PEER-REVIEWED · OPEN ACCESS
Sriwijaya Journal of Forensic and Medicolegal
Vol. 3, No. 2 (2025), pp. 95–110 · e-ISSN 2987-1530 · Article ID: 247
journal homepage: phlox.or.id/index.php/sjfm

Original Research Article

Failure to Palpate and Failure to Image: Quantifying Standard-of-Care Departure in Fatal Delayed Intracranial Foreign Body Diagnosis, Indonesia

Aisyah Andina Rasyid[1]*, Priscilla Kapoor[2], Febria Suryani[3]

[1]*Department of Otorhinolaryngology, CMHC Research Center, Palembang, Indonesia*
[2]*Department of Radiology, Mumbai State Hospital, Mumbai, India*
[3]*Department of Health Sciences, Emerald Medical Center, Jakarta, Indonesia*

ARTICLE INFO

Article history
Received: September 30, 2025
Revised: October 25, 2025
Accepted: November 20, 2025
Available online: November 27, 2025
Published: November 30, 2025

Keywords
Clinical forensic medicine
Delayed diagnosis
Intracranial foreign body
Medicolegal
Mild traumatic brain injury

Corresponding author
Aisyah Andina Rasyid
E-mail:
aisyah.andina.rasyid@cattleyacenter.id
ORCID: 0009-0008-9662-7006

DOI
10.59345/sjfm.v3i2.247

ABSTRACT

Background. Retained intracranial foreign bodies in patients alert at a Glasgow Coma Scale (GCS) of 15 are a rare, lethal and medicolegally exposed failure in Indonesian assault casework, yet the share attributable to specific documentable procedural omissions is unquantified.

Objective. To quantify the association of failure to palpate the wound bed and non-adherence to computed tomography (CT) guidelines with delayed diagnosis, and to derive medicolegal attribution metrics.

Methods. This retrospective multi-centre cohort (STROBE) examined 45 confirmed retained intracranial foreign bodies among 1,240 mild traumatic brain injury presentations (prevalence 3.63%, 95% CI 2.72–4.82) at three North Sumatran trauma centres, 2019–2023. Diagnosis beyond 24 hours was the pre-specified outcome. Analysis of aggregate data used exact conditional and Firth-penalised odds ratios, Holm multiplicity control, and partial identification of the multivariable model across 4,000 admissible joint arrangements.

Results. Diagnosis was delayed in 14 of 45 cases (31.1%, 95% CI 19.5–45.7). Failure to palpate the wound bed carried an exact conditional odds ratio of 29.86 (4.79–281.0; $p < 0.001$) and CT non-adherence 20.88 (3.57–174.7; $p < 0.001$). Palpation failure achieved sensitivity 78.6%, specificity 90.3% and area under the curve 0.844; the attributable fraction was 76.3% (50.8–96.1). Adjusted odds ratios were bounded, not estimated, because aggregate data cannot identify a multivariable model.

Conclusion. Two zero-cost omissions dominate a fatal diagnostic failure; recording wound-bed palpation and the imaging decision is clinically protective and, under KUHAP Articles 133 and 179 and Law No. 17 of 2023, is what a visum et repertum requires.

Copyright. © 2025 The Author(s). Published by Phlox Institute: Indonesian Medical Research Organization, on behalf of the Sriwijaya Journal of Forensic and Medicolegal. The authors retain copyright of their work.

Access licence. This is an open access article distributed under the terms of the Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International Licence (CC BY-NC-SA 4.0), which permits non-commercial use, sharing, adaptation and reproduction in any medium, provided the original author and source are credited and any derivative work is distributed under the same licence.

1. Introduction

Forensic medicine is where a clinical encounter becomes evidence. What the examining physician recorded — and what was never recorded — determines what a court can later establish, and the international evidence is that recording is the weak link. In a direct comparison of 132 clinician-performed clinical forensic examinations against specialist practice, a full-body examination was documented in only 37.9% of cases and inspection of concealed body sites in 9.8%, photodocumentation was incomplete in 62.4%, and roughly one description in three was interpretive rather than objective [1]. In Indonesia the obligation is statutory rather than discretionary: Article 133 of the Kitab Undang-Undang Hukum Acara Pidana (KUHAP) obliges a physician to examine a victim or a body on the written request of an investigator and to issue a visum et repertum, and Article 179 obliges the same physician to give expert evidence when summoned. Professional standards are set by the Perhimpunan Dokter Forensik Indonesia (PDFI), while Law No. 29 of 2004 on Medical Practice and Law No. 17 of 2023 on Health define the substantive standard of care whose breach founds liability — a statute whose criminal provisions are, on survey evidence, still poorly understood by Indonesian health workers [2].

Indonesian audits of the visum et repertum locate the deficiency precisely, and always in the same place. Scoring 127 injury reports from a West Javanese hospital with the Herkutanto instrument returned 100% for the opening and closing sections but only 74% for the findings section and 33% for the conclusion, an overall quality of 54.9% [3]. A series of 612 reports from Binjai in North Sumatra scored 100% on the introduction but 75% on the findings, giving 82.35% overall [4]; 92 reports from Palu scored 75.81% on the reporting section and 50.54% on the conclusion [5]. Across 332 reports from Samarinda, blunt trauma accounted for 78% and minor injury for 48.8% [6] and in 46

reports from Yogyakarta 88% involved mechanical violence and 70% were graded as minor injury [7]. The pattern is consistent: the administrative shell of the Indonesian medicolegal report is complete, and the examination findings inside it are not. Knowledge among the general practitioners who write them is uneven — 78.6% scored as good in one North Sumatran district hospital survey, with younger doctors scoring markedly better [8] — and the services themselves are under strain, with only 7 of 22 quality dimensions meeting patient expectations in one Indonesian clinical forensic unit, the largest deficits being examination facilities and staff readiness to perform the examination [9].

Head injury dominates the casework this system handles. In 73 trauma deaths examined at a police hospital forensic unit in Medan, North Sumatra, blunt force caused 63.01% and the head was the most frequently injured region, in about 59% of cases [10]; an Indonesian autopsy series of 68 violent deaths likewise found the head the commonest site, with 81.5% of victims male [11]. On the clinical side the case-mix is overwhelmingly mild: 88% of Indonesian head-injury admissions in one hospital series presented with a Glasgow Coma Scale (GCS) of 13 to 15, and computed tomography showed no lesion in 68% [12]. Internationally the same shift is documented. In the CENTER-TBI registry, 40% (95% CI 39-41) of traumatic brain injuries across 18 countries followed low-energy mechanisms; those patients less often had impaired consciousness than the high-energy group (7.8% versus 10%), yet died in hospital at the same rate (6.3% versus 6.0%) [13]. Contemporary diagnostic criteria have accordingly moved away from resting on the GCS alone [14].

A normal-looking mild presentation is not a benign one, and this is now well quantified. Among 991 patients with a GCS of 15 and a negative head computed tomography scan, only 27% had recovered function at two weeks and 44% at six months, and 88% of those with incomplete recovery had not returned to their pre-injury life [15]. Where the scan is positive, the cost is measurable at one year: the contusion, subarachnoid and subdural lesion cluster predicted incomplete recovery with an odds ratio of 1.80 (95% CI 1.39-2.33) and unfavourable outcome with 3.23 (1.59-6.58) [16] and complicated mild injury produces persisting deficits in processing speed, memory and executive function beyond three months [17]. Post-concussion syndrome follows about 16% of mild injuries and no validated emergency department stratification tool yet exists [18]. Blood biomarkers do not close the gap — day-of-injury GFAP and UCH-L1 discriminated 6-month death well (area under the curve 0.87 and 0.89) but incomplete recovery poorly (0.62 and 0.61), and performed distinctly worse in the GCS 13-15 band than in more severe injury [19]. Nor does an initial normal scan close it: in a national cohort of 626,695 adults discharged after minor head trauma with a negative scan, 0.4% developed delayed intracranial haemorrhage within 14 days, 64.3% of those within three days, 2.7% required neurosurgery, and one-year mortality more than doubled (adjusted hazard ratio 2.15, 95% CI 1.86-2.48) [20].

Within this population sits a specific and lethal failure mode. Small dense projectiles driven through the scalp during a forceful assault can traverse a localised skull micro-fracture and lodge intracranially beneath a laceration of unremarkable appearance. A systematic review of non-missile penetrating brain injury found retained foreign bodies in 37 of 41 patients (90.2%), all requiring operative removal, with incomplete neurological recovery in 47.4% at last follow-up [21]. When the object is missed, the pathology that follows is recognisable at autopsy: in 76 neuropathologically examined medicolegal autopsies with primary head trauma, hypoxic-ischaemic neuronal injury was present in 81.0% of assault cases and brain oedema in 47.6%, against 7.1% of falls [22] and survival of 24 hours or more after injury distinguishes traumatic axonal injury from cases with no axonal injury at all [23]. Autopsy is also what reveals the omission: injuries present at autopsy but absent from the clinical record were found in 39 of 192 trauma decedents (20.3%), significantly associated with a hospital stay under 48 hours and with not undergoing computed tomography [24].

Both safeguards that should interrupt this sequence are cheap and specific. The secondary survey requires digital or instrumental palpation of every scalp wound to the periosteal level, seeking crepitus, step-off deformity and the tactile signature of embedded material. The Canadian CT Head Rule mandates non-contrast computed tomography for any dangerous mechanism irrespective of the GCS, which makes assault an imaging indication in its own right [25]. Adherence to both falls short in practice, and — importantly for the present question — it fails in both directions. A Canadian trauma-centre audit found the rule followed in 187 of 213 applicable cases (87.8%), but with computed tomography omitted in 13.5% of rule-indicated cases and ordered against the rule in 11.9% of non-indicated ones [26]. In a rural emergency department without on-site scanning, 62.1% of mild head injuries met rule criteria yet only 35.1% of those were transferred for imaging, including just 16.7% of patients meeting a medium-risk criterion [27]. Conversely, of 1,000 scanned head-trauma patients, 80.2% of scans in the mild-trauma group lacked any guideline indication [28]. Rule performance itself is imperfect: across five decision rules applied prospectively to 434 patients, sensitivity for any traumatic lesion ranged from 79.75% to 91.14% [29] and in an Ethiopian validation both the Canadian rule and the New Orleans Criteria reached 100% sensitivity at specificities of only 41.5% and 26.5% [30].

That imaging and examination omissions matter medicolegally is not in doubt. Diagnostic error is estimated to leave 795,000 Americans permanently disabled or dead each year, with a weighted mean error rate of 11.1% across dangerous conditions [31]; missed diagnosis is consistently more likely in vulnerable emergency department populations, with odds ratios between 1.18 and 4.5 [32]. In 5,854 emergency department malpractice cases totalling $1.01 billion in indemnity, errors of clinical judgment were the dominant contributing factor, cited in 76.8% to 89.1% of cases [33] and among 5,367 closed diagnostic-error claims, payment was driven by injury severity and by contributing factors falling outside the standard diagnostic pathway [34]. In trauma specifically, 9.2% of 99,754 multiply injured patients had a delayed diagnosed injury, the head was the single most frequent site at 35.8%, and mortality was higher in the delayed group (10.8% versus 6.5%) [35]. Whether the safeguard can be improved is answered encouragingly: an education-based initiative raised documented tertiary survey performance from 30.1% to 85.1% and more than tripled the missed injuries detected, from 1.7% to 5.7% [36]. Compliance alone is not sufficient, however: a mandated 24-hour tertiary survey raised completion from 80.2% to 86.8% yet left missed injury rates unchanged at 1.6% versus 1.5% [37].

What is missing is quantification in an Indonesian forensic setting. No study has isolated the retained intracranial foreign body subpopulation, and none reports the quantities a court or the Majelis Kehormatan Disiplin

Kedokteran Indonesia (MKDKI) now asks for: how much of an observed diagnostic failure is attributable to a given omission, how probable it is that the omission caused the failure, and how strong an unmeasured confounder would have to be to explain the association away. The present study therefore aimed, first, to quantify the association between specific procedural omissions and delayed diagnosis of retained intracranial foreign bodies using inference valid in small, sparse samples; second, to characterise those omissions as bedside diagnostic markers and to derive their attributable fraction, probability of causation and sensitivity to unmeasured confounding; and third, to state explicitly which quantities the available data can and cannot identify. The three participating centres serve as principal forensic and medicolegal facilities for their catchment in North Sumatra, a province where a recent Hospital Safety Index assessment rated three of five surveyed hospitals at level B, meaning their capacity to function during an emergency is at risk [38].

2. Methods

2.1 Study design

This was a retrospective, multi-centre, observational cohort study in clinical forensic medicine and medicolegal traumatology, reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [39]; a completed checklist is supplied, a declaration made in only 28.6% of original articles for which a guideline is methodologically advisable [40]. The unit of analysis was the individual patient episode. The study addressed a prognostic-factor question — which documented procedural omissions predict delayed diagnosis — and a diagnostic-accuracy question — how those omissions perform as bedside markers.

2.2 Ethical approval

This retrospective study used anonymised clinical, forensic and medicolegal records, including post-mortem examination records, from three tertiary trauma centres in North Sumatra, Indonesia. Ethical approval was obtained from the CMHC Ethics Committee, Indonesia (Approval No. 2024/0193). Individual consent was waived per institutional protocol for anonymised retrospective data analysis. The study was conducted in accordance with the Declaration of Helsinki. No identifying information appears in this report or in any figure. Autopsy material was handled under the same anonymisation protocol as the clinical records.

2.3 Setting and participants

Data were drawn from the trauma registries, emergency department records and medicolegal forensic autopsy records of three tertiary-care and regional general hospitals in North Sumatra, Indonesia, over a five-year period from 1 January 2019 to 31 December 2023. The window was chosen to accumulate a sufficient number of a rare event while remaining inside the coverage of the standardised electronic medical record systems introduced at all three centres from 2018. No change to computed tomography scanner availability or to institutional head-trauma protocols was documented at any centre during the period.

Cases were ascertained from five independent sources in order to minimise both over- and under-identification: ICD-10 discharge diagnostic codes (S09.8 and S06.0–S06.9); surgical operation registers for craniectomy or wound exploration; radiological reports documenting a foreign body on computed tomography or skull radiograph; operative reports documenting foreign body extraction; and forensic autopsy reports recording an intracranial foreign body as the cause of, or a contributor to, death. The last source is essential: without it, cases that were never diagnosed in life would be systematically excluded, and the delayed-diagnosis rate would be biased downwards.

Inclusion criteria were adult patients aged 18 years or older; a history of acute head trauma; an initial emergency department GCS of 13 to 15; and a retained intracranial or deeply embedded cranial foreign body confirmed by any of the ascertainment routes above. Patients with severe traumatic brain injury (GCS ≤ 8 on arrival), those requiring immediate tracheal intubation, and those with massive open cranial vault destruction in which foreign material was visible without deep wound exploration were excluded, so that the cohort was restricted to presentations in which the diagnosis genuinely depended on clinical judgement. The selection cascade is set out in Figure 1.

2.4 Variables and outcome definition

Data were extracted at every site using a single pre-piloted structured instrument covering four domains. Demographics comprised age in years and sex. Injury characteristics comprised mechanism (assault versus road traffic accident), wound location, wound length and depth in centimetres, and surrounding soft-tissue oedema. Clinical presentation comprised the initial emergency department GCS, focal neurological deficit, emesis, severe headache and intoxication status on arrival. Procedural and management variables comprised documentation of digital or instrumental wound-bed palpation to the periosteal level (yes/no); adherence to the Canadian CT Head Rule, defined as ordering a non-contrast computed tomography scan whenever any rule criterion was met (adherent/non-adherent); and the interval from emergency department triage to definitive identification of the foreign body.

The pre-specified primary outcome, delayed diagnosis, was failure to identify the intracranial foreign body within 24 hours of the index emergency department encounter. The definition captures both the patient who was discharged and later re-presented and the patient in whom the diagnosis was first established at forensic autopsy. Palpation was recorded as not performed where the record contained no positive documentation of it. This is a coding convention rather than complete ascertainment, and it has two consequences stated here rather than left to be inferred. First, the variable has no missing values by construction, which must not be read as evidence of complete data. Second, the aggregate records available for this analysis do not permit exposures positively documented as not performed to be separated from those inferred from silence, so the misclassification rate of the instrument is unknown; a quantitative bias analysis addressing this is specified in section 2.7 and reported in section 3.7. The same non-recoverability applies to Canadian CT Head Rule sub-criteria in the non-adherent cases, to whether a visum et repertum was issued and what it recorded, to per-variable missingness, to case counts by calendar year, and to exposure prevalence and outcome rate by centre. Each is listed as a limitation rather than silently omitted.

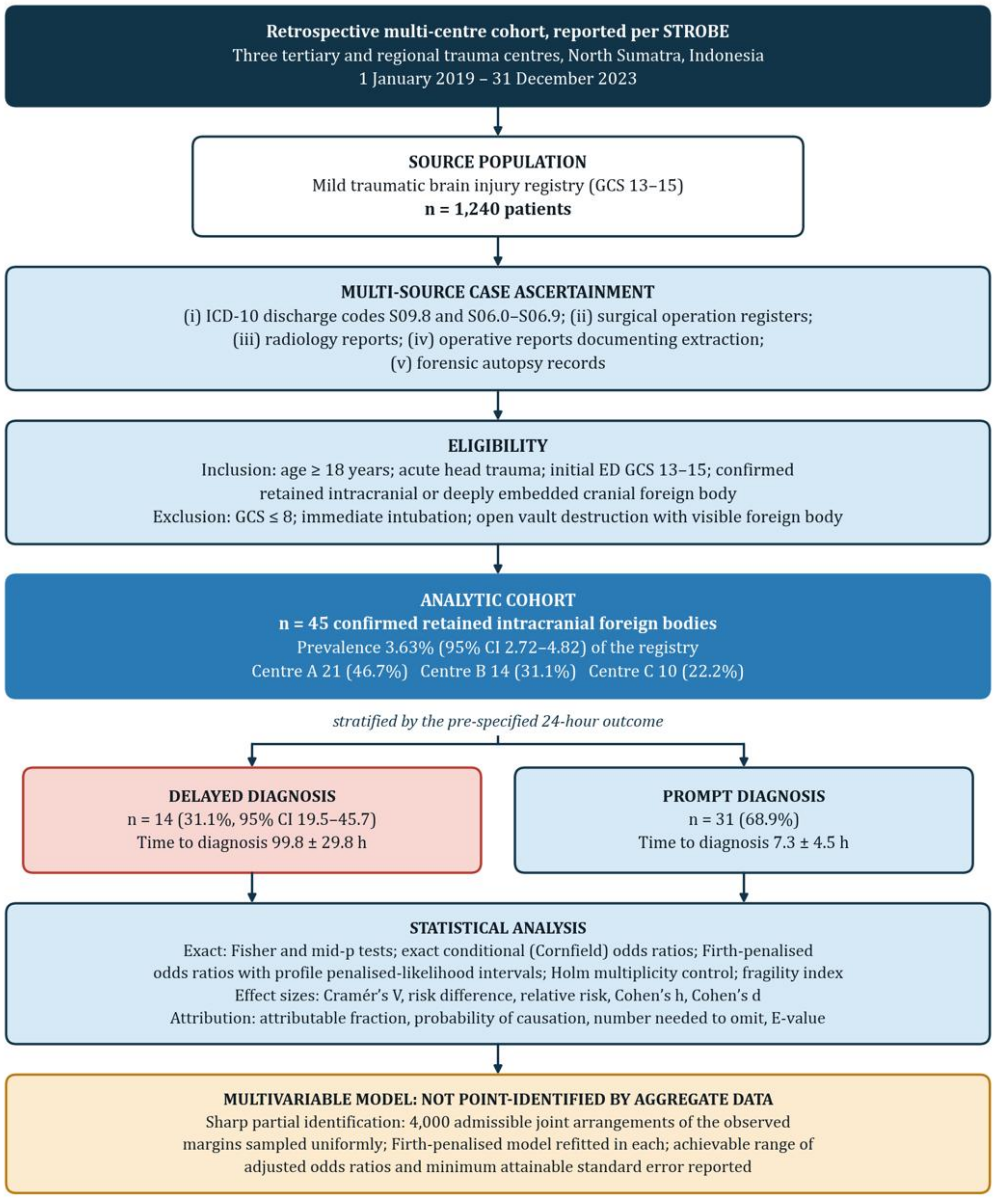


Figure 1. Study flow and analytic pathway. Multi-source ascertainment from a 1,240-patient mild traumatic brain injury registry yielded 45 confirmed retained intracranial foreign bodies, of which 14 were diagnosed beyond the pre-specified 24-hour threshold. The final panel records that the multivariable model is not point-identified by aggregate data and was therefore bounded rather than estimated.

2.5 Sample size

The cohort was a complete five-year census of eligible cases rather than a sample drawn to a target size, so no recruitment-stage calculation was possible. Adequacy was therefore assessed after the fact in two directions. First, the power of Fisher's exact test was approximated by Monte Carlo simulation (6,000 replicates; Monte Carlo standard error ≤ 0.007) at the observed marginal prevalences with $\alpha = 0.05$: with 14 events and 31 non-events the design had 99.8% power against the observed palpation-failure effect, 98.7% against the observed computed tomography effect, and 69.2% against the intoxication effect. Second, the minimum detectable odds ratio at 80% power was derived by the same simulation and lay between 9.3 and 11.6 depending on the exposure prevalence; that variation is driven by baseline exposure prevalence and carries no substantive meaning. The design is therefore well powered for large procedural effects and unable to detect moderate ones, which is stated here rather than left to be inferred: the null results for wound depth and male sex are uninformative, not negative, and no post-hoc power figure should be read as validating any analysis. For the multivariable model the relevant criterion is not events-per-variable but modern shrinkage-based sample-size requirements [41,42] which 14 events cannot satisfy for a model of this complexity; this is the formal reason for the partial-identification approach described in section 2.7.

2.6 Forensic and clinical assessment protocol

Clinical forensic assessment followed the standard Indonesian medicolegal sequence. Living patients were assessed by the attending emergency physician, with wound characteristics, exploration findings and imaging decisions transcribed into the medical record and, where an investigator's written request under KUHAP Article 133 had been received, into a visum et repertum. Post-mortem examinations were performed by forensic pathologists at the participating installations following national autopsy protocol, with external examination, full internal examination, documentation of the entry wound and the intracranial trajectory, retrieval and description of the

foreign body, and cause and manner of death recorded to ICD-10. Adherence to the Canadian CT Head Rule was adjudicated against the published rule criteria [25] by review of the recorded indication, not by the imaging outcome, so that a scan omitted for a patient who turned out to have no lesion still counted as non-adherence if a rule criterion had been met.

2.7 Statistical analysis

All analyses were performed in Python 3.11 with NumPy 2.4 and SciPy 1.17; every stochastic procedure used the fixed seed 20260812. Because the analysis was conducted on aggregate contingency data rather than a case-level file, every method was chosen to be exact or penalised rather than to rely on large-sample approximation.

Proportions are reported with Wilson score intervals, which are preferred to Wald intervals because they retain nominal coverage at the small denominators and extreme proportions that occur throughout this cohort. Continuous comparisons used Welch's t test with Hedges-corrected Cohen's d and its 95% confidence interval; Welch rather than Student was chosen because the variance ratio for time to diagnosis is extreme. Categorical comparisons used Fisher's exact test, pre-specified rather than selected on the basis of observed cell counts [43] with the mid-p variant reported alongside, together with exact conditional (Cornfield) odds ratios and exact limits derived from the non-central hypergeometric distribution, which remain valid where asymptotic intervals fail [44]. Because several strata contain cells of one to three observations, Firth-penalised logistic regression with profile penalised-likelihood intervals was fitted for every predictor as the separation-robust primary estimate; Wald intervals on a penalised estimate are known to behave poorly and were not used [45]; penalised coefficients are not read as calibrated absolute risks [46]. Effect sizes accompany every comparison — Cramér's V, risk difference, relative risk, Cohen's h and Cohen's d — and each is interpreted against its conventional benchmark at first use. Multiplicity across the seven bivariate comparisons was controlled by the Holm procedure, which is nominated as primary; Benjamini–Hochberg values are reported alongside for completeness and are not used for inference. Fragility indices were computed as the number of outcome reassignments in the exposed group required to reverse the significance verdict. Exact p values are reported to three decimal places; α was 0.05, two-tailed. Interval estimates are interpreted throughout as compatibility intervals rather than as statements of significance [47] — a distinction misread in the majority of published non-significant results [48]. Fragility indices are interpreted as a sensitivity analysis for outcome misclassification rather than as a significance metric [49,50].

Diagnostic performance of each procedural marker was characterised by sensitivity, specificity, predictive values with Wilson intervals, positive and negative likelihood ratios, Youden's index, the Matthews correlation coefficient, and the area under the curve. Because each marker is binary, the area under the curve equals (sensitivity + specificity)/2 and is a summary of a single operating point rather than a discrimination measure over thresholds; its variance was therefore computed exactly for a Bernoulli marker rather than by the Hanley–McNeil approximation, which was derived for continuous markers. Likelihood ratios are given prominence because they are prevalence-independent, whereas predictive values are conditional on this cohort's outcome prevalence of 31.1% and will differ at any other. For two-item composite rules the joint distribution of the two markers is not recoverable from marginal tables, so sharp Fréchet–Hoeffding bounds were computed for the 'any failure' and 'both failures' rules and are reported as intervals rather than as point estimates.

Forensic attribution used the case-based population attributable fraction with percentile bootstrap intervals over 20,000 resamples, the probability of causation defined as (RR − 1)/RR, the number-needed-to-omit as the reciprocal of the risk difference, and the E-value for both the point estimate and the confidence limit nearest the null [51] noting that the E-value is not invariant to the adjustment strategy [52]. Attributable fractions computed from aggregate exposure summaries require nonparametric treatment [53] an adjusted risk ratio must not be substituted into Levin's formula [54] and the estimand is sample-specific rather than automatically transportable [55]. These are causal quantities and their interpretation requires exchangeability, positivity and consistency. Positivity holds by construction. Consistency is plausible, since the exposures are discrete acts. Exchangeability is doubtful and is not assumed: a clinician who omits palpation may differ systematically from one who does not in ways correlated with the rest of the encounter, and the attribution quantities are therefore computed from crude rather than adjusted associations and are reported as group-level descriptive quantities. They do not establish causation in any individual case, and the E-value is reported precisely to quantify how strong an unmeasured alternative explanation would have to be.

Because the principal exposure is defined by absent documentation, a quantitative bias analysis was pre-specified. Letting φ denote the fraction of recorded exposures that are false positives — palpation performed but not recorded — the odds ratio and its lower confidence limit were recomputed across φ from 0 to 1 under two scenarios: non-differential misclassification affecting both outcome groups equally, and the adversarial case of differential misclassification affecting only the delayed group. The tipping point is reported as the value of φ at which the lower confidence limit reaches the null. Probabilistic bias analysis of this form is feasible from summary-level data [56] and prior-based parameterisation of the bias parameters outperforms naive correction when validation data are sparse [57].

The multivariable model requires explicit comment. The source data identify only the one-way margin of each predictor within each outcome stratum; they do not identify the joint distribution of predictors, and therefore do not identify any multivariable coefficient. Rather than impute a joint distribution and present the result as an estimate, the admissible set was characterised by sampling from it. Within each outcome stratum the columns were permuted independently; because only column margins are fixed and no row constraint applies, the constraint set factorises across columns and independent permutation is exactly uniform on the admissible set. That set is astronomically large — its cardinality is the product of fourteen binomial coefficients and exceeds ten raised to the fifty-ninth power — so the 4,000 draws sample it rather than enumerate it; aggregated data cannot in principle recover individual-level effects [58] and assumption-free bounds of this kind are characteristically wide [59]. A Firth-penalised model including age as a moment-matched covariate was refitted in each draw. Two consequences follow and are stated explicitly. First, the reported range of adjusted odds ratios is a Monte Carlo approximation to the sharp bounds, not the sharp bounds

themselves. Second, the reported minimum attainable standard error is an upper bound on the true minimum over the admissible set — the true minimum can only be smaller — which is conservative in the direction of the argument in section 3.6, which requires the attainable minimum to be large. To confirm that this conclusion does not depend on the model specification, the sampling was repeated under three specifications: the two primary procedural predictors alone, the five clinical and procedural predictors without age, and the full eight-covariate specification. No point estimate of any adjusted odds ratio is claimed.

3. Results

3.1 Cohort and case ascertainment

Multi-source ascertainment across the five-year period identified 45 patients satisfying the inclusion criteria from a registry of 1,240 mild traumatic brain injury presentations, a prevalence of 3.63% (95% CI 2.72–4.82). The selection cascade and the analytic pathway are shown in Figure 1. Centre A contributed 21 cases (46.7%, 95% CI 32.9–60.9), Centre B 14 (31.1%, 19.5–45.7) and Centre C 10 (22.2%, 12.5–36.3). Exposure prevalence and outcome rate by centre, and case counts by calendar year, are not recoverable from the aggregate records available for this analysis and are declared as limitations.

The cohort was predominantly male (84.4%, 95% CI 71.2–92.3) and assault was the commonest mechanism (62.2%, 95% CI 47.6–74.9). An initial GCS of exactly 15 was recorded in 60.0% of cases. Diagnosis was delayed beyond 24 hours in 14 of 45 patients (31.1%, 95% CI 19.5–45.7). Baseline characteristics stratified by diagnostic timing are presented in Table 1.

Table 1. Baseline demographic, clinical and procedural characteristics of the cohort, stratified by diagnostic timing (N = 45).

Characteristic	Total (N = 45)	Delayed (n = 14)	Prompt (n = 31)	Effect size (95% CI)	p†
Age, mean ± SD, years	39.98 ± 11.51	47.7 ± 9.8	36.5 ± 10.6	d = 1.08 (0.41–1.75)	0.002
Male sex, n (%)	38 (84.4)	13 (92.9)	25 (80.6)	V = 0.156	0.407
Mechanism: assault, n (%)	28 (62.2)	7 (50.0)	21 (67.7)	V = 0.169	0.326
Initial GCS 15, n (%)	27 (60.0)	13 (92.9)	14 (45.2)	V = 0.451	0.003
Initial GCS 13–14, n (%)	18 (40.0)	1 (7.1)	17 (54.8)	—	—
Wound depth > 0.5 cm, n (%)	27 (60.0)‡	11 (78.6)	16 (51.6)	V = 0.255	0.111
Intoxication on arrival, n (%)	11 (24.4)‡	7 (50.0)	4 (12.9)	V = 0.400	0.020
Failure to palpate wound bed, n (%)	14 (31.1)‡	11 (78.6)	3 (9.7)	V = 0.689	< 0.001
Non-adherence to CT guidelines, n (%)	13 (28.9)	10 (71.4)	3 (9.7)	V = 0.631	< 0.001
Time to diagnosis, mean ± SD, hours	36.08 ± 46.39	99.8 ± 29.8	7.3 ± 4.5	d = 5.50 (4.18–6.83)	< 0.001
Centre A / B / C, n	21 / 14 / 10	—	—	—	—

Notes: † Fisher's exact test for categorical variables; Welch's t test for continuous variables. Exact p values to three decimal places. ‡ Total-column counts and pooled means are computed directly from the group-level counts and the weighted-mean identity. CI = confidence interval; CT = computed tomography; d = Hedges-corrected Cohen's d; GCS = Glasgow Coma Scale; SD = standard deviation; V = Cramér's V. Centre contributions were 46.7%, 31.1% and 22.2% respectively.

Patients whose diagnosis was delayed were older than those diagnosed promptly (47.7 ± 9.8 versus 36.5 ± 10.6 years; mean difference 11.2 years, 95% CI 4.6–17.8; Welch t = 3.46, p = 0.002; Cohen's d = 1.08, 95% CI 0.41–1.75). The interval from triage to definitive identification differed by an order of magnitude (99.8 ± 29.8 versus 7.3 ± 4.5 hours; mean difference 92.5 hours, 95% CI 75.2–109.8), a 13.7-fold ratio. No inferential test is reported for this comparison: the outcome is defined by a threshold on the same variable, so any test of it is circular and the difference is presented descriptively.

3.2 Bivariate association with delayed diagnosis

Exact, penalised and large-sample estimates for all seven candidate predictors are set out in Table 2 and displayed in Figure 2. Failure to document wound-bed palpation was present in 11 of 14 delayed cases (78.6%) against 3 of 31 prompt cases (9.7%), giving a crude odds ratio of 34.22 (95% CI 5.97–196.1), an exact conditional odds ratio of 29.86 (4.79–281.0) and a Firth-penalised odds ratio of 26.76 (5.89–162.4); Fisher exact p < 0.001, mid-p < 0.001, Holm-adjusted p < 0.001. The associated Cramér's V was 0.689 — a large association by the conventional benchmark of 0.10, 0.30 and 0.50 for small, medium and large in a 2 × 2 table — with a risk difference of 68.9 percentage points and Cohen's h of 1.55, also large against the 0.20 / 0.50 / 0.80 benchmark.

Table 2. Bivariate association with delayed diagnosis: large-sample, exact conditional and penalised odds ratios, with multiplicity control and effect sizes.

Risk factor	Delayed n/N (%)	Prompt n/N (%)	Crude OR (95% CI)a	Exact conditional OR (95% CI)b	Firth OR (95% CI)c	p (Fisher)	p (Holm)	p (BH)	V / hd
Failure to palpate wound bed	11/14 (78.6)	3/31 (9.7)	34.22 (5.97–196.1)	29.86 (4.79–281.0)	26.76 (5.89–162.4)	< 0.001	< 0.001	< 0.001	0.689 / 1.55
Non-adherence to CT guidelines	10/14 (71.4)	3/31 (9.7)	23.33 (4.43–123.0)	20.88 (3.57–174.7)	19.00 (4.38–105.5)	< 0.001	< 0.001	< 0.001	0.631 / 1.42
Initial GCS 15 (vs 13-14)	13/14 (92.9)	14/31 (45.2)	15.79 (1.83–136.0)	14.95 (1.83–709.3)	10.86 (2.22–108.0)	0.003	0.014	0.007	0.451 / 1.06
Intoxication on arrival	7/14 (50.0)	4/31 (12.9)	6.75 (1.53–29.8)	6.40 (1.23–39.4)	6.11 (1.53–27.4)	0.020	0.082	0.036	0.400 / 0.91
Wound depth >0.5 cm	11/14 (78.6)	16/31 (51.6)	3.44 (0.80–14.8)	3.35 (0.69–22.3)	3.09 (0.83–14.0)	0.111	0.332	0.155	0.255 / 0.54
Male sex	13/14 (92.9)	25/31 (80.6)	3.12 (0.34–28.7)	3.06 (0.32–154.4)	2.29 (0.42–23.7)	0.407	0.652	0.407	0.156 / 0.47
Mechanism: assault	7/14 (50.0)	21/31 (67.7)	0.48 (0.13–1.7)	0.48 (0.11–2.1)	0.49 (0.14–1.7)	0.326	0.652	0.381	0.169 / -0.35

Notes: a Woolf large-sample interval. b Conditional maximum-likelihood estimate with exact limits from the non-central hypergeometric distribution; preferred where any cell is small. c Firth-penalised estimate with profile penalised-likelihood interval; the separation-robust primary estimate. d Cramér's V and Cohen's h. BH = Benjamini–Hochberg; CI = confidence interval; OR = odds ratio.

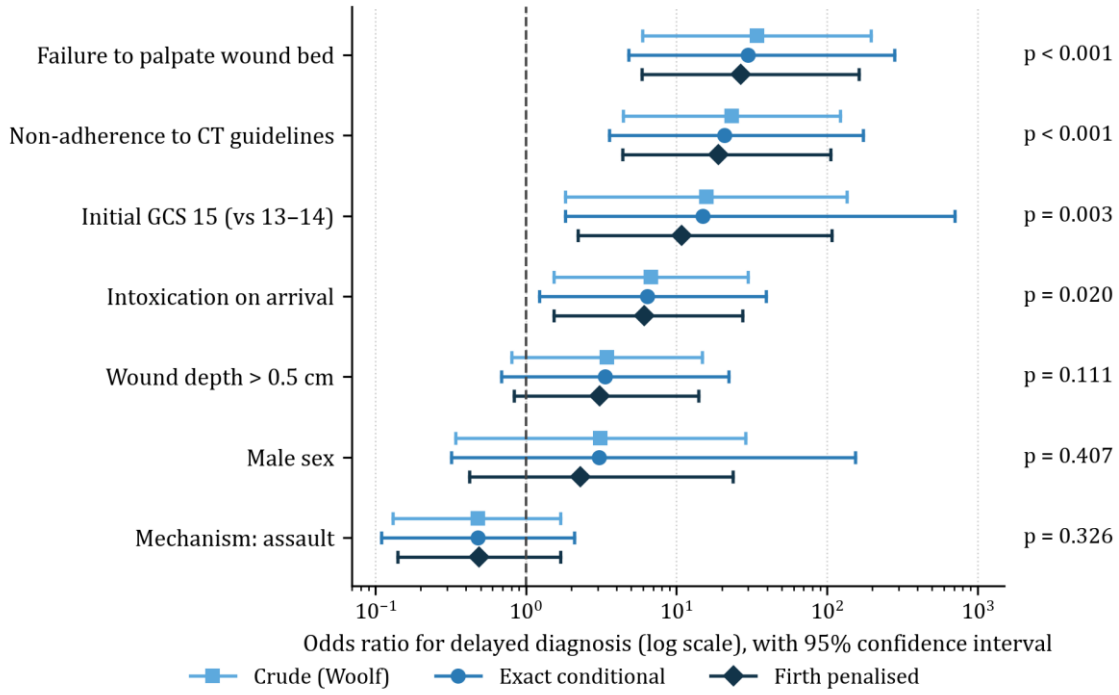


Figure 2. Bivariate odds ratios for delayed diagnosis on a logarithmic scale, showing the large-sample (Woolf), exact conditional and Firth-penalised estimates side by side for each candidate predictor, with 95% confidence intervals shown in full. p values are two-sided Fisher exact tests.

Non-adherence to computed tomography guidelines was documented in 10 of 14 delayed cases (71.4%) versus 3 of 31 prompt cases (9.7%): crude odds ratio 23.33 (4.43–123.0), exact conditional 20.88 (3.57–174.7), Firth 19.00 (4.38–105.5); p < 0.001, Holm-adjusted p < 0.001, Cramér's V 0.631.

An initial GCS of 15 was present in 13 of 14 delayed cases (92.9%) versus 14 of 31 prompt cases (45.2%): crude odds ratio 15.79 (1.83–136.0), Firth 10.86 (2.22–108.0); p = 0.003, Holm-adjusted p = 0.014. Intoxication on arrival reached nominal significance (odds ratio 6.75, p = 0.020) but did not survive Holm adjustment (p = 0.082). Wound depth greater than 0.5 cm (p = 0.111) and male sex (p = 0.407) were not associated with delay.

Assault as the mechanism of injury was, contrary to expectation, less frequent among delayed cases (50.0%) than among prompt cases (67.7%), crude odds ratio 0.48 (0.13–1.73), p = 0.326. Fragility indices for the three predictors surviving multiplicity adjustment were 6 for palpation failure, 5 for computed tomography non-adherence and 2 for GCS 15.

3.3 Multivariable model: what the data can and cannot identify

The aggregate source data identify the marginal distribution of each predictor within each outcome stratum but not their joint distribution, and consequently do not identify any multivariable coefficient. The achievable set was therefore enumerated over 4,000 admissible joint arrangements; results are presented in Table 3 and Figure 3.

Table 3. Sharp partial identification of the multivariable model. Aggregate data identify each predictor's margin within each outcome stratum but not their joint distribution; the Firth-penalised model was therefore refitted across 4,000 admissible joint arrangements rather than estimated once (N = 45, 14 events).

Predictor	Median aOR	Central 95% of admissible configurations	Full admissible range	Minimum attainable SE	Median SE	Configurations with p < 0.05
Non-adherence to CT guidelines	10.32	1.66 – 68.5	0.31 – 67279	0.965	1.388	28.2%
Failure to palpate wound bed	13.80	2.74 – 90.2	0.15 – 103809	0.960	1.368	48.9%
Initial GCS 15 (vs 13-14)	5.71	0.76 – 60.5	0.00 – 4229327	1.016	1.447	3.5%
Intoxication on arrival	3.95	0.47 – 37.1	0.02 – 8433	0.990	1.428	2.9%
Wound depth >0.5 cm	2.37	0.29 – 18.9	0.01 – 2068	0.941	1.345	0.2%

Notes: aOR = adjusted odds ratio; SE = standard error of the log odds ratio. Each arrangement adjusts for all five listed predictors plus male sex, assault mechanism and a moment-matched age covariate. No point estimate of any adjusted odds ratio is claimed, because none is identified by the available data: the median and the central range are summaries of a set of point estimates obtained under different admissible data configurations, not an estimate and an interval. The minimum attainable standard error is an upper bound on the true minimum over the admissible set.

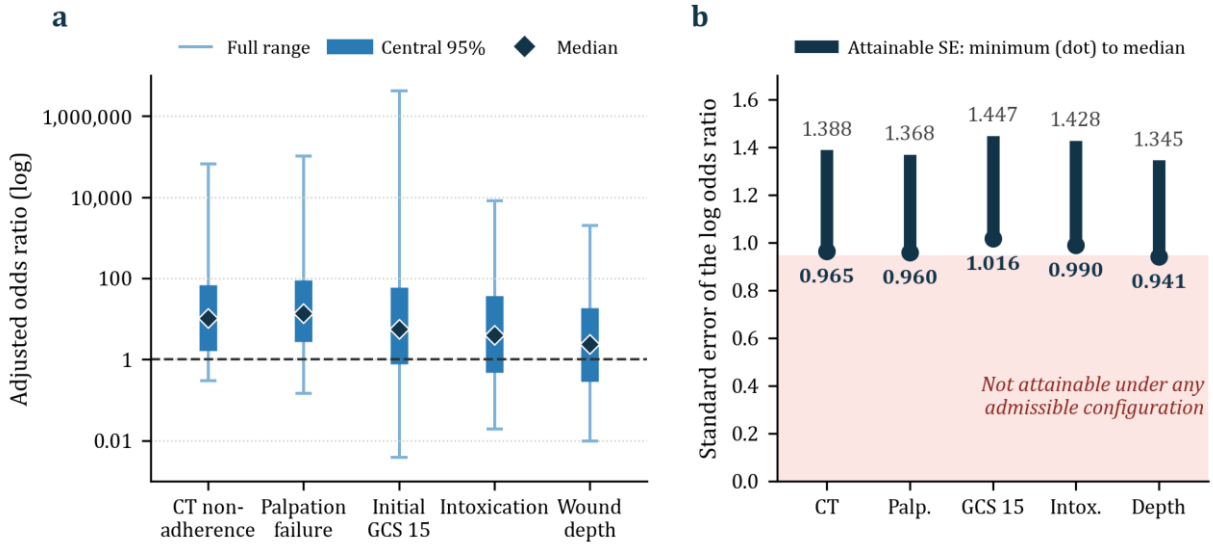


Figure 3. Partial identification of the multivariable model. Panel a: the distribution of adjusted odds ratios achievable across 4,000 joint arrangements consistent with the observed margins, with the median, the central 95% of configurations and the full admissible range. Panel b: the minimum and median standard error attainable anywhere in the admissible set; the shaded band marks precision that no admissible configuration of the data can reach.

The adjusted association reached conventional significance in 48.9% of admissible arrangements for palpation failure, 28.2% for computed tomography non-adherence and 3.5% for an initial GCS of 15 — that is, the adjusted verdict itself depends on which admissible arrangement is assumed, which is the clearest possible statement that the parameter is not identified. Across the admissible set the median adjusted odds ratio for computed tomography non-adherence was 10.32 with a central 95% range of 1.66 to 68.5; for palpation failure 13.80 (2.74 to 90.2); and for an initial GCS of 15 5.71 (0.76 to 60.5). These are the median and the central 95% of admissible configurations, not an estimate and an interval: the median over a set of point estimates obtained under different assumed data configurations has no inferential interpretation.

The minimum standard error attainable anywhere in the admissible set was 0.965 for the computed tomography coefficient, 0.960 for palpation failure and 1.016 for GCS 15, against median values of 1.388, 1.368 and 1.447 respectively. No point estimate of an adjusted odds ratio is reported, because none is identified.

3.4 Diagnostic performance of the procedural markers

Treated as bedside markers of subsequent diagnostic delay, the two procedural omissions performed as set out in Table 4 and Figure 4. Failure to palpate the wound bed achieved a sensitivity of 78.6% (95% CI 52.4–92.4), a specificity of 90.3% (75.1–96.7), a positive predictive value of 78.6% and a negative predictive value of 90.3%, with a positive likelihood ratio of 8.12, a negative likelihood ratio of 0.24, Youden index 0.69 and area under the curve 0.844 (0.725–0.964, exact Bernoulli variance). Because the marker is binary this last quantity summarises a single operating point rather than discrimination across thresholds; the Youden index and the likelihood ratios are the more defensible summaries, and the predictive values are conditional on this cohort's 31.1% outcome prevalence and will differ at any other.

Table 4. Diagnostic performance of the procedural markers for subsequent delayed diagnosis, with sharp bounds for the two-item composite rules.

Marker	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	LR+	LR−	Youden J	MCC	AUC (95% CI)
Failure to palpate wound bed	78.6 (52.4–92.4)	90.3 (75.1–96.7)	78.6 (52.4–92.4)	90.3 (75.1–96.7)	8.12	0.24	0.689	0.689	0.844 (0.706–0.983)
Non-adherence to CT guidelines	71.4 (45.4–88.3)	90.3 (75.1–96.7)	76.9 (49.7–91.8)	87.5 (71.9–95.0)	7.38	0.32	0.618	0.631	0.809 (0.658–0.959)
Initial GCS 15	92.9 (68.5–98.7)	54.8 (37.8–70.8)	48.1 (30.7–66.0)	94.4 (74.2–99.0)	2.06	0.13	0.477	0.451	0.738 (0.571–0.906)
ANY procedural failure (Fréchet bounds)e	78.6 – 100.0	80.6 – 90.3	not identified	not identified	—	—	—	—	OR 15.3 – 121.3
BOTH procedural failures (Fréchet bounds)e	50.0 – 71.4	90.3 – 100.0	not identified	not identified	—	—	—	—	OR 9.3 – 75.0

Notes: Wilson score intervals for proportions; analytic standard error for the area under the curve of a binary marker. e The joint distribution of the two procedural markers is not identified by marginal tables, so composite rules are reported as sharp Fréchet–Hoeffding bounds rather than point estimates. AUC = area under the receiver operating characteristic curve; LR = likelihood ratio; MCC = Matthews correlation coefficient; NPV = negative predictive value; PPV = positive predictive value.

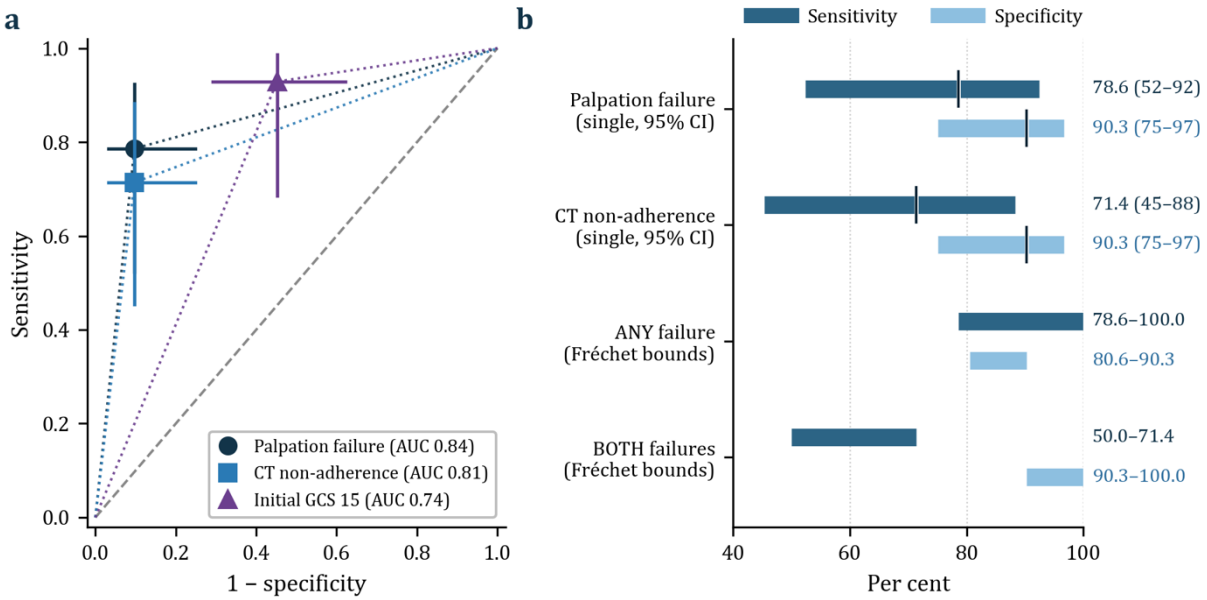


Figure 4. Diagnostic performance of the procedural markers. Left panel: receiver operating characteristic geometry for each single binary marker with its area under the curve. Right panel: sensitivity and specificity for the single markers (point estimate with 95% Wilson interval) and for the two-item composite rules, where the joint distribution is not identified and sharp Fréchet–Hoeffding bounds are shown instead.

Non-adherence to computed tomography guidelines achieved sensitivity 71.4% (45.4–88.3), specificity 90.3% (75.1–96.7), positive likelihood ratio 7.38 and area under the curve 0.809 (0.679–0.938). An initial GCS of 15 was highly sensitive (92.9%) but poorly specific (54.8%), with a negative likelihood ratio of 0.13 and an area under the curve of 0.738; its negative predictive value of 94.4% is computed at this cohort's outcome prevalence and is not transportable.

Because the joint distribution of the two procedural markers is not identified, composite rules are reported as sharp Fréchet–Hoeffding bounds. A rule triggered by either omission would have sensitivity between 78.6% and 100.0% at specificity between 80.6% and 90.3%, with an odds ratio between 15.3 and 121.3. A rule requiring both omissions would have sensitivity between 50.0% and 71.4% at specificity between 90.3% and 100.0%. Over the same admissible set, the two-item ordinal score achieved a median area under the curve of 0.923 (range 0.849–0.952), rising to 0.949 when GCS 15 was added as a third item; both are ranges over admissible configurations rather than estimates with intervals. The composite bounds are too wide to define a screening rule — a sensitivity anywhere between 78.6% and 100.0% is not actionable — and the composite analysis is therefore inconclusive. Narrowing it would require the case-level cross-tabulation of the two markers, which is a single 2 × 2 table that future studies should report.

3.5 Forensic attribution metrics

Attribution quantities are presented in Table 5. The case-based population attributable fraction for failure to palpate the wound bed was 76.3% (bootstrap 95% CI 50.8–96.1), meaning that on these data most delayed diagnoses in the cohort are statistically attributable to that single omission. For computed tomography non-adherence the fraction was 68.4% (40.4–92.4) and for an initial GCS of 15 87.0% (55.7–95.0).

Table 5. Forensic attribution metrics for the procedural and clinical predictors of delayed diagnosis.

Factor	Risk if exposed vs unexposed (%)	Risk difference (pp)	Relative risk	Attributable fraction % (95% CI)f	Probability of causation (%)g	Number needed to omith	E-value: point / limiti	Fragility indexj	Exact power (%)k
Failure to palpate wound bed	78.6 vs 9.7	68.9	8.12	76.3 (50.8–96.1)	87.7	1.45	11.18 / 3.80	6	99.8
Non-adherence to CT guidelines	76.9 vs 12.5	64.4	6.15	68.4 (40.4–92.4)	83.8	1.55	9.13 / 3.19	5	98.7
Initial GCS 15 (vs 13-14)	48.1 vs 5.6	42.6	8.67	87.0 (55.7–95.0)	88.5	2.35	7.41 / 2.04	2	90.6
Intoxication on arrival	63.6 vs 20.6	43.0	3.09	42.6 (9.8–74.5)	67.6	2.32	4.64 / 1.46	2	69.2

Notes: f Case-based population attributable fraction with percentile bootstrap interval over 20,000 resamples. g (RR − 1)/RR; values above 50% meet the balance-of-probabilities threshold. h Reciprocal of the risk difference: the number of patients in whom the step must be omitted to generate one additional delayed diagnosis. i Minimum strength of association, on the risk-ratio scale, that an unmeasured confounder would need with both exposure and outcome to explain away the point estimate and the confidence limit nearest the null. j Outcome reassignments required to reverse the significance verdict. k Monte Carlo power of Fisher's exact test at the observed marginal prevalences, α = 0.05. pp = percentage points; RR = relative risk.

The probability of causation, the quantity that maps most directly onto the balance-of-probabilities test, was 87.7% for palpation failure and 83.8% for computed tomography non-adherence; both exceed the 50% threshold. The number-needed-to-omit — the number of patients in whom the step must be omitted to generate one additional delayed diagnosis — was 1.45 for palpation and 1.55 for imaging. These are group-level descriptive quantities computed from crude associations under assumptions set out in section 2.7; they characterise the cohort and do not establish causation in any individual case. Attributable fractions whose bootstrap intervals include negative values, as for wound depth and male sex in Table 5, are not interpretable as protection.

Sensitivity to unmeasured confounding was substantial. An unmeasured confounder would need to be associated with both palpation failure and delayed diagnosis by a risk ratio of at least 11.18 to explain away the point estimate, and at least 3.80 to shift the confidence limit to the null; the corresponding E-values for computed tomography non-adherence were 9.13 and 3.19, and for GCS 15 7.41 and 2.04.

3.6 Sensitivity analyses

Because the principal exposure is defined by absent documentation, its misclassification was addressed quantitatively; results are in Table 6. Under non-differential misclassification the lower 95% confidence limit for palpation failure reaches the null only when 67.7% of recorded exposures are false positives. Under the adversarial assumption that misclassification affects only the delayed group, the tipping point is 55.3%. In other words, more than half of the recorded palpation failures would have to be cases in which palpation was actually performed but not written down before the association became compatible with the null — and if that were true, the manuscript's practical recommendation would be unchanged, because an unrecorded examination is itself a departure from the documentation standard.

Table 6. Sensitivity analyses. Panel A: quantitative bias analysis for misclassification of the palpation exposure, which is defined by absent documentation. Panel B: dependence of the attainable-precision conclusion on the model specification.

φ: fraction of recorded exposures that are false positives	OR (non-differential)	Lower 95% limit	OR (differential, delayed group only)	Lower 95% limit	Verdict at this φ
0%	34.22	5.97	34.22	5.97	excludes the null
10%	25.31	4.63	22.54	4.30	excludes the null
20%	20.17	3.66	15.79	3.16	excludes the null
30%	16.82	2.92	11.41	2.33	excludes the null
40%	14.47	2.31	8.32	1.70	excludes the null
50%	12.73	1.79	6.04	1.22	excludes the null
60%	11.38	1.33	4.28	0.83	compatible with the null
70%	10.31	0.91	2.88	0.52	compatible with the null
80%	9.45	0.50	1.74	0.27	compatible with the null
Tipping point	not reached	67.7%	87.7%	55.3%	value of φ at which the limit reaches 1.0

Notes: Panel A. Non-differential misclassification applies the same false-positive fraction to both outcome groups; the differential scenario is adversarial, applying it only to the delayed group, and is therefore the conservative reading. The association remains incompatible with the null until more than half of the recorded palpation failures are assumed to be misclassified. φ = fraction misclassified; OR = odds ratio.

The conclusion of section 3.3 about attainable precision does not depend on the model specification. Repeating the sampling under the two primary procedural predictors alone, under the five clinical and procedural predictors without age, and under the full eight-covariate specification gave minimum attainable standard errors for the palpation coefficient of 0.937, 0.930 and 0.975 respectively, and for the computed tomography coefficient of 0.962, 0.947 and 0.970 (Table 7). The attainable minimum is therefore essentially invariant to the covariate set: the imprecision is imposed by the event count, not by the model. All fits converged: 3,000 of 3,000 draws in the full specification, and 4,000 of 4,000 in the primary analysis.

Table 7. Panel B of the sensitivity analysis: minimum attainable standard error by model specification.

Specification	Covariates (k)	Draws converged	Minimum SE, palpation	Minimum SE, CT non-adherence	Median SE, palpation	Median SE, CT non-adherence
Two primary procedural predictors only	2	3,000/3,000	0.937	0.962	1.034	1.057
Five clinical and procedural predictors, no age	5	3,000/3,000	0.930	0.947	1.339	1.358
Full specification: eight covariates including age	8	3,000/3,000	0.975	0.970	1.369	1.394

Notes: SE = standard error of the log odds ratio. The attainable minimum is essentially unchanged across specifications, so the imprecision is imposed by the event count rather than by the covariate set. Reported minima are upper bounds on the true minimum over the admissible set.

4. Discussion

This multi-centre forensic cohort quantifies the contribution of two specific, documentable procedural omissions to a fatal diagnostic failure, using methods valid in sparse data. Failure to palpate the wound bed (exact conditional odds ratio 29.86, 95% CI 4.79–281.0) and non-adherence to computed tomography guidelines (20.88, 3.57–174.7) dominated every other candidate factor, and an initial GCS of 15 was itself associated with delay (14.95). The full set of bivariate estimates is set out in Table 2 and Figure 2, and the cohort from which they derive in Table 1. The forensic significance is that neither omission is a resource problem: palpation requires nothing but thirty seconds and a gloved finger, and the imaging indication was already met by the mechanism. This reframes a failure often attributed to scanner scarcity as a failure of protocol compliance — a distinction with direct consequences for quality improvement, expert testimony and liability.

The magnitude of the computed tomography effect is far larger than anything reported for general mild traumatic brain injury populations, and the comparison is instructive because rule non-adherence in the wider literature is bidirectional. A Canadian trauma-centre audit found overall adherence of 87.8%, with imaging omitted in 13.5% of indicated cases and ordered against the rule in 11.9% of non-indicated ones [26]; in a scanned Turkish cohort, 80.2% of scans performed in mild trauma carried no guideline indication at all [28]. Over-imaging of the well and under-imaging of the indicated therefore coexist. In this cohort non-adherence reached 71.4% among delayed cases, which is closer to the rural pattern described by Lavery and colleagues, where only 35.1% of rule-positive patients — and just 16.7% of those meeting a medium-risk criterion, the category into which assault falls — were transferred for imaging [27]. Two features of decision rules make such omission easy to rationalise: their specificity is low, at 41.5% for the Canadian rule in an Ethiopian validation [30] and between 39.42% and 66.67% across five rules applied prospectively [29] so clinicians who override them are frequently vindicated on the individual scan; even a machine-learning model built to predict deterioration in CT-positive GCS 13-15 patients reached 99% sensitivity at only 7% specificity [60]. That interventions can shift the behaviour in either direction is established: multimodal packages have reduced low-yield head computed tomography without changing the positive-scan rate or 72-hour returns [61,62].

The association between an intact GCS of 15 and delayed diagnosis gives empirical form to the deceptive clinical mask. The wider literature establishes that a reassuring presentation does not exclude injury: a negative scan in a fully alert patient left only 27% functionally recovered at two weeks and 44% at six months [15] and 0.4% of such patients developed delayed intracranial haemorrhage, 64.3% of them within three days, with one-year mortality more than doubled [20]. Meta-analysed predictors of traumatic intracranial haemorrhage in the GCS 13-15 band are dominated by skull base fracture signs (OR 11.71, 95% CI 5.51-24.86), precisely the finding that wound-bed palpation is performed to detect, while intoxication and headache — the features clinicians often weigh instead — were not predictive [63]. What the present data add is that a normal presentation independently suppresses the clinician's own threshold to look. The operative numbers for practice are the negative likelihood ratio of GCS 15 in this population, 0.13, against a specificity of only 54.8% (Table 4, Figure 4) — a marker that reassures far more often than it discriminates.

Against the medicolegal literature the concordance is close. Errors of clinical judgment were the dominant contributing factor in 76.8% to 89.1% of 5,854 emergency department malpractice cases [33] and payment in closed diagnostic-error claims was driven by injury severity and by contributing factors falling outside the standard diagnostic pathway [34] — which is an exact description of an omitted examination step. Diagnostic error on the scale Newman-Toker and colleagues estimate, 795,000 Americans seriously harmed annually at a weighted mean error rate of 11.1% [31] is not a marginal phenomenon, and trigger-based detection shows how poorly it is captured prospectively: positive predictive values of 5.4% to 8.9% against a negative predictive value of 100% [64]. In trauma specifically, 9.2% of 99,754 multiply injured patients had a delayed diagnosed injury, the head was the commonest site at 35.8%, and mortality was 10.8% against 6.5% [35]. Autopsy-based series make the same point from the other end of the pathway: injuries present at autopsy but absent from the record were found in 20.3% of trauma decedents and were significantly associated with not undergoing computed tomography [24]. Without the autopsy arm of the ascertainment strategy used here, the cases with the worst outcomes would have been systematically absent from the denominator. Whether post-mortem computed tomography was available at the participating installations is not recorded in the data available for this analysis; it would have mattered, because pooled sensitivity for cranial vault and skull base fracture is 0.89 and 0.87 [65] and post-mortem imaging identifies traumatic brain injury as the mortal injury in 40% of blunt-trauma decedents [66] so a service with routine access would probably ascertain more cases than were captured here. That autopsy remains the arbiter in care-related death is confirmed nationally: of 350 Finnish deaths meeting criteria for an iatrogenic death, 264 (75%) were clarified by medicolegal autopsy [67].

The mechanism connecting omission to death is specific. Low-velocity penetration by a small dense projectile concentrates kinetic energy over a minute surface area, producing focal stress at the inner table and a localised fracture beneath a scalp wound of unremarkable

dimensions — consistent with the absence of any difference in wound depth between the groups in this cohort (p = 0.111), and with the systematic review finding that 90.2% of non-missile penetrating brain injuries harbour a retained foreign body requiring operative removal.[21] Once the dura is breached, low-flow subarachnoid bleeding accumulates gradually and intracranial pressure rises over 24 to 72 hours while the patient remains ambulatory; the population data on delayed intracranial haemorrhage place 64.3% of events inside the first three days [20]. The neuropathology of the endpoint is well characterised in forensic series: hypoxic-ischaemic neuronal injury in 81.0% of assault-related autopsy cases and brain oedema in 47.6% against 7.1% of falls [22] with survival of 24 hours or more after injury the temporal signature distinguishing traumatic axonal injury [23]. The temporal profile is the point: the 24-hour outcome threshold used here sits inside the window in which intervention remains possible.

For forensic physicians the practical implications are three. First, wound-bed palpation should be a required structured field in the emergency record and in the visum et repertum, recorded as performed or not performed together with the finding, because on these data it carries a positive likelihood ratio of 8.12 for subsequent diagnostic delay and a number-needed-to-omit of 1.45 (Table 5). A three-state field — performed and negative, performed and positive, not performed with reason — is more implementable than a general instruction, and documentation adequacy can be scored reproducibly for peer review, with inter-rater agreement of kappa 0.91 for documentation quality and 0.84 for clinical care [68]. Second, the dangerous-mechanism criterion should be enforced irrespective of GCS, and a decision not to image recorded together with its reason, since an undocumented omission is indistinguishable at law from an unconsidered one. A caveat belongs with both recommendations: this cohort establishes that omitting palpation predicts delay, but contains no direct evidence on the detection yield of palpation itself, which the causal chain assumes and which should be measured prospectively. Third, as Figure 5 sets out, any assault victim with a head laceration discharged without imaging should receive a documented 24-to-72-hour review, which is the interval the pathophysiology dictates.

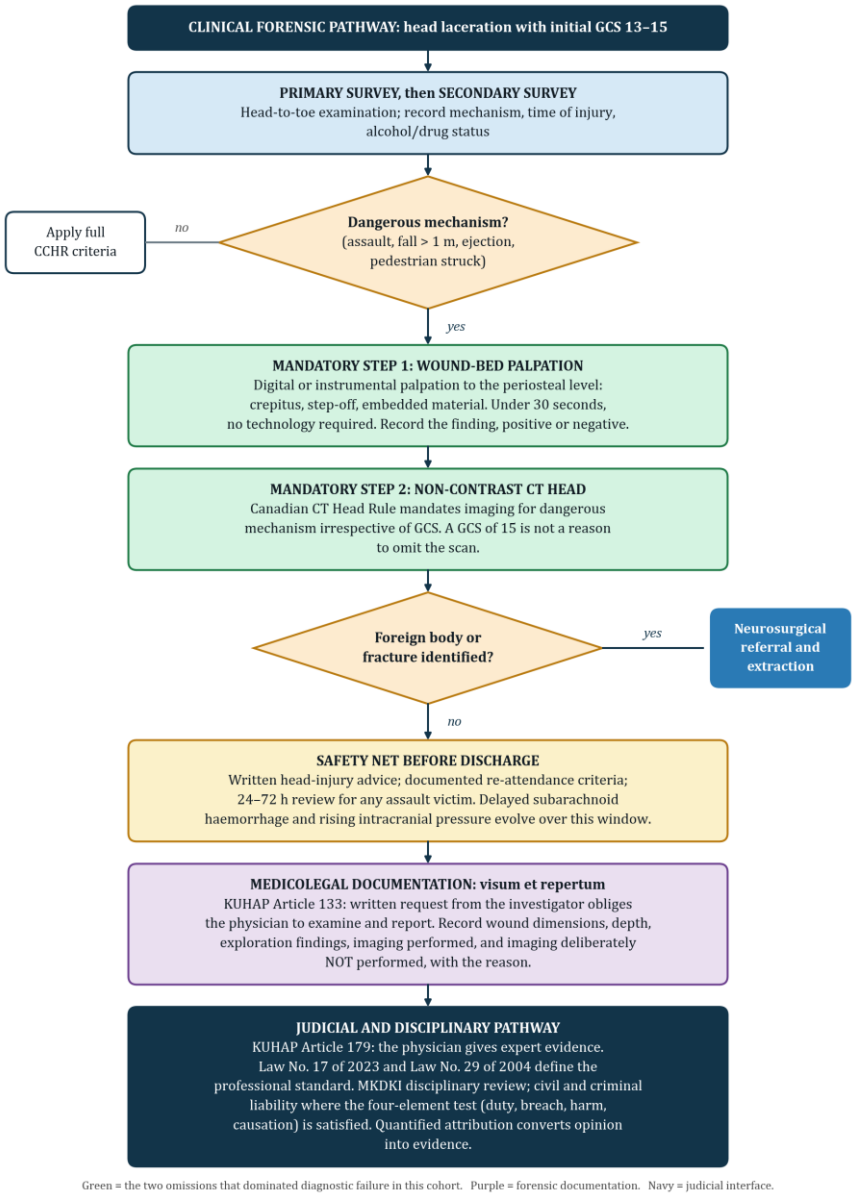


Figure 5. Proposed clinical forensic and medicolegal algorithm for the head-injured patient presenting with a Glasgow Coma Scale of 13 to 15. The two green steps are the omissions that dominated diagnostic failure in this cohort; the purple and navy steps map the pathway onto visum et repertum practice under KUHAP Article 133 and expert evidence under Article 179.

For the Indonesian judicial system the contribution is evidentiary. Under Article 133 of KUHAP a written investigator's request obliges examination and reporting, and under Article 179 the physician must give expert evidence; Law No. 29 of 2004 and Law No. 17 of 2023 supply the substantive standard whose breach founds liability, and knowledge of the latter remains limited among Indonesian health workers [2]. The four-element negligence structure — duty, breach, harm and causation — is where the attribution metrics reported here do work an odds ratio alone cannot. A probability of causation of 87.7% for palpation failure and 83.8% for imaging omission speaks to the balance of probabilities; an E-value of 11.18 states how strong an unmeasured alternative explanation would need to be before the inference fails. Two qualifications must travel with these numbers into any courtroom. They are group-level quantities: an attributable fraction describes what proportion of delayed diagnoses in this cohort is statistically attributable to an omission, not whether a particular omission caused a particular death, and the attributable fraction's target of causation is the exposed group rather than the whole population [54]. They are also sample-specific — transporting one published attributable fraction to its intended target population moved it from 0.21 (95% CI 0.13-0.29) to 0.13 (0.065-0.19) [55] — and estimates from aggregate exposure summaries require nonparametric treatment [53]. The E-value carries its own caveat, being non-invariant to the adjustment strategy used [52]. Stated with those limits, these are terms in which a PDFI-accredited expert can testify without overstating the evidence, and in which POLRI investigators and prosecutors can calibrate a case. Indonesian audits showing that the findings section of the visum et repertum scores 74% and its conclusion 33% [3] and 75% and 50.54% respectively in other series [4,5] indicate that the documentation reform proposed here would also raise the general evidentiary quality of Indonesian forensic reporting.

The study has real strengths. Ascertainment drew on five independent sources including forensic autopsy, which captures cases that never reached a clinical diagnosis and that single-source registry studies systematically lose. The cohort spans three geographically distinct centres over five years with no documented change in scanner availability or protocol. Every inferential statement is exact or penalised rather than asymptotic, which matters when strata contain single observations, and pre-specifying an exact test rather than switching on observed cell counts is the less ambiguous design [43]. The analysis states explicitly which parameters the data identify and which they do not, and the analysis code is deposited publicly. The fragility indices deserve emphasis on the same principle: an index of 6 for palpation failure means the verdict would survive 6 reassignments of outcome, whereas the index of 2 for an initial GCS of 15 means that verdict turns on two patients out of 45 and should be treated as provisional. For context, the median fragility index across 3,758 Cochrane meta-analyses is 5, and 54% have an index of 5 or less [50]; the measure is best read as a sensitivity analysis for outcome misclassification rather than as a significance-threshold metric [49].

The limitations are equally real and several are severe. The design is retrospective, and palpation status was inferred from absent documentation in an unknown number of cases, which will inflate the association if the manoeuvre was more often performed than recorded. It can, however, be bounded: the association remains incompatible with the null until more than half of the recorded exposures are assumed misclassified, even under the adversarial differential scenario (Table 6), and the attainable-precision conclusion holds under every model specification tested (Table 7). Probabilistic bias analysis of this kind is feasible from summary-level data [56] and prior-based parameterisation outperforms naive correction when validation data are sparse [57]. Three further sources of selection bias are unaddressed by the design. Patients assessed and discharged without ever generating a trauma-registry record are absent from the denominator, and because registry entry probably correlates with documentation thoroughness, the exposure prevalence in the source population is likely misestimated in a direction that inflates the association. The three centres are tertiary and regional facilities with electronic records since 2018 and are not representative of peripheral Indonesian practice, where capacity is measurably strained [9,38]. The study window spans the COVID-19 pandemic; case counts by calendar year are not recoverable from the available aggregates, so this cannot be tested and must be treated as an open confounder, as must between-centre variation.

With 14 events the study is powered only for large effects; the minimum detectable odds ratio at 80% power was approximately 9 to 12, so the null findings for wound depth and sex are uninformative rather than negative — a distinction misstated in 56% of published trials with non-significant primary outcomes [48] and one reason interval estimates are better read as compatibility indices [47]. Most importantly, the multivariable question cannot be answered from aggregate data at all: aggregated data cannot recover individual-level effects even in principle [58] the adjusted odds ratios reported here are bounds, and the width of those bounds — from 2.74 to 90.2 for palpation failure, shown in Table 3 and Figure 3 — is the honest measure of what is not known; assumption-free bounds are characteristically wide, which is itself informative about how far conventional point estimates rest on unverifiable assumptions [59]. Penalised estimates should not be read as calibrated absolute risks [46] and no model of this complexity could satisfy modern shrinkage-based sample-size criteria on 14 events [41,42]. Six further variables are not recoverable from the available aggregates and are declared rather than omitted: the split between documented and inferred palpation status, which Canadian CT Head Rule criteria were met in the non-adherent cases, whether a visum et repertum was issued and what it recorded, per-variable missingness, case counts by year, and exposure prevalence by centre. Neither inter-rater agreement between data extractors nor blinding to outcome was assessed. The direction of expected transportability error can nonetheless be stated: in peripheral facilities the exposure prevalence and outcome rate would be higher but the odds ratio probably lower, because near-universal non-palpation leaves the exposure with little variance. Transportability to the fall-dominated elderly cohorts of high-income settings is untested; this population is 84% male with a mean age of 40 and an assault-dominated mechanism.

5. Conclusion

Delayed diagnosis of a retained intracranial foreign body in mild traumatic brain injury is a preventable, protocol-dependent failure with fatal consequences. In 45 cases from three North Sumatran trauma centres, failure to palpate the wound bed (exact conditional odds ratio 29.86, 95% CI 4.79–281.0; attributable fraction 76.3%; E-value 11.18) and non-adherence to computed tomography guidelines (20.88, 3.57–174.7) were the dominant

determinants, with areas under the curve of 0.844 and 0.809 as bedside markers. Indonesian emergency departments should make wound-bed palpation and the imaging decision mandatory structured fields in the record and in the visum et repertum, and should enforce the dangerous-mechanism criterion irrespective of GCS. For PDFI standard-setting and for POLRI-facing expert evidence under KUHAP Articles 133 and 179, quantified attribution converts opinion into testable evidence, and the pathway in Figure 5 states where in the encounter each obligation falls. Prospective, case-level, multi-centre data — ideally with day-of-injury biomarker stratification [19] — are required before any adjusted effect can be estimated rather than bounded.

Declarations

Use of generative artificial intelligence. The authors declare that no generative artificial intelligence (AI) tools were used in the preparation, drafting, analysis, or editing of this manuscript.

Author contributions. All authors contributed to the conception and design of the study, the acquisition and interpretation of the data, and the drafting and critical revision of the manuscript. All authors have reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding. This study received no specific grant from any funding agency.

Conflict of interest. The authors declare no conflict of interest.

Ethical approval. This retrospective study used anonymised clinical, forensic and medicolegal records, including post-mortem examination records, from three tertiary trauma centres in North Sumatra, Indonesia. Ethical approval was obtained from the CMHC Ethics Committee, Indonesia (Approval No. CMHC/EC/2024/0193). Individual consent was waived per institutional protocol for anonymised retrospective data analysis. The study was conducted in accordance with the Declaration of Helsinki.

Consent for publication. Individual informed consent was waived for anonymised retrospective record analysis.

Data availability. Aggregate data supporting the findings are contained within the article. The analysis code reproducing every reported statistic, including the partial-identification sampler and the bias analysis, is deposited in a public repository; the fixed random seed is stated in section 2.7.

Reporting guideline. STROBE. A completed checklist is supplied as a supplementary file.

Acknowledgements. The authors thank the forensic medicine and medical records staff of the participating centres.

References

1. Walz C, Schwarz CS, Imdahl K, et al. Comparison of the quality of clinical forensic examination of victims of physical violence conducted by clinicians and forensic examiners. Int J Legal Med. 2023; 137(6): 1777-1786. doi:10.1007/s00414-023-02985-4
2. Heriyanto H. Tingkat pengetahuan tenaga kesehatan dan masyarakat terhadap undang-undang kesehatan yang baru (Undang-Undang Nomor 17 Tahun 2023 tentang Kesehatan). J Huk HAM Wara Sains. 2024; 3(1): 138-145. doi:10.58812/jhhws.v3i01.1021
3. Pranoto AST, Wibisono B, Rivani R. Evaluasi kualitas visum et repertum perlukaan pada korban hidup di RSUD Waled, Cirebon (2021-2023). Soepra J Huk Kesehat. 2025; 11(1): 93-100. doi:10.24167/sjhk.v11i1.12531
4. Suhatril GA. Kualitas visum et repertum korban hidup kasus perlukaan di RSUD Dr. RM Djoelham Binjai dari tahun 2019 sampai 2021. COMSERVA. 2025; 4(9): 2802. doi:10.59141/comserva.v4i9.2802
5. Furqani S, Muthaher AS, Nasrun. Kualitas laporan visum et repertum kasus perlukaan korban hidup di Rumah Sakit Bhayangkara Palu tahun 2017. Medika Alkhairaat. 2021; 3(3): 117-123. doi:10.31970/ma.v3i3.87
6. Wijaya A, Umar D, Nugroho H. Gambaran visum et repertum (VeR) perlukaan di Instalasi Kedokteran Forensik dan Medikolegal RSUD Abdul Wahab Sjahranie Samarinda tahun 2015-2019. J Sains Kesehat. 2021; 3(4): 417-423. doi:10.25026/jsk.v3i4.322
7. Eko A, Utami HD, Siswoputro AY. Characteristics visum et repertum of live victims at Bhayangkara POLDA DIY Hospital (2014-2016). J Forensik Medikoleg Indones. 2023; 4(2): 371. doi:10.20884/jfmi.v4i2.9429
8. Gulo N, Harianja DD. Gambaran pengetahuan dokter umum tentang visum et repertum perbuatan cabul dan dugaan perkosaan di RSUD Gunungsitoli Nias. J Kedokt STM. 2023; 6(1): 17-22. doi:10.30743/stm.v6i1.352
9. Afandi D, Arini M. Importance-performance analysis of clinical forensic services quality at Bhayangkara Hospital Pekanbaru, Indonesia. Malays J Med Sci. 2024; 31(1): 103-113. doi:10.21315/mjms2024.31.1.9
10. Marissha ED, Ismurrizal. Gambaran jenis trauma penyebab kematian di Bagian Forensik Rumah Sakit Bhayangkara Medan 2021. J Kedokt STM. 2022; 5(2): 164-173. doi:10.30743/stm.v5i2.341
11. Saputra S. Variasi pola trauma pada kasus kematian akibat tindak kekerasan berdasarkan data autopsi di Instalasi Forensik RSUD Dr. Doris Sylvanus tahun 2016-2020. Indones J Leg Forensic Sci. 2022; 12(2): 84. doi:10.24843/ijlfs.2022.v12.i02.p04
12. Narti W, Gani AB, Makmun A, et al. Karakteristik pasien cedera kepala di Rumah Sakit Ibnu Sina Makassar tahun 2022. Fakumi Med J. 2024; 3(11): 814-821. doi:10.33096/fmj.v3i11.413
13. Lecky FE, Otesile O, Marincowitz C, et al. The burden of traumatic brain injury from low-energy falls among patients from 18 countries in the CENTER-TBI Registry: a comparative cohort study. PLoS Med. 2021; 18(9): e1003761. doi:10.1371/journal.pmed.1003761
14. Silverberg ND, Iverson GL, Cogan A, et al. The American Congress of Rehabilitation Medicine diagnostic criteria for mild traumatic brain injury. Arch Phys Med Rehabil. 2023; 104(8): 1343-1355. doi:10.1016/j.apmr.2023.03.036
15. Madhok DY, Rodriguez RM, Barber J, et al. Outcomes in patients with mild traumatic brain injury without acute intracranial traumatic injury. JAMA Netw Open. 2022; 5(8): e2223245. doi:10.1001/jamanetworkopen.2022.23245
16. Yuh EL, Jain S, Sun X, et al. Pathological computed tomography features associated with adverse outcomes after mild traumatic brain injury: a TRACK-TBI study with external validation in CENTER-TBI. JAMA Neurol. 2021; 78(9): 1137-1148. doi:10.1001/jamaneurol.2021.2120
17. Hacker D, Jones CA, Yasin E, et al. Cognitive outcome after complicated mild traumatic brain injury: a literature review and meta-analysis. J Neurotrauma. 2023; 40(19-20): 1995-2014. doi:10.1089/neu.2023.0020
18. Lubbers V, van den Hoven DJ, van der Naalt J, et al. Emergency department risk factors for post-concussion syndrome after mild traumatic brain injury: a systematic review. J Neurotrauma. 2024; 41(11-12): 1253-1270. doi:10.1089/neu.2023.0302
19. Korley FK, Jain S, Sun X, et al. Prognostic value of day-of-injury plasma GFAP and UCH-L1 concentrations for predicting functional recovery after traumatic brain injury in patients from the US TRACK-TBI cohort: an observational cohort study. Lancet Neurol. 2022; 21(9): 803-813. doi:10.1016/S1474-4422(22)00256-3
20. Kwon H, Kim YJ, Lee JH, et al. Incidence and outcomes of delayed intracranial hemorrhage: a population-based cohort study. Sci Rep. 2024; 14(1): 19502. doi:10.1038/s41598-024-70553-w
21. Czyżewski W, Szymoniuk M, Litak J, et al. Management and outcomes of non-missile penetrating brain injury involving the anterior skull base: a case report and systematic review. J Clin Med. 2025; 14(16): 5731. doi:10.3390/jcm14165731
22. Laaksonen J, Mäkinen H, Oura P. Prevalence of secondary brain injuries and association with trauma circumstances in neuropathologically examined medico-legal autopsy cases with primary head trauma. Leg Med (Tokyo). 2024; 71: 102502. doi:10.1016/j.legalmed.2024.102502
23. Narayan G, Oura P. Comparison of background characteristics and neuropathology findings between medico-legal autopsy cases with traumatic axonal injury, vascular axonal injury, or absence of axonal injury in β-

amyloid precursor protein stain. Int J Legal Med. 2025; 139(3): 1335-1342. doi:10.1007/s00414-025-03415-3
24. Gonçalves AC, Parreira JG, Gianvecchio VAP, et al. The role of autopsy on the diagnosis of missed injuries and on the trauma quality program goal definitions: study of 192 cases. Rev Col Bras Cir. 2022; 49: e20223319. doi:10.1590/0100-6991e-20223319_en
25. Stiell IG, Wells GA, Vandemheen K, et al. The Canadian CT Head Rule for patients with minor head injury. Lancet. 2001; 357(9266): 1391-1396. doi:10.1016/S0140-6736(00)04561-X
26. Sampalli A, Kang J, Campbell SG, et al. Adherence to the Canadian CT Head Rule in a Nova Scotian emergency and trauma center. Cureus. 2023; 15(5): e39484. doi:10.7759/cureus.39484
27. Lavery MD, Williamson RA, Curran J, et al. Canadian CT head rule adherence in a rural hospital without in-house computed tomography. Can J Rural Med. 2024; 29(4): 167-172. doi:10.4103/cjrm.cjrm_1_24
28. Gurbuz H, Demir M. Evaluation of cranial computed tomography use and guideline compliance in head trauma patients presenting to the emergency department. Niger J Clin Pract. 2025; 28(9): 1027-1031. doi:10.4103/njcp.njcp_472_25
29. Forouzannia SM, Najafimehr H, Oskooi RK, et al. Clinical decision rules in predicting computed tomography scan findings and need for neurosurgical intervention in mild traumatic brain injury: a prospective observational study. Eur J Trauma Emerg Surg. 2024; 50(4): 1199-1207. doi:10.1007/s00068-023-02373-y
30. Habte YW, Pajer HB, Abicho TB, et al. Validation of the Canadian CT Head Rule and the New Orleans Criteria for mild traumatic brain injury in Ethiopia. World Neurosurg. 2023; 173: e600-e605. doi:10.1016/j.wneu.2023.02.108
31. Newman-Toker DE, Nassery N, Schaffer AC, et al. Burden of serious harms from diagnostic error in the USA. BMJ Qual Saf. 2024; 33(2): 109-120. doi:10.1136/bmjqs-2021-014130
32. Herasevich S, Soleimani J, Huang C, et al. Diagnostic error among vulnerable populations presenting to the emergency department with cardiovascular and cerebrovascular or neurological symptoms: a systematic review. BMJ Qual Saf. 2023; 32(11): 676-688. doi:10.1136/bmjqs-2022-015038
33. Antkowiak PS, Lai SY, Burke RC, et al. Characterizing malpractice cases involving emergency department advanced practice providers, physicians in training, and attending physicians. Acad Emerg Med. 2023; 30(12): 1237-1245. doi:10.1111/acem.14800
34. Grenon V, Szymonifka J, Adler-Milstein J, et al. Factors associated with diagnostic error: an analysis of closed medical malpractice claims. J Patient Saf. 2023; 19(3): 211-215. doi:10.1097/PTS.0000000000001105
35. Gümbel D, Matthes G, Ekkernkamp A, et al. Influencing factors for delayed diagnosed injuries in multiple trauma patients - introducing the 'Risk for Delayed Diagnoses Score' (RIDD-Score). Eur J Trauma Emerg Surg. 2024; 50(5): 2199-2207. doi:10.1007/s00068-024-02571-2
36. Holmstrom AL, Ott KC, Weiss HK, et al. Improving trauma tertiary survey performance and missed injury identification using an education-based quality improvement initiative. J Trauma Acute Care Surg. 2021; 90(6): 1048-1053. doi:10.1097/TA.0000000000003152
37. Wilbers A, DeHoet CA, Sliter RJ, et al. An analysis of missed injuries at a level 1 trauma center with a tertiary survey protocol. Am J Surg. 2022; 224(1): 131-135. doi:10.1016/j.amjsurg.2022.04.010
38. El-Matury HJ, Lestari F, Sunindijo RY. Survey data set on the readiness of hospitals for disaster and emergency in north Sumatra Indonesia. Data Brief. 2024; 54: 110358. doi:10.1016/j.dib.2024.110358
39. 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): 1453-1457. doi:10.1016/S0140-6736(07)61602-X
40. Cisneros-Barrios A, Hernández-Delgado P, Gradilla-Magaña P, et al. EQUATOR reporting guideline use and editorial enforcement in four medical specialties: a cross-sectional study of 20 high-impact journals. BMC Med Res Methodol. 2026; 26(1): 171. doi:10.1186/s12874-026-02901-5
41. Pate A, Riley RD, Collins GS, et al. Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. Stat Methods Med Res. 2023; 32(3): 555-571. doi:10.1177/09622802231151220
42. Riley RD, Whittle R, Sadatsafavi M, et al. A general sample size framework for developing or updating a predictive algorithm: with application to clinical prediction models. BMC Med Res Methodol. 2026; 26(1): 121. doi:10.1186/s12874-026-02856-7
43. Neuhäuser M, Ruxton GD. The choice between Pearson's chi-squared test and Fisher's exact test for 2 x 2 tables. Pharm Stat. 2025; 24(3): e70012. doi:10.1002/pst.70012
44. Hansen S, Rice K. Exact inference for fixed effects meta-analysis of 2 x 2 tables. Stat Med. 2023; 42(19): 3333-3352. doi:10.1002/sim.9808
45. Šinkovec H, Heinze G, Blagus R, et al. To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets. BMC Med Res Methodol. 2021; 21(1): 199. doi:10.1186/s12874-021-01374-y
46. Hashibe S, Hongo W, Shinozaki T. Caveats on using Firth's penalization in the model-based regression standardization for rare diseases. Stat Med. 2026; 45(15-17): e70644. doi:10.1002/sim.70644
47. Cole SR, Edwards JK, Greenland S. Surprise! Am J Epidemiol. 2021; 190(2): 191-193. doi:10.1093/aje/kwaa136
48. Hemming K, Javid I, Taljaard M. A review of high impact journals found that misinterpretation of non-statistically significant results from randomized trials was common. J Clin Epidemiol. 2022; 145: 112-120. doi:10.1016/j.jclinepi.2022.01.014
49. VanderWeele TJ. Reframing the fragility index as bias analysis: sensitivity analysis, P-values, parameterizations, and confidence intervals. Am J Epidemiol. 2026; 195(4): 1182-1184. doi:10.1093/aje/kwaf276
50. Xing A, Xing X, Murad MH, et al. Evaluating the properties of the fragility index of meta-analyses. BMC Med Res Methodol. 2025; 25(1): 212. doi:10.1186/s12874-025-02648-5
51. Sjölander A, Ciocănea-Teodorescu I, Gabriel EE. Bounds and E-values for marginal causal effects. Epidemiology. 2026; 37(1): 5-15. doi:10.1097/EDE.0000000000001919
52. Barrette E, Higuera L, Wherry K. Assessment of the E-value in the presence of bias amplification: a simulation study. BMC Med Res Methodol. 2024; 24(1): 79. doi:10.1186/s12874-024-02196-4
53. Chan CE, Zepeda-Tello R, Camacho-García-Formentí D, et al. Nonparametric estimation of the potential impact fraction and the population attributable fraction with individual-level and aggregated data. Stat Med. 2025; 44(18-19): e70214. doi:10.1002/sim.70214
54. Suzuki E, Yamamoto E. Errors in the calculation of the population attributable fraction. Epidemiology. 2024; 35(4): 469-472. doi:10.1097/EDE.0000000000001731
55. Klose M, Zivich PN, Cole SR. Revisiting the population attributable fraction. Epidemiology. 2025; 36(4): 482-486. doi:10.1097/EDE.0000000000001867
56. Fox MP, MacLehose RF, Lash TL. SAS and R code for probabilistic quantitative bias analysis for misclassified binary variables and binary unmeasured confounders. Int J Epidemiol. 2023; 52(5): 1624-1633. doi:10.1093/ije/dyad053
57. Zhang Q, MacLehose RF, Collin LJ, et al. Parameterization of beta distributions for bias parameters of binary exposure misclassification in probabilistic bias analysis. Epidemiology. 2025; 36(2): 237-244. doi:10.1097/EDE.0000000000001818
58. Kakampakou L, Stokes J, Hoehn A, et al. Simulating hierarchical data to assess the utility of ecological versus multilevel analyses in obtaining individual-level causal effects. BMC Med Res Methodol. 2025; 25(1): 79. doi:10.1186/s12874-025-02504-6
59. Diemer EW, Shi J, Swanson SA. Partial identification of the effects of sustained treatment strategies. Epidemiology. 2024; 35(3): 308-312. doi:10.1097/EDE.0000000000001721
60. Marincowitz C, Paton L, Lecky F, et al. Predicting need for hospital admission in patients with traumatic brain injury or skull fractures identified on CT imaging: a machine learning approach. Emerg Med J. 2022; 39(5): 394-401. doi:10.1136/emermed-2020-210776

61. Luke JC, Kubick RN, Sucharew HJ, et al. A quality improvement campaign reduces head and cervical spine imaging in low-risk trauma patients. West J Emerg Med. 2026; 27(4): 947-955. doi:10.5811/westjem.50707

62. Berlyand Y, Smith Shain R, Savarino JR, et al. A multimodal intervention reduces computed tomography use in blunt head injury. Am J Emerg Med. 2026; 109: 148-154. doi:10.1016/j.ajem.2026.07.005

63. Yang LJ, Lassaren P, Londi F, et al. Risk factors for traumatic intracranial hemorrhage in mild traumatic brain injury patients at the emergency department: a systematic review and meta-analysis. Scand J Trauma Resusc Emerg Med. 2024; 32(1): 91. doi:10.1186/s13049-024-01262-6

64. Khalili M, Enayati M, Patel S, et al. Identifying diagnostic errors in the emergency department using trigger-based strategies. BMJ Open Qual. 2025; 14(3): e003389. doi:10.1136/bmjoq-2025-003389

65. Henningsen MJ, Larsen ST, Jacobsen C, et al. Sensitivity and specificity of post-mortem computed tomography in skull fracture detection - a systematic review and meta-analysis. Int J Legal Med. 2022; 136(5): 1363-1377. doi:10.1007/s00414-022-02803-3

66. Levin JH, Pecoraro A, Ochs V, et al. Characterization of fatal blunt injuries using postmortem computed tomography. J Trauma Acute Care Surg. 2023; 95(2): 186-190. doi:10.1097/TA.0000000000004012

67. Kuvaja P, Keinänen N, Pakanen L. The incidence of iatrogenic deaths in the Finnish cause-of-death statistics; a retrospective study. J Forensic Leg Med. 2022; 86: 102302. doi:10.1016/j.jflm.2021.102302

68. Baer H, Satnick D, McHugh J, et al. A new teaching tool for peer review of charting and care in the emergency department. Jt Comm J Qual Patient Saf. 2023; 49(2): 105-110. doi:10.1016/j.jcjq.2022.10.007