Biological and Clinical Sciences Research Journal
eISSN: 2708-2261; pISSN: 2958-4728
www.bcsrj.com
DOI: https://doi.org/10.54112/bcsrj.v6i6.2003
*Biol. Clin. Sci. Res. J., Volume 6(6), **2025**: 2003*
Original Research Article

Evaluation of Possum Scoring System in Patients Undergoing Laparotomy for Risk Assessment

Syeda Mahjabeen[*1], Farhan Zaheer[2], Tayram Khalid[1], Hafiz Yahya Iftikhar[1], Syeda Zubaria Qamar[1], Amna Khan[1]

[1]*Department of Surgery, Dr Ruth K M Pfau CHK Karachi, Pakistan*
[2]*Department of Surgery, Dow University of Health Sciences, Karachi, Pakistan*
**Corresponding author`s email address:* syeda.mahjaben95@gmail.com

(Received, 14th March 2025, Accepted 2nd June 2025, Published 30th June 2025)

Abstract: *Accurate perioperative risk stratification is essential for consent, triage, and audit in general surgery. The Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM) and its Portsmouth modification (P-POSSUM) are widely used, but external performance varies by setting.* ***Objective:*** *To evaluate discrimination and calibration of POSSUM (mortality and morbidity) and P-POSSUM (mortality) in patients undergoing midline laparotomy, overall and within key subgroups.* ***Methods:*** *We conducted a prospective observational study of consecutive adults undergoing elective or emergency midline laparotomy at Surgical Unit 5, Civil Hospital Karachi, from 1st August 2024 to 31st January 2025. POSSUM physiological (PS) and operative (OS) scores were computed; predicted risks were derived using standard logistic equations. Primary outcomes were 30-day mortality and 30-day morbidity (any postoperative complication ≥Clavien–Dindo II). Discrimination (AUC) and calibration (Hosmer–Lemeshow, observed: expected [O: E] ratios, decile plots) were assessed, including subgroup analyses by urgency and ASA class.* ***Results:*** *Among 150 patients (60% male; mean age 56.6±15.9 years), 65% underwent emergency surgery. Thirty-day outcomes were: complications 55%, ICU admission 32%, mortality 9%; mean length of stay 14.6±9.1 days. Mortality discrimination was moderate for P-POSSUM (AUC 0.652, 95% CI 0.484–0.820) and POSSUM (AUC 0.629, 95% CI 0.460–0.798); morbidity discrimination was acceptable for POSSUM (AUC 0.704, 95% CI 0.622–0.786). Calibration indicated overall over-prediction: O: E 0.888 for P-POSSUM mortality (13 observed/14.64 expected), 0.421 for POSSUM mortality (13/30.89), and 0.895 for POSSUM morbidity (82/91.57); Hosmer–Lemeshow was significant for mortality and morbidity ($p \leq 0.005$). Subgroups showed under-prediction in elective mortality (O: E 1.405) and over-prediction in emergency mortality (O: E 0.763); ASA I–II and III–V strata were closer to unity.* ***Conclusions:*** *POSSUM/P-POSSUM provided moderate discrimination but clinically relevant miscalibration, most pronounced at risk extremes. For decision support and benchmarking, local intercept/slope recalibration and, where appropriate, context-specific models (e.g., NELA or CR-POSSUM) are recommended to improve accuracy.*

Keywords: POSSUM; P-POSSUM; Emergency laparotomy; Risk stratification; Calibration

[***How to Cite:*** Mahjabeen S, Zaheer F, Khalid T, Iftikhar HY, Qamar SZ, Khan A. Evaluation of possum scoring system in patients undergoing laparotomy for risk assessment. *Biol. Clin. Sci. Res. J.,* **2025;** 6(6): 489-493. doi: https://doi.org/10.54112/bcsrj.v6i6.2003

Introduction

Accurate peri-operative risk stratification underpins informed consent, peri-operative optimization, level-of-care decisions, and fair benchmarking. The Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM) combines a 12-item physiological score with a 6-item operative severity score to generate predicted risks of morbidity and mortality via logistic equations, originally derived and prospectively validated by Copeland et al. in general surgical patients (1). Early external evaluations noted that the original mortality equation tended to over-predict deaths, particularly in low-risk groups, prompting the Portsmouth modification (P-POSSUM), which recalibrated the intercept and slopes to improve calibration across risk strata (2).

Emergency laparotomy (EL) remains one of the highest-risk general surgical procedures; contemporary series report 30-day mortality between ~5–10% in unselected cohorts, and substantially higher in older or frail populations. In a geriatric EL cohort (n=157), 90-day mortality reached 29.3%, and models incorporating P-POSSUM demonstrated AUCs ≈0.88–0.93 for mortality discrimination (6). In a prospective EL study from a developing-country referral center (n=120), POSSUM's linear analysis markedly over-predicted mortality (O: E 0.39) and morbidity (O: E 0.68), whereas exponential analysis improved calibration (mortality O: E 0.62; morbidity O: E 0.91). P-POSSUM achieved closer mortality prediction (O: E 0.88). (3) Similar signals—good discrimination but imperfect calibration—have been reported in large peri-operative cohorts, with fewer observed than expected deaths across deciles of risk (8).

Beyond classic POSSUM/P-POSSUM, several risk tools (e.g., NELA, NSQIP, SORT) are now used clinically. Comparative work increasingly suggests that the National Emergency Laparotomy Audit (NELA) model equals or outperforms P-POSSUM for 30-day mortality prediction. In 830 EL patients from Singapore, both scores discriminated well (AUC 0.86 vs 0.84), but NELA demonstrated superior overall performance and calibration (O: E 0.58 vs 0.34). (5) Additional multicenter studies and critiques have highlighted practical challenges of using POSSUM for comparative audit (missing data burden, population drift), even while acknowledging its enduring value for risk-adjusted audit when applied appropriately (4,8).

In abdominal surgery more broadly, recent studies continue to find that POSSUM often overestimates both morbidity and mortality in moderate/high-risk patients. In contrast, P-POSSUM is usually a closer fit for mortality, but it may still require local recalibration (7). Methodologically, contemporary evaluations emphasize reporting both discrimination (e.g., AUC) and calibration (e.g., Hosmer–Lemeshow, calibration plots), as well as observed-to-expected (O: E) ratios within clinically relevant risk bands.

Objective. To evaluate the performance of POSSUM (morbidity and mortality equations) and P-POSSUM (mortality) in patients undergoing midline laparotomy for 30-day risk assessment.

Methodology

We conducted a prospective observational study of 150 consecutive adults (≥18 years) undergoing midline laparotomy (elective or emergency) performed at Surgical Unit 5, Civil Hospital Karachi, between 1st August 2024 and 31st January 2025. Patients who had laparoscopic-only procedures, minor diagnostic laparotomies without therapeutic intent, trauma laparotomies, or missing key POSSUM variables were excluded. Demographic data (age, sex), urgency (elective/emergency), indication for surgery, American Society of Anesthesiologists (ASA) class, and pre-operative physiology (heart rate, systolic blood pressure, hemoglobin, white blood cell count, serum creatinine, ECG) were abstracted from charts and electronic records. Intra-operative variables (operative magnitude, multiple procedures, blood loss, peritoneal contamination, malignancy, mode of surgery) were recorded at the end of the procedure.

For each patient, we calculated the POSSUM physiological score (PS) and operative severity score (OS) using the original 12 and 6 variable frameworks. Predicted risks were generated with the standard logistic equations: POSSUM mortality $\ln(R/(1-R)) = -7.04 + 0.13\cdot PS + 0.16\cdot OS$, POSSUM morbidity $\ln(R/(1-R)) = -5.91 + 0.16\cdot PS + 0.19\cdot OS$, and P-POSSUM mortality $\ln(R/(1-R)) = -9.065 + 0.1692\cdot PS + 0.1550\cdot OS$; predicted probabilities were obtained by inverse-logit transformation. (1,2,10) The primary outcomes were 30-day all-cause mortality and 30-day postoperative morbidity (any complication fulfilling ≥Clavien–Dindo grade II). Secondary outcomes included ICU admission and postoperative length of stay.

We followed patients for 30 days post-operation through in-hospital surveillance and phone reviews as needed. Data quality checks were performed before analysis; continuous variables were summarized as mean ± SD, and categorical variables were summarized as frequency (percentage). We assessed discrimination using ROC curves (AUC) and calibration using calibration plots, Hosmer–Lemeshow goodness-of-fit, and overall observed-to-expected (O: E) ratios with 95% CIs. Prespecified subgroup analyses compared elective vs emergency cases and ASA I–II vs ASA III–V. All analyses were performed with two-sided $\alpha=0.05$.

Results

In this cohort of 150 patients, males comprised 60% (n=90) and females 40% (n=60). The mean age was 56.58 ± 15.86 years.

Clinically, 65% (n=97) underwent emergency surgery and 35% (n=53) were elective. The ASA distribution was 7% ASA I (n=11), 29% ASA II (n=44), 40% ASA III (n=60), 20% ASA IV (n=30), and 3% ASA V (n=5). Indications included bowel obstruction in 52 patients (25%), perforation in 30 (20%), peritonitis in 25 (17%), ischemia in 10 (7%), trauma in 10 (7%), malignancy in 18 (12%), and other causes in 5 (3%). Sepsis was present in 60 patients (40%). ECG findings showed normal tracings in 105 (70%), atrial fibrillation in 26 (17%), and ischaemic changes in 19 (13%). A malignancy Diagnosis was recorded in 32 patients (21%). Laboratory and physiological means (±SD) were: hemoglobin 12.15 ± 1.82 g/dL, WBC 11.34 ± 4.14 × 10^9/L, creatinine 1.22 ± 0.47 mg/dL, systolic blood pressure 120.77 ± 18.23 mmHg, and heart rate 93.5 ± 17.8 bpm. The mean POSSUM Physiological Score was 24.27 ± 6.33, and Operative Score 13.91 ± 4.81. Predicted risks averaged 20.59 ± 16.28% for POSSUM mortality, 61.05 ± 24.59% for POSSUM morbidity, and 9.76 ± 10.63% for P-POSSUM mortality.

Outcomes showed ICU admission in 48 patients (32%), 30-day complications in 82 (55%), and 30-day mortality in 13 (9%). Mean length of stay was 14.56 ± 9.1 days.

Observed-to-expected (O: E) performance overall indicated a tendency toward over-prediction (O: E < 1) for two of the three models: for mortality, P-POSSUM observed 13 deaths versus 14.64 expected (O: E = 0.888; 95% CI 0.473–1.518), and POSSUM observed 13 versus 30.89 expected (O: E = 0.421; 95% CI 0.224–0.720). For morbidity, POSSUM observed 82 complications versus 91.57 expected (O: E = 0.895; 95% CI 0.712–1.112).

Stratified by urgency, elective cases (Emergency = 0) showed P-POSSUM mortality O: E = 1.405 (4 observed vs 2.85 expected; 95% CI 0.383–3.599), suggesting under-prediction in this subgroup, whereas emergency cases (Emergency = 1) had O: E = 0.763 (9 vs 11.8; 95% CI 0.349–1.448), consistent with over-prediction. For morbidity, elective cases had an O: E ratio of 0.968 (25 vs 25.83; 95% CI 0.626–1.428), and emergency cases had an O: E ratio of 0.867 (57 vs 65.73; 95% CI 0.657–1.123), both close to—but below—unity.

By ASA, lower-risk patients (ASA high = 0; I–II) had P-POSSUM mortality O: E = 1.004 (5 vs 4.98; 95% CI 0.326–2.344) and morbidity O: E = 0.904 (28 vs 30.96; 95% CI 0.601–1.307), while higher-risk patients (ASA high = 1; III–V) had mortality O: E = 0.828 (8 vs 9.67; 95% CI 0.357–1.631) and morbidity O: E = 0.891 (54 vs 60.61; 95% CI 0.669–1.162). Overall, these patterns suggest that calibration varies by risk stratum and urgency, with modest over-prediction in the aggregate and near-unity O: E in several clinically relevant subgroups.

This plot compares how well the two mortality models separate patients who died within 30 days from those who survived. The P-POSSUM curve sat modestly above the diagonal, giving AUC = 0.652 (95% CI 0.484–0.820), while POSSUM-mortality was similar at AUC = 0.629 (95% CI 0.460–0.798). In practical terms, both models showed moderate discrimination—better than chance, but with clear room to improve if you need high-confidence individual risk stratification. The two curves were close, so on this dataset, P-POSSUM did not dramatically outperform the original POSSUM mortality equation.

The morbidity ROC curve demonstrates the model's ability to distinguish patients with any postoperative complication by day 30. The curve tracked further from the diagonal than the mortality curves, yielding AUC = 0.704 (95% CI 0.622–0.786)—acceptable discrimination for clinical audit and cohort-level planning. You would expect better separation at higher thresholds (e.g., to "rule in" high-risk), but it is still imperfect for individual prediction.

This Figure contrasts the mean predicted risk with the observed event rate within ten equal-sized risk groups. Overall, the model over-predicted deaths (Observed: Expected O: E = 0.89, 13 observed vs 14.64 expected), and the Hosmer–Lemeshow test indicated miscalibration ($\chi^2 = 21.74$, df = 8, $p = 0.005$). The shape is informative: in the lowest-risk decile, the observed rate exceeded the mean predicted by about +5.5 percentage points, while in the highest-risk decile, the model overshot observed risk by –8.7 percentage points. This "under at the low end, over at the high end" pattern suggests the predictions are too extreme (a calibration-slope issue), i.e., risks may need recalibration (shrinkage) for this population.

The morbidity calibration told a similar story. Overall, there were fewer complications than expected (O: E ≈ 0.90, 82 observed vs 91.57 expected), and the Hosmer–Lemeshow statistic again flagged miscalibration ($\chi^2 = 28.51$, df = 8, $p < 0.001$). Decile-wise, the model under-estimated risk in the lowest-risk group by roughly +9.4 percentage points and over-estimated in the highest-risk group by –14.6 percentage points. This mirrors the mortality pattern and implies the model would benefit from local recalibration (adjusting intercept/slope) before being used for individual clinical decisions, while remaining serviceable for risk-adjusted audit with appropriate caveats.

Table 1: Demographic variables

Variable	Frequency (%)
Gender	
Male	90 (60%)
Female	60 (40%)
Age (years)	56.58±15.86

Table 2: Clinical variables

Variable	Frequency and Mean
Urgency	
Elective	53 (35%)
Emergency	97 (65%)
ASA	
1	11 (7%)
2	44 (29%)
3	60 (40%)
4	30 (20%)
5	5 (3%)
Indications	
Bowel obstruction	52 (25%)
Ischemia	10 (7%)
Malignancy	18 (12%)
Other	5 (3%)
Perforation	30 (20%)
Peritonitis	25 (17%)
Trauma	10 (7%)
Sepsis	60 (40%)
ECG	
Atrial fibrillation	26 (17%)
Ischaemic changes	19 (13%)
Normal	105 (70%)
Malignancy	32 (21%)
Hb (g/dL)	12.15±1.82
Wbc (10e9/L)	11.34±4.14
Creatinine (mg/dL)	1.22±0.47
Systolic bp (mmHg)	120.77±18.23
Heart rate (bpm)	93.5±17.8
POSSUM Physiological Score	24.27±6.33
POSSUM Operative Score	13.91±4.81
PredMort POSSUM	20.59±16.28
PredMorb POSSUM	61.05±24.59
PredMort PPOSsum	9.76±10.63

Table 3: Output Variables

Variable	Frequency and mean
ICU admission	48 (32%)
Complication 30d	82 (55%)
Death 30d	13 (9%)
Length of stay (days)	14.56±9.1

Table 4: Observed to expected (O: E) overall

Model/Outcome	Observed events	Expected events	O: E ratio	95% CI lower	95% CI upper
Mortality (P-POSSUM)	13	14.64	0.888	0.473	1.518
Mortality (POSSUM)	13	30.89	0.421	0.224	0.72
Morbidity (POSSUM)	82	91.57	0.895	0.712	1.112

Table 5: OE by urgency

Subgroup	Model/Outcome	Observed events	Expected events	O: E ratio	95% CI lower	95% CI upper
Emergency=0	Mortality (P-POSSUM)	4	2.85	1.405	0.383	3.599
Emergency=1	Mortality (P-POSSUM)	9	11.8	0.763	0.349	1.448
Emergency=0	Morbidity (POSSUM)	25	25.83	0.968	0.626	1.428
Emergency=1	Morbidity (POSSUM)	57	65.73	0.867	0.657	1.123

Table 6: OE by ASA

Subgroup	Model/Outcome	Observed events	Expected events	O: E ratio	95% CI lower	95% CI upper
ASA high=0	Mortality (P-POSSUM)	5	4.98	1.004	0.326	2.344
ASA high=1	Mortality (P-POSSUM)	8	9.67	0.828	0.357	1.631

ASA high=0	Morbidity (POSSUM)	28	30.96	0.904	0.601	1.307
ASA high=1	Morbidity (POSSUM)	54	60.61	0.891	0.669	1.162

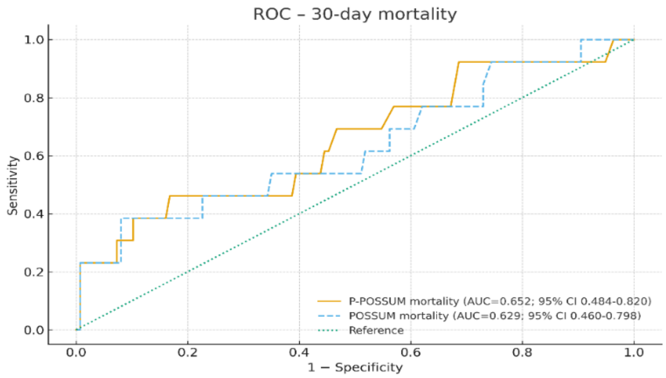

Figure 1 – ROC (30-day mortality, P-POSSUM vs POSSUM).

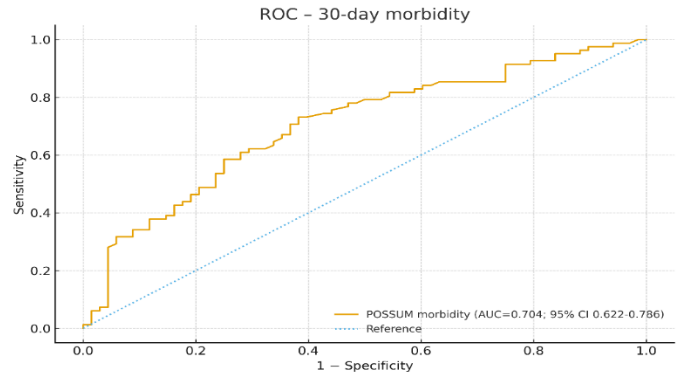

Figure 2 – ROC (30-day morbidity, POSSUM).

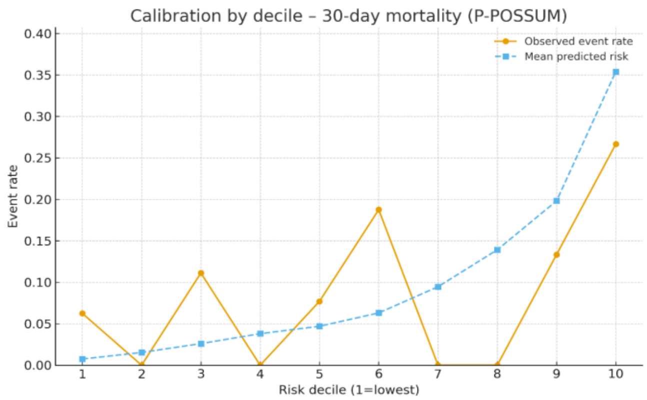

Figure 3 – Calibration by deciles (30-day mortality, P-POSSUM).

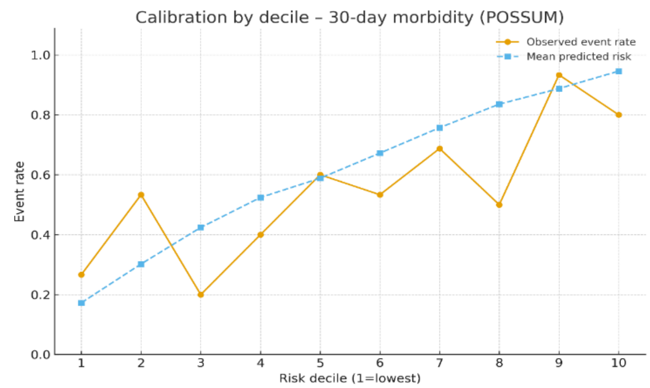

Figure 4 – Calibration by deciles (30-day morbidity, POSSUM).

.

Discussion

In this cohort, POSSUM/P-POSSUM showed moderate discrimination (mortality AUC ~0.63–0.65; morbidity AUC ~0.70) with systematic over-prediction overall (O: E <1) and calibration drift at the extremes of risk. This pattern—reasonable rank-ordering but imperfect calibration—aligns with external reports across general and abdominal surgery, where original equations often overestimate risk outside their derivation context and benefit from local recalibration or specialty-specific variants (16,17). The development of CR-POSSUM exemplifies this response to miscalibration in colorectal populations, where a domain-specific model yielded a better fit (11, 19).

Our mortality AUCs were lower than several series reporting stronger discrimination for P-POSSUM (often ≥0.80), underscoring how case-mix, endpoint definitions, and data completeness can depress performance in heterogeneous or smaller samples (15,17,20). A systematic review similarly highlighted recurrent over-prediction and variable calibration of POSSUM family models across subspecialties, advocating recalibration rather than abandonment (16).

The decile calibration curves here (under-prediction at the low end, over-prediction at the high end) suggest a calibration slope <1, a signal of coefficient over-fitting to other populations or temporal shifts in baseline risk. Comparable decile-based miscalibration for POSSUM/P-POSSUM has been documented in external validations from general-surgery cohorts, with O: E for morbidity commonly ~0.88–0.95 and frequent Hosmer–Lemeshow significance (15,17).

Subgroup analyses provide clinical context. Elective cases showed under-prediction of mortality (O: E >1) while emergency cases showed over-prediction (O: E <1). Prior work indicates that emergency-specific models calibrate better within that stratum, and adopting context-specific models improves fit. In emergency laparotomy, the NELA risk model achieved good discrimination with better calibration for provider-level audit. The multiple comparisons report shows that NELA matches or outperforms P-POSSUM for 30-day mortality. However, some cohorts still observe under-estimation—emphasizing the need for local checks (12–14,18).

Practically, our findings support two steps when these tools inform decisions or benchmarking: (1) simple local recalibration (adjusted intercept/slope) to align predicted with observed risks; (2) use of domain-specific or contemporary models (e.g., CR-POSSUM for colorectal, NELA for emergency laparotomy) when case-mix is concentrated in those areas (11,12,16). Finally, our single-center scope, modest sample size (event rate 9%), and potential measurement variability widen confidence intervals and increase hazard ratio sensitivity—everyday realities in surgical audit that reinforce periodic model updating and prospective validation within institutions (16).

Conclusion

POSSUM and P-POSSUM demonstrated moderate discrimination but systematically over-predicted risk, with miscalibration most evident at low and high risk deciles. Subgroup patterns (elective vs emergency; ASA strata) showed variable fit, underscoring the influence of case-mix on model performance. These tools remain useful for risk-adjusted audit, but individual clinical decisions should be cautious without local recalibration. We recommend intercept/slope recalibration and, where applicable, adoption of context-specific models (e.g., NELA or CR-POSSUM) to improve accuracy.

Declarations

Data Availability statement

All data generated or analysed during the study are included in the manuscript.

Ethics approval and consent to participate

Approved by the department concerned. (IRBEC-24)

Consent for publication

Approved

Funding

Not applicable

Conflict of interest

The authors declared the absence of a conflict of interest.

Author Contribution

SM (Surgical Resident)
Manuscript drafting, Study Design,
FZ (Professor)
Review of Literature, Data entry, Data analysis, and drafting an article.
TK (Surgical Resident Associate Professor)
Conception of Study, Development of Research Methodology Design,
HYI (Surgical Resident)
Study Design, manuscript review, and critical input.
SZQ (Surgical Resident)
Manuscript drafting, Study Design,
AK (Surgical Resident)
Review of Literature, Data entry, Data analysis, and drafting an article.

All authors reviewed the results and approved the final version of the manuscript. They are also accountable for the integrity of the study.

References

1. Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg.* 1991;78(3):355–360. https://doi.org/10.1002/bjs.1800780327
2. Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. *Br J Surg.* 1998;85(9):1217–1220. https://doi.org/10.1046/j.1365-2168.1998.00840.x
3. Mohil RS, et al. POSSUM and P-POSSUM for risk-adjusted audit of patients undergoing emergency laparotomy. *Br J Surg.* 2004;91(2):195–199. https://doi.org/10.1002/bjs.4468
4. Bann SD, Sarin S. Comparative audit: the trouble with POSSUM. *J R Soc Med.* 2001;94(12):632–634. https://doi.org/10.1258/jrsm.94.12.632
5. Lai CPT, Goo TT, Ong MW, Prakash PS, Lim WW, Drakeford PA. A comparison of the P-POSSUM and NELA risk score for patients undergoing emergency laparotomy. *World J Surg.* 2021;45(8):2439–2446. https://doi.org/10.1007/s00268-021-06120-5
6. Cao Y, et al. The statistical importance of P-POSSUM scores after emergency laparotomy in geriatric patients. *BMC Med Inform Decis Mak.* 2020;20:86. https://doi.org/10.1186/s12911-020-01100-9
7. Valenzuela S, et al. Morbidity-mortality assessment in abdominal surgery: are we predicting or overreacting? *BMC Surg.* 2022;22:19. https://doi.org/10.1186/s12893-021-01455-1
8. Scott S, et al. An evaluation of POSSUM and P-POSSUM scoring in predicting postoperative mortality in a level-1 critical care setting. *BMC Anesthesiol.* 2014;14:104. https://doi.org/10.1186/1471-2253-14-104
9. Copeland GP. The POSSUM system of surgical audit. *JAMA Surg.* 2002;137(1):15–19. https://doi.org/10.1001/archsurg.137.1.15
10. Kim SH, et al. Risk assessment of mortality following intraoperative cardiac arrest using POSSUM and P-POSSUM. *Yonsei Med J.* 2015;56(5):1401–1407. https://doi.org/10.3349/ymj.2015.56.5.1401
11. Tekkis PP, Prytherch DR, Kocher HM, et al. Development of a dedicated risk-adjustment scoring system for colorectal surgery (CR-POSSUM). *Br J Surg.* 2004;91(9):1174–1182. https://doi.org/10.1002/bjs.4614
12. Eugene N, Oliver CM, Bassett MG, et al. A novel risk adjustment model for adult patients undergoing emergency laparotomy: the NELA risk model. *Br J Anaesth.* 2018;121(4):739–748. https://doi.org/10.1016/j.bja.2018.06.026
13. Hansted AK, Daugaard CL, Kähler P, et al. Validation of the updated NELA risk prediction model in emergency laparotomy. *Acta Anaesthesiol Scand.* 2023;67(6):725–735. https://doi.org/10.1111/aas.14294
14. Alabbasy MM, Elsisy AAE, Mahmoud A, Alhanafy SS. Comparison between P-POSSUM and NELA risk score for patients undergoing emergency laparotomy. *BMC Surg.* 2023;23:325. https://doi.org/10.1186/s12893-023-02189-y
15. Ngulube A, Yates J, Walker K, et al. Validation of POSSUM, P-POSSUM, and the Surgical Risk Scale in major general surgical procedures. *Ann Med Surg (Lond).* 2019;41:33–39. https://doi.org/10.1016/j.amsu.2019.03.007
16. Wang H, Chen T, Wang H, Song Y, Li X, Wang J. A systematic review of the Physiological and Operative Severity Score for the enumeration of Mortality and morbidity and its Portsmouth modification as predictors of post-operative morbidity and mortality in patients undergoing pancreatic surgery. *Am J Surg.* 2013;205(4):466–472. https://doi.org/10.1016/j.amjsurg.2012.06.011
17. González-Martínez S, Martín-Baranera M, Martí-Saurí I, Borrell-Grau N, Pueyo-Zurdo JM. Comparison of the risk prediction systems POSSUM and P-POSSUM with the Surgical Risk Scale: a prospective cohort study of 721 patients. *Int J Surg.* 2016;29:19–24. https://doi.org/10.1016/j.ijsu.2016.03.005
18. Chughlay Z, et al. Predictive performance of NELA versus P-POSSUM mortality scores after emergency laparotomy. *Cureus.* 2022;14(12):e32979. https://doi.org/10.7759/cureus.32979
19. Horžić M, Kopljar M, Cupurdija K, Bielen I, Lacković Ž, Šošić V. Comparison of P-POSSUM and CR-POSSUM in colorectal surgery. *Arch Surg.* 2007;142(2):172–177. https://doi.org/10.1001/archsurg.142.2.172
20. Ramakrishnan P, et al. Evaluation of POSSUM scoring systems in predicting post-operative mortality and morbidity. *Bahrain J Anaesthesiol.* 2020;2(2):[pagination not available]. [No DOI assigned].