

ORIGINAL ARTICLE

Discrimination and Calibration of the ICH Score for 30-Day Mortality After Surgical Evacuation of Spontaneous Intracerebral Hemorrhage at an Indonesian Tertiary Referral Center

Grady Christian¹, Trijoso Permono², Ziske Maritska³¹ Department of Surgery, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia² Department of Neurosurgery, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia³ Department of Biology Medicine, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia

ARTICLE INFO

Article history:

Received: April 18, 2026

Revised: May 21, 2026

Accepted: June 17, 2026

Available Online: June 25, 2026

Published: June 28, 2026

Keywords:

Cerebral hemorrhage; Intracerebral hemorrhage score; Prognosis; Calibration; Neurosurgical procedures

Corresponding author:

Grady Christian

Email:

grady_christian@yahoo.com

DOI:

<https://doi.org/10.37275/sjs.v9i1.171>

Access licence:

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0).

Copyright:

© 2026 The Authors. Authors retain copyright and grant the journal the right of first publication. Published by HM Publisher.

ABSTRACT

Background: The ICH Score is the most widely used prognostic grading scale for spontaneous intracerebral hemorrhage (ICH), but its accuracy after surgical hematoma evacuation has not been evaluated in Indonesia.

Objective: To evaluate the discrimination and calibration of the ICH Score for 30-day mortality in surgically treated patients.

Methods: Retrospective cohort of 81 consecutive adults with spontaneous ICH who underwent hematoma evacuation at Dr. Mohammad Hoesin General Hospital, Palembang (July 2024–June 2025). Discrimination was summarized by the area under the receiver operating characteristic curve (AUC, DeLong 95% confidence interval [CI]) and calibration against the original stratum risks; component associations were tested exactly, and adjusted component effects were bounded across all datasets consistent with the observed tables.

Results: Twenty-nine (35.8%) died within 30 days: none of 20 with scores 0–1, 4 of 36 (11.1%) with a score of 2, and all 25 with scores 3–4. The AUC was 0.958 (95% CI 0.917–0.998). A cutoff of ≥ 3 gave 86.2% sensitivity and 100% specificity. Total deaths were close to expectation (observed/expected 0.94), yet stratum-level calibration was poor, with an exact Hosmer–Lemeshow-type p of 0.011 and a recalibration slope of 3.00 (95% CI 1.88–6.21). Age ≥ 80 years (odds ratio [OR] 26.84) and intraventricular hemorrhage (OR 15.24) were associated with death; only the latter remained significant in every admissible dataset.

Conclusion: After surgical evacuation the ICH Score ranked 30-day mortality well, but its original risk estimates were miscalibrated. Locally recalibrated risks may suit counseling better than the original percentages but need external validation; a high score should not define futility.

1. Introduction

Stroke remained the third leading cause of death worldwide in 2021, accounting for 7.3 million deaths, 160.5 million disability-adjusted life-years and 11.9 million incident events, and spontaneous intracerebral hemorrhage (ICH) contributes a disproportionate share of these deaths relative to its incidence.¹ The burden is concentrated in Asia, where part of the ICH burden is attributable to ambient particulate pollution in countries that include Indonesia,² and where individual-patient data from blood-pressure trials show that outcomes after ICH differ systematically between Asian and non-Asian patients (inverse-probability-weighted odds ratio for 90-day death or dependency 0.67).³ In Indonesia, national insurance records place stroke at the downstream end of a syndemic of diabetes, chronic kidney disease and poverty,⁴ and hospital series report in-hospital ICH mortality of 17.8% in a tertiary center and 36.2% in hypertensive ICH in Bandung, and a 30-day mortality of 39.9% in Bali.^{5–7} Comparable figures from Malaysia, Vietnam and Thailand, and from patients aged ≥ 80 years in the Philippines, confirm that ICH remains among the most lethal emergencies in Southeast Asia.^{8–11}

Mortality after ICH reflects two overlapping injury phases. The expanding hematoma causes immediate mechanical disruption and raises intracranial pressure, whereas blood-breakdown products drive secondary injury over the following

days: perihematomal edema grows from a median of 14 mL to 21 mL over two weeks and lesion volume at day 7 predicts poor outcome,¹² residual edema remains prognostic even after endoscopic evacuation,¹³ and anemia promotes hematoma expansion and unfavorable recovery.¹⁴ Extension into the ventricles adds a separate mechanism: intraventricular hemorrhage (IVH) expands in about a quarter of patients,¹⁵ bilateral lateral-ventricle involvement carries roughly twice the risk of unilateral involvement,¹⁶ and IVH accompanied by hydrocephalus carries roughly twice the 30-day mortality of IVH alone.¹⁷ A useful prognostic instrument must therefore capture both the initial lesion and the routes to secondary deterioration.

The ICH Score, derived by Hemphill and colleagues from five admission variables — Glasgow Coma Scale (GCS), hematoma volume, IVH, infratentorial origin and age ≥ 80 years — remains the reference instrument for predicting 30-day mortality.¹⁸ Contemporary validations report areas under the receiver operating characteristic curve (AUCs) between 0.80 and 0.96,^{9,19–22} and head-to-head comparisons of 27 published models place most of them between 0.60 and 0.86.²³ Systematic reviews, however, document frequent methodological weaknesses in ICH prognostic modeling, identify GCS, IVH, volume and location as the robust hyperacute predictors, and note that external validation of newer machine-learning models is rare.^{24–26} Discrimination has also proved more transportable than calibration: the original stratum risks overestimate

mortality in young adults (observed 0%, 3% and 41% versus predicted 13%, 26% and 72% at scores 1–3) and after stereotactic aspiration (10.4% observed versus 35.1% predicted).^{27,28} Two Indonesian studies concluded that modified versions of the score — age-modified or point-modified — outperform the original in Indonesian patients.^{5,7}

Whether the ICH Score remains accurate after surgical evacuation is a distinct question, because surgery is intended to remove the very mass effect that the volume component measures. In operated cohorts the score has reached AUCs of 0.71 in Peru, 0.757 in Poland and 0.80 in Denmark, lower than in several mixed populations,^{29–31} and a Japanese surgical series observed a 30-day mortality of 7% against 55% predicted.³² Combining the score with the modified Graeb ventricular score improved prediction in operated patients in China,³³ and in cerebellar ICH surgery reduced mortality only above a score of 3.³⁴ Trials now support selective evacuation — early minimally invasive removal in ENRICH and a pooled mortality benefit in the 2025 Cochrane review, with a borderline result for decompressive craniectomy in SWITCH — while MIND showed no superiority of minimally invasive surgery over medical management.^{35–38} Surgical patients in INTERACT3 had lower 6-month mortality than comparable non-operated patients.³⁹ Prognostic scores also interact with treatment intensity: withdrawal of life-sustaining treatment mediates much of the mortality attributed to impaired consciousness,⁴⁰ higher ICH Scores are strongly associated with withdrawal decisions,⁴¹ and early do-not-resuscitate orders are followed by less escalation of care;⁴² together, these patterns can make poor predictions self-fulfilling.

Dr. Mohammad Hoesin General Hospital (RSMH) in Palembang is the tertiary neurosurgical referral center for South Sumatra, and its patients reach surgery after referral pathways, intensive-care constraints and family decision-making that differ from those of the derivation cohort. To our knowledge, no study has evaluated the ICH Score exclusively in an operated Indonesian population, and none in the region has examined

whether the original stratum risks remain calibrated after evacuation. This study therefore aimed to evaluate the discrimination and calibration of the admission ICH Score for 30-day mortality among patients undergoing surgical evacuation of spontaneous ICH at RSMH, to estimate the association of each score component with death, and to determine which component-level conclusions the available data can and cannot support.

2. Methods

Study design and setting

This retrospective cohort study was conducted at RSMH, Palembang, South Sumatra, Indonesia, a central general hospital that serves as the tertiary neurosurgical referral center for the province. It is reported according to the STROBE statement for cohort studies⁴³ and, because it evaluates an existing prognostic score in a new population, according to the TRIPOD recommendations for external validation of prediction models (TRIPOD+AI);⁴⁴ the completed checklist is provided as Supplementary File 1. The study period covered patients admitted between July 2024 and June 2025, and outcome follow-up was completed from the hospital record within 30 days of surgery.

Participants and eligibility

All patients with spontaneous ICH who underwent surgical hematoma evacuation during the study period were identified from hospital records ($n = 92$). Inclusion criteria were age ≥ 18 years, spontaneous ICH confirmed on non-contrast computed tomography (CT), surgical evacuation of the hematoma, and a complete record of admission GCS, CT findings and 30-day vital status. Hemorrhage secondary to a brain tumor ($n = 4$) or a vascular malformation ($n = 5$) was excluded because these lesions follow different natural histories, and two patients were excluded because the admission record was incomplete. All 81 remaining eligible patients were analyzed (consecutive, total sampling), as shown in the participant flow in Figure 1; no analyzed patient had a missing predictor or outcome value.

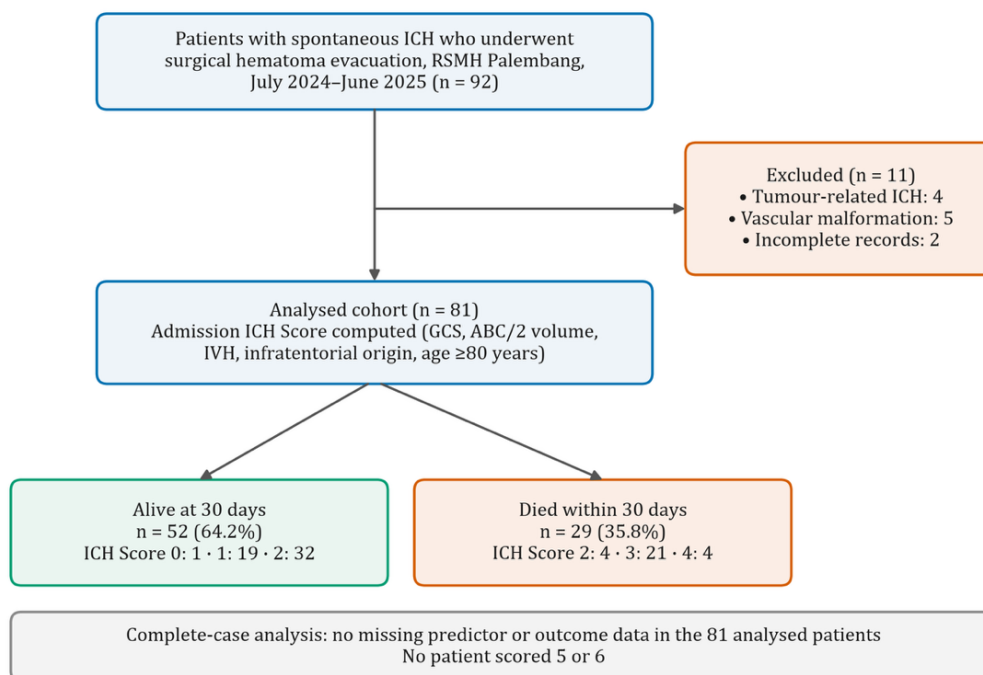


Figure 1. Participant flow. Patients with spontaneous intracerebral hemorrhage (ICH) who underwent surgical hematoma evacuation at Dr. Mohammad Hoesin General Hospital, Palembang, between July 2024 and June 2025, exclusions, and 30-day outcome by admission ICH Score.

Variables and data collection

The exposure was the admission ICH Score, calculated from the GCS and the initial CT scan recorded at hospital admission as the sum of five component scores:¹⁸

$$\text{ICH Score} = S_{\text{GCS}} + S_{\text{vol}} + S_{\text{IVH}} + S_{\text{IT}} + S_{\text{age}} \quad (1)$$

where S_{GCS} is 2 points for a GCS of 3–4, 1 point for 5–12 and 0 for 13–15, and S_{vol} , S_{IVH} , S_{IT} and S_{age} each add 1 point for hematoma volume ≥ 30 mL, the presence of IVH, infratentorial origin (pons or cerebellum) and age ≥ 80 years, respectively, giving a possible range of 0–6.

Hematoma volume was estimated with the ABC/2 method, in which A is the largest axial diameter, B the diameter perpendicular to A on the same slice, and C the number of slices showing hemorrhage multiplied by slice thickness; ABC/2 slightly overestimates planimetric volume in contemporary comparisons, which matters for patients close to the 30-mL threshold.⁴⁵ Sex was recorded as a demographic covariate. The primary outcome was 30-day mortality, defined as death from any cause within 30 days after the surgical procedure and ascertained from inpatient records and follow-up documentation. Data were extracted from the medical records into a digital database, and identifying information was removed before analysis.

Bias and missing data

Consecutive inclusion of every eligible operated patient avoided sampling bias within the surgical population. The score components were recorded at admission, before the outcome occurred, which limits information bias in the predictors, and the outcome — death from any cause — is objective and was ascertained for every analyzed patient. The two patients with incomplete admission records were excluded rather than imputed, so the analysis is a complete-case analysis; because they represent 2.4% of eligible patients (2/83), their exclusion is unlikely to change any estimate materially. Selection into surgery is the main source of bias that cannot be addressed within the cohort: patients judged too ill to benefit, those who died before operation and those whose families declined surgery were not eligible, so the results describe the score's performance only among patients selected for evacuation.

Sample size

The planned minimum sample was 80 patients, calculated for an expected sensitivity of 75% with an absolute precision of $\pm 15\%$ at an anticipated 30-day mortality of 40%:

$$n = \frac{z^2 \cdot \text{Sn}(1 - \text{Sn})}{d^2 \cdot P} = \frac{1.96^2 \times 0.75 \times 0.25}{0.15^2 \times 0.40} \approx 80 \quad (2)$$

where z is the standard normal deviate (1.96), Sn the expected sensitivity, d the precision and P the anticipated outcome proportion. Current guidance for external validation recommends sizing the sample on the precision of calibration rather than on rules of thumb such as 100–200 events, which are often inadequate,^{46,47} so the 29 events available here were treated as sufficient for discrimination but limited for calibration. The precision achieved (a half-width of about 0.04 for the 95% CI of the AUC and an interval of 1.88–6.21 for the calibration slope) is reported with every estimate; a calibration-focused validation would require considerably more deaths.

Statistical analysis

Analyses used Python 3 (NumPy 2.4, SciPy 1.18 and statsmodels 0.15) with purpose-written exact routines. Categorical variables are reported as counts and percentages with 95% Wilson score intervals, which retain close-to-nominal coverage at small denominators:

$$\text{CI} = \frac{x + \frac{z^2}{2} \pm z \sqrt{\frac{x(n-x)}{n} + \frac{z^2}{4}}}{n + z^2} \quad (3)$$

where x is the number of events and n the denominator; a Clopper–Pearson interval was used where every patient in a stratum died. Associations between each score component and death were estimated as odds ratios (OR) with Woolf 95% CIs (Haldane–Anscombe correction when a cell was zero), risk differences with Newcombe intervals and Cramér's V ; Fisher's exact test was used for every 2×2 table and the Fisher–Freeman–Halton exact test and the Cochran–Armitage trend test for the three GCS bands; the six tests of components and sex were additionally Holm-adjusted as one family. Mortality across ICH Score strata was tested for trend with the Cochran–Armitage statistic and its exact conditional permutation p value, and the score distributions of survivors and non-survivors were compared with the exact Mann–Whitney U test with mid-ranks for ties.

Discrimination was quantified by the AUC with the DeLong variance, a logit-transformed DeLong interval and a stratified bootstrap with 20,000 resamples. For every possible cutoff we calculated sensitivity (Sn), specificity (Sp), predictive values, likelihood ratios (with a 0.5 continuity correction when a cell was zero), accuracy and the Youden index:

$$J = \text{Sn} + \text{Sp} - 1 \quad (4)$$

Stratum-specific likelihood ratios were computed for scores 0–1, 2 and 3–4. Calibration was assessed against the original 30-day mortality of each ICH Score stratum (0%, 13%, 26%, 72%, 97% and 100% for scores 0–5),¹⁸ which Maljaars et al. restated for scores 1–3.²⁷ Overall agreement was summarized by the ratio of observed to expected deaths with an exact Poisson interval and by the Brier score. Agreement within strata was tested with a Hosmer–Lemeshow-type χ^2 over the informative strata (degrees of freedom equal to the number of strata, because no parameters were estimated in this cohort), whose exact p value was obtained by enumerating every possible outcome configuration under the original risks because several expected counts were below five, and with exact two-sided binomial tests of each stratum with Holm adjustment and Cohen's h . Calibration slope and intercept were estimated by Firth-penalized logistic recalibration, because the outcome was almost completely separated by the score:

$$\text{logit}(\text{Pr}(\text{death})) = a + b \cdot \text{logit}(p_H) \quad (5)$$

where p_H is the original stratum risk (bounded at 0.005 and 0.995 for the logit; bounds of 0.01 and 0.99 gave an identical slope) and $b = 1$ indicates a gradient identical to that of the derivation cohort; profile penalized-likelihood intervals were used for b . A Firth logistic model of death on the score (per point) provided locally recalibrated risks, Nagelkerke R^2 and a Hosmer–Lemeshow statistic; its optimism was estimated by bootstrap internal validation (2,000 resamples, refitting the model in each). Clinical utility was summarized by decision-curve analysis for a specific action — triggering a structured multidisciplinary review of goals of care, ventricular drainage and intensive-care needs within 24 hours of surgery — in which the apparent net benefit of classifying patients as high risk at a threshold probability p_t is:

$$\text{Net benefit} = \frac{\text{TP}}{N} - \frac{\text{FP}}{N} \times \frac{p_t}{1 - p_t} \quad (6)$$

Thirty-day mortality was compared descriptively with verified published cohorts, grouped by outcome definition (30-day or in-hospital), using the difference in proportions and Cohen's h .

The hospital records provided the joint distribution of each component with the outcome and of the total score with the

outcome, but not the cross-classification of the components with one another. Adjusted component effects were therefore estimated as a sensitivity analysis over the space of patient-level datasets consistent with the observed data, without imputing any patient-level value. An admissible dataset assigned the GCS band and the four binary items to each patient within each outcome group so that both the observed component-by-outcome tables and the observed score distribution in survivors and non-survivors were reproduced exactly. We drew 20,000 admissible datasets with a Markov chain whose moves preserve every observed margin and every patient's score (200 chains started from random integer-programming solutions, 4,000 burn-in moves and 400 moves between samples) and searched for the attainable extremes of each adjusted coefficient by simulated annealing. Firth logistic regression (Model A: death on GCS points, age ≥ 80 years, volume ≥ 30 mL, IVH and infratentorial origin; Model B: Model A plus sex) with penalized-likelihood-ratio tests was fitted to every dataset; non-converging Newton fits were refitted by direct maximization of the penalized likelihood, so no dataset was discarded. Because the percentiles describe the sampled space under a uniform measure over labeled configurations, they are not probabilities of the true effect, and inference rests on the extremes. Simulated annealing was run with eight restarts per coefficient and direction for Model A and six for Model B. A component was considered robustly associated with death only if its adjusted OR exceeded 1 with $p < 0.05$ in every dataset examined; if the attainable range spanned 1, the data were considered unable to identify the adjusted association. The same admissible datasets were used to bound the AUC of the score with each component removed, and integer programming gave the exact minimum

and maximum number of patients with each feature within each score stratum. Two-sided $\alpha=0.05$ was used throughout and p values are reported to three decimals.

Ethical approval

The protocol received formal approval from the Research Ethics Committee of Dr. Mohammad Hoesin General Hospital Palembang, under document number DP.04.03/D.XVIII.06.08/ETIK/014/2026. The study was conducted in accordance with the Declaration of Helsinki, the 2016 CIOMS International Ethical Guidelines and the WHO 2011 standards for research ethics review. Because only existing medical records were used and all data were anonymized before analysis, the requirement for individual informed consent was waived.

3. Results

Participants

Of the 92 patients who underwent surgery for ICH, 81 met the eligibility criteria, as shown in Figure 1. The cohort was 56.8% male (46/81; exact binomial test against parity $p = 0.266$), and 11 patients (13.6%, 95% CI 7.8–22.7) were aged ≥ 80 years. Most patients were admitted with a GCS of 5–12 (69/81, 85.2%), six (7.4%) with a GCS of 3–4 and six (7.4%) with a GCS of 13–15. Hematoma volume was ≥ 30 mL in 56 patients (69.1%, 95% CI 58.4–78.1), IVH was present in 17 (21.0%, 95% CI 13.5–31.1) and the hemorrhage was infratentorial in five (6.2%). Twenty-nine patients died within 30 days of surgery, a mortality of 35.8% (95% CI 26.2–46.7). The characteristics of survivors and non-survivors are detailed in Table 1.

Table 1. Characteristics of the 81 operated patients overall and by 30-day vital status.

Characteristic	All patients ($n = 81$), n (%; 95% CI)	Alive at 30 days ($n = 52$), n (%)	Died within 30 days ($n = 29$), n (%)	p
Sex				0.820
Male	46 (56.8; 45.9–67.0)	29 (55.8)	17 (58.6)	
Female	35 (43.2; 33.0–54.1)	23 (44.2)	12 (41.4)	
Age				<0.001
<80 years	70 (86.4; 77.3–92.2)	51 (98.1)	19 (65.5)	
≥ 80 years	11 (13.6; 7.8–22.7)	1 (1.9)	10 (34.5)	
Admission GCS				<0.001
13–15 (0 points)	6 (7.4; 3.4–15.2)	6 (11.5)	0 (0.0)	
5–12 (1 point)	69 (85.2; 75.9–91.3)	46 (88.5)	23 (79.3)	
3–4 (2 points)	6 (7.4; 3.4–15.2)	0 (0.0)	6 (20.7)	
Hematoma volume (ABC/2)				0.078
<30 mL	25 (30.9; 21.9–41.6)	20 (38.5)	5 (17.2)	
≥ 30 mL	56 (69.1; 58.4–78.1)	32 (61.5)	24 (82.8)	
Intraventricular hemorrhage				<0.001
Absent	64 (79.0; 68.9–86.5)	49 (94.2)	15 (51.7)	
Present	17 (21.0; 13.5–31.1)	3 (5.8)	14 (48.3)	
Origin				0.053
Supratentorial	76 (93.8; 86.4–97.3)	51 (98.1)	25 (86.2)	
Infratentorial	5 (6.2; 2.7–13.6)	1 (1.9)	4 (13.8)	
ICH Score, points				<0.001
0	1 (1.2; 0.2–6.7)	1 (1.9)	0 (0.0)	
1	19 (23.5; 15.6–33.8)	19 (36.5)	0 (0.0)	
2	36 (44.4; 34.1–55.3)	32 (61.5)	4 (13.8)	
3	21 (25.9; 17.6–36.4)	0 (0.0)	21 (72.4)	
4	4 (4.9; 1.9–12.0)	0 (0.0)	4 (13.8)	
Median (IQR)	2 (2–3)	2 (1–2)	3 (3–3)	

Values are n (column %); 95% Wilson intervals for the whole cohort. p values: Fisher's exact test for 2×2 tables, Fisher–Freeman–Halton exact test for GCS and exact Mann–Whitney test for the ICH Score ($U = 64$, $z = 7.24$). Sex ratio against parity: exact binomial $p = 0.266$. ABC/2, ellipsoid volume estimate; GCS, Glasgow Coma Scale; IQR, interquartile range.

Score components and mortality

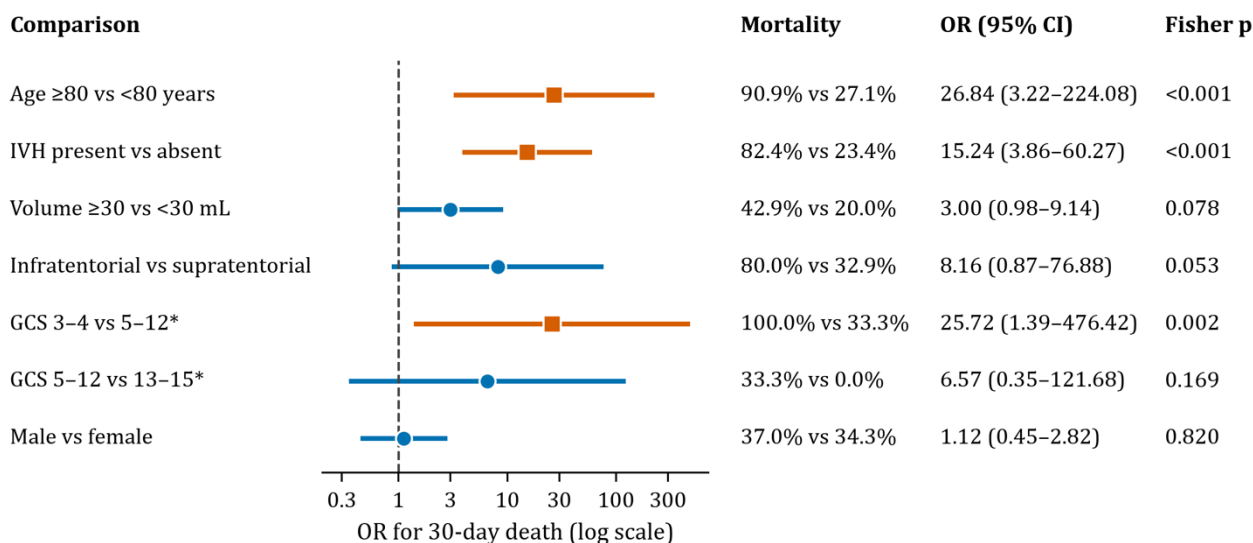
Mortality differed most sharply by age and ventricular extension. Ten of the 11 patients aged ≥ 80 years died (90.9%, 95% CI 62.3–98.4) compared with 19 of 70 younger patients (27.1%), giving an OR of 26.84 (95% CI 3.22–224.08), a risk difference of 63.8 percentage points and Fisher $p < 0.001$. IVH was associated with a mortality of 82.4% (14/17) against 23.4% (15/64) without it (OR 15.24 [95% CI 3.86–60.27]; Cramér's V 0.500; $p < 0.001$). Mortality rose across the GCS bands from 0/6 (GCS 13–15) to 23/69 (33.3%, GCS 5–12) and 6/6 (GCS 3–4)

(Fisher–Freeman–Halton $p < 0.001$; exact trend $p < 0.001$). Hematoma volume ≥ 30 mL (42.9% versus 20.0%; OR 3.00 [95% CI 0.98–9.14]; $p = 0.078$) and infratentorial origin (80.0% versus 32.9%; OR 8.16 [95% CI 0.87–76.88]; $p = 0.053$) showed point estimates in the expected direction with odds-ratio intervals that included 1, whereas sex was unrelated to death (37.0% in men versus 34.3% in women; $p = 0.820$). After Holm adjustment across the six tests, age, IVH and GCS remained significant (Holm $p < 0.001$, $p < 0.001$ and $p = 0.002$), whereas volume and infratentorial origin did not (Holm $p = 0.159$ for both). These associations are detailed in Table 2 and displayed in Figure 2.

Table 2. Association of each ICH Score component and sex with 30-day mortality.

Comparison	Mortality, exposed % (95% CI)	Mortality, reference % (95% CI)	OR (95% CI)	Risk difference, points (95% CI)	Cramér's V	p	Holm p
Age ≥ 80 vs < 80 years	90.9 (62.3–98.4)	27.1 (18.1–38.5)	26.84 (3.22–224.08)	63.8 (32.9 to 75.5)	0.456	< 0.001	< 0.001
IVH present vs absent	82.4 (59.0–93.8)	23.4 (14.7–35.1)	15.24 (3.86–60.27)	58.9 (32.8 to 73.3)	0.500	< 0.001	< 0.001
Volume ≥ 30 vs < 30 mL	42.9 (30.8–55.9)	20.0 (8.9–39.1)	3.00 (0.98–9.14)	22.9 (0.2 to 40.0)	0.220	0.078	0.159
Infratentorial vs supratentorial	80.0 (37.6–96.4)	32.9 (23.4–44.1)	8.16 (0.87–76.88)	47.1 (3.2 to 66.0)	0.236	0.053	0.159
Male vs female	37.0 (24.5–51.4)	34.3 (20.8–50.8)	1.12 (0.45–2.82)	2.7 (–18.0 to 22.4)	0.028	0.820	0.820
GCS 3–4 vs 5–12 ^a	100.0 (61.0–100.0)	33.3 (23.4–45.1)	25.72 (1.39–476.42)	66.7 (25.9 to 76.6)	—	0.002	—
GCS 5–12 vs 13–15 ^a	33.3 (23.4–45.1)	0.0 (0.0–39.0)	6.57 (0.35–121.68)	33.3 (–7.0 to 45.1)	—	0.169	—
GCS, three bands (overall)	—	—	—	—	0.420	$< 0.001^b$	0.002

Mortality percentages with 95% Wilson intervals; OR with Woolf intervals; risk difference with Newcombe intervals; p from Fisher's exact test; near the null, the Newcombe risk-difference interval and Fisher's test can disagree (volume, infratentorial origin). ^aHaldane–Anscombe correction (zero cell). ^bFisher–Freeman–Halton exact test; exact Cochran–Armitage trend $p < 0.001$. Holm p : Holm adjustment across the six component and sex tests. IVH, intraventricular hemorrhage; OR, odds ratio.



Square = $p < 0.05$; circle = $p \geq 0.05$. *Haldane–Anscombe correction for a zero cell.

Figure 2. Odds ratios (log scale) with 95% CIs for 30-day death by ICH Score component and sex, with mortality in the exposed and reference groups and Fisher's exact p values. Squares denote $p < 0.05$. GCS, Glasgow Coma Scale; IVH, intraventricular hemorrhage; OR, odds ratio.

Score distribution and stratum mortality

The admission ICH Score ranged from 0 to 4 (median 2, interquartile range [IQR] 2–3); no patient scored 5 or 6. Thirty-six patients (44.4%) scored 2, 21 (25.9%) scored 3, 19 (23.5%) scored 1, four (4.9%) scored 4 and one (1.2%) scored 0, as shown in Figure 3. Non-survivors had a median score of 3 (IQR 3–3) and survivors a median of 2 (IQR 1–2) (exact Mann–Whitney $U = 64$, $z = 7.24$, $p < 0.001$). No patient with a score of 0

or 1 died (0/20; 95% CI for score 1, 0–16.8%), four of 36 patients with a score of 2 died (11.1%, 95% CI 4.4–25.3), and all 25 patients with a score of 3 or 4 died (Clopper–Pearson 95% CI 86.3–100%). The gradient was monotonic (Cochran–Armitage $z = 7.08$, exact $p < 0.001$), and stratum-specific likelihood ratios were 0.043 for scores 0–1, 0.224 (95% CI 0.088–0.571) for score 2 and 90.1 (continuity-corrected) for scores 3–4, as summarized in Table 3.

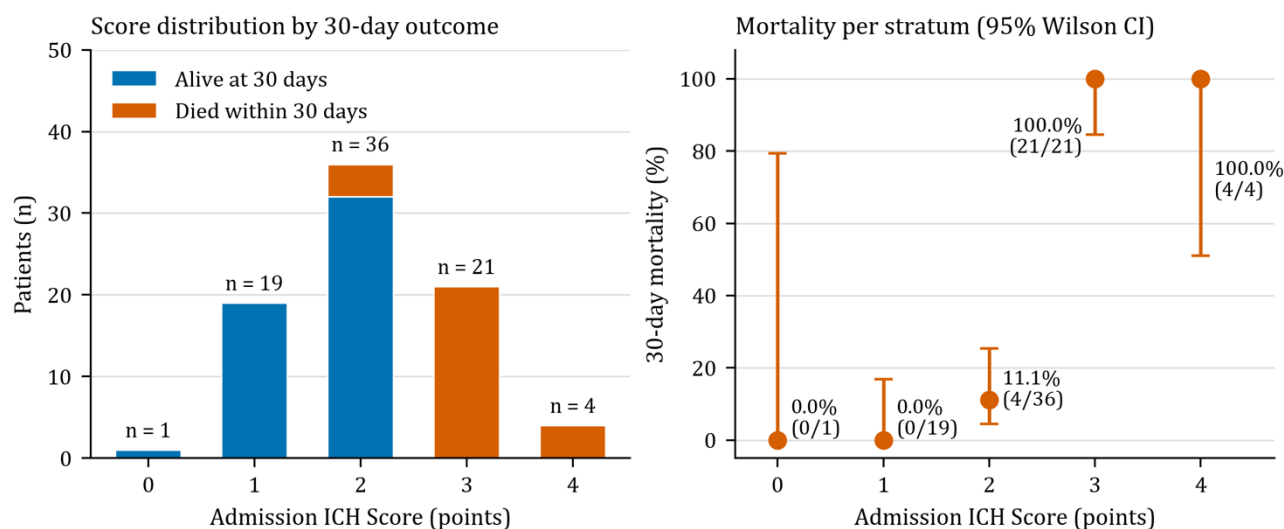


Figure 3. Admission ICH Score by 30-day outcome (left) and 30-day mortality in each score stratum with 95% Wilson intervals (right). No patient scored 5 or 6.

Table 3. Thirty-day mortality by admission ICH Score compared with the original derivation-cohort risks.

ICH Score	Patients, n (%)	Deaths, n	Observed mortality, % (95% CI)	Hemphill 2001 mortality, %	Exact binomial p (Holm)	Cohen's h	Locally recalibrated risk, %
0	1 (1.2)	0	0.0 (0.0–79.3)	0	—	0.00	0.0
1	19 (23.5)	0	0.0 (0.0–16.8)	13	0.161 (0.323)	–0.74	0.0
2	36 (44.4)	4	11.1 (4.4–25.3)	26	0.055 (0.165)	–0.39	12.2
3	21 (25.9)	21	100.0 (84.5–100.0)	72	0.002 (0.009)	1.12	97.7
4	4 (4.9)	4	100.0 (51.0–100.0)	97	1.000 (1.000)	0.35	100.0
5–6	0 (0.0)	0	—	100	—	—	—
All	81 (100.0)	29	35.8 (26.2–46.7)	38.1 (expected)	—	—	35.8

Observed mortality with 95% Wilson intervals (Clopper–Pearson 95% CI for scores 3–4 combined, 25/25; 86.3–100). Cochran–Armitage trend $z = 7.08$, exact $p < 0.001$. Exact two-sided binomial tests against the Hemphill 2001 stratum risk,¹⁸ Holm-adjusted across scores 1–4. Locally recalibrated risk from the Firth model of death on the score (per point). Stratum-specific likelihood ratios: scores 0–1, 0.043 (95% CI 0.003–0.69); score 2, 0.224 (0.088–0.571); scores 3–4, 90.1 (5.7–1,427.6) (0.5 continuity correction for the zero cells).

Discrimination

The AUC of the ICH Score for 30-day mortality was 0.958 (DeLong 95% CI 0.917–0.998; logit-transformed 0.893–0.984; bootstrap 0.914–0.990), as shown in Figure 4. The Youden index was maximal at a cutoff of ≥ 3 ($J = 0.862$), where sensitivity was 86.2% (95% CI 69.4–94.5), specificity 100% (95% CI 93.1–

100.0), positive predictive value 100% (95% CI 86.7–100.0), negative predictive value 92.9% (95% CI 83.0–97.2) and accuracy 95.1%. A cutoff of ≥ 2 identified every death (sensitivity 100%, 95% CI 88.3–100.0) at the cost of a specificity of 38.5% (95% CI 26.5–52.0). The performance at every cutoff is detailed in Table 4.

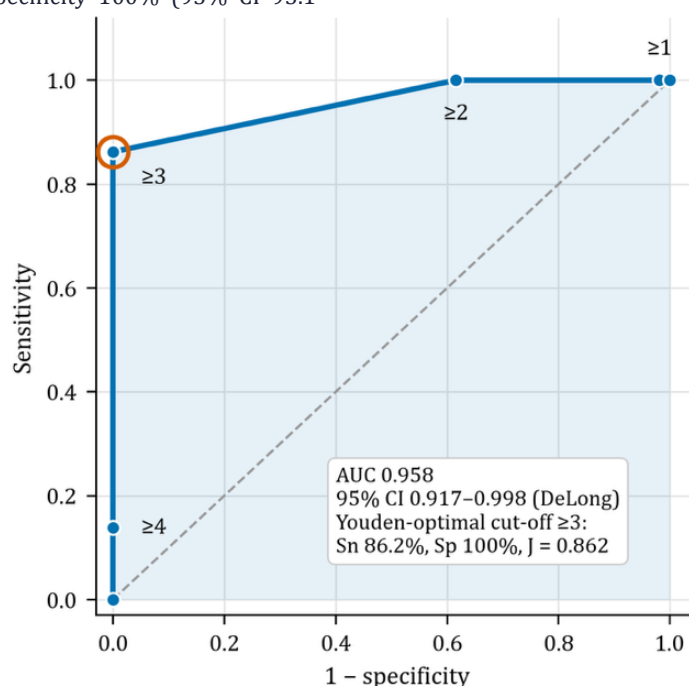


Figure 4. Empirical receiver operating characteristic curve of the admission ICH Score for 30-day mortality. Points are labeled with the cutoff; the circled point marks the Youden-optimal cutoff (≥ 3). AUC, area under the curve; CI, confidence interval.

Table 4. Prognostic performance of the ICH Score for 30-day mortality at each cutoff.

Cutoff	TP/FP/FN/TN	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	LR+ (95% CI)	LR- (95% CI)	Youde n /	Accurac y, %
≥1	29/51/0/1	100.0 (88.3–100.0)	1.9 (0.3–10.1)	36.2 (26.6–47.2)	100.0 (20.7–100.0)	1.01 (0.95–1.08) ^c	0.589 (0.025–14.008) ^c	0.019	37.0
≥2	29/32/0/20	100.0 (88.3–100.0)	38.5 (26.5–52.0)	47.5 (35.5–59.8)	100.0 (83.9–100.0)	1.60 (1.29–2.00) ^c	0.043 (0.003–0.687) ^c	0.385	60.5
≥3	25/0/4/52	86.2 (69.4–94.5)	100.0 (93.1–100.0)	100.0 (86.7–100.0)	92.9 (83.0–97.2)	90.10 (5.69–1,427.55) ^c	0.151 (0.065–0.355) ^c	0.862	95.1
≥4	4/0/25/52	13.8 (5.5–30.6)	100.0 (93.1–100.0)	100.0 (51.0–100.0)	67.5 (56.5–76.9)	15.90 (0.89–285.30) ^c	0.858 (0.737–1.000) ^c	0.138	69.1

AUC 0.958 (DeLong 95% CI 0.917–0.998; logit-transformed 0.893–0.984; stratified bootstrap 0.914–0.990). Proportions with 95% Wilson intervals. ^c0.5 continuity correction because of a zero cell (uncorrected LR+ at ≥3 is infinite). FN, false negative; FP, false positive; LR, likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive.

Calibration

Applying the original stratum risks predicted 30.83 deaths against the 29 observed, an observed/expected ratio of 0.94 (95% CI 0.63–1.35). This overall agreement concealed strong stratum-level miscalibration (Hosmer–Lemeshow $\chi^2=15.28$; exact $p = 0.011$; joint exact test of the four informative strata $p = 0.001$). Observed mortality was lower than predicted at score 1 (0% versus 13%; $p = 0.161$) and score 2 (11.1% versus 26%; $p = 0.055$, Holm $p = 0.165$), and higher at score 3 (100% versus 72%; $p = 0.002$, Holm $p = 0.009$; Cohen's $h = 1.12$). The Firth recalibration slope was 3.00 (95% CI 1.88–6.21) with an

intercept of 1.11, indicating a risk gradient about three times steeper than in the derivation cohort, as shown in Figure 5. The Brier score of the original risks was 0.078, against 0.230 for a model assigning every patient the cohort mortality. A local Firth model gave an OR of 312 per score point (95% CI 33.9–42,181), with apparent predicted mortality of 12.2% at score 2 and 97.7% at score 3, a Nagelkerke R^2 of 0.858 and good internal fit (Hosmer–Lemeshow $p = 0.912$); bootstrap internal validation showed little optimism (optimism-corrected calibration slope 0.99; corrected Brier score 0.047). The calibration statistics are presented in Table 5.

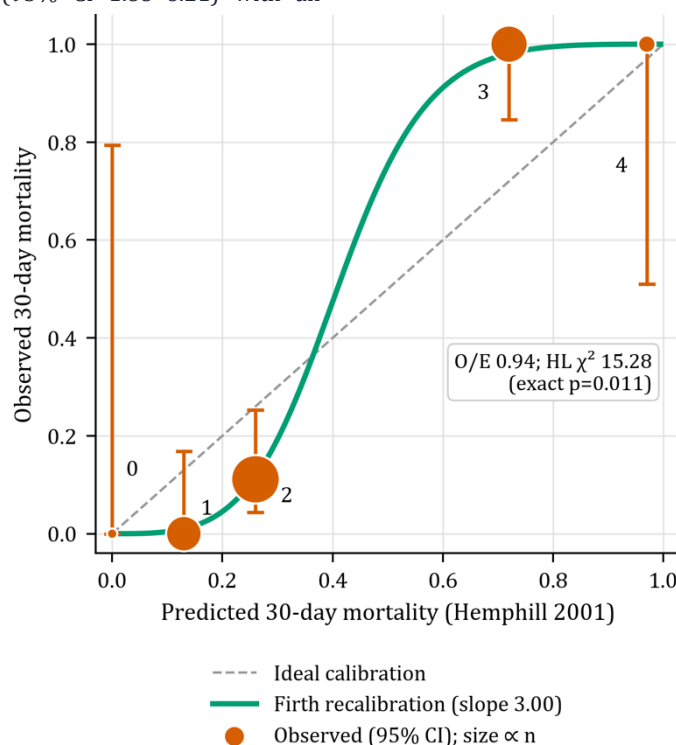


Figure 5. Calibration of the original ICH Score risks. Observed 30-day mortality (95% Wilson intervals; marker area proportional to stratum size; labels give the score) against the risk predicted by the derivation cohort, with the Firth recalibration curve and the line of ideal calibration. HL, Hosmer–Lemeshow; O/E, observed/expected.

Table 5. Calibration of the original ICH Score risks and performance of the locally recalibrated model.

Measure	Estimate (95% CI)	Interpretation
Expected deaths with original stratum risks	30.83 (observed 29)	—
Observed/expected deaths	0.94 (0.63–1.35)	Overall mortality reproduced
Hosmer–Lemeshow-type χ^2 (4 df)	15.28; exact $p = 0.011$	Stratum-level miscalibration
Brier score, original risks	0.078	Reference (no-information) 0.230
Recalibration slope b	3.00 (1.88–6.21)	Gradient steeper than derivation cohort ($b > 1$)
Recalibration intercept a	1.11	Joint with slope
Calibration-in-the-large (slope fixed at 1)	–0.13	Close to 0
Local Firth model: OR per score point	312 (33.9–42,181)	Quasi-complete separation; Firth penalization
Local Firth model: Nagelkerke R^2	0.858	—
Local Firth model: Hosmer–Lemeshow χ^2 (3 df)	0.53; $p = 0.912$	Good internal fit
Local Firth model: Brier score	0.044	—

Measure	Estimate (95% CI)	Interpretation
Joint exact test, strata 1–4 (probability ordering)	0.001	Confirms miscalibration
Local model, bootstrap optimism-corrected calibration slope	0.99	Minimal optimism (2,000 resamples)
Local model, bootstrap optimism-corrected Brier score	0.047	—

Original risks: 0%, 13%, 26%, 72%, 97% and 100% for scores 0–5. The recalibration slope and intercept were estimated by Firth-penalized logistic regression of death on the logit of the original risk (bounded at 0.005–0.995), with a profile penalized-likelihood interval for the slope. df, degrees of freedom.

Comparison with published cohorts

Among cohorts reporting 30-day mortality, the 35.8% observed here was similar to that of the Bali cohort (39.9%) and the Peruvian surgical cohort (27.4%) but higher than in the Danish (23%) and Japanese (7%) surgical cohorts. Among cohorts reporting in-hospital mortality, which is not directly comparable, it was similar to the Bandung cohort (36.2%) and

higher than an Indonesian tertiary-center series (17.8%). At the stratum level, mortality at score 2 matched the Danish operated cohort exactly (11.1% versus 11%), whereas mortality at score 3 far exceeded it (100% versus 33%; $h = 1.92$) and exceeded the 41.5% in-hospital mortality recorded for scores 3–4 in a 10-year Florida registry. These comparisons are detailed in Table 6, and the stratum-level contrast is shown in Figure 6.

Table 6. Thirty-day mortality in the present cohort compared with published ICH cohorts.

Comparator	Setting	n	Outcome	Comparator mortality, %	Present cohort, %	Cohen's h
Rodriguez-Calienes et al. 2022 ²⁹	Peru; operated	73	30-day	27.4	35.8	0.18
Haga et al. 2021 ³¹	Denmark; operated, supratentorial	126	30-day	23.0	35.8	0.28
Hong et al. 2021 ³²	Japan; operated	203	30-day	7.0	35.8	0.75
Widyadharma et al. 2021 ⁷	Indonesia (Bali); all ICH	311	30-day	39.9	35.8	−0.08
Amalia et al. 2026 ⁶	Indonesia (Bandung); hypertensive ICH	177	In-hospital	36.2	35.8	−0.01
Sasongko et al. 2025 ⁵	Indonesia; all ICH	225	In-hospital	17.8	35.8	0.41
Haga et al. 2021 ³¹ , score 2	Denmark; operated, supratentorial	126	30-day	11.0	11.1	0.00
Haga et al. 2021 ³¹ , score 3	Denmark; operated, supratentorial	126	30-day	33.0	100.0	1.92
Haga et al. 2021 ³¹ , score 4	Denmark; operated, supratentorial	126	30-day	76.0	100.0	1.02
Massad et al. 2025 ⁴¹ , scores 3–4	USA (Florida registry)	12,426	In-hospital	41.5	100.0	1.74
Massad et al. 2025 ⁴¹ , scores 0–2	USA (Florida registry)	12,426	In-hospital	6.6	7.1	0.02

Comparisons are descriptive; in-hospital comparators are not directly comparable with 30-day mortality. Comparator values as reported in the cited studies; n is the comparator cohort size. Present cohort: 29/81 overall and stratum counts for scores. Cohen's h is the arcsine difference (present minus comparator). Significance tests against published cohorts were removed at the Editor's request.

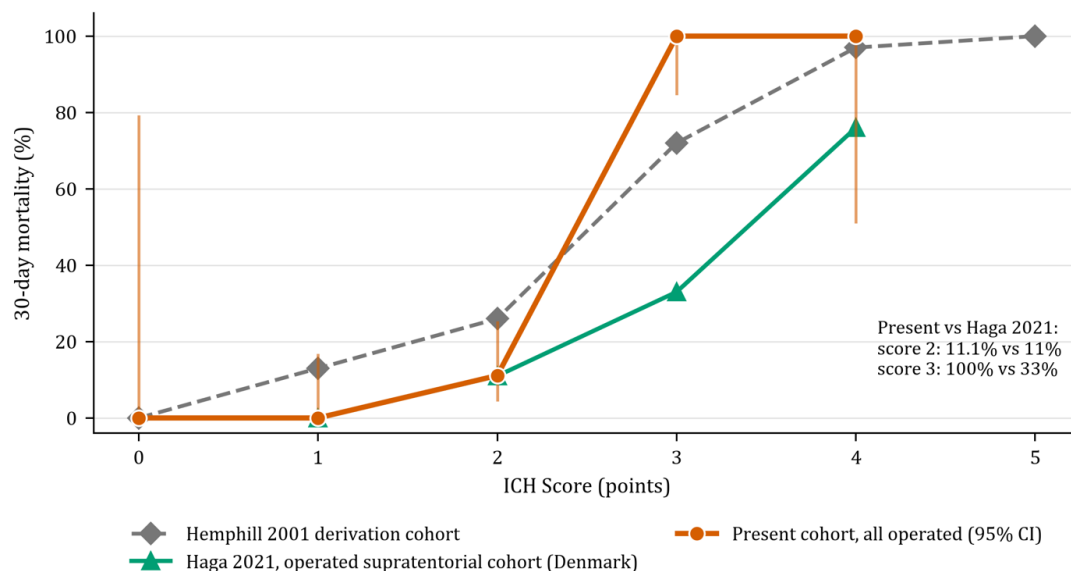


Figure 6. Thirty-day mortality by ICH Score in the present all-operated cohort (95% Wilson intervals), the Hemphill 2001 derivation cohort and an operated Danish cohort of supratentorial ICH (Haga et al. 2021³¹).

Clinical utility

In decision-curve analysis, using a score ≥ 3 to trigger the structured review yielded a constant apparent net benefit of 0.309, because no survivor was misclassified. This rule exceeded the net benefit of classifying every patient as high risk at threshold probabilities above 7.1% and that of the ≥ 2 rule above 11.1%, while the original stratum risks matched the ≥ 3 rule only

at thresholds between 26% and 72%, as shown in Figure 7. At a threshold of 20%, for example, the net benefit was 0.309 for the ≥ 3 rule, 0.259 for the ≥ 2 rule and 0.198 for classifying every patient as high risk, equivalent to five more net true-positive classifications per 100 patients than the ≥ 2 rule, and 11 more than classifying every patient as high risk, once false-positive classifications are weighted by the threshold odds.

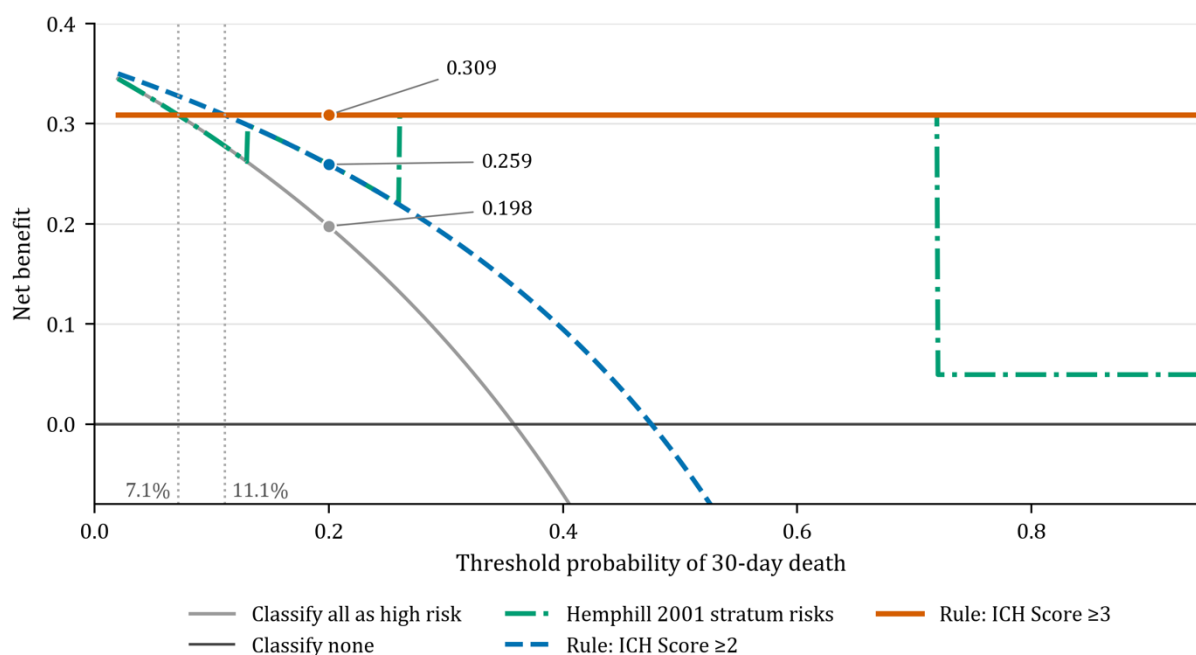


Figure 7. Decision curves for classifying patients as at high risk of 30-day death: ICH Score ≥ 3 and ≥ 2 rules, the original stratum risks, and the default strategies of classifying all or no patients as high risk.

Adjusted component effects: bounds analysis

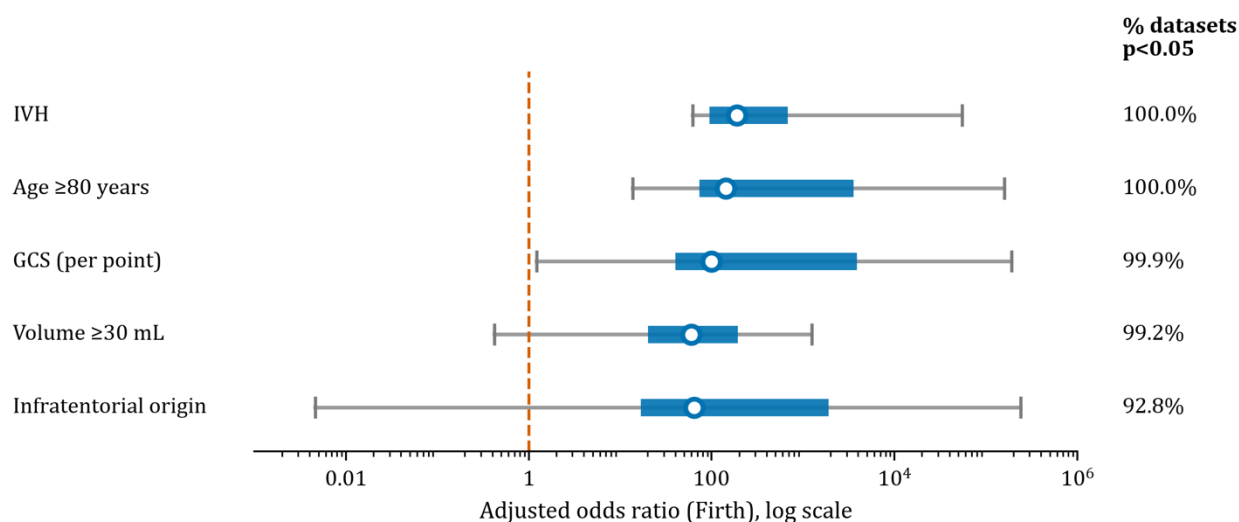
Across 20,000 admissible datasets, every adjusted OR was very large because death was almost a deterministic function of the summed score; the magnitudes were therefore not identifiable, whereas direction and robustness could be assessed, as detailed in Table 7 and shown in Figure 8. IVH was the only component that remained positive and significant in every dataset examined (median adjusted OR 189, 2.5th–97.5th percentile 94.5–678; smallest value found 62.4, $p < 0.001$). Age ≥ 80 years and GCS points kept a positive direction in every dataset (smallest adjusted ORs found 13.7 and 1.22), but significance was not guaranteed ($p = 0.080$ and $p = 0.896$ at those extremes). For hematoma volume ≥ 30 mL (attainable range 0.42–1,241)

and infratentorial origin (0.005–241,749) the attainable adjusted ORs spanned 1, so the data cannot identify their adjusted associations, even though they reached significance in 99.2% and 92.8% of the sampled datasets, which describes the sampled space rather than a probability. Across the eight annealing restarts the smallest adjusted OR found for IVH ranged from 62.6 to 76.9, and no sampled or annealed dataset produced a value below 62.4. When sex was added (Model B), IVH and age kept a positive direction in every dataset examined but significance was no longer guaranteed (p up to 0.071 and 0.238 at the smallest ORs found), the GCS effect was no longer identifiable (smallest OR 0.15), and sex itself was significant in only 0.9% of datasets.

Table 7. Adjusted component associations bounded across all patient-level datasets consistent with the observed data (Firth logistic regression).

Model / predictor	Median adjusted OR [2.5th–97.5th percentile]	Most extreme adjusted OR found (min–max)	p at minimum	Datasets with $p < 0.05$, %	Datasets with OR > 1 , %	Verdict
Model A (score components)						
GCS (per point)	99.1 [40.0–3,894]	1.22 to 1.9×10^5	0.896	99.9	100.0	Direction robust; significance not guaranteed
Age ≥ 80 years	143 [73.3–3,559]	13.7 to 1.6×10^5	0.080	100.0	100.0	Direction robust; significance not guaranteed
Volume ≥ 30 mL	59.8 [20.1–194]	0.42 to 1,241	0.596	99.2	100.0	Not identifiable (range spans 1)
IVH	189 [94.5–678]	62.4 to 54,907	< 0.001	100.0	100.0	Robust: positive and significant in every dataset
Infratentorial origin	64.7 [16.9–1,896]	4.6×10^{-3} to 2.4×10^5	0.112	92.8	99.7	Not identifiable (range spans 1)
Model B (Model A + sex)						
GCS (per point)	109 [36.4–6,679]	0.15 to 1.9×10^8	0.580	99.9	100.0	Not identifiable (range spans 1)
Age ≥ 80 years	158 [66.2–5,102]	5.73 to 1.3×10^7	0.238	100.0	100.0	Direction robust; significance not guaranteed
Volume ≥ 30 mL	62.9 [17.5–314]	0.14 to 4.4×10^5	0.599	98.6	100.0	Not identifiable (range spans 1)
IVH	188 [85.3–1,058]	10.6 to 2.0×10^7	0.071	100.0	100.0	Direction robust; significance not guaranteed
Infratentorial origin	66.2 [14.3–3,294]	6.6×10^{-8} to 1.3×10^8	0.014	90.5	99.7	Not identifiable (range spans 1)
Male sex	1.09 [0.14–7.47]	2.8×10^{-3} to 4,477	< 0.001	0.9	54.1	Not identifiable; no association

Admissible datasets reproduce exactly the observed component-by-outcome tables and the score distribution among survivors and non-survivors. Percentiles and percentages from 20,000 sampled datasets per model; minima and maxima are the most extreme values found by sampling and simulated annealing and are lower bounds on the attainable range. *p* values from penalized-likelihood-ratio tests. Across the eight annealing restarts of Model A, the smallest IVH OR found ranged from 62.6 to 76.9.



Grey line: most extreme values found; blue bar: 2.5th–97.5th percentile of 20,000 admissible datasets; circle: median.

Figure 8. Adjusted odds ratios (Firth logistic regression, log scale) of the ICH Score components across 20,000 patient-level datasets consistent with the observed data (Model A). The dashed line marks an odds ratio of 1; the right column gives the percentage of datasets in which each component reached $p < 0.05$.

Removing IVH, age or GCS from the score lowered the AUC in every admissible dataset (attainable ranges 0.791–0.865, 0.834–0.897 and 0.852–0.926, respectively, against 0.958 for the full score). Removing infratentorial origin lowered it in 99.9% of datasets (0.898–0.959), whereas removing volume could either lower or raise discrimination (0.903–0.992; lower in 94.8%), as detailed in Table 8 and shown in Figure 9. Integer

programming showed that 16–21 of the 21 deaths at score 3 had a volume ≥ 30 mL and 6–14 had IVH, while 27–32 of the 32 survivors at score 2 had a volume ≥ 30 mL and at most three had IVH; which features distinguished the four deaths at score 2 could not be determined; these exact bounds are detailed in Table 9.

Table 8. Discrimination of the ICH Score with one component removed, bounded across the admissible datasets.

Score version	Median AUC [2.5th–97.5th percentile]	Attainable AUC range found	Datasets with AUC below full score, %	Interpretation
Full ICH Score (observed)	0.958 (DeLong 95% CI 0.917–0.998)	—	—	Reference
Without IVH	0.848 [0.828–0.865]	0.791–0.865	100.0	Removal always lowers AUC
Without age ≥ 80 years	0.877 [0.856–0.897]	0.834–0.897	100.0	Removal always lowers AUC
Without GCS points	0.890 [0.867–0.914]	0.852–0.926	100.0	Removal always lowers AUC
Without infratentorial origin	0.928 [0.911–0.949]	0.898–0.959	99.9	Removal lowers AUC in almost every dataset
Without volume ≥ 30 mL	0.931 [0.910–0.961]	0.903–0.992	94.8	Contribution not identifiable; removal can raise AUC

AUC of the summed score after removing one component, calculated in each of 20,000 admissible datasets; ranges include the extremes found by simulated annealing.

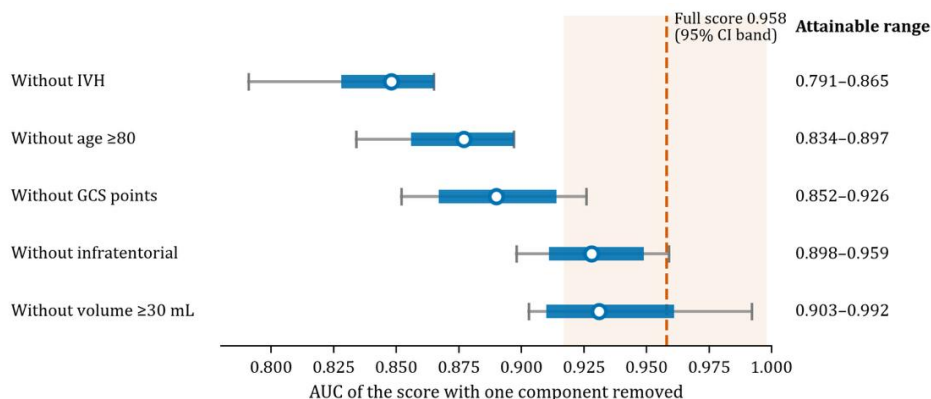


Figure 9. AUC of the ICH Score with one component removed across the admissible datasets, compared with the observed AUC of the full score (dashed line; shaded band, DeLong 95% CI).

Table 9. Exact bounds on the number of patients with each feature within each score stratum.

Stratum	GCS 3–4 (deaths) or 13–15 (survivors)	Age ≥ 80 years	Volume ≥ 30 mL	IVH	Infratentorial origin
Deaths, score 2 ($n = 4$)	0–4	0–4	0–4	0–4	0–4

Stratum	GCS 3–4 (deaths) or 13–15 (survivors)	Age ≥80 years	Volume ≥30 mL	IVH	Infratentorial origin
Deaths, score 3 (<i>n</i> = 21)	0–6	2–10	16–21	6–14	0–4
Deaths, score 4 (<i>n</i> = 4)	0–4	0–4	0–4	0–4	0–4
Survivors, score 0 (<i>n</i> = 1)	1	0	0	0	0
Survivors, score 1 (<i>n</i> = 19)	0–5	0–1	0–5	0–3	0–1
Survivors, score 2 (<i>n</i> = 32)	0–5	0–1	27–32	0–3	0–1

Minimum–maximum number of patients with the feature, obtained by integer programming over all patient-level datasets consistent with the observed data; a single value indicates that the number is fixed.

4. Discussion

In this cohort of 81 patients who underwent surgical evacuation of spontaneous ICH at RSMH, the admission ICH Score discriminated 30-day survivors from non-survivors with near-complete separation (AUC 0.958, 95% CI 0.917–0.998): no patient scoring 0–1 died and no patient scoring 3–4 survived. The score was nonetheless poorly calibrated to its original risk estimates. The risk gradient was about three times steeper than in the derivation cohort, so the original percentages overstated the danger at scores 1–2 and understated it at score 3. Among the components, age ≥80 years and IVH had the strongest unadjusted associations with death, IVH was the only component whose adjusted association remained significant across every dataset examined in the component model, and the adjusted contribution of hematoma volume could not be identified from these data.

Discrimination in context

The discrimination observed here is at the upper end of published values. A Vietnamese single-center cohort of 598 patients reported a virtually identical AUC (95.75%) with a 30-day mortality of 18.4%,⁹ whereas the multinational RES-Q registry, which included Indonesian centers, reported 0.82 for discharge mortality in 8,450 patients,¹⁹ a Dallas cohort 0.890 at admission and 0.914 after 6-hour reassessment,²¹ a Singaporean cohort 0.862,²² and a trial-derived North American cohort 0.80.²⁰ For functional outcome rather than death, the score achieved 0.819 in India,⁴⁸ and a Chinese comparison of 27 models found a ceiling of 0.856 for 3-month mortality.²³ Cohorts from Bangladesh, Saudi Arabia and Iran report a monotonic rise in mortality with each score point and 30-day or in-hospital mortality between 30% and 44%, comparable with the 35.8% recorded here.^{49–51} Within surgical cohorts specifically, the score has generally discriminated less well — 0.71 in Peru, 0.757 in Poland and 0.80 in Denmark — and an inflammation-based nomogram reached 0.915–0.936 in 587 operated Chinese patients.^{29–31,52} The present AUC therefore reflects an unusually clean separation of outcomes at the extremes of the score rather than a property that can be expected in every operated population. Selection into surgery shapes this spectrum: patients judged too ill to benefit, who are likely to have high scores and to die, were not eligible, and patients with small hemorrhages and good consciousness, who are likely to survive, are rarely operated on. Excluding both groups narrows the range over which the score must discriminate; in a population that included non-operated patients the AUC could be either higher, because the extremes add easily classified patients, or lower, because medical management and earlier deaths would change mortality within each stratum. Two features probably contribute: the cohort contained almost no patients at the uninformative extremes (one patient scored 0 and none scored 5–6), and the score-2 stratum, the only one with both outcomes, contained 44% of patients but only four deaths. Because no parameter was estimated, the AUC carries no optimism from model fitting, but its interval (0.917–0.998) is wide enough to be compatible with the values reported in the larger surgical series.

Calibration and the steep risk gradient

The main finding is the shape of the risk gradient. Hemphill and colleagues observed 30-day mortality rising progressively through the strata,¹⁸ and the original percentages have since been shown to overestimate risk in young adults²⁷ and after stereotactic aspiration, where observed mortality was 10.4% against 35.1% predicted.²⁸ A Japanese surgical series reported an even larger gap, 7% observed against 55% predicted,³² and in a Canadian cohort managed with full medical and surgical support 30-day mortality was only 16%.⁵³ Our results, detailed in Table 3, reproduce the over-prediction at low scores — no deaths at score 1 against 13% predicted and 11.1% at score 2 against 26% predicted — and agree exactly with the 11% observed at score 2 in a Danish series of operated supratentorial hemorrhage.³¹ They diverge sharply at score 3, where every patient died, as shown in Figure 6. In the Danish cohort only 33% of operated patients with a score of 3 died, and in the Florida Stroke Registry, which followed 12,426 patients over 10 years, in-hospital mortality for scores 3–4 combined was 41.5%.^{31,41} Even if the true risk had been 33%, 21 deaths among 21 patients would have occurred with a probability below 0.001; chance is therefore an unlikely explanation.

Several explanations are plausible and none can be confirmed with the present data. One is limitation of care. In the Florida registry the odds of withdrawal of life-sustaining treatment were 9.35 times higher for scores 3–4 than for scores 0–2,⁴¹ in a registry of 37,613 patients withdrawal mediated a substantial part of the excess mortality associated with impaired consciousness,⁴⁰ and early do-not-resuscitate orders in the United Kingdom were associated with much lower odds of critical-care escalation (OR 0.07) and higher odds of palliation within 72 hours (OR 8.76).⁴² If such decisions concentrated in patients with high scores, the score would have predicted the decisions as much as the disease. Another is case mix within the score-3 stratum: integer programming showed that at least 16 of the 21 deaths at score 3 had a volume ≥30 mL and at least six had IVH, a pathway that parenchymal evacuation does not treat; the bounds do not show how many patients had both features. A further possibility is system capacity. Access to neurosurgical intensive care, the timing of referral and the interval from ictus to surgery all influence survival, and time to evacuation was associated with outcome in endoscopic surgery (OR 0.95 per hour of delay).⁵⁴ Finally, mortality was counted from surgery rather than from ictus. Patients who died before surgery were excluded by design, and the 30-day window ends later after the ictus than in the derivation cohort, which tends to add deaths; because most deaths after ICH occur early, this shift is probably small, but it acts in the direction of higher observed mortality and a steeper apparent gradient. None of these factors — care limitation, operative timing, intensive-care access or the interval from ictus to surgery — was recorded in the present cohort, so their contributions cannot be separated.

The steep gradient has a practical consequence. As detailed in Table 3, a clinician using the original percentages would tell the family of a patient with a score of 2 that the risk of death is about one in four, when in this setting it was about one in nine, and

would describe a score of 3 as carrying a 72% risk when every such patient died. The locally recalibrated model gives 12.2% at score 2 and 97.7% at score 3; bootstrap internal validation showed little optimism, but with only 29 deaths these estimates are imprecise and should be regarded as a local benchmark to be updated rather than as fixed values. The finding should not be translated into a futility threshold. The perfect specificity at a cutoff of ≥ 3 rests on 25 patients; its lower confidence bound for the positive predictive value is 86.7%, and a Danish operated cohort in which one-third of score-3 patients died shows that survival at this level is possible when care is continued. In cerebellar hemorrhage, surgery reduced mortality specifically above a score of 3, which argues against using a high score to withhold operation.³⁴

Score components

The association between age ≥ 80 years and death was the strongest unadjusted signal (OR 26.84), with 10 of 11 octogenarians dying. This contrasts with an Indonesian tertiary-center cohort in which age ≥ 80 years was not significant ($p = 0.546$) and a lower threshold of 59 years was proposed,⁵ and with a Portuguese cohort of patients aged ≥ 75 years in which age was not an independent predictor once volume and GCS were considered.⁵⁵ Very old patients nevertheless carry high case fatality — 40.6% in a Portuguese series in which the ICH Score still achieved an AUC of 0.882,⁵⁶ and 20% in-hospital fatality among Filipino ICH patients aged ≥ 80 years.¹¹ After decompressive craniectomy, mortality in patients older than 65 years was 33.3% and no patient in that group achieved good recovery at 12 months,⁵⁷ whereas in Japanese patients aged ≥ 75 years endoscopic surgery and craniotomy produced similar

functional outcomes.⁵⁸ In the present cohort the adjusted age effect was positive in every admissible dataset but not always significant, so the very high unadjusted OR partly reflects co-occurring severity rather than age alone.

IVH was the most robust component. It was associated with an absolute increase in mortality of 58.9 percentage points, it remained positive and significant in every admissible dataset of the component model (including the most extreme datasets found in every annealing restart), as detailed in Table 7, its direction was preserved when sex was added, and removing it from the score caused the largest loss of discrimination. This agrees with a meta-analysis of hyperacute predictors (OR 2.05 for IVH),²⁵ with Spanish data showing that bilateral lateral-ventricle extension carries an OR of 3.93,¹⁶ with the CLEAR III observation that IVH expands in 27% of patients,¹⁵ and with an Indonesian-led meta-analysis showing that IVH with hydrocephalus roughly doubles 30-day mortality compared with IVH alone.¹⁷ The mechanism is relevant to surgical planning: parenchymal evacuation removes the mass lesion but does not clear ventricular clot or relieve obstructive hydrocephalus, as summarized in the conceptual model in Figure 10. In a Boston cohort of 2,486 patients, external ventricular drainage in patients with IVH was associated with lower 90-day mortality (adjusted OR 0.19),⁵⁹ and in operated Chinese patients adding the modified Graeb score to the ICH Score improved prediction.³³ Because the use of external ventricular drainage was not recorded here, we cannot tell whether the excess mortality with IVH reflected untreated hydrocephalus or the severity of the hemorrhage itself.

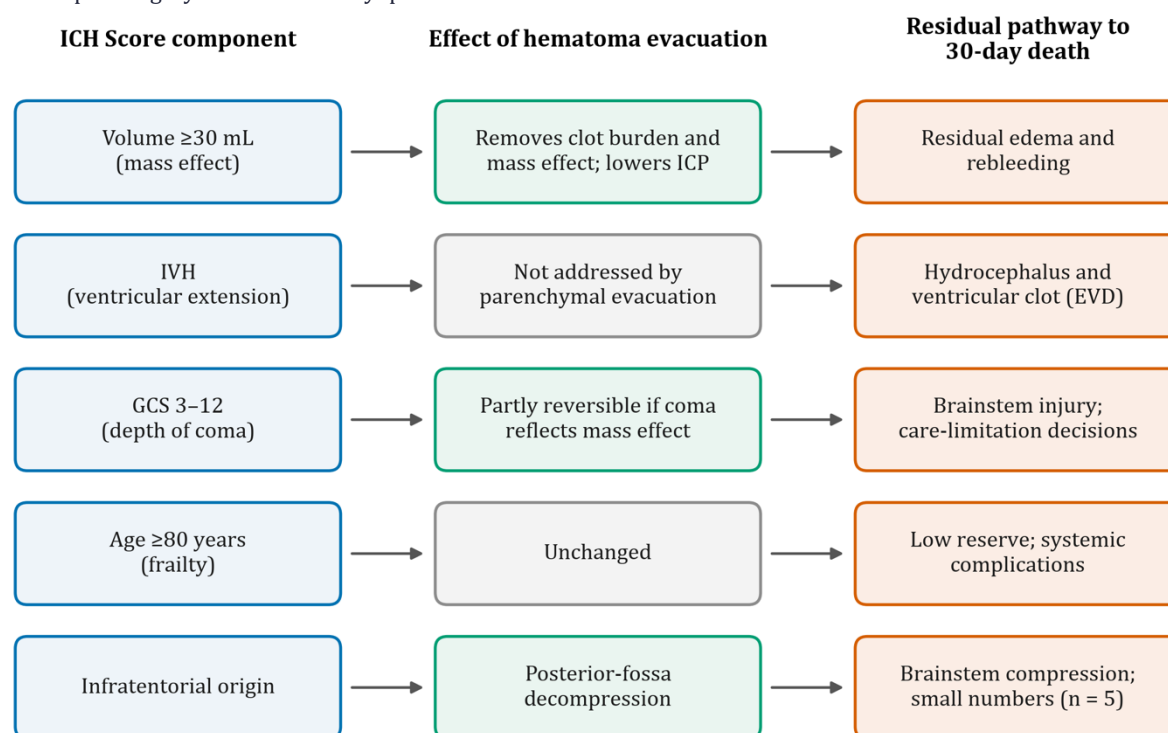


Figure 10. Conceptual, hypothesis-generating model of how surgical hematoma evacuation modifies the pathways captured by each ICH Score component, and the residual pathways to 30-day death that evacuation does not address. EVD, external ventricular drain; ICP, intracranial pressure; WLST, withdrawal of life-sustaining treatment.

GCS behaved as expected: mortality rose from 0% at 13–15 to 33.3% at 5–12 and 100% at 3–4. In adjusted analyses its direction was positive in every dataset but its significance was not guaranteed, consistent with the overlap between coma depth, IVH and age in severely affected patients. Withdrawal of treatment also mediates much of the mortality associated with

depressed consciousness,⁴⁰ so its prognostic weight in any single cohort reflects local practice as well as biology.

Hematoma volume deserves particular attention because it is the component most directly modified by surgery. Its unadjusted association was borderline (42.9% versus 20.0%; OR 3.00, 95% CI 0.98–9.14), its adjusted effect could not be

identified, and removing it from the score could either lower or raise the AUC, as detailed in Table 8. In non-surgical or mixed cohorts volume is a consistent predictor — volume >30 mL carried an OR of 2.68 for 30-day death among 964 West Africans and 8.82 among elderly Portuguese patients^{55,60} — and these data are consistent with the hypothesis that evacuation attenuates the prognostic weight of the initial volume. They do not prove it: removing volume lowered the AUC in 94.8% of the sampled datasets, but the attainable range (0.903–0.992) includes values above the AUC of the full score, so a residual contribution cannot be established. More than two-thirds of the cohort had large hematomas, which limits contrast; ABC/2 overestimates planimetric volume by about 4 mL on average and discriminates less well than volumetry (AUC 0.690 versus 0.739),⁴⁵ so threshold misclassification is likely; and perihematomal edema remains prognostic both without surgery¹² and after evacuation, where postoperative edema of 75–100 mL was associated with an OR of 6.75 for functional dependence.¹³ Lower hemoglobin, which is associated with hematoma expansion, is a further, potentially modifiable contributor that the score does not capture.¹⁴

Only five patients had infratentorial hemorrhage, and four of them died. The attainable adjusted ORs spanned 1, so these data cannot characterize the infratentorial component. Cerebellar hemorrhage behaves differently from supratentorial hemorrhage after surgery: in a nine-center German cohort of 403 patients, surgery reduced mortality only when the ICH Score exceeded 3,³⁴ and long-term survival after cerebellar hemorrhage is limited, with 34% of patients alive at 10 years in an international cohort.⁶¹ Posterior-fossa hemorrhage should be studied separately in larger samples.

Surgical evidence and regional context

The present study was not designed to test whether surgery benefits patients, because every patient was operated. The surrounding evidence nevertheless frames the results. ENRICH showed that early minimally invasive evacuation improved utility-weighted modified Rankin scores (0.458 versus 0.374), SWITCH showed a borderline reduction in death or severe dependency after decompressive craniectomy for deep hemorrhage (adjusted risk ratio 0.77, 95% CI 0.59–1.01), and MIND found no superiority of minimally invasive surgery.^{35,36,38} The 2025 Cochrane review estimated a risk ratio of 0.79 for death with surgery,³⁷ and in INTERACT3, which recruited across Chinese hospitals, operated patients had lower 6-month mortality (OR 0.55).³⁹ A Thai randomized trial favored endoscopic surgery over craniotomy (favorable outcome 48.4% versus 35.5%),⁶² a United States registry associated minimally invasive surgery with lower in-hospital death (13.5% versus 23.5%; adjusted OR 0.50) without functional gain,⁶³ and in Chinese patients with a National Institutes of Health Stroke Scale (NIHSS) score ≥10 surgery improved 1-year functional independence (49.7% versus 38.3%).⁶⁴ By contrast, an Indian cohort of 661 patients found no difference in long-term functional outcome once the ICH Score was accounted for.⁶⁵ Against this background, the low mortality at scores 1–2 and the uniform mortality at scores 3–4 describe outcomes after surgery at one Indonesian center; they cannot be read as evidence for or against the operation, but they explain why calibration is likely to drift as surgical selection and technique change.

The regional context matters for transportability. Asian patients in blood-pressure trials had lower 90-day death or dependency than non-Asian patients,³ whereas short-term mortality in Kuala Lumpur (29.1%), Vietnam (18.4% at 30 days), Thailand (47.1% at discharge among operated patients) and Indonesia (17.8% in hospital to 39.9% at 30 days) varies

widely between centers.^{5,7–10} In Bandung, the ICH Score remained an independent predictor of death in hypertensive ICH (OR 2.757),⁶ and single-center Indonesian data link hemodynamic markers such as wide pulse pressure to in-hospital mortality after hemorrhagic stroke.⁶⁶ Indonesian electronic health records have begun to support locally derived stroke prognostic models,⁶⁷ and national insurance data highlight the metabolic and socioeconomic burden that shapes stroke outcomes.⁴ Features of Indonesian referral care — inter-hospital transfer times, the availability of neurosurgical intensive-care beds and family-centered decisions about the intensity of treatment — could each shift calibration, but none has been measured in relation to ICH outcome, and they remain hypotheses. These sources suggest that prognostic tools for Indonesian neurosurgical practice should be recalibrated, and probably extended, rather than imported with their original risk tables.

Implications for practice

Three implications follow. First, the ICH Score ranked operated patients by risk well at RSMH and can help structure triage and family discussions, pending external validation. Second, the original percentages should not be quoted to families of operated patients in this setting; locally recalibrated estimates, presented with their uncertainty, may be more appropriate, although they are in-sample estimates that require external validation, and the decision-curve analysis shows that a ≥3 rule provided at least as much apparent net benefit as the original stratum risks at every threshold and more below 26% or above 72%. Third, a high score should prompt a structured multidisciplinary review of goals of care, ventricular drainage and intensive-care capacity rather than an automatic conclusion of futility — whether earlier ventricular drainage changes outcome in these patients is a hypothesis for prospective evaluation, not a finding of this study — because the observed 100% mortality at scores 3–4 may partly reflect care decisions that the score itself influenced. At the institutional level, documenting the timing and reasons for any limitation of treatment would allow future analyses to separate prognosis from practice.

Methodological contribution

The hospital records cross-tabulated each component and the total score against the outcome but not the components against one another. Instead of imputing patient-level data, we bounded the adjusted analyses across every dataset consistent with the observations. This converts an unanswerable question — what are the adjusted ORs? — into answerable ones: which conclusions hold in every compatible dataset, which hold in none, and which depend on information that was not recorded. The approach addresses a concern raised by systematic reviews of ICH prognostic models, which frequently find prognostic models developed in small samples at high risk of overfitting,^{24,26} and it makes explicit that adjusted component effects in a cohort whose outcome is almost determined by the summed score are not interpretable as magnitudes.

Research agenda

Four studies would resolve the questions left open here. First, a prospective multicenter registry of operated ICH in South Sumatra and other Indonesian provinces should enroll enough patients to estimate calibration precisely; current guidance favors precision-based calculations over rules of thumb, and even 100–200 deaths, or roughly 280–560 operated patients at the mortality observed here, may be insufficient.^{46,47} Second, that registry should record the timing and reasons for any limitation of care, the use of external ventricular drainage, the

type of operation and the interval from ictus to surgery, so that the self-fulfilling-prophecy hypothesis and the IVH mechanism can be tested directly. Third, it should compare the original ICH Score with the age-modified and point-modified scores proposed for Indonesian patients, with the max-ICH, FUNC and ICH Functional Outcome Score (ICH-FOS) scores, and with extensions that add inflammatory or nutritional markers or serial reassessment.^{5,7,19,21,52,68,69} Machine-learning models, which reach a pooled mortality C-index of 0.860 but are seldom externally validated,²⁶ should be compared on the same data.²² Fourth, outcome should be anchored at ictus and extended to 90-day and 12-month functional status, because mortality alone cannot distinguish good survival from survival with severe disability.^{53,68}

Strengths and limitations

The study has several strengths. To our knowledge, it is the first evaluation of the ICH Score restricted to operated patients in Indonesia, it reports discrimination together with calibration and clinical utility, every unadjusted statistic was computed exactly from fully observed tables, and the adjusted analyses avoid imputing patient-level data while stating precisely which conclusions the data support.

The limitations are substantial. This was a retrospective single-center study with 29 deaths, so estimates are imprecise and calibration in particular should be revalidated. Patients who were not operated on, including those considered too ill for surgery or whose families declined it, were not included, so the results apply only to patients selected for surgery. The cross-classification of the score components was not available, which is why adjusted effects were reported as bounds; exact GCS values, continuous age and volume, the timing of surgery, the operative technique, external ventricular drainage and decisions to limit treatment were not recorded. ABC/2 volumes were measured by the clinical team without assessment of inter-rater reliability. Mortality was measured from surgery rather than from ictus and functional outcome was not available. Finally, the comparisons with published cohorts used percentages reported in those studies and are descriptive.

5. Conclusion

Among patients undergoing surgical evacuation of spontaneous ICH at a tertiary referral center in South Sumatra, the admission ICH Score discriminated 30-day mortality well (AUC 0.958, 95% CI 0.917–0.998), with no deaths at scores 0–1, 11.1% at score 2 and 100% at scores 3–4. Its original risk estimates, however, were miscalibrated, with a gradient about three times steeper than in the derivation cohort (recalibration slope 3.00). IVH was the most robust component, whereas the adjusted contributions of hematoma volume and infratentorial origin could not be identified. The score can be used to rank risk and, once locally recalibrated estimates are validated, to structure counseling with them, but not to define a futility threshold. For neurosurgical teams at RSMH and comparable Indonesian referral centers, the practical message is to retain the ICH Score as a common language for severity, to consider locally observed risks, with their uncertainty, rather than the original percentages when counseling families, and to treat a high score as a trigger for structured review of ventricular drainage, intensive-care needs and goals of care. A prospective multicenter Indonesian registry that records care-limitation decisions, ventricular drainage, operative technique and functional outcome is needed to confirm these findings and to establish a recalibrated, surgically relevant prognostic tool.

6. Declarations

Ethics Approval and Consent to Participate

The protocol received formal approval from the Research Ethics Committee of Dr. Mohammad Hoesin General Hospital Palembang, under document number DP.04.03/D.XVIII.06.08/ETIK/014/2026. The study was conducted in accordance with the Declaration of Helsinki. Individual informed consent was waived because only existing, anonymized medical records were analyzed.

Consent for Publication

Not applicable; the article contains no individual patient data, images or other identifying information.

Availability of Data and Materials

All aggregate data analyzed are contained in the tables of this article. The anonymized patient-level dataset is held by the Department of Neurosurgery, Dr. Mohammad Hoesin General Hospital Palembang, and is available from the corresponding author on reasonable request, subject to approval by the hospital's research ethics committee. Analysis code is available from the corresponding author.

Artificial Intelligence Use Statement

None.

Author Contribution Statement

G.C.: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft. T.P.: Conceptualization, Supervision, Validation, Writing – review & editing. Z.M.: Methodology, Formal analysis, Supervision, Writing – review & editing. All authors read and approved the final manuscript.

Funding Statement

None.

Conflict of Interest / Competing Interests

The authors declare that they have no competing interests.

Acknowledgements

None.

7. References

1. Feigin VL, Abate MD, Abate YH, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Neurol.* 2024;23(10):973–1003. doi: 10.1016/S1474-4422(24)00369-7.
2. Yang R, Wang X, Hao P. The burden of intracerebral hemorrhage attributable to ambient particulate matter pollution in five Asian countries: a 32-year comparative analysis. *BMC Public Health.* 2025;25(1):2467. doi: 10.1186/s12889-025-23531-4.
3. You S, Ouyang M, Moullaali TJ, et al. Clinical outcome differences in Asia versus non-Asia intracerebral hemorrhage: IPD analysis from BP-lowering clinical trials. *Stroke.* 2026. Epub ahead of print. doi: 10.1161/STROKEAHA.126.055677.
4. Zainuddin AA, Kadir RRA, Qalby N, et al. Syndemic analysis of stroke in Indonesia. *Lancet Reg Health West Pac.* 2026;69:101849. doi: 10.1016/j.lanwpc.2026.101849.
5. Sasongko AB, Perdana Wahjoepramono PO, Halim D, et al. Age-modified and point-modified intracerebral hemorrhage (ICH) scores are more reliable alternatives to the original ICH score for Indonesian patients: a retrospective cohort study at a tertiary center in Indonesia. *World Neurosurg X.* 2025;26:100444. doi: 10.1016/j.wnsx.2025.100444.
6. Amalia L, Fadhillah ST, Pusparini I, et al. Predictors of mortality of hypertensive intracerebral hemorrhage (HICH) patients: a single centre study. *BMC Neurol.* 2026;26(1):333. doi: 10.1186/s12883-026-04847-z.
7. Widyadharma IPE, Krishna A, Soejitno A, et al. Modified ICH score was superior to original ICH score for assessment of 30-day

- mortality and good outcome of non-traumatic intracerebral hemorrhage. *Clin Neurol Neurosurg.* 2021;209:106913. doi: 10.1016/j.clineuro.2021.106913.
8. Ramalu RR, Kandasamy R, Raffiq A, et al. Risk factors associated with outcome of spontaneous intracerebral haemorrhage: Hospital Kuala Lumpur experience. *Malays J Med Sci.* 2022;29(1):76-90. doi: 10.21315/mjms.2022.29.1.8.
 9. Ba Lan T, Hoai Tuan Anh C, Lan Anh Nguyen T, et al. Short-term mortality prognosis in spontaneous intracranial hemorrhage: a retrospective study at 115 People's Hospital, HCMC, Vietnam. *J Public Health Res.* 2025;14(2):22799036251334178. doi: 10.1177/22799036251334178.
 10. Kittiwattananagul W, Namwaing P, Khamsai S, et al. Clinical characteristics and outcomes of patients with intracerebral hemorrhage: nonsurgical versus surgical treatment. *J Emerg Trauma Shock.* 2023;16(4):145-9. doi: 10.4103/jets.jets_55_23.
 11. Cañete MT, Im S, Pascual JL, et al. Stroke in the very old: a study from the Philippine Neurological Association One Database. *Australas J Ageing.* 2024;43(2):409-14. doi: 10.1111/ajag.13271.
 12. Samarasekera N, Ferguson K, Parry-Jones AR, et al. Perihematomal oedema evolution over 2 weeks after spontaneous intracerebral haemorrhage and association with outcome: a prospective cohort study. *Cerebrovasc Dis.* 2025;54(3):298-307. doi: 10.1159/000540099.
 13. Wu X, Liu H, Zhang R, et al. Prognostic significance of perihematomal edema in basal ganglia hemorrhage after minimally invasive endoscopic evacuation. *J Neurosurg.* 2023;139(6):1784-91. doi: 10.3171/2023.4.JNS222910.
 14. Cottarelli A, Mamoon R, Ji R, et al. Low hemoglobin causes hematoma expansion and poor intracerebral hemorrhage outcomes. *Stroke.* 2025;56(5):1234-42. doi: 10.1161/STROKEAHA.124.049499.
 15. Roh DJ, Asonye IS, Carvalho Poyraz F, et al. Intraventricular hemorrhage expansion in the CLEAR III trial: a post hoc exploratory analysis. *Stroke.* 2022;53(6):1847-53. doi: 10.1161/STROKEAHA.121.037438.
 16. García-Pérez D, Campos J, Ayo-Martín Ó, et al. Topographic patterns of intraventricular hemorrhage extension and their prognostic significance in intracerebral hemorrhage. *Neurol Sci.* 2026;47(3):288. doi: 10.1007/s10072-026-08915-7.
 17. Wahjoepramono POP, Sasongko AB, Halim D, et al. Hydrocephalus is an independent factor affecting morbidity and mortality of ICH patients: systematic review and meta-analysis. *World Neurosurg X.* 2023;19:100194. doi: 10.1016/j.wnsx.2023.100194.
 18. Hemphill JC, Bonovich DC, Besmertis L, et al. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. *Stroke.* 2001;32(4):891-7. doi: 10.1161/01.str.32.4.891.
 19. de Souza AC, Neto G, Thang NH, et al. Comparative external validation of ICH prognostic scores for discharge outcomes in the RES-Q multinational quality registry. *Eur Stroke J.* 2026;11(9):aakag105. doi: 10.1093/esj/aakag105.
 20. Kazaryan SA, Shkirkova K, Saver JL, et al. The National Institutes of Health Stroke Scale is comparable to the ICH score in predicting outcomes in spontaneous acute intracerebral hemorrhage. *Front Neurol.* 2024;15:1401793. doi: 10.3389/fneur.2024.1401793.
 21. Meyrat R, Vivian E, Dulaney B, et al. Enhanced prediction of in-hospital mortality in intracerebral hemorrhage: impact of serial neurological and radiological reassessment with the ICH score at 6 hours postadmission. *J Neurosurg.* 2026;144(2):442-51. doi: 10.3171/2025.7.JNS251312.
 22. Lim MJR, Quek RHC, Ng KJ, et al. Machine learning models prognosticate functional outcomes better than clinical scores in spontaneous intracerebral haemorrhage. *J Stroke Cerebrovasc Dis.* 2022;31(2):106234. doi: 10.1016/j.jstrokecerebrovasdis.2021.106234.
 23. Ji R, Wang W, Liu X, et al. Head-to-head comparison of prognostic models of spontaneous intracerebral hemorrhage: tools for personalized care and clinical trial in ICH. *Neurol Res.* 2022;44(2):146-55. doi: 10.1080/01616412.2021.1967678.
 24. Simon-Pimmel J, Foucher Y, Léger M, et al. Methodological quality of multivariate prognostic models for intracranial haemorrhages in intensive care units: a systematic review. *BMJ Open.* 2021;11(9):e047279. doi: 10.1136/bmjopen-2020-047279.
 25. Hammerbeck U, Abdulle A, Heal C, et al. Hyperacute prediction of functional outcome in spontaneous intracerebral haemorrhage: systematic review and meta-analysis. *Eur Stroke J.* 2022;7(1):6-14. doi: 10.1177/23969873211067663.
 26. Deng Q, Cheng W, Lu H, et al. Machine learning-based prediction of poor outcomes in intracerebral hemorrhage: a systematic review and meta-analysis. *Brain Behav.* 2026;16(8):e71572. doi: 10.1002/brb3.71572.
 27. Maljaars J, Garg A, Molian V, et al. The intracerebral hemorrhage score overestimates mortality in young adults. *J Stroke Cerebrovasc Dis.* 2021;30(9):105963. doi: 10.1016/j.jstrokecerebrovasdis.2021.105963.
 28. Shapiro SD, Alkayali M, Reynolds A, et al. Stereotactic IntraCerebral Underwater Blood Aspiration (SCUBA) improves survival following intracerebral hemorrhage as compared with predicted mortality. *World Neurosurg.* 2022;161:e289-94. doi: 10.1016/j.wneu.2022.01.123.
 29. Rodriguez-Calienes A, Malaga M, Alva-Diaz C, et al. Validation of the ICH score and ICH-GS in a Peruvian surgical cohort: a retrospective study. *Neurosurg Rev.* 2022;45(1):763-70. doi: 10.1007/s10143-021-01605-2.
 30. Wiśniewski K, Zaczekowski K, Podstawka M, et al. Predictors of 30-day mortality for surgically treated patients with spontaneous supratentorial intracerebral hemorrhage and validation of the Surgical Swedish Intracerebral Hemorrhage Score: a retrospective single-center analysis of 136 cases. *World Neurosurg.* 2024;186:e539-51. doi: 10.1016/j.wneu.2024.03.172.
 31. Haga JA, Poulsen FR, Forsse A. Validation of the newly conceived Surgical Swedish ICH grading scale for surgically treated patients with intracerebral hemorrhage: patient series. *J Neurosurg Case Lessons.* 2021;1(1):CASE2044. doi: 10.3171/CASE2044.
 32. Hong S, Maruyama K, Noguchi A, et al. Is using intracerebral hemorrhage scoring systems valid for mortality prediction in surgically treated patients? *Neurosurg Rev.* 2021;44(5):2747-53. doi: 10.1007/s10143-020-01451-8.
 33. Wang S, Xu X, Yu Q, et al. Combining modified Graeb score and intracerebral hemorrhage score to predict poor outcome in patients with spontaneous intracerebral hemorrhage undergoing surgical treatment. *Front Neurol.* 2022;13:915370. doi: 10.3389/fneur.2022.915370.
 34. Won SY, Walter J, Hernandez-Duran S, et al. Reappraisal of intracerebral hemorrhages and intracerebral hemorrhage grading scale score in surgically and medically managed cerebellar intracerebral hemorrhage. *Neurosurgery.* 2023;92(5):1021-8. doi: 10.1227/neu.0000000000002318.
 35. Pradilla G, Ratcliff JJ, Hall AJ, et al. Trial of early minimally invasive removal of intracerebral hemorrhage. *N Engl J Med.* 2024;390(14):1277-89. doi: 10.1056/NEJMoa2308440.
 36. Beck J, Fung C, Strbian D, et al. Decompressive craniectomy plus best medical treatment versus best medical treatment alone for spontaneous severe deep supratentorial intracerebral haemorrhage: a randomised controlled clinical trial. *Lancet.* 2024;403(10442):2395-404. doi: 10.1016/S0140-6736(24)00702-5.
 37. Wilting FN, Sondag L, Schreuder FH, et al. Surgery for spontaneous supratentorial intracerebral haemorrhage. *Cochrane Database Syst Rev.* 2025;2025(9):CD015387. doi: 10.1002/14651858.CD015387.pub2.
 38. Arthur AS, Jahromi BS, Saphier PS, et al. Minimally invasive surgery vs medical management alone for intracerebral hemorrhage: the MIND randomized clinical trial. *JAMA Neurol.* 2025;82(11):1113-21. doi: 10.1001/jamaneurol.2025.3151.
 39. Hu X, Ouyang M, Xu J, et al. Surgical outcomes from haematoma evacuation for intracerebral haemorrhage in the INTERACT3 study. *Lancet Reg Health West Pac.* 2025;62:101669. doi: 10.1016/j.lanwpc.2025.101669.
 40. Alkhachroum A, Bustillo AJ, Asdagh N, et al. Withdrawal of life-sustaining treatment mediates mortality in patients with intracerebral hemorrhage with impaired consciousness. *Stroke.* 2021;52(12):3891-8. doi: 10.1161/STROKEAHA.121.035233.

41. Massad N, Zhou L, Manolovitz B, et al. Association of the ICH score with withdrawal of life-sustaining treatment over a 10-year period. *Ann Clin Transl Neurol.* 2025;12(10):1992-2001. doi: 10.1002/acn3.70136.
42. Minhas JS, Sammut-Powell C, Birlleson E, et al. Are do-not-resuscitate orders associated with limitations of care beyond their intended purpose in patients with acute intracerebral haemorrhage? Analysis of the ABC-ICH study. *BMJ Open Qual.* 2021;10(1):e001113. doi: 10.1136/bmj-2020-001113.
43. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007;370(9596):1453-7. doi: 10.1016/S0140-6736(07)61602-X.
44. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. doi: 10.1136/bmj-2023-078378.
45. Zaczkowski K, Podstawka M, Dębska A, et al. The difference between the ABC/2 method and volumetric assessment in evaluating the volume of intracerebral hematomas and its impact on prognostic scale outcomes. *J Clin Neurosci.* 2025;142:111703. doi: 10.1016/j.jocn.2025.111703.
46. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med.* 2021;40(19):4230-51. doi: 10.1002/sim.9025.
47. Snell KIE, Archer L, Ensor J, et al. External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. *J Clin Epidemiol.* 2021;135:79-89. doi: 10.1016/j.jclinepi.2021.02.011.
48. Babu S, Pulicken M, Thazhathuveedu AK. Peripheral blood neutrophil-to-lymphocyte ratio as a predictor of functional outcomes in patients with hemorrhagic stroke. *Indian J Crit Care Med.* 2022;26(1):18-22. doi: 10.5005/jp-journals-10071-24074.
49. Ray SK, Sadekur Rahman Sarkar M, Ahmed KMA, et al. Predicting 30-day outcomes in primary intracerebral hemorrhage using the intracerebral hemorrhage score: a study in Bangladesh. *Cureus.* 2024;16(11):e73227. doi: 10.7759/cureus.73227.
50. Albakr AI. Spontaneous intracerebral hemorrhage among hypertensive patients in Saudi Arabia: study from a tertiary center. *Saudi J Med Med Sci.* 2022;10(2):139-45. doi: 10.4103/sjmms.sjmms_73_22.
51. Faghih-Jouybari M, Taghi Raof M, Abdollahzade S, et al. Mortality and morbidity in patients with spontaneous intracerebral hemorrhage: a single-center experience. *Curr J Neurol.* 2021;20(1):32-6. doi: 10.18502/cjn.v20i1.6377.
52. Luo S, Li H, Miao Y, et al. Predictive value of the systemic immune-inflammation index in 28-day survival after surgery for intracerebral hemorrhage. *Eur J Med Res.* 2025;30(1):858. doi: 10.1186/s40001-025-03067-4.
53. Abulhasan YB, Teitelbaum J, Al-Ramadhani K, et al. Functional outcomes and mortality in patients with intracerebral hemorrhage after intensive medical and surgical support. *Neurology.* 2023;100(19):e1985-95. doi: 10.1212/WNL.0000000000207132.
54. Kellner CP, Song R, Ali M, et al. Time to evacuation and functional outcome after minimally invasive endoscopic intracerebral hemorrhage evacuation. *Stroke.* 2021;52(9):e536-9. doi: 10.1161/STROKEAHA.121.034392.
55. Batista A, Osório R, Varela A, et al. Prediction of short-term prognosis in elderly patients with spontaneous intracerebral hemorrhage. *Eur Geriatr Med.* 2021;12(6):1267-73. doi: 10.1007/s41999-021-00529-w.
56. Batista R, Pereira M, Vaz DC, et al. Prognostic accuracy of common mortality prognostic scales in very old patients with intracerebral haemorrhage. *Ann Neurosci.* 2025;32(3):161-6. doi: 10.1177/09727531231185200.
57. Kapapa T, Jesuthasan S, Schiller F, et al. Outcome after intracerebral haemorrhage and decompressive craniectomy in older adults. *Neurol Int.* 2024;16(3):590-604. doi: 10.3390/neurolint16030044.
58. Tahara S, Hattori Y, Aso S, et al. Endoscopic surgery versus craniotomy for spontaneous intracerebral hemorrhage in the late elderly patients. *J Stroke Cerebrovasc Dis.* 2023;32(11):107327. doi: 10.1016/j.jstrokecerebrovasdis.2023.107327.
59. Warren AD, Li Q, Schwab K, et al. External ventricular drain placement is associated with lower mortality after intracerebral hemorrhage with intraventricular hemorrhage. *Int J Emerg Med.* 2022;15(1):51. doi: 10.1186/s12245-022-00450-4.
60. Komolafe MA, Sunmonu T, Akinyemi J, et al. Clinical and neuroimaging factors associated with 30-day fatality among indigenous West Africans with spontaneous intracerebral hemorrhage. *J Neurol Sci.* 2024;456:122848. doi: 10.1016/j.jns.2023.122848.
61. Senff JR, Singh SD, Pasi M, et al. Long-term outcomes in patients with spontaneous cerebellar hemorrhage: an international cohort study. *Stroke.* 2024;55(5):1210-7. doi: 10.1161/STROKEAHA.123.044622.
62. Noiphithak R, Yindeedee V, Ratanavinitkul W, et al. Treatment outcomes between endoscopic surgery and conventional craniotomy for spontaneous supratentorial intracerebral hemorrhage: a randomized controlled trial. *Neurosurg Rev.* 2023;46(1):136. doi: 10.1007/s10143-023-02035-y.
63. Murthy SB, Zhang C, Garton ALA, et al. Outcomes following minimally invasive surgery for intracerebral hemorrhage in the AHA Get With The Guidelines-Stroke Registry. *Stroke.* 2025;56(6):1441-50. doi: 10.1161/STROKEAHA.124.048650.
64. Li G, Kang K, Ju Y, et al. Surgical treatment could improve long-term functional outcomes in intracerebral hemorrhage patients with NIHSS ≥ 10 . *CNS Neurosci Ther.* 2025;31(6):e70458. doi: 10.1111/cns.70458.
65. Sharma A, Agarwal A, Garg A, et al. Comparison of long-term outcomes in patients with supratentorial spontaneous intracerebral hemorrhage treated with and without surgical intervention. *Ann Indian Acad Neurol.* 2025;28(2):220-6. doi: 10.4103/aian.aian_497_24.
66. Makkiyah FA, Dicha S, Nurizzka RH. A single-center experience of correlation of pulse pressure to mortality of stroke hemorrhage patients in Indonesia. *ScientificWorldJournal.* 2023;2023:1-5. doi: 10.1155/2023/5517493.
67. Yamanie N, Felistia Y, Susanto NH, et al. Prognostic model of in-hospital ischemic stroke mortality based on an electronic health record cohort in Indonesia. *PLoS One.* 2024;19(6):e0305100. doi: 10.1371/journal.pone.0305100.
68. Neves Briard J, Kansara V, Shen Q, et al. Extension of the FUNC score for prediction of 12-month functional independence after primary intracerebral hemorrhage. *Neurocrit Care.* 2026. Epub ahead of print. doi: 10.1007/s12028-026-02661-6.
69. Haferkorn K, Bender M, Uhl E, et al. The ICH score combined with the CRP/albumin ratio as an improved early prognostic tool for in-hospital mortality in spontaneous intracerebral hemorrhage. *Brain Spine.* 2026;6:106084. doi: 10.1016/j.bas.2026.106084.