

ORIGINAL ARTICLE

Diagnostic Accuracy of the Blunt Abdominal Trauma Scoring System for Intra-Abdominal Injury in a High-Prevalence Indonesian Trauma Cohort

Akbar Rizky Wicaksana^{1*}, Alsen Arlan², Theodorus³¹ Specialized Residency Training Program, Department of Surgery, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia² Division of Digestive Surgery, Medical Staff Group/Department of Surgery, Dr. Mohammad Hoesin General Hospital, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia³ Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia

ARTICLE INFO

Article history:

Received: August 29, 2025

Revised: September 25, 2025

Accepted: October 27, 2025

Available online: November 2, 2025

Published: November 5, 2025

Keywords:

Abdominal injuries; Indonesia; Sensitivity and specificity; Trauma severity indices; Wounds, nonpenetrating

Corresponding author:

Akbar Rizky Wicaksana

Email:

akbar.rizkyw@gmail.com

DOI:

<https://doi.org/10.37275/sjs.v8i2.164>

Access licence:

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

Copyright:

© 2025 The Authors. Authors retain copyright and grant the journal the right of first publication. Published by HM Publisher.

ABSTRACT

Background: The Blunt Abdominal Trauma Scoring System (BATSS) guides imaging after blunt abdominal trauma, but its accuracy in Indonesian patients already selected for computed tomography (CT) or laparotomy is uncertain.**Objective:** To evaluate BATSS diagnostic accuracy for intra-abdominal injury (IAI) in a high-prevalence trauma cohort.**Methods:** This retrospective study included patients aged ≥ 17 years who underwent contrast-enhanced CT or laparotomy in Palembang, Indonesia, from July 2021 to July 2025. Analyses estimated discrimination, threshold accuracy, stratum-specific likelihood ratios (SSLRs), and arithmetic bounds on component frequencies.**Results:** Among 37 patients, 34 (91.9%) had IAI. The area under the receiver operating characteristic curve (AUC) was 0.525 (95% CI 0.053–0.996; permutation $p = 0.906$). At the protocol threshold of ≥ 12 , sensitivity was 64.7% (95% CI 46.5–80.3%) and specificity 66.7% (9.4–99.2%); at ≥ 10 , they were 82.4% and 33.3%. Risk-band SSLRs (0.53, 0.53, 1.94) had confidence intervals spanning 1. Six injured patients scored 2–4. One uninjured patient scored 19, which requires a positive focused assessment with sonography for trauma under BATSS weights; removing that patient raised the AUC to 0.757.**Conclusion:** In patients already selected for CT or laparotomy, BATSS did not discriminate IAI, and a low score did not exclude injury. The estimates were fragile. BATSS should not be used to withhold imaging in this setting; prospective multicentre validation including unimaged patients is needed.

1. Introduction

Injury remains a leading cause of death and disability in Southeast Asia. In 2021 the countries of the Association of Southeast Asian Nations recorded an estimated 35.5 million injury cases and about 317,000 injury deaths, and road injury was the leading cause of injury death in most member states, including Indonesia¹. Globally, road-injury incidence exceeded 100 million cases in 2019 and peaks in early adulthood². In Indonesia, road traffic crashes are common in both urban and rural districts, helmet use is protective, and motorcycle risk-taking begins in adolescence: about 40% of surveyed Jakarta schoolchildren reported a motorcycle crash in the previous year^{3,4}. Motorcycles accounted for more than 90% of traffic-related head-injury admissions in North Maluku, and more than half of those patients reach hospital more than four hours after injury⁵. Blunt abdominal trauma is a frequent component of this burden. Series from tertiary centres in South Asia describe predominantly male, road-traffic-injured patients in whom the liver, spleen and bowel are frequently injured^{6,7}. In an Indian level I trauma-centre series, 37.0% underwent surgery and mortality was 11.7%⁷.

The central difficulty after blunt abdominal trauma is that injury may be clinically silent at presentation. Physician gestalt predicted significant abdominal or pelvic injury with an area under the receiver operating characteristic curve (AUC) of only 0.70 in a multicentre cohort of more than 2,000 adults⁸; abdominal signs had a sensitivity below 50% in polytrauma⁹; examination missed nearly a third of injuries in older patients

after low-energy falls¹⁰; and the absence of abdominal tenderness alone does not rule out intra-abdominal injury¹¹. Focused assessment with sonography for trauma (FAST) detects haemoperitoneum with high specificity but a pooled sensitivity of only about 60%^{12,13}, its sensitivity improves with serial examination¹⁴ and is lower in haemodynamically stable patients¹⁵, and in pelvic fracture retroperitoneal haematoma produces both false-negative and false-positive scans^{16–18}. Hollow viscus and mesenteric injuries, which account for roughly one in ten blunt abdominal injuries and carry high mortality when diagnosis is delayed, are the injuries most likely to escape early detection¹⁹; the abdominal seat-belt sign marks patients at particular risk^{20,21}.

Contrast-enhanced computed tomography (CT) is the reference investigation for haemodynamically stable patients. In a 2024 cohort of 480 haemodynamically stable patients with high-energy blunt trauma and a normal initial CT, one patient (0.2%) had an intra-abdominal injury identified during observation²². CT is nevertheless not free: whole-body imaging finds abdominal injury in fewer than a quarter of the patients scanned, without a mortality benefit, while increasing radiation dose, incidental findings and emergency-department length of stay^{23–25}. Low-severity patients with a normal examination and FAST rarely have actionable CT findings²⁶. An August 2025 preprint described a decision instrument with the potential to reduce abdominopelvic CT use in minor blunt trauma by nearly half, but prospective validation and peer review remained necessary²⁷. In settings where CT is scarce, the trade-off runs the

other way: adding CT to FAST in stable patients in Egypt halved unnecessary laparotomies and reduced cost²⁸, and negative laparotomy carries excess mortality²⁹. In Indonesia, frontline clinicians report weak prehospital handover and diagnostic delays³⁰, many patients reach tertiary care late after referral from district hospitals³¹, and a reliable bedside triage tool would therefore be valuable.

The Blunt Abdominal Trauma Scoring System (BATSS) was derived in Tehran from 261 adults with suspected blunt abdominal trauma, of whom 18.4% had IAI on CT³². It assigns weighted points to seven bedside findings — tachycardia, hypotension, abdominal pain, abdominal tenderness, lower chest-wall signs, pelvic fracture and a positive FAST — and divides patients into low (<8), moderate (8–11) and high (≥12) risk bands. The derivation study reported that no low-risk patient had a positive CT and that every high-risk patient did, but those '100%' figures compared only the extreme bands and set the moderate band aside³²; in a later cohort of 1000 patients from the same group, two-thirds scored in the low band³³. Subsequent single-centre studies from India, Pakistan, Egypt and Indonesia reported differing diagnostic performance at different thresholds and for different target conditions^{34–42}; one prospective Indian study concluded that the score was not useful in its setting⁴³. Competing tools such as the Clinical Abdominal Scoring System (CASS), the FASILA score and prehospital laparotomy scores have also been evaluated^{44–47}. This variation makes local validation and explicit reporting of thresholds and outcomes essential.

Two methodological issues complicate the interpretation of this literature. First, a score's sensitivity and specificity are not fixed properties: they shift between referred and non-referred populations, and higher prevalence tends to accompany higher sensitivity and lower specificity^{48,49}. Second, BATSS is usually validated only in patients who received the reference standard, and when the decision to obtain CT or operate is itself influenced by findings that the score contains, partial verification bias distorts both sensitivity and specificity⁵⁰. In Indonesia, the only published BATSS validation, from Bali, analysed 44 patients and reported a sensitivity of 91.4% and a specificity of 77.8% at ≥12³⁴, and a Balinese evaluation of CASS in a cohort with 94% positive laparotomies found a negative predictive value of 16.6%⁵¹. No study has reported the full threshold behaviour of BATSS, its stratum-specific likelihood ratios, or what its fixed point weights imply about the patients it misclassifies, in the spectrum in which surgeons actually encounter the score — patients already selected for definitive investigation at a provincial trauma centre.

This study therefore aimed to evaluate the diagnostic accuracy of BATSS for IAI among adults with blunt abdominal trauma who underwent CT or laparotomy at Dr. Mohammad Hoesin General Hospital, Palembang. The specific objectives were to estimate discrimination and accuracy at every observed threshold, including the original risk bands; to compare these estimates with published validations; to use the arithmetic of the score to determine what the observed totals imply about the components present in correctly and incorrectly classified patients; and to quantify the precision of the estimates and the sample size that a definitive validation would require.

2. Methods

Study design and reporting

This was a retrospective, single-centre diagnostic accuracy study of the Blunt Abdominal Trauma Scoring System (BATSS) as the index test and intra-abdominal injury (IAI) as the target condition. It is reported in accordance with the Standards for Reporting of Diagnostic Accuracy Studies (STARD 2015)⁵². The intended role of the index test was triage: to identify, at the first emergency-department assessment, the adults with blunt abdominal trauma who need definitive investigation for IAI. Because only patients who underwent such an investigation could be verified against the reference standard, the study estimates accuracy within the spectrum of patients whom clinicians had already selected for computed tomography (CT) or laparotomy; this spectrum is stated explicitly throughout and its consequences were examined quantitatively.

Setting and participants

The study was performed at Dr. Mohammad Hoesin General Hospital, Palembang, the tertiary referral hospital and designated trauma centre for South Sumatra province, Indonesia, and the main teaching hospital of the Faculty of Medicine, Universitas Sriwijaya. The hospital receives trauma patients directly from the city and by referral from district hospitals across the province, many of which lack continuous CT services. Patients were identified from the hospital's medical-record archive and the surgical department's trauma register for admissions between 1 July 2021 and 31 July 2025, and all eligible patients in that period were included consecutively.

Eligible patients were aged 17 years or older, presented to the emergency department with blunt abdominal trauma as the primary injury, and completed a definitive investigation for IAI, defined as contrast-enhanced CT of the abdomen and pelvis or exploratory laparotomy. Patients were excluded if the record did not allow the BATSS to be calculated or the final diagnosis to be ascertained, if the mechanism was penetrating or mixed, or if they were younger than 17 years. The decision to obtain CT or to operate was made by the treating team as part of routine care and was not influenced by the study. The number of records screened and excluded was not retained in the study database.

Index test

The BATSS was derived in Iran by multivariable logistic regression in adults with blunt abdominal trauma and assigns weighted points to seven findings available at the bedside³². The components, definitions and weights applied are listed in Table 1. Each component was abstracted from the first documented emergency-department assessment — the first recorded heart rate and systolic blood pressure, the documented history and abdominal examination, the documented chest-wall findings, the radiological report of the pelvis and the result of the initial focused assessment with sonography for trauma (FAST) — and the total score (0–24) was calculated by a single investigator using a standardised extraction form. The original risk bands were used: low (<8), moderate (8–11) and high (≥12) risk³². The study protocol pre-specified the high-risk threshold (≥12) for the primary 2 × 2 analysis; all other observed thresholds, including the cut-off of 8.5 that favoured sensitivity, were examined as secondary analyses. The abstractor was not blinded to the reference-standard result, which was documented in the same record.

Table 1. Components of the Blunt Abdominal Trauma Scoring System (BATSS).

Component	Definition used (first emergency-department documentation)	Points	Class
Tachycardia	Heart rate >100 beats/min at first recording	1	Soft
Hypotension	Systolic blood pressure <100 mmHg at first recording	4	Major
Abdominal pain	Patient-reported abdominal pain	2	Soft
Abdominal tenderness	Tenderness, guarding or rigidity on examination	3	Soft
Lower chest-wall sign	Bruising, tenderness or fracture over ribs 6–12	1	Soft
Pelvic fracture	Radiologically confirmed pelvic fracture	5	Major
Positive FAST	Free intraperitoneal fluid on the initial FAST examination	8	Major
Total	Range 0–24; risk bands: low <8, moderate 8–11, high ≥12	24	

FAST, focused assessment with sonography for trauma. 'Major' and 'soft' are descriptive labels used in this analysis for the three objective high-weight signs (17 points in total) and the four low-weight signs (7 points in total); they are not part of the original score.

Records that did not allow every component to be determined were excluded at eligibility, so all seven components were available for each included patient and none was imputed; no index-test result was indeterminate. FAST was performed during the primary survey by the emergency or surgical team on duty as part of routine care, and the operator and scanning protocol were not recorded.

For interpretation, the three objective, high-weight components — hypotension (4 points), pelvic fracture (5 points) and a positive FAST (8 points) — are referred to as major signs, and the four low-weight components — tachycardia (1 point), abdominal pain (2 points), abdominal tenderness (3 points) and a lower chest-wall sign (1 point) — as soft signs. These labels are descriptive and are not part of the original score. The total score of each patient was available for analysis; the presence or absence of each of the seven components was not retained in the analysis dataset.

Target condition and reference standard

IAI was defined as any traumatic injury to a solid organ, hollow viscus, the diaphragm, the mesentery or a major abdominal vessel. It was considered present when the final, official report of the contrast-enhanced CT scan described such an injury or when it was identified and documented in the operative report of an exploratory laparotomy. It was considered absent when the definitive CT scan was reported as showing no acute traumatic intra-abdominal finding or when laparotomy found no injury requiring repair. For patients with IAI, the organs involved were recorded and classified as solid organ only, hollow viscus only, or mixed solid and hollow injury. The route of verification (CT, laparotomy or both) was determined clinically and was not recorded separately for each patient.

The reference standard was obtained during the index admission; the interval between the emergency-department assessment and CT or laparotomy was not recorded. Because CT and laparotomy differ in their sensitivity for hollow-viscus injury and in their false-positive findings, the use of two reference standards (differential verification) is acknowledged as a potential source of bias, and the target condition was any IAI rather than IAI requiring therapeutic intervention.

Other variables

Age at admission, sex and mechanism of injury (road traffic crash, fall or other) were abstracted from the same records. Age was analysed as a continuous variable and in three groups (<19, 19–40 and >40 years) as in the original study protocol.

Sample size

The a priori sample size was calculated for the precision of sensitivity: assuming an expected sensitivity of 95%, a precision of 10 percentage points, $\alpha = 0.05$ and a prevalence of 50%, 37 patients were required, increased to 41 to allow for 10% incomplete records. All 37 eligible patients identified during the

study period were analysed. Because the achieved number of patients without IAI determines the precision of specificity, the precision and power actually attained were quantified, together with the sample sizes that a definitive validation would require.

Statistical analysis

Categorical variables are presented as counts and percentages with 95% Wilson score confidence intervals (CIs), and continuous variables as mean $\pm$ standard deviation and median with interquartile range (IQR). The sex ratio was tested against parity with an exact binomial test. The distribution of the total score was compared between patients with and without IAI with the Mann–Whitney U statistic, the Hodges–Lehmann shift, Cliff's δ and Hedges' g ; because only three patients had no IAI, the p value was obtained by exact permutation over all 7,770 possible allocations of three negative labels among the 37 observed scores.

Discrimination was summarised by the area under the empirical receiver operating characteristic (ROC) curve (AUC), with a 95% CI by the DeLong method, which was pre-specified as the primary interval; the nonparametric asymptotic interval implemented in SPSS, a logit-transformed DeLong interval and a stratified bootstrap (20,000 resamples) were reported as sensitivity analyses, and the AUC was also bounded by breaking tied scores against and in favour of the score. At every observed threshold, sensitivity, specificity and predictive values were calculated with exact Clopper–Pearson 95% CIs, positive and negative likelihood ratios (LR+, LR–) with log-method CIs (adding 0.5 to every cell when a cell was zero), the diagnostic odds ratio and the Youden index ($J = \text{sensitivity} + \text{specificity} - 1$). For the three original risk bands, stratum-specific likelihood ratios (SSLRs) were calculated with log-method CIs, and trend across bands was tested with an exact permutation version of the Cochran–Armitage test and the Fisher–Freeman–Halton test. The association between the score and IAI was also modelled with Firth penalised-likelihood logistic regression, which remains finite with sparse outcomes, reporting the odds ratio (OR) per point with a profile penalised-likelihood 95% CI and the Brier score.

Because the score is a sum of fixed integer weights, each observed total is compatible with only a limited set of component combinations. All 128 possible combinations of the seven components were enumerated, and for each observed total the components that were certainly present, possibly present or certainly absent were identified. Summing over patients gave exact lower and upper bounds for the number of patients in each reference-standard group who could have had each component, and hence for the sensitivity and specificity of each component; each bound is attainable for that component. The same enumeration identified the minimum number of major signs that every attainable total requires.

Sex, age group and mechanism were available only as marginal frequencies, without cross-classification against IAI. Their associations with IAI were therefore examined as a sensitivity analysis over the space of datasets consistent with the observed marginals rather than estimated. Every contingency table compatible with the margins was enumerated, and the attainable ranges of the OR, Cramér's V , risk difference and Fisher exact p value were reported together with the proportion of admissible tables reaching $p < 0.05$; the same approach bounded the sensitivity of the score for solid-organ versus hollow-viscus injuries. To bound the score-IAI association after adjustment, Firth models of IAI on the score plus one covariate were fitted to every admissible patient-level allocation of that covariate (62,029 allocations for sex and 1,806 for mechanism) and, for the three-level age group, to 4,000 random allocations (of 81,516,171 distinct admissible allocations) supplemented by hill-climbing towards the extreme coefficients; models that did not converge by Newton–Raphson were refitted by direct maximisation of the penalised likelihood rather than discarded. A range that spans the null indicates that the data cannot identify the association.

Literature comparisons were restricted to sources publicly available by 20 September 2025. Sensitivity and specificity at the reported thresholds were compared with the present cohort using two-sided Fisher's exact tests on available published counts, with Holm adjustment within three sensitivity and two specificity comparisons. The single AUC benchmark used this cohort's DeLong standard error while treating the external AUC as fixed; it was not a two-sample DeLong test. Cohen's h was calculated as this cohort minus the published value, and, because the studies differ in prevalence, threshold, outcome definition and reference standard, these tests were interpreted as descriptions of between-study heterogeneity rather than as tests of a common accuracy. Predictive values were calculated across pre-test probabilities from 2% to 98% using Jeffreys posterior distributions for sensitivity and specificity (20,000 draws) to obtain 95% credible bands, and posterior probabilities that specificity exceeded clinically relevant values were reported. Decision curve analysis compared the net benefit of imaging or exploring according to the score with imaging all and imaging none, first at the observed prevalence and then, as an illustration, at the 18.4% prevalence of the derivation cohort of unselected emergency-department patients^{32,53}. The power of the present design to reject an AUC of 0.50 was computed from the Hanley–McNeil variance, and sample sizes for prespecified precision of sensitivity, specificity and the AUC were derived for a future validation study^{54,55}.

Three further sensitivity analyses addressed the small number of patients without IAI and the restriction to verified

patients. First, each uninjured patient was removed in turn (leave-one-out) and the AUC and specificity recalculated. Second, to quantify partial verification, hypothetical uninjured patients who would not have been imaged or operated on — with scores of 2, 4 or 7 — were added one at a time to find how many would be needed for the AUC to reach 0.70 and 0.80; these scenarios describe the unobserved population and are not data. Third, Bayesian summaries of specificity were repeated under a uniform Beta(1, 1) prior, a near-Haldane Beta(0.01, 0.01) prior and an informative Beta(8, 2) prior centred on the 80% specificity typical of published validations.

All tests were two-sided with $\alpha = 0.05$, and p values are reported to three decimal places. The original data were held in IBM SPSS Statistics 26 (IBM Corp., Armonk, NY, USA). The source manuscript reports that the primary cohort analyses used Python 3.14 with NumPy 2.4, SciPy 1.18 and matplotlib 3.10, with source figures produced at 600 dpi; these original software versions could not be verified because the original code was not provided. The authorised literature comparisons were recalculated using Python 3.12.14, NumPy 2.3.5 and SciPy 1.18.1, and Figure 8 was redrawn in matplotlib 3.11.2 at 400 dpi. The other ten source figures were retained.

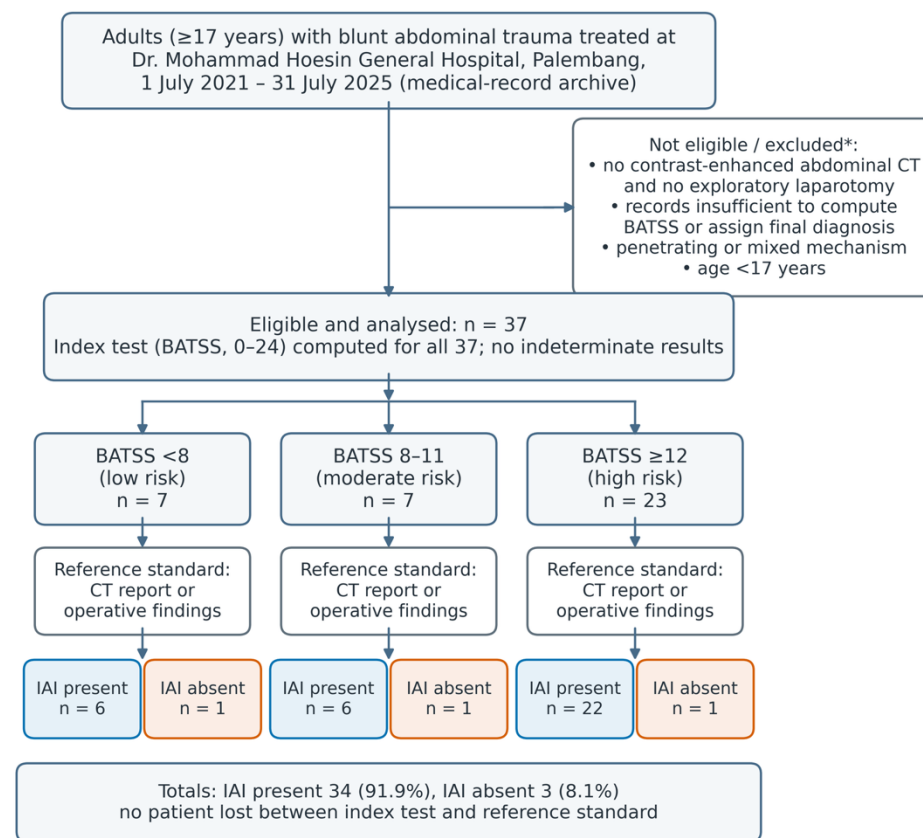
Ethical approval

The study was conducted in accordance with the Declaration of Helsinki. The complete research protocol was submitted to the institutional review board, the Ethics Committee of Dr. Mohammad Hoesin General Hospital Palembang, and was granted a formal ethical exemption (No. DP.04.03/D.XVIII.06.08/ETIK/202/2025). As the study relied exclusively on the retrospective review of de-identified data from existing medical records, the committee waived the need for individual patient informed consent. No identifiable patient information is reported.

3. Results

Participants

Thirty-seven patients met the eligibility criteria; all had both the index test and the reference standard, and no index-test result was missing or indeterminate (Figure 1). Their characteristics are summarised in Table 2 and Figure 2. The mean age was 29.8 ± 11.5 years (median 27, range 17–55); 25 patients (67.6%; 95% CI 51.5–80.4%) were aged 19–40 years, and six (16.2%) were 17 or 18 years old. Twenty-nine were men (78.4%; 95% CI 62.8–88.6%), a male-to-female ratio of 3.6:1 that departed from parity (exact binomial $p < 0.001$). Road traffic crashes accounted for 33 injuries (89.2%; 95% CI 75.3–95.7%), falls for two and other mechanisms for two. The reference standard identified IAI in 34 patients, a prevalence of 91.9% (95% CI 78.7–97.2%); three patients had no IAI.



*Counts of screened and excluded records were not retained in the study database.

Figure 1. STARD 2015 flow of participants. All 37 eligible patients received both the index test (BATSS, scored from the first emergency-department documentation) and the reference standard (contrast-enhanced CT report or operative findings); the boxes give the reference-standard result within each original BATSS risk band. IAI, intra-abdominal injury.

Table 2. Characteristics of the 37 patients with blunt abdominal trauma.

Characteristic	n (%) or summary	95% CI, %
Age, years	mean 29.8 ± 11.5; median 27 (range 17–55)	
Sex		
Male	29 (78.4)	62.8–88.6
Female	8 (21.6)	11.4–37.2
Age group		
17–18 years	6 (16.2)	7.7–31.1
19–40 years	25 (67.6)	51.5–80.4
>40 years	6 (16.2)	7.7–31.1
Mechanism of injury		
Road traffic crash	33 (89.2)	75.3–95.7
Fall	2 (5.4)	1.5–17.7
Other	2 (5.4)	1.5–17.7
Reference standard		
Intra-abdominal injury present	34 (91.9)	78.7–97.2
Intra-abdominal injury absent	3 (8.1)	2.8–21.3
Injury pattern		
Solid organ only	16 (43.2)	28.7–59.1
Hollow viscus only	15 (40.5)	26.3–56.5
Mixed solid and hollow	3 (8.1)	2.8–21.3
No intra-abdominal injury	3 (8.1)	2.8–21.3
BATSS risk band		
Low (<8)	7 (18.9)	9.5–34.2
Moderate (8–11)	7 (18.9)	9.5–34.2
High (≥12)	23 (62.2)	46.1–75.9
BATSS total score	mean 11.84 ± 4.75; median 13 (IQR 10–14; range 2–20)	

CI, Wilson score confidence interval; IQR, interquartile range (SPSS weighted-average definition). Male:female ratio 3.6:1 (exact binomial test against 1:1, $p < 0.001$).

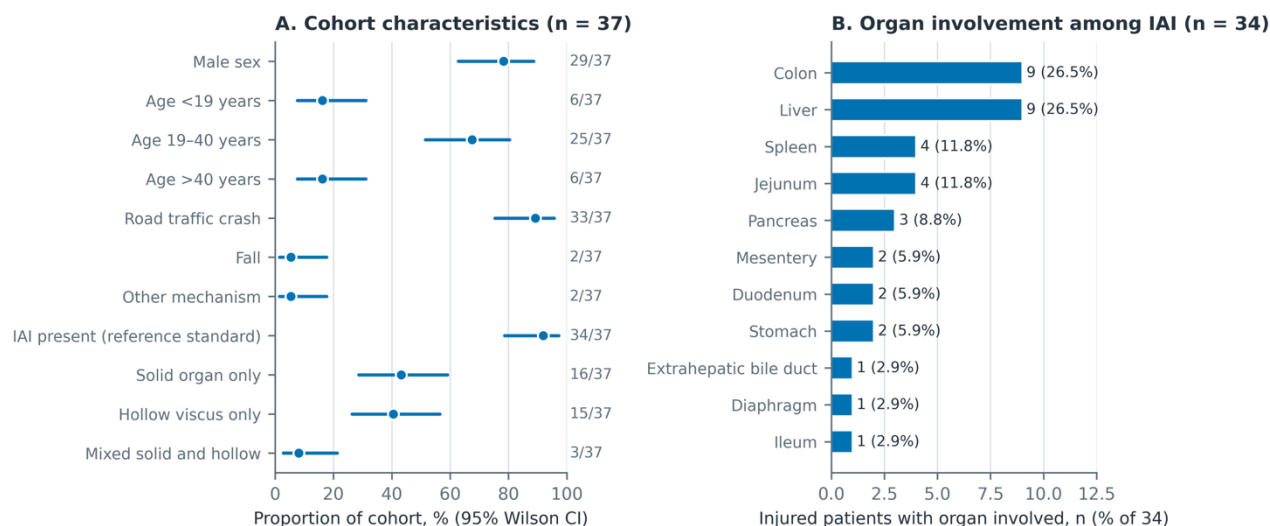


Figure 2. Cohort profile. (A) Demographic, mechanism and injury-pattern proportions with 95% Wilson confidence intervals (n = 37). (B) Organ involvement among the 34 patients with intra-abdominal injury, counting each organ once per patient (combined injuries contribute to every organ involved).

Injury pattern

Among the 34 patients with IAI, 16 (47.1% of injured patients; 43.2% of the cohort) had solid-organ injury only, 15 (44.1%; 40.5% of the cohort) had hollow-viscus injury only and three (8.8%) had combined solid and hollow injuries (Table 3). When each organ was counted once per patient, the liver and the colon were each involved in nine patients (26.5% of injured patients;

95% CI 14.6–43.1%), followed by the spleen and jejunum (four each), the pancreas (three) and the stomach and duodenum (two each); one patient each had injuries of the ileum, diaphragm or extrahepatic bile duct, and two had mesenteric injuries (Figure 2B). Eighteen of the 34 injured patients (52.9%; 95% CI 36.7–68.5%) therefore had at least one hollow-viscus injury.

Table 3. Organs injured among the 34 patients with intra-abdominal injury.

Organ(s) injured (as recorded)	n (% of 34)	95% CI, %
Colon	8 (23.5)	
Liver	6 (17.6)	
Spleen	4 (11.8)	
Jejunum	3 (8.8)	
Pancreas	3 (8.8)	
Duodenum	2 (5.9)	
Stomach	1 (2.9)	
Ileum	1 (2.9)	
Mesentery	1 (2.9)	
Ileal mesentery	1 (2.9)	
Diaphragm and stomach	1 (2.9)	
Liver and extrahepatic bile duct	1 (2.9)	
Liver and colon	1 (2.9)	
Liver and jejunum	1 (2.9)	
Organ involved (any)		
Liver	9 (26.5)	14.6–43.1
Colon	9 (26.5)	14.6–43.1
Jejunum	4 (11.8)	4.7–26.6
Spleen	4 (11.8)	4.7–26.6
Pancreas	3 (8.8)	3.0–23.0
Stomach	2 (5.9)	1.6–19.1
Duodenum	2 (5.9)	1.6–19.1
Mesentery	2 (5.9)	1.6–19.1
Ileum	1 (2.9)	0.5–14.9
Diaphragm	1 (2.9)	0.5–14.9
Extrahepatic bile duct	1 (2.9)	0.5–14.9
Any hollow-viscus involvement	18 (52.9)	36.7–68.5

Upper block: mutually exclusive injury categories as recorded. Lower block: each organ counted once per patient, so combined injuries contribute to every organ involved. CI, Wilson score interval. Hollow-viscus involvement = hollow viscus only (15) plus mixed solid and hollow injuries (3).

Distribution of the score

The total score ranged from 2 to 20 (mean 11.84 ± 4.75 ; median 13, IQR 10–14), and only 12 distinct values were observed (Table 4, Figure 3). Twenty-three patients (62.2%) were in the high-risk band, seven (18.9%) in the moderate band and seven (18.9%) in the low band. Patients with IAI had a median score of

13 (IQR 10–14; mean 11.79 ± 4.73); the three patients without IAI scored 7, 11 and 19 (median 11; mean 12.33). Six injured patients scored 2, 3 or 4, lower than every uninjured patient, whereas the uninjured patient with a score of 19 ranked above 31 of the 34 injured patients. The distributions were not distinguishable: Mann-Whitney $U = 53.5$, Hodges-Lehmann

shift 0.5 point, Cliff's $\delta = 0.049$ and Hedges' $g = -0.11$ (95% CI -1.26 to 1.05), with an exact permutation p of 0.906. In Firth logistic regression, the OR for IAI per BATSS point was 0.997

(95% CI 0.75–1.24; $p = 0.977$), and the Brier score of the fitted model (0.0749) did not improve on that of the prevalence alone (0.0745).

Table 4. Distribution of the BATSS total score by reference-standard status.

BATSS score	Original band	IAI present (n = 34)	IAI absent (n = 3)	Sensitivity at $\geq$ score, %	1 – specificity at $\geq$ score, %
20	High	1	0	2.9	0.0
19	High	2	1	8.8	33.3
15	High	3	0	17.6	33.3
14	High	10	0	47.1	33.3
13	High	5	0	61.8	33.3
12	High	1	0	64.7	33.3
11	Moderate	3	1	73.5	66.7
10	Moderate	3	0	82.4	66.7
7	Low	0	1	82.4	100.0
4	Low	2	0	88.2	100.0
3	Low	2	0	94.1	100.0
2	Low	2	0	100.0	100.0
Summary		median 13 (IQR 10–14); mean 11.79 ± 4.73	median 11 (range 7–19); mean 12.33 ± 6.11		

IAI, intra-abdominal injury; IQR, interquartile range. Only the 12 listed values were observed. Mann–Whitney $U = 53.5$, $AUC = 0.525$, Cliff's $\delta = 0.049$, Hodges–Lehmann shift (present – absent) = 0.5 point; exact permutation $p = 0.906$ (all 7,770 allocations of three negatives among 37 patients).

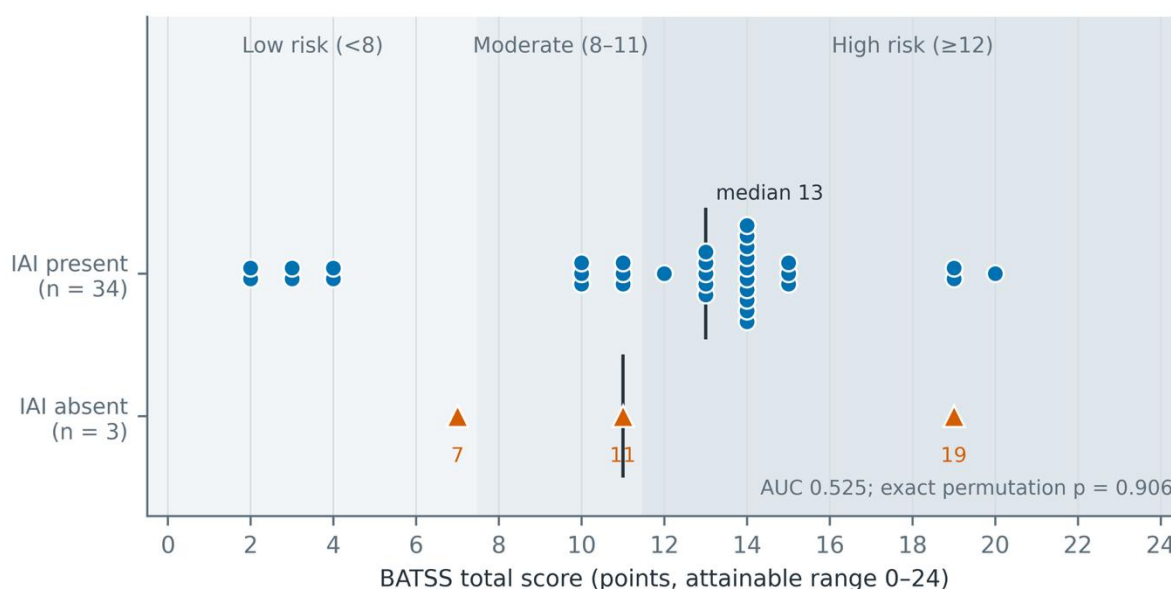


Figure 3. Observed BATSS total score of every patient by reference-standard status. Each circle is one patient with intra-abdominal injury (IAI) and each triangle one patient without IAI (scores 7, 11 and 19); vertical bars mark group medians; shading shows the original risk bands. No patient scored 0–1, 5–6, 8–9 or 16–18. $AUC = 0.525$; exact permutation $p = 0.906$.

Discrimination and accuracy across thresholds

The empirical ROC curve lay close to the line of no discrimination (Figure 4). The AUC was 0.525 (DeLong 95% CI 0.053–0.996; SPSS nonparametric 95% CI 0.133–0.916; logit-

DeLong 0.143–0.880; bootstrap 0.088–0.863), and the exact permutation test gave $p = 0.906$. Breaking tied scores against or in favour of the score bounded the AUC between 0.500 and 0.549.

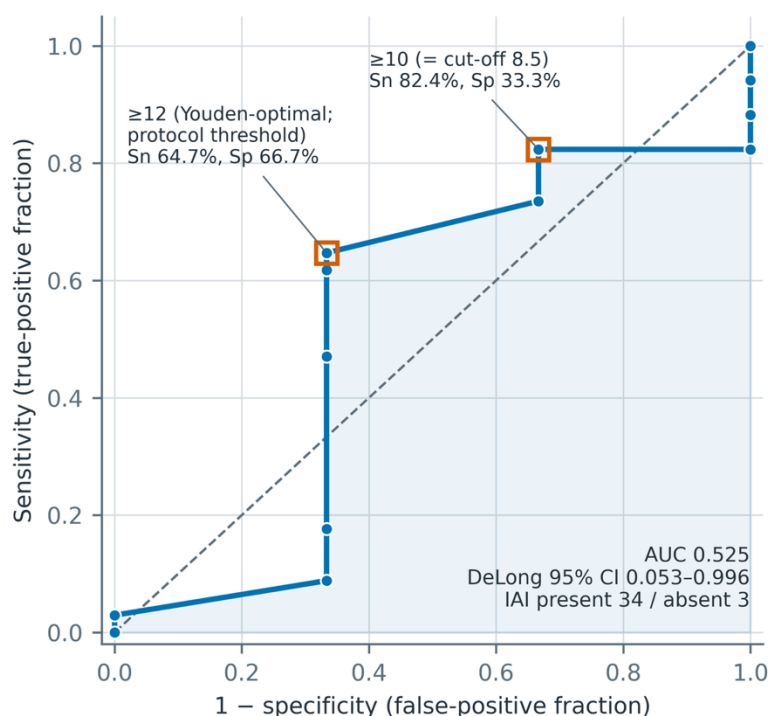


Figure 4. Empirical receiver operating characteristic curve of BATSS for intra-abdominal injury (34 with and 3 without injury). Each point is an observed cut-off; diagonal segments reflect tied scores across reference-standard groups. Squares mark the protocol threshold ≥ 12 , which also maximised the Youden index, and the threshold ≥ 10 (SPSS cut-off 8.5). AUC 0.525 (DeLong 95% CI 0.053–0.996).

Accuracy at every observed threshold is shown in Table 5 and Figure 5. At the protocol threshold of ≥ 12 , the score classified 22 of 34 injured patients as positive (sensitivity 64.7 (46.5–80.3%)) and two of three uninjured patients as negative (specificity 66.7 (9.4–99.2%)); LR+ was 1.94 (0.38–9.80) and LR– 0.53 (0.21–1.33). This threshold gave the highest Youden index (0.314). At the sensitivity-oriented cut-off of 8.5 — equivalent to ≥ 10 , and identical in classification to ≥ 8 because

no patient scored 8 or 9 — sensitivity was 82.4 (65.5–93.2%) and specificity 33.3 (0.8–90.6%), with a Youden index of 0.157, LR+ 1.24 (0.55–2.79) and LR– 0.53 (0.09–3.07). The corresponding positive predictive value was 93.3 (77.9–99.2%) and the negative predictive value 14.3 (0.4–57.9%). Every threshold at or below 7 classified all three uninjured patients as positive, and no threshold produced an LR+ whose 95% CI excluded 1 or an LR– whose 95% CI lay entirely below 1.

Table 5. Diagnostic accuracy of BATSS for intra-abdominal injury at every observed threshold.

Threshold	TP/FP/FN/TN	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	LR+ (95% CI)	LR– (95% CI)	Youden J
≥ 3	32/3/2/0	94.1 (80.3–99.3)	0.0 (0.0–70.8)	91.4 (76.9–98.2)	0.0 (0.0–84.2)	1.06 (0.72–1.55)	0.57 (0.03–9.93)	-0.059
≥ 4	30/3/4/0	88.2 (72.5–96.7)	0.0 (0.0–70.8)	90.9 (75.7–98.1)	0.0 (0.0–60.2)	1.00 (0.67–1.47)	1.03 (0.07–15.81)	-0.118
≥ 7	28/3/6/0	82.4 (65.5–93.2)	0.0 (0.0–70.8)	90.3 (74.2–98.0)	0.0 (0.0–45.9)	0.93 (0.62–1.39)	1.49 (0.10–21.76)	-0.176
≥ 10 (= 8.5; $\equiv \geq 8$)	28/2/6/1	82.4 (65.5–93.2)	33.3 (0.8–90.6)	93.3 (77.9–99.2)	14.3 (0.4–57.9)	1.24 (0.55–2.79)	0.53 (0.09–3.07)	0.157
≥ 11	25/2/9/1	73.5 (55.6–87.1)	33.3 (0.8–90.6)	92.6 (75.7–99.1)	10.0 (0.3–44.5)	1.10 (0.48–2.52)	0.79 (0.15–4.33)	0.069
≥ 12 (protocol)	22/1/12/2	64.7 (46.5–80.3)	66.7 (9.4–99.2)	95.7 (78.1–99.9)	14.3 (1.8–42.8)	1.94 (0.38–9.80)	0.53 (0.21–1.33)	0.314
≥ 13	21/1/13/2	61.8 (43.6–77.8)	66.7 (9.4–99.2)	95.5 (77.2–99.9)	13.3 (1.7–40.5)	1.85 (0.37–9.38)	0.57 (0.23–1.42)	0.284
≥ 14	16/1/18/2	47.1 (29.8–64.9)	66.7 (9.4–99.2)	94.1 (71.3–99.9)	10.0 (1.2–31.7)	1.41 (0.27–7.27)	0.79 (0.34–1.88)	0.137
≥ 15	6/1/28/2	17.6 (6.8–34.5)	66.7 (9.4–99.2)	85.7 (42.1–99.6)	6.7 (0.8–22.1)	0.53 (0.09–3.07)	1.24 (0.55–2.79)	-0.157
≥ 19	3/1/31/2	8.8 (1.9–23.7)	66.7 (9.4–99.2)	75.0 (19.4–99.4)	6.1 (0.7–20.2)	0.26 (0.04–1.83)	1.37 (0.61–3.06)	-0.245
≥ 20	1/0/33/3	2.9 (0.1–15.3)	100.0 (29.2–100.0)	100.0 (2.5–100.0)	8.3 (1.8–22.5)	0.34 (0.02–7.09)	1.09 (0.75–1.59)	0.029

Test positive if score $\geq$ threshold; no patient scored 8 or 9, so ≥ 8 classifies identically to ≥ 10 . Proportions with exact Clopper–Pearson CIs; likelihood ratios with log-method CIs (0.5 added to every cell when a cell was zero). Predictive values apply only to the observed prevalence of 91.9%. TP, true positive; FP, false positive; FN, false negative; TN, true negative; PPV/NPV, positive/negative predictive value; LR, likelihood ratio.

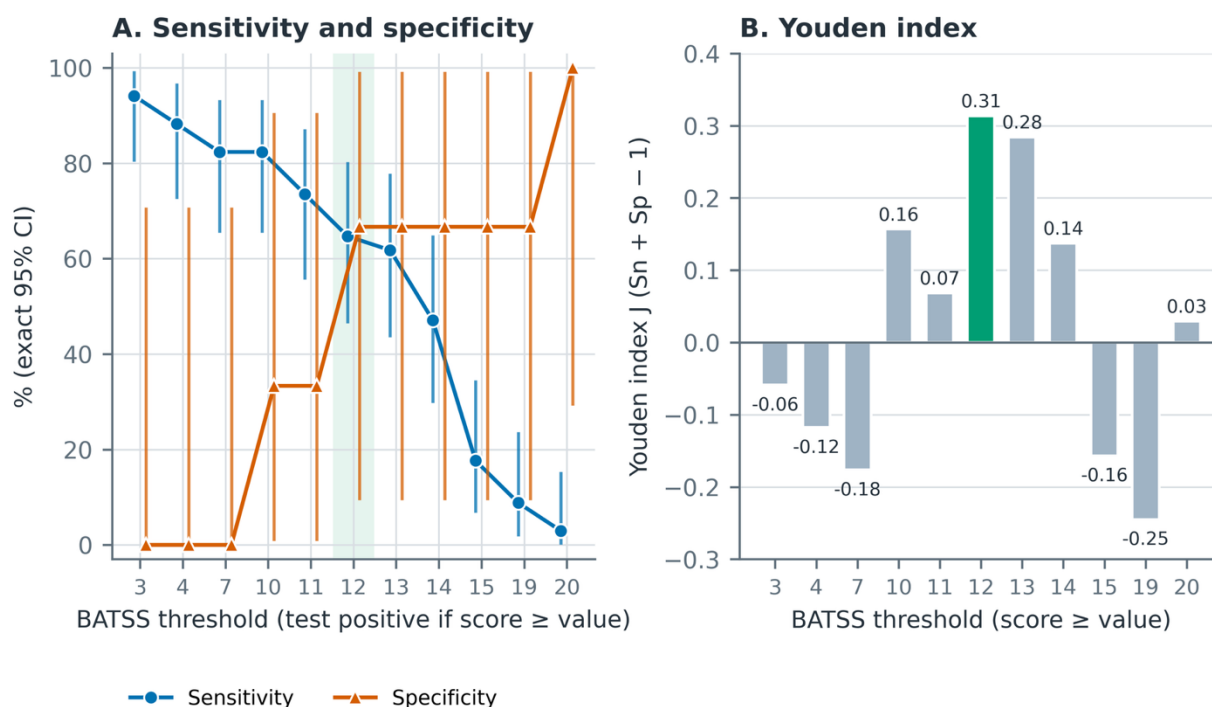


Figure 5. Diagnostic performance at every observed BATSS threshold. (A) Sensitivity (circles) and specificity (triangles) with exact Clopper-Pearson 95% confidence intervals; the shaded column marks ≥ 12 . (B) Youden index $J = \text{sensitivity} + \text{specificity} - 1$; the maximum (green) occurred at ≥ 12 . Because only three patients had no injury, specificity can take only the values 0%, 33.3%, 66.7% and 100%.

Original risk bands

IAI was present in six of seven low-risk patients (85.7%; 95% CI 48.7–97.4%), six of seven moderate-risk patients (85.7%) and 22 of 23 high-risk patients (95.7%; 95% CI 79.0–99.2%) (Table 6, Figure 6). The SSLR was 0.53 (95% CI 0.09–3.07) for both the

low and moderate bands and 1.94 (0.38–9.80) for the high band; every interval included 1. Neither the exact trend test ($p = 0.482$) nor the Fisher-Freeman-Halton test ($p = 0.316$) showed a gradient. Applied to the 18.4% pre-test probability of the derivation cohort³², the point SSLRs would move the probability of IAI to 10.7% in the low band and 30.4% in the high band.

Table 6. Intra-abdominal injury and stratum-specific likelihood ratios across the original BATSS risk bands.

Risk band	n	IAI present/absent	IAI, % (95% CI)	SSLR (95% CI)	Post-test probability at 50% pre-test, %	Post-test probability at 18.4% pre-test, %
Low risk (<8)	7	6/1	85.7 (48.7–97.4)	0.53 (0.09–3.07)	34.6	10.7
Moderate risk (8–11)	7	6/1	85.7 (48.7–97.4)	0.53 (0.09–3.07)	34.6	10.7
High risk (≥ 12)	23	22/1	95.7 (79.0–99.2)	1.94 (0.38–9.80)	66.0	30.4

IAI, intra-abdominal injury; SSLR, stratum-specific likelihood ratio = (proportion of injured patients in the band)/(proportion of uninjured patients in the band), log-method CI. Exact permutation test for trend across bands $p = 0.482$ (Cochran-Armitage $z = 0.99$); Fisher-Freeman-Halton exact $p = 0.316$. Post-test probabilities apply the point SSLR to illustrative pre-test probabilities. With a single uninjured patient in each band, the log-method intervals are approximate.

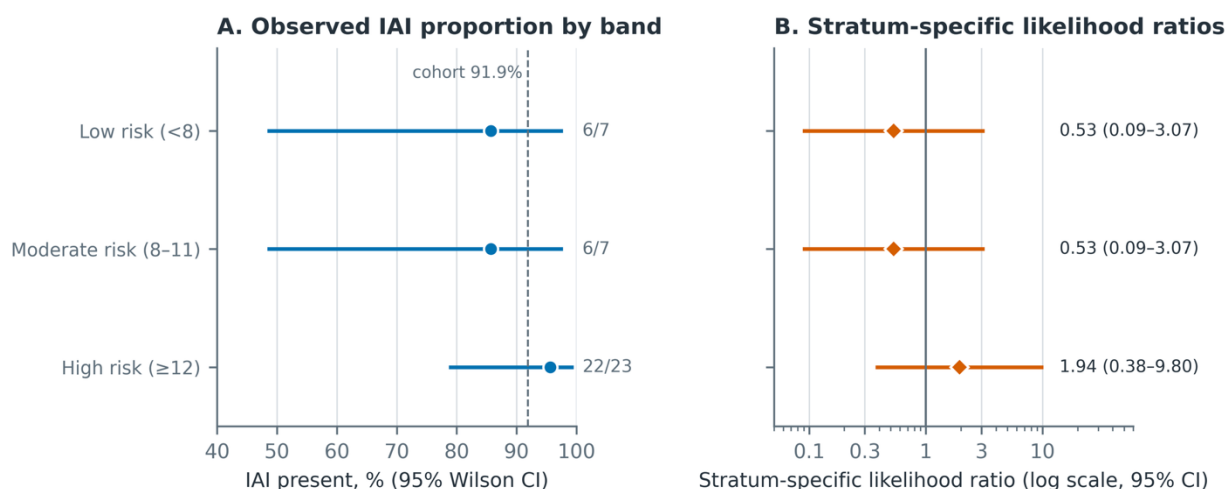


Figure 6. Original BATSS risk bands. (A) Proportion with intra-abdominal injury (IAI) in each band with 95% Wilson confidence intervals; the dashed line is the cohort prevalence. (B) Stratum-specific likelihood ratios (SSLR) with 95% confidence intervals on a log scale; an SSLR of 1 (vertical line) leaves the pre-test probability unchanged.

Arithmetic architecture of the score

Enumeration of all 128 component combinations showed that the four soft signs together contribute at most 7 points, so no total of 8 or more can be reached without at least one major sign, a total of 16 or more requires at least two, and a total of 17 or more cannot be reached without a positive FAST (Figure 7A). Applied to the observed totals (Table 7, Figure 7B), 28 of the 34 injured patients (82.4%) and two of the three uninjured patients had scores that obligatorily included at least one major sign; three injured patients (scores 19–20) and one uninjured patient

(score 19) had scores that obligatorily included a positive FAST together with hypotension or a pelvic fracture. Consequently, the number of uninjured patients with a positive FAST lay between 1 and 2 of three (FAST specificity 33.3–66.7%), and the number of injured patients with a positive FAST between 3 and 28 of 34. The six injured patients with totals of 2–4 cannot have had a positive FAST or a pelvic fracture; at most one major sign, hypotension, is arithmetically possible only for the two who scored 4. The bounds for the soft signs were uninformative, spanning 0–100% in both groups.

Table 7. Item frequencies compatible with the observed BATSS totals (exact arithmetic bounds).

Item	Points	Class	Positive among IAI present (of 34)	Positive among IAI absent (of 3)	Item sensitivity range, %	Item specificity range, %
Tachycardia (HR >100/min)	1	Soft	0–34	0–3	0.0–100.0	0.0–100.0
Hypotension (SBP <100 mmHg)	4	Major	0–30	0–3	0.0–88.2	0.0–100.0
Abdominal pain	2	Soft	0–34	0–3	0.0–100.0	0.0–100.0
Abdominal tenderness/guarding	3	Soft	0–32	0–3	0.0–94.1	0.0–100.0
Lower chest wall sign	1	Soft	0–34	0–3	0.0–100.0	0.0–100.0
Pelvic fracture	5	Major	1–28	0–3	2.9–82.4	0.0–100.0
Positive FAST	8	Major	3–28	1–2	8.8–82.4	33.3–66.7

For each patient, every one of the 128 possible item combinations whose points sum to the observed total was enumerated; an item is 'certain' if present in all compatible combinations and 'possible' if present in at least one. Group bounds sum these over patients; each bound is attainable for that item, although not necessarily simultaneously for all items. Totals ≥ 17 cannot be reached without a positive FAST (maximum without FAST = 16), totals ≥ 8 require at least one major sign (soft signs sum to at most 7), and totals ≤ 7 exclude a positive FAST. The uninjured patient who scored 19 therefore had a positive FAST together with hypotension or a pelvic fracture.

A. Minimum number of major signs needed to reach each score (F = positive FAST obligatory)



B. Patients by obligatory major-sign content implied by their observed score

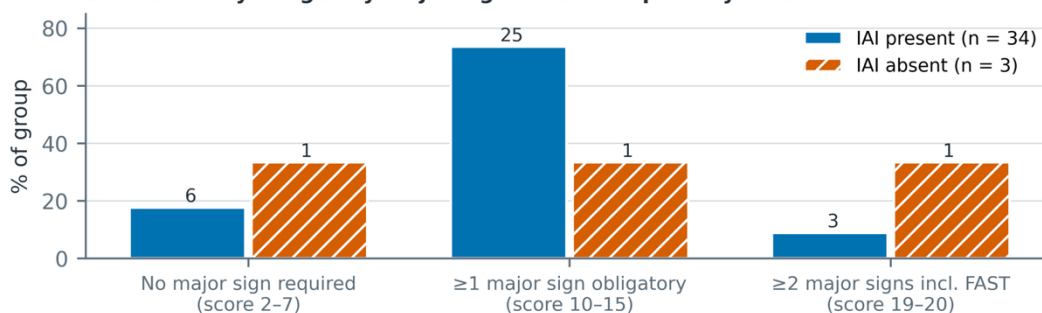


Figure 7. Arithmetic architecture of BATSS. (A) For every attainable total (0–24), the minimum number of major signs (hypotension, pelvic fracture, positive FAST) that any item combination reaching that total must contain; 'F' marks totals that cannot be reached without a positive FAST. The four soft signs (abdominal pain, tenderness, lower chest-wall sign, tachycardia) sum to at most 7, so they cannot by themselves lift a patient out of the low-risk band. (B) Patients grouped by the major-sign content that their observed score makes obligatory, by reference-standard status.

Comparison with published studies

Table 8 and Figure 8 place the estimates beside validations available by 20 September 2025. At ≥ 12 , sensitivity was lower than in the Bali validation³⁴ (64.7% vs 91.4%; Fisher $p = 0.009$, Holm $p = 0.013$; Cohen's $h = -0.68$). At the lower threshold (≥ 8 to ≥ 10), sensitivity (82.4%) was below the 100% reported from Chennai³⁵ (Fisher $p = 0.007$, Holm $p = 0.013$) and from Kozhikode⁴⁰, where the prevalence of injury (90.3%) resembled that of the present cohort (Fisher $p < 0.001$, Holm $p < 0.001$). Specificity at ≥ 12 did not differ detectably from the Bali estimate (66.7% vs 77.8%; $p = 1.000$), whereas specificity at ≥ 10 (33.3%)

was lower than the 98.3% reported from Chennai (Fisher $p = 0.005$, Holm $p = 0.009$; $h = -1.65$). The AUC of 0.525 was below the 0.992 reported from Chennai³⁵, but a benchmark z-test using only this cohort's DeLong standard error gave $p = 0.052$; this calculation treats the published AUC as fixed and is not a two-sample comparison. The case mix also differed sharply from the derivation group's 1000-patient emergency-department cohort³³: 62.2% of the present patients were in the high-risk band compared with 23.0% (exact binomial $p < 0.001$; $h = 0.82$), the mean score was 11.84 versus 6.29 ($p < 0.001$), and the prevalence of injury was 91.9% compared with 18.4% in the derivation cohort³² ($p < 0.001$). Because these studies differ in

prevalence, threshold, outcome definition and reference standard, the contrasts describe heterogeneity between study populations rather than differences in a single underlying

accuracy; the spectrum contrast with the derivation cohort is the most informative of them.

Table 8. BATSS performance in this cohort compared with published studies.

Study (setting)	Measure	Threshold	Published, %	This cohort, %	Test	p	Holm p	Cohen's h
Karjosukarso 2019 (Indonesia, Bali) ³⁴	Sensitivity	≥12	91.4 (32/35)	64.7 (22/34)	Fisher	0.009	0.013	−0.68
Subbiah 2021 (India) ³⁵	Sensitivity	≥8	100 (41/41)	82.4 (28/34)	Fisher	0.007	0.013	−0.87
Manjush 2024 (India; prevalence 90.3%) ⁴⁰	Sensitivity	8.5	100 (140/140)	82.4 (28/34)	Fisher	<0.001	<0.001	−0.87
Karjosukarso 2019 (Indonesia, Bali) ³⁴	Specificity	≥12	77.8 (7/9)	66.7 (2/3)	Fisher	1.000	1.000	−0.25
Subbiah 2021 (India) ³⁵	Specificity	≥8	98.3 (58/59)	33.3 (1/3)	Fisher	0.005	0.009	−1.65
Subbiah 2021 (India) ³⁵	AUC	—	0.992	0.525	z (fixed benchmark)	0.052	0.052	—
Shojaee 2020 derivation-group cohort, n = 1000 ³³	High-risk band (≥12)	—	23.0	62.2 (23/37)	Exact binomial	<0.001	—	0.82
Shojaee 2020 derivation-group cohort, n = 1000 ³³	Mean BATSS	—	6.29 ± 5.80	11.84 ± 4.75	One-sample t	<0.001	—	d = 1.17
Shojaee 2014 derivation cohort, n = 261 ³²	IAI prevalence	—	18.4 (48/261)	91.9 (34/37)	Fisher	<0.001	—	1.68

Comparisons use the threshold each study reported (this cohort: ≥12 = protocol threshold; ≥8 and 8.5 classify identically to ≥10 here). The literature cutoff was 20 September 2025. Two-sided Fisher's exact tests use the published counts; Holm adjustment is applied within three sensitivity and two specificity comparisons. The single AUC benchmark uses this cohort's unrounded DeLong standard error (0.2407487) and treats the published AUC as fixed; no uncertainty from the external AUC is propagated. Cohen's h is this cohort minus the published value (0.2 small, 0.5 medium, 0.8 large). Case-mix comparisons are descriptive and outside these accuracy families. The derivation study compared only the extreme bands³². Manjush 2024 reported specificity of 97.3% with 15 patients without injury, which is arithmetically inconsistent (14/15 = 93.3%); its specificity is not tested. Other reported point estimates without usable threshold-specific denominators include Sadia 2024 (sensitivity 65.9%, specificity 95%)⁴¹, Shah 2023 (100%, 87.1%)³⁹, and Sivarajan 2021 (specificity 100% above 12)³⁶. IAI, intra-abdominal injury.

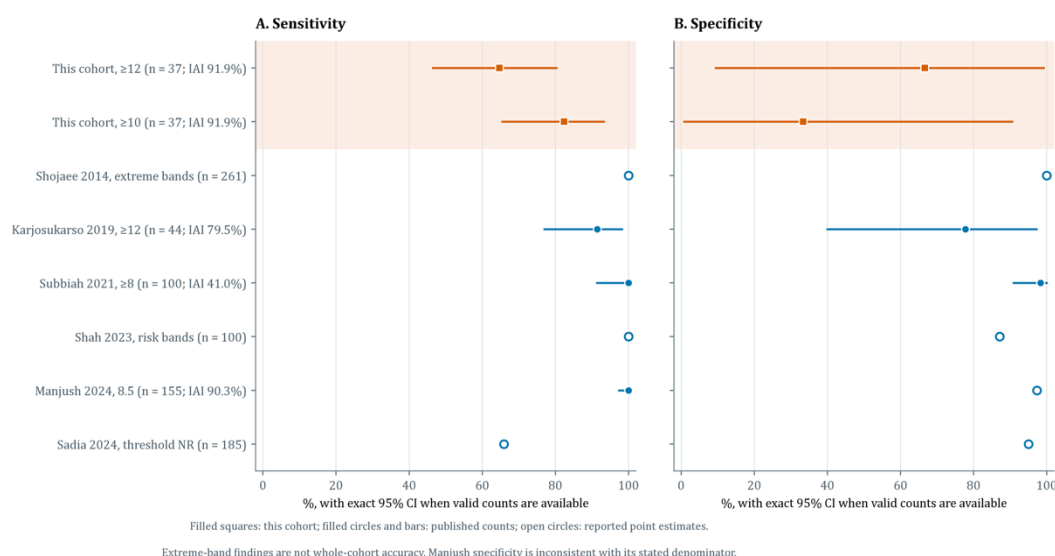


Figure 8. Sensitivity and specificity of BATSS in this cohort and selected validations publicly available by 20 September 2025. Filled markers with bars show exact Clopper–Pearson 95% confidence intervals from reported counts; open markers show published point estimates without usable denominators. Shojaee's extreme-band statements are not whole-cohort sensitivity or specificity. The reported Manjush specificity is inconsistent with its denominator and is shown without an interval. No pooled estimate is displayed.

Predictive values across pre-test probabilities and decision curves

Because sensitivity and specificity were both moderate or poor, predictive values depended almost entirely on the pre-test probability (Figure 9). At ≥12, the negative predictive value rose from 14.3% at the observed prevalence to 65.4% at a pre-test probability of 50% and 89.3% at 18.4% (95% credible band 68.2–94.2%), whereas the positive predictive value fell to 30.4% (95% credible band 14.3–78.9%) at 18.4%. The posterior probability that specificity at ≥12 exceeded 80% was 25.1%, and the posterior probability that sensitivity at ≥10 exceeded 90% was 7.7%. In the observed cohort, imaging or exploring every patient yielded a higher net benefit than either score-guided

strategy at all threshold probabilities below about 85%, and neither BATSS strategy exceeded the net benefit of imaging all within the clinically plausible range (Figure 10A). When the observed sensitivity and specificity were transported to the 18.4% pre-test probability of the derivation cohort, the point-estimate curves of the score-guided strategies lay slightly above both default strategies only within a narrow band of low threshold probabilities (Figure 10B); with specificity estimated from three patients, this pattern is illustrative only. Under alternative priors, the posterior probability that specificity at ≥12 exceeded 80% ranged from 18.1% (uniform prior) to 44.2% (informative prior centred on 80%), and at ≥10 it did not exceed 20.5%.

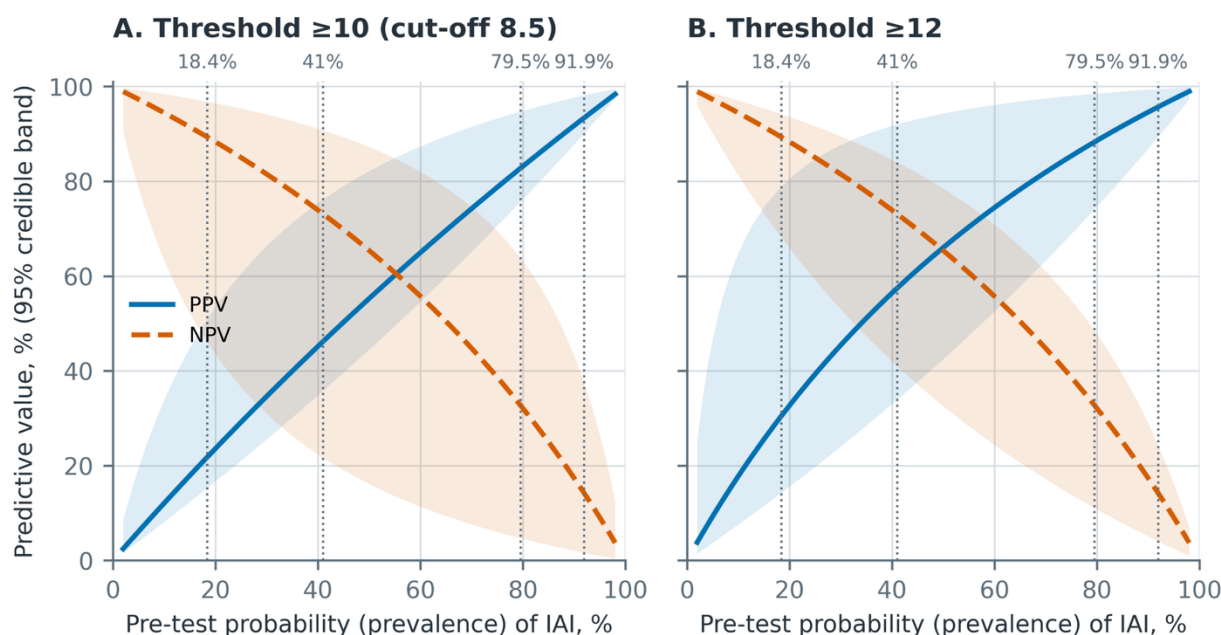


Figure 9. Positive (PPV) and negative (NPV) predictive values of BATSS as a function of pre-test probability, at the thresholds ≥ 10 (A) and ≥ 12 (B). Lines use the observed sensitivity and specificity; shaded bands are 95% credible intervals from Jeffreys posteriors (20,000 draws). Dotted lines mark the prevalence of injury in the score's derivation cohort (18.4%), the Chennai (41%) and Bali (79.5%) validation cohorts and the present cohort (91.9%).

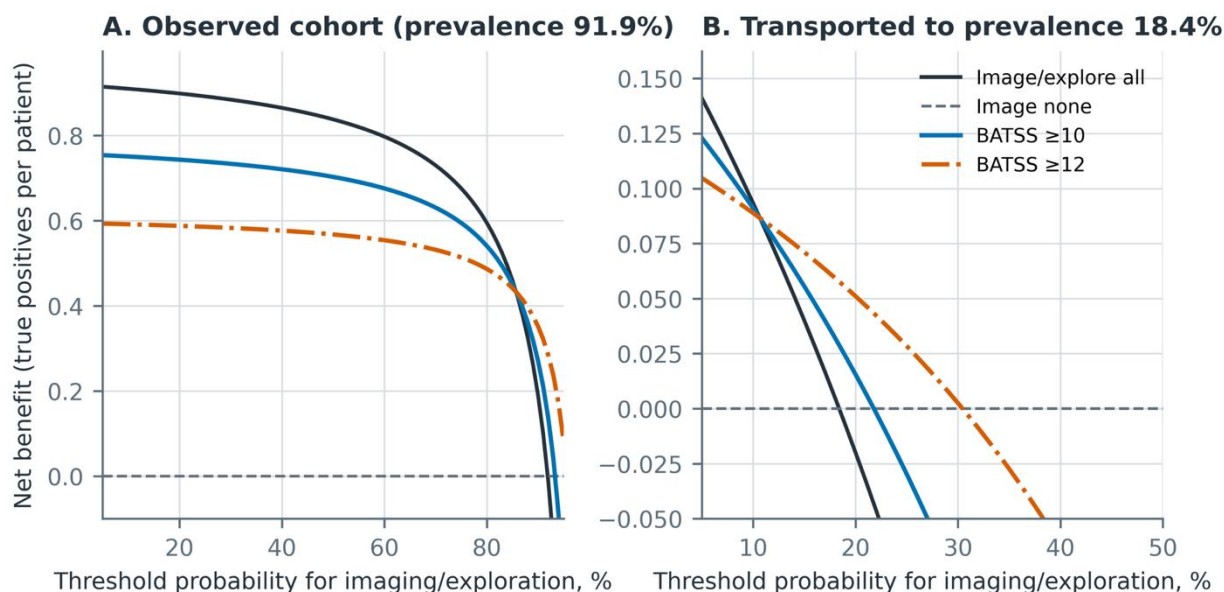


Figure 10. Decision curves for imaging or exploring patients according to BATSS. Net benefit is plotted against the threshold probability at which a clinician would image or explore. (A) The observed cohort (prevalence 91.9%). (B) The same sensitivity and specificity transported to an 18.4% pre-test probability, the prevalence in the score's derivation cohort of unselected emergency-department patients; panel B uses point estimates only and specificity rests on three patients.

Sensitivity analysis over datasets consistent with the observed marginals

Associations of sex, age group and mechanism with IAI could not be identified from the data (Table 9). Across the four tables consistent with the sex and IAI margins, the OR ranged from 0.45 to 37.5, and only the single table that placed all three uninjured patients among the eight women reached $p < 0.05$; for mechanism the OR ranged from 0.97 to 156 (two of four tables significant), and for age group Cramér's V ranged from 0.17 to 0.68 (four of ten tables significant). Whether the score was more

sensitive for solid-organ or hollow-viscus injury was also not identifiable: sensitivity at ≥ 10 could lie anywhere between 62.5% and 100% for solid-organ injury and between 60.0% and 100% for hollow-viscus injury. In contrast, the absence of a score-IAI association was robust to adjustment: the adjusted OR per point ranged from 0.51 to 1.25 across all 62,029 sex allocations (significant in 0.02%), from 0.56 to 1.21 across all 1,806 mechanism allocations (0.33%) and from 0.55 to 1.39 for age group (no sampled allocation significant). All 6 non-converged sex models were refitted by direct maximisation and retained.

Table 9. Sensitivity analysis over all datasets consistent with the observed marginal distributions.

Association	Admissible tables/datasets	Attainable range of effect	Fisher/LR p range	Tables/datasets with p < 0.05	Interpretation
Sex × IAI	4	OR 0.45–37.5	0.007–1.000	1 of 4	Not identifiable
Age group (3 levels) × IAI	10	Cramér's V 0.17–0.68	0.005–1.000	4 of 10	Not identifiable
Mechanism (road traffic crash vs other) × IAI	4	OR 0.97–156	<0.001–1.000	2 of 4	Not identifiable
Injury type × BATSS ≥10 (sensitivity, injured only)	22	Sn solid 62.5–100.0%; hollow 60.0–100.0%	<0.001–1.000	10 of 22	Not identifiable
BATSS per point, adjusted for sex (Firth)	62,029	aOR 0.51–1.25 (median 0.99)	—	0.02%	Null-compatible in >99.9%
BATSS per point, adjusted for mechanism (Firth)	1,806	aOR 0.56–1.21 (median 1.00)	—	0.33%	Null-compatible in >99%
BATSS per point, adjusted for age group (Firth)	4,000 sampled + hill-climbing	aOR 0.55–1.39 (median 0.99)	—	0.0%	Null-compatible

Sex, age group, mechanism and injury type were recorded only as marginal frequencies, so their joint distribution with injury status (or with the score) is unknown. Every contingency table (or, for multivariable models, every patient-level allocation; age group: 4,000 random allocations plus hill-climbing towards the extremes, so the range is a lower bound on the attainable range) consistent with the observed margins was analysed. Unadjusted Firth OR per BATSS point 0.997 (95% profile CI 0.75–1.24; $p = 0.977$). aOR, adjusted odds ratio; IAI, intra-abdominal injury; OR, odds ratio; Sn, sensitivity.

Precision, power and sample size

With three uninjured patients, specificity could take only four values (0%, 33.3%, 66.7% or 100%), and the 95% CI at ≥ 12 spanned 9.4–99.2% (Figure 11A). The design had 15% power to reject an AUC of 0.50 if the true AUC were 0.70, 34% if it were

0.80 and 80% if it were 0.90 (Figure 11B). Estimating a specificity of 80% to within ± 10 percentage points would require 62 patients without IAI — 766 patients at the prevalence observed here, but 76 at the 18.4% prevalence of the derivation cohort (Table 10).

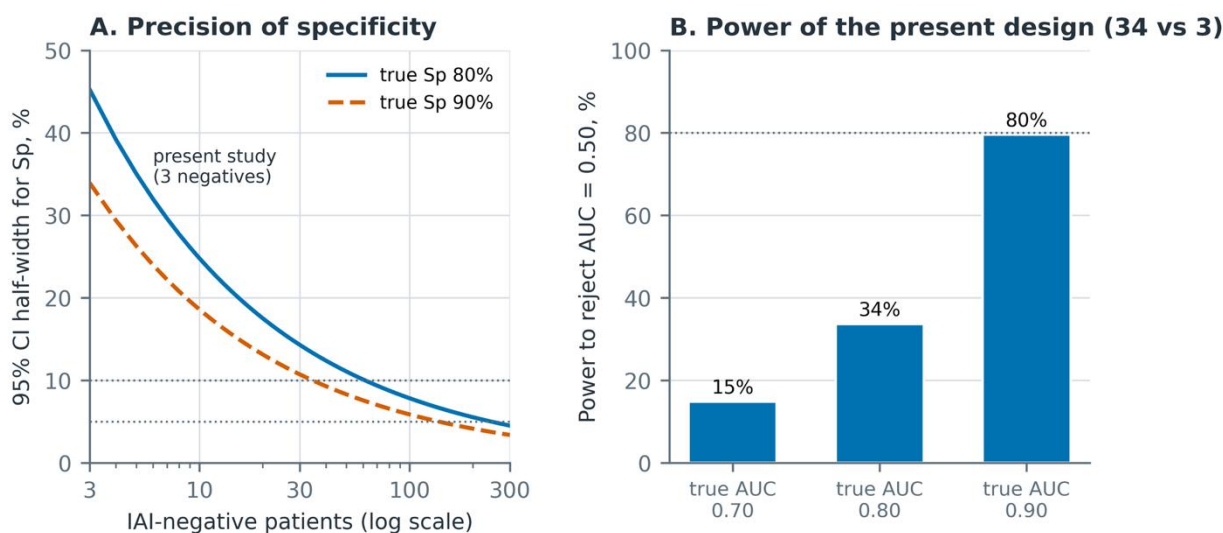


Figure 11. Precision and power. (A) Half-width of the 95% confidence interval for specificity as a function of the number of patients without intra-abdominal injury, for true specificities of 80% and 90%; the present study had three. (B) Power of the present design (34 with and 3 without injury, $\alpha = 0.05$, Hanley–McNeil variance) to reject an AUC of 0.50 if the true AUC were 0.70, 0.80 or 0.90.

Table 10. Sample sizes needed for a definitive validation of BATSS at different pre-test probabilities.

Target precision	Patients needed in the relevant group	Total at prevalence 91.9% (imaged/operated spectrum)	Total at prevalence 50%	Total at prevalence 18.4% (derivation cohort)
Specificity 80% $\pm$ 10 points	62 without IAI	766	124	76
Specificity 80% $\pm$ 5 points	246 without IAI	3,038	492	302
Specificity 90% $\pm$ 10 points	35 without IAI	433	70	43
Specificity 90% $\pm$ 5 points	139 without IAI	1,717	278	171
Sensitivity 90% $\pm$ 10 points	35 with IAI	39	70	191
Sensitivity 90% $\pm$ 5 points	139 with IAI	152	278	756
Sensitivity 95% $\pm$ 10 points	19 with IAI	21	38	104
Sensitivity 95% $\pm$ 5 points	73 with IAI	80	146	397
AUC 0.80, 95% CI half-width ≤ 0.10	—	160 (13 without IAI)	78	161

Normal-approximation sample size $n = z^2 p(1-p)/d^2$ for the relevant group, divided by the expected proportion of that group; AUC precision from the Hanley–McNeil variance. The present study (37 patients; a priori target 41) had 3 patients without injury, giving a power of 34% to reject an AUC of 0.50 if the true AUC were 0.80. IAI, intra-abdominal injury.

Influence of individual patients and partial-verification scenarios

Removing a single uninjured patient changed the AUC from 0.525 to between 0.375 (removing the patient who scored 7) and 0.757 (removing the patient who scored 19), and specificity at ≥ 12 from 66.7% to 50.0% or 100% (Table 11). Removing any single injured patient changed the AUC only within 0.510–0.540.

In the partial-verification scenarios, adding 5 hypothetical uninjured patients with a score of 2 — the kind of patient who is typically observed without imaging — would raise the AUC to at least 0.80, whereas 16 patients scoring 4 or 36 scoring 7 would be required; however many such patients were added, the AUC could not exceed 0.85 if they scored 4 or 0.82 if they scored 7, because six injured patients scored 4 or less.

Table 11. Influence of individual patients, tied scores, partial-verification scenarios and prior choice.

Analysis	AUC	Specificity at ≥ 12 , %	Specificity at ≥ 10 , %
Leave-one-out (uninjured patients)			
Without the uninjured patient scoring 19	0.757	100.0	50.0
Without the uninjured patient scoring 11	0.441	50.0	50.0
Without the uninjured patient scoring 7	0.375	50.0	0.0
Without any one injured patient (range)	0.510–0.540	66.7	33.3
Tied scores	0.500–0.549	—	—
Hypothetical unverified uninjured patients added			
Score 2: number needed for AUC ≥ 0.70 / ≥ 0.80	2 / 5	—	—
Score 2: AUC with 20 added; upper limit	0.912; 0.971	95.7	—
Score 4: number needed for AUC ≥ 0.70 / ≥ 0.80	4 / 16	—	—
Score 4: AUC with 20 added; upper limit	0.810; 0.853	95.7	—
Score 7: number needed for AUC ≥ 0.70 / ≥ 0.80	5 / 36	—	—
Score 7: AUC with 20 added; upper limit	0.785; 0.824	95.7	—
Prior sensitivity: P(specificity $\geq 80\%$)		at ≥ 12	at ≥ 10
Jeffreys Beta(0.5, 0.5) (primary)	—	25.1	3.4
Uniform Beta(1, 1)	—	18.1	2.7
Near-Haldane Beta(0.01, 0.01)	—	35.7	4.0
Informative Beta(8, 2), centred on 80%	—	44.2	20.5

Observed AUC 0.525; specificity 66.7% at ≥ 12 and 33.3% at ≥ 10 . Hypothetical patients represent uninjured adults who would not have been imaged or operated on and are not data; they quantify how strongly partial verification could have depressed the AUC. The upper limit is the AUC approached as the number of added patients grows without bound. AUC, area under the receiver operating characteristic curve.

4. Discussion

Principal findings

In this cohort of adults with blunt abdominal trauma who underwent CT or laparotomy at a provincial Indonesian trauma centre, BATSS did not discriminate between patients with and without IAI within this spectrum. The AUC was 0.525, the scores of injured and uninjured patients overlapped almost completely, the protocol threshold of ≥ 12 achieved a sensitivity of 64.7% and a specificity of 66.7%, and none of the stratum-specific likelihood ratios differed from 1. Six of 34 injured patients scored 4 or less, below every uninjured patient, whereas the uninjured patient with the highest score (19) ranked above 31 of the 34 injured patients. These findings are not an artefact of a single threshold: no cut-off yielded a likelihood ratio whose confidence interval excluded 1, and the absence of association persisted after adjustment for sex, age group or mechanism across every dataset consistent with the recorded margins.

Comparison with published validations

The present estimates are lower than several earlier single-centre reports. The derivation study reported perfect separation between its extreme bands in an emergency-department cohort with an IAI prevalence of 18.4%³². Chennai reported sensitivity of 100%, specificity of 98.3% and AUC of 0.992³⁵, while Kozhikode reported sensitivity of 100%⁴⁰. At ≥ 12 , sensitivity in Palembang was approximately 27 percentage points lower than in the Bali validation³⁴, with Holm-adjusted $p = 0.013$, whereas specificity could not be distinguished from the Bali estimate. Some favourable reports contain internal arithmetical inconsistencies: specificity of 97.3% is not attainable with 15 uninjured patients⁴⁰, and a positive predictive value of 100% is incompatible with specificity of 87.1%³⁹. Outcome definitions also vary between any IAI, need for laparotomy and need for surgical intervention^{36,41}. These differences and reporting limitations constrain transportability of the published estimates. The selected earlier studies do not provide a pooled benchmark or formal publication-bias assessment; the present study should therefore be interpreted directly against the individual pre-cutoff studies and its own substantial uncertainty.

The contrast with the only other Indonesian validation deserves particular attention. In Bali, 35 of 44 patients (79.5%) had organ injury; at ≥ 12 the score identified 32 of the 35 injured patients and correctly classified seven of the nine uninjured ones³⁴. Three features may explain the higher sensitivity there.

A lower prevalence implies a less filtered population in which injured patients more often presented with major signs; nine uninjured patients give a more stable specificity than three; and the Bali analysis classified patients by organ rupture, which favours injuries accompanied by haemorrhage, whereas the present target condition included hollow-viscus and mesenteric injuries that rarely raise the score. Neither study published item-level data, so these explanations remain hypotheses that a pooled analysis of Indonesian centres could test.

The target condition also differs across studies. Here any IAI on CT or at laparotomy counted as positive, so a minor solid-organ contusion managed without intervention and a mesenteric avulsion requiring resection carried equal weight. The derivation study used CT-positive injury³², whereas several validations used the need for laparotomy or surgical intervention^{36,37,41,42}. If BATSS identifies injuries that need intervention better than injuries in general, its accuracy for clinically important IAI could be higher than estimated here; because injury grades and interventions were not linked to the score, this question remains open.

The pattern of a high positive predictive value with a collapsed negative predictive value is not unique to BATSS. In a Balinese cohort in which 75 of 80 laparotomies were positive, CASS above 12 had a sensitivity of 80%, a specificity of 60%, a positive predictive value of 96.7% and a negative predictive value of 16.6%⁵¹ — values almost identical to those of BATSS at ≥ 10 here. Serial FAST in an Indian cohort retained positive predictive values above 93% while negative predictive values ranged only from 17.6% to 46.2%¹⁴. Where nearly every patient tested is injured, any test will appear to 'rule in' well and 'rule out' badly; the informative quantities are the likelihood ratios, and for BATSS in this spectrum they were close to 1.

Why the score did not discriminate: spectrum and verification

BATSS was designed for the undifferentiated emergency-department population, in which most patients are uninjured and the score's task is to identify the minority who need imaging or operation³². The present cohort is a different population — the residue of clinical triage — and the findings describe how the score behaves after clinicians have already acted on its components. They do not show that the score lacks discrimination in unselected patients; they show that, once a surgeon has decided to image or operate, the score adds little.

The most plausible explanation for the gap between this cohort and the derivation study is the spectrum of patients. Only patients who underwent CT or laparotomy could be included, and in this hospital the decision to image or operate was made by surgeons who had already integrated the same haemodynamic, examination and sonographic findings that BATSS sums. The resulting cohort differed sharply from the derivation group's emergency-department population: 62.2% were in the high-risk band rather than 23.0%, the mean score was 11.84 rather than 6.29, and the prevalence of injury was 91.9% rather than 18.4%^{32,33}. Within such a filtered population the low-scoring uninjured patients who give a score its specificity — the patients with abdominal pain but normal vital signs and a negative FAST who are observed and discharged — are largely absent. Partial verification bias of exactly this kind is expected when a finding that influences referral for the reference standard is then evaluated only in verified patients⁵⁰, and meta-epidemiological work confirms that sensitivity and specificity move in opposite directions as prevalence rises^{48,49}.

The scenarios in Table 11 make the magnitude of this bias concrete. Had as few as five uninjured patients with a score of 2 been verified — patients who would normally be observed without imaging — the AUC would have exceeded 0.80, and omitting the single uninjured patient who scored 19 would have raised it to 0.76. The low AUC is therefore compatible with a score that discriminates reasonably in unselected patients but cannot do so among those already selected. It also illustrates how fragile small validation studies are: every published BATSS validation with fewer than 20 uninjured patients is exposed to the same instability, in either direction.

A second contributor is incorporation. FAST, hypotension and pelvic fracture are both components of the score and triggers for laparotomy or CT, and positive FAST findings in patients with pelvic fractures arise from retroperitoneal blood that does not represent IAI^{16–18}. The uninjured patient who scored 19 illustrates the consequence: that total cannot be reached without a positive FAST, and together with hypotension or a pelvic fracture it would have led any surgeon to investigate, yet no IAI was found. Alternative explanations cannot be excluded — an injury missed by the reference standard (one injury was identified during observation among 480 stable patients with a negative initial CT in an earlier cohort²²), a documentation or scoring error, or free fluid from a non-traumatic source — but whichever applies, the case shows how one patient can dominate the specificity of a small study. More generally, 30 of the 37 patients (81.1%) had totals that obligatorily included at least one major sign, so the score and the decision to verify drew on largely the same information, which inflates apparent sensitivity and deflates apparent specificity. Differential verification may have acted in either direction: laparotomy is usually prompted by major signs and detects hollow-viscus injuries that CT can miss, whereas low scorers verified by CT alone may harbour injuries that laparotomy would have shown; without the per-patient route, the net direction cannot be determined. Pelvic fracture is common in high-energy road trauma⁵⁶, yet fewer than one in five patients with pelvic fracture have abdominal co-injuries⁵⁷; the five points it adds may therefore raise the score of uninjured patients as readily as of injured ones.

The arithmetic of the score

Because BATSS is a sum of fixed integer weights, its totals carry information about which components could have been present even when the components themselves were not recorded. The four soft signs — abdominal pain, tenderness, a lower chest-wall sign and tachycardia — contribute at most 7 points, so a patient

cannot leave the low-risk band on symptoms and examination alone; every total of 8 or more requires at least one of hypotension, pelvic fracture or a positive FAST, and every total of 17 or more requires a positive FAST. In practice, therefore, the moderate and high bands are largely defined by the three major signs, and the low band is the band of patients without them. A worked example makes the logic concrete. A haemodynamically normal patient with abdominal pain, tenderness, a lower chest-wall bruise and a heart rate of 110 beats/min scores $2 + 3 + 1 + 1 = 7$ and remains in the low band however impressive the examination; adding a positive FAST lifts the same patient to 15, whereas adding hypotension instead lifts them only to 11. Conversely, a total of 19 can arise only from a positive FAST (8 points) plus at least 11 of the remaining 16 points, which requires hypotension or a pelvic fracture in addition to most of the soft signs. The same reasoning applies to any additive clinical score with integer weights, and it can be used to check reported component frequencies against reported totals. This structure explains both failure modes observed here. Injured patients without haemodynamic compromise, pelvic fracture or haemoperitoneum — exactly the patients with contained solid-organ, retroperitoneal or early hollow-viscus injuries — are confined by arithmetic to the low band however tender their abdomen, and the six injured patients who scored 2–4 cannot have had a positive FAST or a pelvic fracture. Conversely, uninjured patients with a false-positive FAST or a pelvic fracture are lifted into the higher bands. Abdominal pain and tenderness can shift every total upwards by up to 5 points, but clinical examination alone has limited diagnostic value and the absence of tenderness does not exclude IAI¹¹.

This analysis replaces the usual component-by-component comparison, which requires item-level data. It is deliberately conservative: the bounds in Table 7 show what the totals make certain, not what is likely, and for the soft signs they are uninformative. Its value lies in identifying which misclassifications are structural. No reweighting of the soft signs could move the six low-scoring injured patients into the high band without also moving uninjured patients with similar findings, and no threshold could separate the uninjured patient with a total of 19 from the injured patients with totals of 12–15. A score that is dominated by FAST and pelvic fracture will inherit their limitations in the injuries they miss and the false positives they generate.

Injuries the score cannot see

More than half of the injured patients in this cohort had at least one hollow-viscus injury, and the colon was involved as often as the liver. This distribution differs from earlier South Asian cohorts in which solid-organ injuries predominated^{6,7}. Hollow viscus and mesenteric injuries may produce little haemoperitoneum early, and their diagnosis can be delayed despite initial FAST or CT. CT-based scores for bowel injury have reported sensitivities of only 39–57%⁵⁸, while approaches that incorporate clinical findings are limited when patients cannot be reliably examined⁵⁹. Clinical-radiological approaches combining features such as the seat-belt sign, tenderness, guarding and CT findings have reported AUCs of 0.87–0.96^{21,60}. The Bowel Injury Prediction Score combines mesenteric CT injury grade, leucocyte count and abdominal tenderness to identify patients at high risk of bowel injury⁶¹. BATSS includes tenderness, but does not incorporate the seat-belt sign, guarding, CT injury grades or leucocyte count as separate items³². Whether the score performed differently for solid-organ and hollow-viscus injuries could not be determined here, because injury type was not linked to the score; sensitivity at ≥ 10 could have been anywhere between 60% and 100% for

either group. The high share of hollow-viscus injuries may itself reflect referral. Patients bleeding from solid organs may be operated on locally or die before transfer, whereas bowel injuries may declare themselves more slowly and reach the provincial centre after the delays described in Indonesian referral pathways^{5,31}. These referral mechanisms remain hypotheses because injury-specific transfer data were not available.

Solid-organ injuries that are contained by their capsule may also be missed by the score at presentation. Such injuries are often managed non-operatively, with a success rate of 97% in a Malaysian series⁶², but non-operative management depends on accurate CT staging and close monitoring, transfusion requirement and liver injury predict its failure, and pancreatic injuries are frequently under-staged. A score whose low band contains injured patients is therefore not merely inaccurate; if used to withhold CT, it would deprive precisely these patients of the imaging on which non-operative care depends.

Implications for practice in Indonesian trauma centres

For patients who are already being considered for CT or laparotomy, these data do not support using a low BATSS to withhold imaging, and they suggest that a high score adds little to the clinical findings that generated it. Applied to the 18.4% prevalence of the derivation cohort, the point likelihood ratios observed here would move the probability of IAI only to about 11% in the low band and 30% in the high band — too little to change management. The decision curves point in the same direction: at the prevalence seen in referred patients, imaging everyone produced more net benefit than any score-guided strategy at every plausible threshold probability. In a resource-constrained system the more useful question may be how to reduce unnecessary CT in low-risk patients before they enter this filtered spectrum; earlier selective-imaging research and an August 2025 preprint suggest potential for this approach^{26,27}, but they require prospective evaluation in Indonesian emergency departments.

Pending prospective evidence, the practical implication for an Indonesian trauma centre is to anchor diagnostic decisions in established pathways rather than in the score. Haemodynamically unstable patients with a positive FAST proceed to operation; stable patients with abdominal findings, a seat-belt sign, a distracting injury or a high-energy mechanism undergo contrast-enhanced CT when it is available; and stable patients in whom CT is unavailable or deferred receive serial clinical examination and repeated FAST, which improves sensitivity over a single scan^{14,20,59}. BATSS may still serve as a concise way to document and communicate the presence of major signs, but it should not replace these steps.

The case mix of this cohort — young men injured in road traffic crashes, often referred from district hospitals, with a high proportion of hollow-viscus injuries — is typical of Indonesian provincial trauma practice. Referral delays are common in Indonesia, as shown by a median of 12.5 hours to reach a Jakarta trauma centre after pelvic fracture³¹ and more than four hours in over half of motorcycle-related head injuries in North Maluku⁵, and prehospital systems are fragmented³⁰. By the time patients reach a provincial centre, physiology has had time to declare itself, and the score is effectively applied to patients who have already been triaged by distance and time. A triage score intended for use at first contact should therefore be validated where it would be used — in district emergency departments, before referral — rather than in the residual population that reaches a tertiary centre. For South Sumatra, this implies a

provincial network study rather than another single-centre series.

Research agenda

Three studies would resolve the uncertainties identified here. First, a prospective multicentre validation in Indonesian emergency departments should enrol consecutive adults with blunt abdominal trauma regardless of whether they undergo CT, record each BATSS component at first assessment, and verify the outcome by CT, laparotomy or structured follow-up for patients who are not imaged. Estimating a specificity of 80% to within ± 10 percentage points would require about 62 patients without IAI; at the 18.4% prevalence of the derivation cohort this corresponds to roughly 76 patients, whereas an AUC of 0.80 with a confidence-interval half-width of 0.10 would require about 161 patients (Table 10); formal external validation of calibration would require several hundred events^{54,55}. Second, the same cohort should compare BATSS head-to-head with clinician gestalt, serial FAST, CASS and candidate laboratory markers such as lactate and transaminases, informed by earlier studies of clinical assessment and diagnostic predictors^{8,9,14,44}, and should report net benefit rather than accuracy alone⁵³. Third, any redesign of the score — for example adding the seat-belt sign and abdominal wall findings^{20,21} or reconsidering the weight given to pelvic fracture — should be treated as a hypothesis for a new derivation study with item-level data and then validated externally, so that the arithmetic constraints described here can be tested directly.

Strengths and limitations

The study has several strengths. Every patient's total score and reference-standard result were available, so all accuracy measures, including the full ROC curve, every threshold and the stratum-specific likelihood ratios, were computed from observed data with exact methods suited to small samples. The analysis was reported to STARD 2015⁵², compared against selected pre-cutoff validations with usable published counts, and complemented by decision curves, Bayesian summaries and precision analysis that make the limits of the evidence explicit. Where data were missing, the analysis used assumption-free bounds over all datasets consistent with what was recorded rather than imputation, and the arithmetic of the score was used to extract information about components that were not recorded.

The limitations are substantial and largely shared with the published literature. Only three patients had no IAI, so specificity could take only four values and its confidence intervals spanned most of the possible range; the study had 34% power to detect a true AUC of 0.80. The cohort was restricted to patients who underwent CT or laparotomy, and the number of patients with blunt abdominal trauma who were not investigated was not recorded, so partial verification bias could not be corrected and the results apply only to this filtered spectrum. Verification differed between patients (CT or laparotomy), the route was not recorded per patient, and the abstractor was not blinded to the reference standard, which could have biased the scoring of subjective items towards agreement with the outcome and would tend to overstate rather than understate accuracy. The interval between the index test and the reference standard was not recorded, and the target condition was any IAI, so minor injuries that required no intervention counted as positives. Scoring was retrospective from emergency-department documentation by a single investigator, without an assessment of inter-rater reliability, and FAST results were taken from routine documentation without a record of the operator or protocol. Item-level BATSS components, injury grades and the need for therapeutic

intervention were not retained in the analysis dataset, which prevented component-level and outcome-specific analyses; sex, age group and mechanism were available only as marginal frequencies. Finally, this is a single-centre study at a provincial referral hospital, and its case mix may not represent district emergency departments.

5. Conclusion

Among adults with blunt abdominal trauma who had already been selected for CT or laparotomy at an Indonesian provincial trauma centre, BATSS did not discriminate between patients with and without intra-abdominal injury (AUC 0.525; 95% CI 0.053–0.996), and none of its original risk bands shifted the probability of injury. A low score did not exclude injury: one in six injured patients scored 4 or less. The arithmetic of the score — the soft signs cannot exceed 7 points, and the highest totals require a positive FAST — explains why patients without haemodynamic compromise, pelvic fracture or haemoperitoneum are under-scored and why false-positive major signs over-score uninjured patients. The estimates are fragile, because only three patients were uninjured, and they apply to the verified, high-prevalence spectrum rather than to unselected emergency-department patients. Within that spectrum, BATSS should not be used to withhold CT from patients whom clinicians are already considering for imaging. Before the score is adopted for triage in Indonesia, it should be validated prospectively across referring and receiving hospitals, including patients who are not imaged, with item-level data and enough uninjured patients to estimate specificity precisely.

6. Declarations

Ethics Approval and Consent to Participate

The study was conducted in accordance with the Declaration of Helsinki. The complete research protocol was submitted to the institutional review board, the Ethics Committee of Dr. Mohammad Hoesin General Hospital Palembang, and was granted a formal ethical exemption (No. DP.04.03/D.XVIII.06.08/ETIK/202/2025). As the study relied exclusively on the retrospective review of de-identified data from existing medical records, the committee waived the need for individual patient informed consent.

Consent for Publication

Not applicable. The manuscript contains no individual patient data, images or other information from which a patient could be identified.

Availability of Data and Materials

The de-identified score-level dataset (BATSS total and reference-standard status for each patient) and the summary data are fully reported in Table 4. Further de-identified data are available from the corresponding author on reasonable request, subject to approval by the hospital's ethics committee.

Artificial Intelligence Use Statement

None declared.

Author Contribution Statement

ARW: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft. AA: Conceptualization, Supervision, Validation, Writing – review & editing. T: Methodology, Formal analysis, Supervision, Writing – review & editing. All authors approved final version of manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of Interest

The authors declare no competing interests.

Acknowledgements

The authors thank the staff of the medical records unit and the emergency department of Dr. Mohammad Hoesin General Hospital, Palembang, for their assistance with data retrieval.

7. References

- GBD 2021 ASEAN Injuries Collaborators. The epidemiology and burden of injury in countries of the Association of Southeast Asian Nations (ASEAN), 1990-2021: findings from the Global Burden of Disease Study 2021. *Lancet Public Health*. 2025;10(6):e456-e466. doi: 10.1016/S2468-2667(25)00069-6
- Xu Y, Chen M, Yang R, et al. Global, Regional, and National Burden of Road Injuries from 1990 to 2019. *Int J Environ Res Public Health*. 2022;19(24):16479. doi: 10.3390/ijerph192416479
- Zainafree I, Syukria N, Addina S, et al. Risk factors of road traffic accidents in Rural and Urban areas of Indonesia based on the national survey of year 2018. *Niger Postgrad Med J*. 2022;29(2):82-8. doi: 10.4103/npmj.npmj_777_21
- Umnayatun Y, Nurmansyah MI, Farradika Y, et al. Motorcycle risky behaviours and road accidents among adolescents in Jakarta metropolitan area, Indonesia. *Int J Inj Contr Saf Promot*. 2021;28(3):339-46. doi: 10.1080/17457300.2021.1928229
- Nugroho AW, Masrika NUE. The epidemiology of motorcycle-related acute traumatic brain injury: The NOMADEN study. *Surg Neurol Int*. 2025;16:27. doi: 10.25259/SNI.929_2024
- Naeem BK, Iqbal M, Perveen S, et al. Blunt abdominal trauma patients presenting to a tertiary care facility of Pakistan: A three years' experience. *J Pak Med Assoc*. 2022;72(10):2093-6. doi: 10.47391/JPMA.3825
- Singh A, Prasad G, Mishra P, et al. Lessons learned from blunt trauma abdomen: Surgical experience in level I trauma centre. *Turk J Surg*. 2021;37(3):277-85. doi: 10.47717/turksurg.2021.4886
- Rodriguez NM, Mower WR, Raja AS, et al. Accuracy of physician gestalt in prediction of significant abdominal and pelvic injury in adult blunt trauma patients. *Acad Emerg Med*. 2023;30(10):1039-46. doi: 10.1111/acem.14768
- Fabig S, Weigert N, Migliorini F, et al. Predictive parameters for early detection of clinically relevant abdominal trauma in multiple-injury or polytraumatised patients: a retrospective analysis. *Eur J Med Res*. 2024;29(1):394. doi: 10.1186/s40001-024-01969-3
- Kania T, Pandya S, Demissie S, et al. Physical exam is not an accurate predictor of injury in geriatric patients with low-energy blunt trauma - A retrospective cohort study. *Ann Med Surg (Lond)*. 2022;81:104503. doi: 10.1016/j.amsu.2022.104503
- Nishijima DK, Simel DL, Wisner DH, et al. Does this adult patient have a blunt intra-abdominal injury? *JAMA*. 2012;307(14):1517-27. doi: 10.1001/jama.2012.422
- Lin KT, Lin ZY, Huang CC, et al. Prehospital ultrasound scanning for abdominal free fluid detection in trauma patients: a systematic review and meta-analysis. *BMC Emerg Med*. 2024;24(1):7. doi: 10.1186/s12873-023-00919-2
- Gamberini L, Scquizzato T, Tartaglione M, et al. Diagnostic accuracy for hemoperitoneum, influence on prehospital times and time-to-definitive treatment of prehospital FAST: A systematic review and individual participant data meta-analysis. *Injury*. 2023;54(6):1421-31. doi: 10.1016/j.injury.2023.03.024
- Kakran A, Singh R, Chaudhary P, et al. Sensitivity and Specificity of Serial Focused Assessment With Sonography in Trauma (FAST) in Patients With Blunt Abdominal Trauma. *Cureus*. 2025;17(9):e92103. doi: 10.7759/cureus.92103
- Natarajan B, Gupta PK, Cemaj S, et al. FAST scan: is it worth doing in hemodynamically stable blunt trauma patients? *Surgery*. 2010;148(4):695-700. doi: 10.1016/j.surg.2010.07.032

16. Chaijareenont C, Krutsri C, Sumpritpradit P, et al. FAST accuracy in major pelvic fractures for decision-making of abdominal exploration: Systematic review and meta-analysis. *Ann Med Surg (Lond)*. 2020;60:175-81. doi: 10.1016/j.amsu.2020.10.018
17. Jensen LR, Possfelt-Møller E, Nielsen AE, et al. Accuracy of FAST in detecting intraabdominal bleeding in major trauma with pelvic and/or acetabular fractures: a retrospective cohort study. *Eur J Orthop Surg Traumatol*. 2024;34(3):1479-86. doi: 10.1007/s00590-023-03813-6
18. Kumar P, Mishra SS, Jain M, et al. Limitations of Focused Assessment with Sonography for Trauma in Pelvic Trauma: A Retrospective Review of False-Positive Results and Non-Therapeutic Laparotomies. *J Orthop Case Rep*. 2025;15(4):245-51. doi: 10.13107/jocr.2025.v15.i04.5512
19. Wadhwa M, Kumar R, Trehan M, et al. Blunt Abdominal Trauma With Hollow Viscus and Mesenteric Injury: A Prospective Study of 50 Cases. *Cureus*. 2021;13(2):e13321. doi: 10.7759/cureus.13321
20. Pourmand A, Terrebonne E, Shapovalov V, et al. Prevalence of intra-abdominal injury among patients with seatbelt signs, a systematic review and meta-analysis. *Am J Emerg Med*. 2024;76:199-206. doi: 10.1016/j.ajem.2023.11.058
21. Delaplain PT, Tay-Lasso E, Biffi WL, et al. Excluding Hollow Viscus Injury for Abdominal Seat Belt Sign Using Computed Tomography. *JAMA Surg*. 2022;157(9):771-8. doi: 10.1001/jamasurg.2022.2770
22. Zarama V, Torres N, Duque E, et al. Incidence of intra-abdominal injuries in hemodynamically stable blunt trauma patients with a normal computed tomography scan admitted to the emergency department. *BMC Emerg Med*. 2024;24(1):103. doi: 10.1186/s12873-024-01014-w
23. Fathi M, Mirjafari A, Yaghoobpoor S, et al. Diagnostic utility of whole-body computed tomography/pan-scan in trauma: a systematic review and meta-analysis study. *Emerg Radiol*. 2024;31(2):251-68. doi: 10.1007/s10140-024-02213-5
24. Kerslake M, Chui Y, Pham TP, et al. Rethinking the pan scan in stable trauma: A comparison of whole-body computed tomography and selective imaging in clinically stable blunt force trauma. *Emerg Med Australas*. 2025;37(1):e14552. doi: 10.1111/1742-6723.14552
25. Bunn C, Ringhouse B, Patel P, et al. Trends in utilization of whole-body computed tomography in blunt trauma after MVC: Analysis of the Trauma Quality Improvement Program database. *J Trauma Acute Care Surg*. 2021;90(6):951-8. doi: 10.1097/TA.0000000000003129
26. Wasfie T, Rivera D, Naisan M, et al. The Value of Abdominal and Pelvic CT Scan in Trauma Patients With Low Injury Severity Score and High Glasgow Coma Scale. *Am Surg*. 2023;89(4):821-4. doi: 10.1177/00031348211047496
27. Shopen N, Lahav Z, Kanani F, et al. A Clinical Decision Instrument to Safely Reduce Abdominopelvic CT Use in Minor Blunt Trauma Patients [preprint]. Research Square. Posted 4 August 2025. doi: 10.21203/rs.3.rs-7156788/v1
28. Abdel Hamid MA, Abd-erRazik MA, Nagy M, et al. Computed tomography benefits and cost in hemodynamically stable patients with blunt abdominal trauma at an Egyptian University Hospital. *Afr J Emerg Med*. 2024;14(2):96-9. doi: 10.1016/j.afjem.2023.11.006
29. Atkins K, Schneider A, Charles A. Negative laparotomy rates and outcomes following blunt traumatic injury in the United States. *Injury*. 2023;54(8):110894. doi: 10.1016/j.injury.2023.110894
30. Tokita A, Nunokawa H, Liu K, et al. Key challenges in prehospital and emergency care in Indonesia and Malaysia: a survey of frontline clinicians. *BMC Res Notes*. 2024;17(1):287. doi: 10.1186/s13104-024-06916-3
31. Dilogio IH, Gunawan B, Oesman I, et al. Management of pelvic ring fractures in limited resources country: a retrospective study. *Eur J Orthop Surg Traumatol*. 2023;33(3):465-72. doi: 10.1007/s00590-022-03457-y
32. Shojae M, Faridaalae G, Yousefifard M, et al. New scoring system for intra-abdominal injury diagnosis after blunt trauma. *Chin J Traumatol*. 2014;17(1):19-24. PMID: 24506918
33. Shojae M, Sabzghabaei A, Heidari A. Efficacy of new scoring system for diagnosis of abdominal injury after blunt abdominal trauma in patients referred to emergency department. *Chin J Traumatol*. 2020;23(3):145-8. doi: 10.1016/j.cjtee.2020.03.003
34. Karjosukarso AS, Wiargitha IK, Mahadewa TGB. [Diagnostic validity of the Blunt Abdominal Trauma Scoring System (BATSS) in blunt abdominal trauma at Sanglah General Hospital, Denpasar, Bali]. *Medicina (Bali)*. 2019;50(2). Indonesian. doi: 10.15562/medicina.v50i1.181
35. Subbiah TS, Gopinathan A, Ramachandran B. Comparison of Blunt Abdominal Trauma Scoring System (BATSS) versus FAST in assessment of Blunt Trauma Abdomen. *Biomed Pharmacol J*. 2021;14(4):2183-9. doi: 10.13005/bpj/2316
36. Sivarajan N. Prospective study comparing clinical abdominal scoring system(CASS) with blunt abdominal trauma severity scoring(BATSS). *MedPulse Int J Surg*. 2021;19(2):53-7. doi: 10.26611/1061925
37. Beltagy IRE, Abu Al-Yazid MM, Ghalwash ET. Evaluation of blunt abdominal trauma severity score (BATSS) in predicting the necessity of laparotomy for admitted cases with blunt abdominal trauma. *Al-Azhar Med J*. 2021;50(4):2531-42. doi: 10.21608/amj.2021.196395
38. Fahmi KSE, Refaat DO, Ali KA, et al. Diagnostic Evaluation of Blunt Abdominal Trauma Scoring System (BATSS). *Egypt J Hosp Med*. 2022;86(1):871-6. doi: 10.21608/ejhm.2022.218041
39. Shah JD, Shah TA, Modi JB. Blunt Abdominal Trauma Severity Scoring System: Exceptional Score in Blunt Abdominal Trauma Management. *Int J Recent Surg Med Sci*. 2023;9:71-6. doi: 10.25259/ijrsm-2022-4-20
40. Manjush E, Sathish Kumar R. Comparative study between blunt abdominal trauma scoring system (BATSS) and CECT abdomen to detect blunt trauma abdomen in adult patients. *Basrah J Surg*. 2024;30(2):35-43. doi: 10.33762/bsurg.2024.153210.1090
41. Sadia, Khalid O, Zulfikar I, et al. Comparison of blunt abdominal trauma severity scoring (BATSS) and clinical abdominal scoring system (CASS) in blunt abdominal trauma for need of surgical intervention. *Biol Clin Sci Res J*. 2024;2024(1):777. doi: 10.54112/bcsrj.v2024i1.777
42. El-Meligy MH, Aboshosha AS, Amer NF. Assessment of Blunt Abdominal Trauma Severity Score in Predicting the Necessity of Laparotomy. *Menoufia Med J*. 2024;37(2). doi: 10.59204/2314-6788.3054
43. Shah DK, Patel KM, Patel S, et al. Efficacy of blunt abdominal trauma scoring system in management of blunt abdominal trauma. *Int Surg J*. 2022;9(10):1726. doi: 10.18203/2349-2902.isj20222597
44. Jang Y, Jun H. Analysis of trauma scoring system for patients with abdominal trauma. *Ulus Travma Acil Cerrahi Derg*. 2022;29(1):68-72. doi: 10.14744/tjtes.2022.94475
45. Alam A, Gupta A, Gupta N, et al. Evaluation of ISS, RTS, CASS and TRISS scoring systems for predicting outcomes of blunt trauma abdomen. *Pol Przegl Chir*. 2021;93(2):9-15. doi: 10.5604/01.3001.0014.7394
46. Gutierrez A, Matsushima K, Grigorian A, et al. Derivation and Validation of a Score Using Prehospital Data to Identify Adults With Trauma Requiring Early Laparotomy. *JAMA Netw Open*. 2022;5(1):e2145860. doi: 10.1001/jamanetworkopen.2021.45860
47. El-Menyar A, Ramzee AF, Elmegabar BH, et al. Validation of the FASILA Score for Predicting Interventions and Outcomes in Traumatic Abdominal and Pelvic Injuries: A Prospective Clinical Study. *World J Surg*. 2025;49(7):151-9. doi: 10.1002/wjs.12632
48. Vijfschagt ND, Burger H, Berger MY, et al. Variation in sensitivity and specificity of diverse diagnostic tests across health-care settings: a meta-epidemiological study. *J Clin Epidemiol*. 2025;184:111816. doi: 10.1016/j.jclinepi.2025.111816
49. Murad MH, Lin L, Chu H, et al. The association of sensitivity and specificity with disease prevalence: analysis of 6909 studies of diagnostic test accuracy. *CMAJ*. 2023;195(27):E925-31. doi: 10.1503/cmaj.221802
50. Kohn MA. Studies of diagnostic test accuracy: Partial verification bias and test result-based sampling. *J Clin Epidemiol*. 2022;145:179-82. doi: 10.1016/j.jclinepi.2022.01.022
51. Yordanian NO, Wiargitha IK, Golden N, et al. Evaluation of clinical abdominal scoring system for predicting outcomes of blunt

- abdominal trauma. *Univ Med.* 2023;42(2):173-81. doi: 10.18051/UnivMed.2023.v42.173-181
52. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527. doi: 10.1136/bmj.h5527
 53. Vickers AJ, Van Calster B, Wynants L, et al. Decision curve analysis: confidence intervals and hypothesis testing for net benefit. *Diagn Progn Res.* 2023;7(1):11. doi: 10.1186/s41512-023-00148-y
 54. 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
 55. 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
 56. Singh A, Min Lim AS, Huh Lau BP, et al. Epidemiology of pelvic and acetabular fractures in a tertiary hospital in Singapore. *Singapore Med J.* 2022;63(7):388-93. doi: 10.11622/smedj.2021024
 57. Fokin AA, Wycech Knight J, Tharp ME, et al. Pelvic fractures in blunt trauma patients: A comparative study. *World J Orthop.* 2024;15(5):418-34. doi: 10.5312/wjo.v15.i5.418
 58. Sethi I, Aicher AE, Zawin M, et al. Index CT-Based Scoring Systems in Operative Blunt Bowel and Mesenteric Injury Identification. *J Surg Res.* 2024;294:240-6. doi: 10.1016/j.jss.2023.10.001
 59. Smyth L, Bendinelli C, Lee N, et al. WSES guidelines on blunt and penetrating bowel injury: diagnosis, investigations, and treatment. *World J Emerg Surg.* 2022;17:13. doi: 10.1186/s13017-022-00418-y
 60. Santos J, Delaplain PT, Tay-Lasso E, et al. Development and Validation of a Novel Hollow Viscus Injury Prediction Score for Abdominal Seatbelt Sign: A Pacific Coast Surgical Association Multicenter Study. *J Am Coll Surg.* 2023;237(6):826-33. doi: 10.1097/XCS.0000000000000863
 61. McNutt MK, Chinapuvvula NR, Beckmann NM, et al. Early surgical intervention for blunt bowel injury: the Bowel Injury Prediction Score (BIPS). *J Trauma Acute Care Surg.* 2015;78(1):105-11. doi: 10.1097/TA.0000000000000471
 62. Ismail MI, Ramli NS, Tan JH, et al. Non-operative management of solid organ injuries in a middle-income country, how does it stack up? *Injury.* 2022;53(9):2992-7. doi: 10.1016/j.injury.2022.03.036