traumatic SDH selected for initial nonoperative management. Its performance should be interpreted as prediction of a local, decision-mediated treatment endpoint rather than intervention-independent biological progression. The model may support targeted surveillance and timely reassessment, but it should not be used as a stand-alone determinant of operative indication. External validation across centers with differing decision thresholds, together with formal assessment of measurement reliability and SDH-thickness functional form, is required before clinical implementation.

Declarations

Clinical trial number

not applicable.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics approval

The Institutional Review Board of the Chang Gung Medical Foundation approved this retrospective study (IRB No. 202600525B0C5001). The study was conducted in accordance with the Declaration of Helsinki and applicable local regulations.

Consent to participate

The Institutional Review Board waived the requirement for individual informed consent because this was a retrospective analysis of routinely collected, de-identified clinical data.

Consent for publication

Not applicable. This report does not contain identifiable data from an individual person.

Protocol and registration

No separate study protocol was prepared, and the study was not prospectively registered.

Patient and public involvement

Patients and members of the public were not involved in the design, conduct, reporting, or dissemination planning of this retrospective study.

Author Contribution

Po-Hsin Lin contributed to investigation, data curation, interpretation of the results, and writing of the original draft. Zhuo-Hao Liu, Ching-Chang Chen, Yi-Ting Chen, Chun-Ting Chen, Ting-Wei Chang, and Yu-Chi Wang contributed to data acquisition, data interpretation, and critical revision of the manuscript. Shuo-Chi Chien contributed to conceptualization, methodology, formal analysis, data curation, investigation, visualization, writing of the original draft, project administration, and supervision. All authors reviewed and edited the manuscript, approved the final version, and agreed to be accountable for all aspects of the work.

Acknowledgments:

None.

Data Availability

The patient-level data underlying this study are not publicly available because they contain potentially identifiable clinical information. De-identified data may be made available by the corresponding author upon reasonable request, subject to Institutional Review Board approval, institutional data-governance requirements, and execution of an appropriate data-sharing agreement.

Code availability

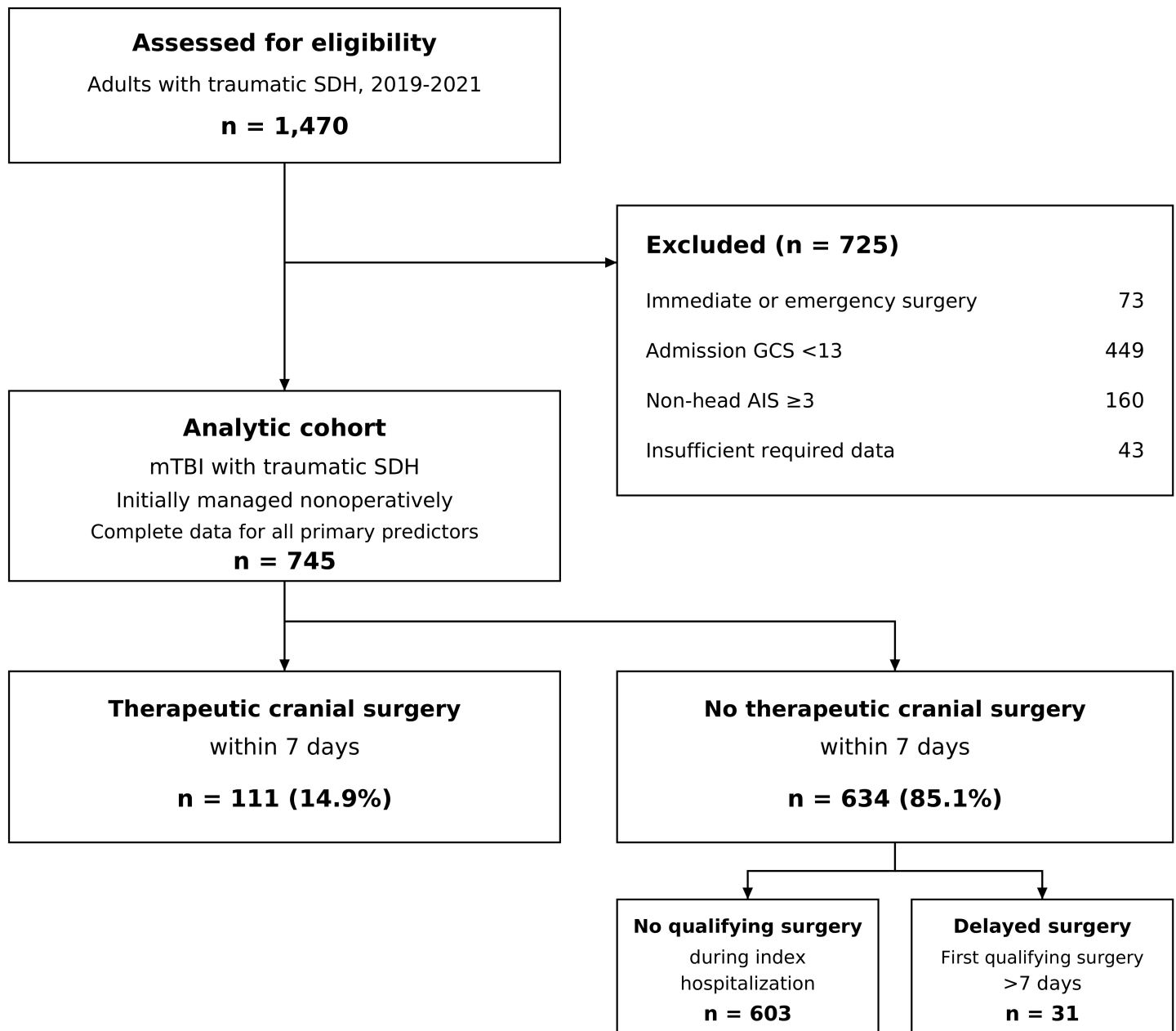
The analytical code is available from the corresponding author upon reasonable request.

References

1. Bajsarowicz P, Prakash I, Lamoureux J et al (2015) Nonsurgical acute traumatic subdural hematoma: what is the risk? *J Neurosurg* 123:1176–1183. <https://doi.org/10.3171/2014.10.JNS141728>
2. Bullock MR, Chesnut R, Ghajar J et al (2006) Surgical management of acute epidural hematomas. *Neurosurgery* 58(Suppl 3). <https://doi.org/10.1227/01.NEU.0000210363.91172.A8>. S7-S15
3. Bullock MR, Chesnut R, Ghajar J et al (2006) Surgical management of acute subdural hematomas. *Neurosurgery* 58(Suppl 3). <https://doi.org/10.1227/01.NEU.0000210364.29290.C9>. S16-S24
4. Collins GS, Moons KGM, 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. <https://doi.org/10.1136/bmj-2023-078378>
5. Marincowitz C, Lecky FE, Townend W et al (2018) The risk of deterioration in GCS 13–15 patients with traumatic brain injury identified by computed tomography imaging: a systematic review and meta-analysis. *J Neurotrauma* 35:703–718. <https://doi.org/10.1089/neu.2017.5259>
6. Nassiri F, Badhiwala JH, Witiw CD et al (2017) The clinical significance of isolated traumatic subarachnoid hemorrhage in mild traumatic brain injury: a meta-analysis. *J Trauma Acute Care Surg* 83:725–731. <https://doi.org/10.1097/TA.0000000000001617>
7. Orlando A, Levy AS, Rubin BA et al (2019) Isolated subdural hematomas in mild traumatic brain injury. Part 1: the association between radiographic characteristics and neurosurgical intervention. *J Neurosurg* 130:1616–1625. <https://doi.org/10.3171/2018.1.JNS171884>
8. Orlando A, Coresh J, Carrick MM et al (2023) Characterizing interhospital variability in neurosurgical interventions for patients with mild traumatic brain injury and intracranial hemorrhage. *Neurotrauma Rep* 4:149–158. <https://doi.org/10.1089/neur.2022.0078>
9. Orlando A, Panchal RR, Mellor L et al (2025) Risk factors for neurosurgical intervention within 48 hours of admission for patients with mild traumatic brain injury and isolated subdural hematoma. *J Neurosurg* 142:547–560. <https://doi.org/10.3171/2024.5.JNS232476>

10. Orlando A, Panchal RR, Quan G et al (2025) Creating and validating a neurosurgical intervention rule-out tool for patients with mild traumatic brain injury and isolated subdural hematoma: a 5-year, six-center retrospective cohort study. *J Neurosurg* 142:839–850. <https://doi.org/10.3171/2024.5.JNS232478>
11. Patel PD, Broadwin M, Choudhry AJ et al (2023) Risk factors associated with neurosurgical intervention in patients with mild traumatic intracranial hemorrhage. *J Surg Res* 283:137–145. <https://doi.org/10.1016/j.jss.2022.10.042>
12. Riley RD, Snell KIE, Ensor J et al (2019) Minimum sample size for developing a multivariable prediction model: Part II-binary and time-to-event outcomes. *Stat Med* 38:1276–1296. <https://doi.org/10.1002/sim.7992>
13. Steyerberg EW, Harrell FE Jr (2016) Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 69:245–247. <https://doi.org/10.1016/j.jclinepi.2015.04.005>
14. Sweeney TE, Salles A, Harris OA, Spain DA, Staudenmayer KL (2015) Prediction of neurosurgical intervention after mild traumatic brain injury using the National Trauma Data Bank. *World J Emerg Surg* 10:23. <https://doi.org/10.1186/s13017-015-0017-6>
15. Taddei G, Caproni S, Pietrantonio A et al (2025) Evidence-based indications for repeat head CT after mild traumatic brain injury: a systematic review and meta-analysis. *Neurosurg Rev* 48:397. <https://doi.org/10.1007/s10143-025-03549-3>
16. Tourigny JN, Boucher V, Paquet V et al (2021) Predictors of neurosurgical intervention in complicated mild traumatic brain injury patients: a retrospective cohort study. *Brain Inj* 35:1267–1274. <https://doi.org/10.1080/02699052.2021.1972147>
17. Tourigny JN, Boucher V, Paquet V et al (2022) External validation of the updated Brain Injury Guidelines for complicated mild traumatic brain injuries: a retrospective cohort study. *J Neurosurg* 137:782–788. <https://doi.org/10.3171/2021.10.JNS211794>
18. Tourigny JN, Boucher V, Dubucs X et al (2024) Derivation of the Quebec Brain Injury Categories for complicated mild traumatic brain injuries. *J Neurosurg* 141:1730–1738. <https://doi.org/10.3171/2024.4.JNS24183>
19. Valente JH, Anderson JD, Paolo WF et al (2023) Clinical policy: critical issues in the management of adult patients presenting to the emergency department with mild traumatic brain injury. *Ann Emerg Med* 81:e63–e105. <https://doi.org/10.1016/j.annemergmed.2023.01.014>
20. Van Calster B, McLernon DJ, van Smeden M et al (2019) Calibration: the Achilles heel of predictive analytics. *BMC Med* 17:230. <https://doi.org/10.1186/s12916-019-1466-7>
21. Vickers AJ, Elkin EB (2006) Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak* 26:565–574. <https://doi.org/10.1177/0272989X06295361>

Figures



Time zero: emergency department arrival at the treating neurosurgical center.
Qualifying surgery excludes monitoring-only, diagnostic-only, and drainage-only procedures.
AIS, Abbreviated Injury Scale; GCS, Glasgow Coma Scale; mTBI, mild traumatic brain injury;
SDH, subdural hematoma.

Figure 1

Cohort flow diagram. Among 1,470 adult patients with traumatic subdural hematoma assessed during 2019-2021, 725 were excluded because of immediate or emergency surgery at the initial neurosurgical evaluation (n = 73), admission GCS <13 (n = 449), major extracranial injury defined as non-head Abbreviated Injury Scale ≥3 (n = 160), or insufficient required data (n = 43). The analytic cohort comprised 745 patients with mTBI and traumatic SDH who were initially managed nonoperatively with clinical and radiographic observation. Subsequent qualifying therapeutic cranial surgery within 7 days occurred in 111 patients; 634 did not meet this endpoint, including 603 with no surgery during

hospitalization and 31 whose first qualifying operation occurred more than 7 days after time zero. Time zero was emergency-department arrival at the treating neurosurgical center. AIS, Abbreviated Injury Scale; GCS, Glasgow Coma Scale; mTBI, mild traumatic brain injury; SDH, subdural hematoma.

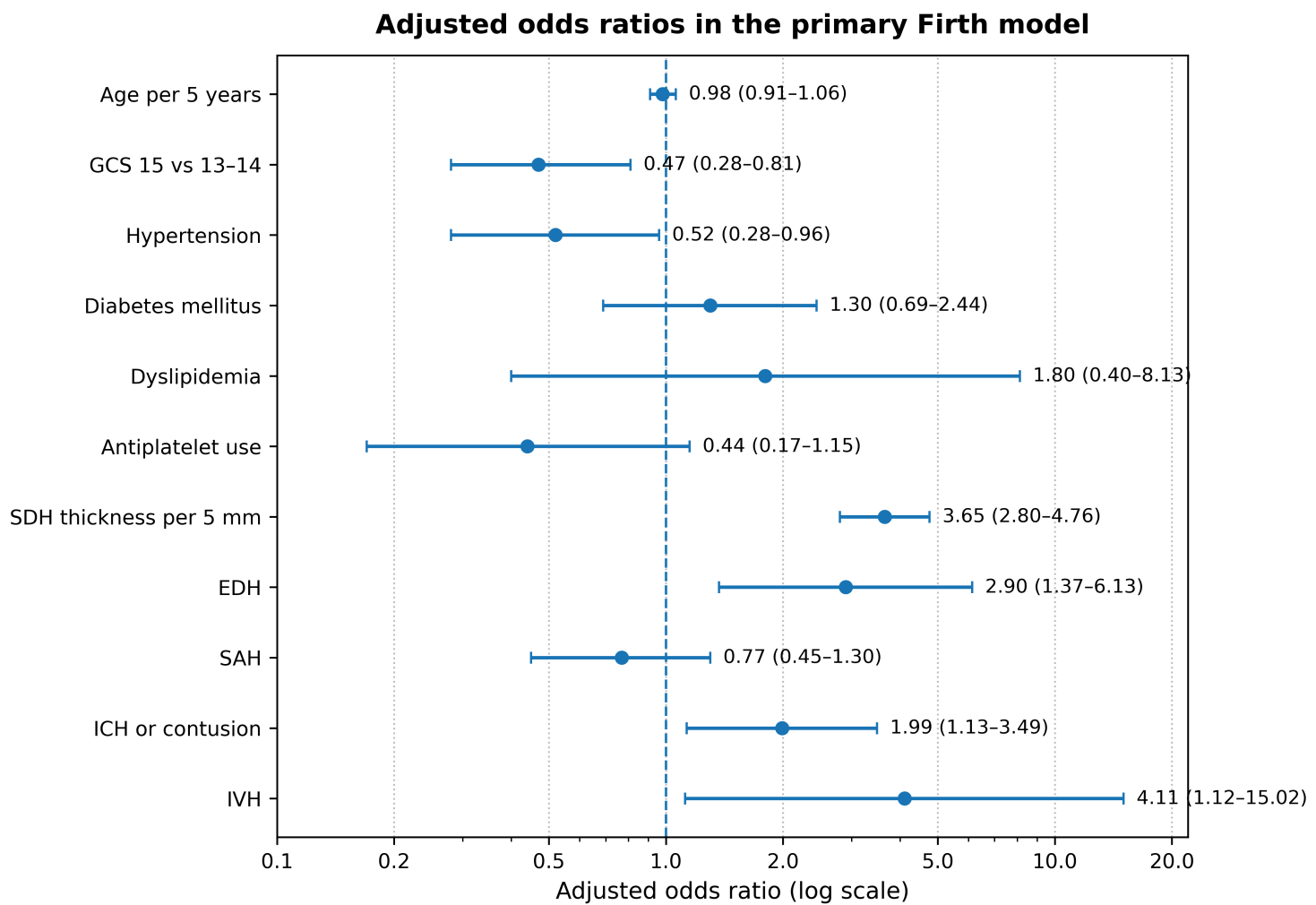


Figure 2

Adjusted odds ratios from the primary Firth penalized logistic regression model for therapeutic cranial neurosurgical intervention within 7 days. Points denote adjusted odds ratios and horizontal lines denote 95% confidence intervals. The vertical line indicates an adjusted odds ratio of 1. Age was modeled per 5-year increase and maximum SDH thickness per 5-mm increase. EDH, epidural hematoma; GCS, Glasgow Coma Scale; ICH, intracerebral hemorrhage; IVH, intraventricular hemorrhage; SAH, subarachnoid hemorrhage; SDH, subdural hematoma.

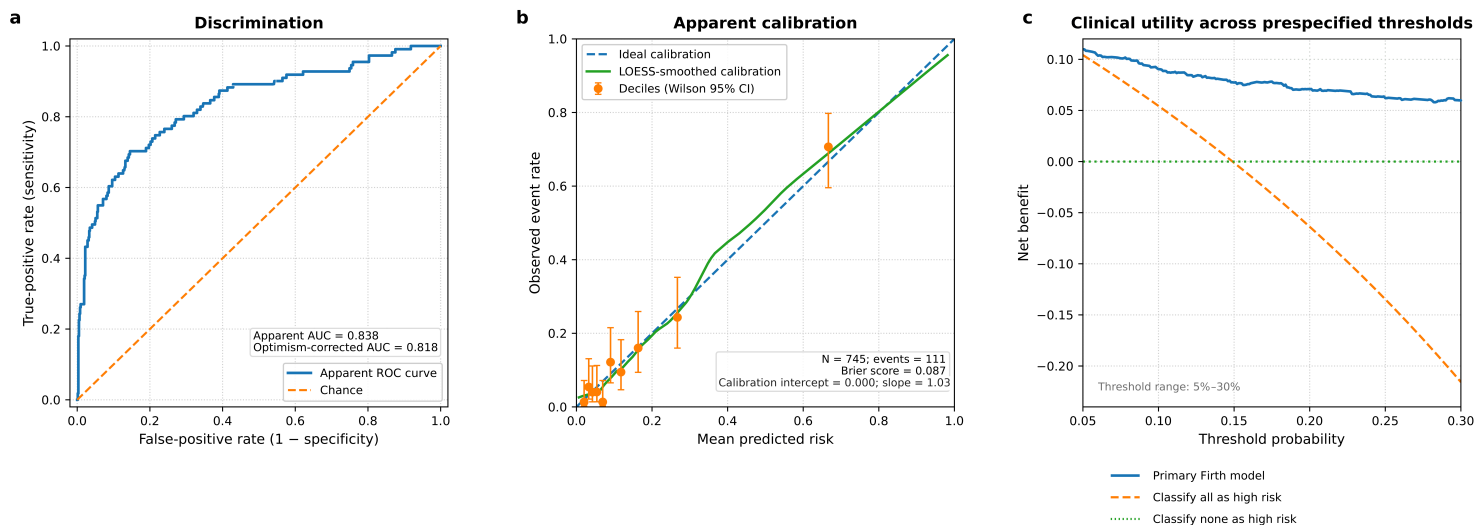


Figure 3

Performance and potential clinical utility of the primary Firth model. a Receiver operating characteristic curve showing apparent AUC 0.838 and optimism-corrected AUC 0.818. b Apparent calibration, including the ideal line, LOESS-smoothed calibration curve, and observed event rates by decile with Wilson 95% confidence intervals; apparent Brier score was 0.087, joint calibration intercept 0.000, and joint calibration slope 1.03; calibration-in-the-large was -0.041 . c Decision-curve analysis across prespecified threshold probabilities of 5%–30%. The modeled action is risk-directed enhanced observation, repeat imaging, or expedited neurosurgical reassessment; thresholds are not operative indications. AUC, area under the curve; CI, confidence interval; LOESS, locally estimated scatterplot smoothing.