Volume 3, Issue 5, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

A COMPARATIVE STUDY OF APACHE-II AND P-POSSUM SCORING SYSTEMS IN PREDICTING POSTOPERATIVE MORTALITY FOR PATIENTS UNDERGOING NON-TRAUMATIC EMERGENCY LAPAROTOMY

Dr Mehreen Ansari[*1], Dr Anum Fatima[2], Dr Barkha Kumari[3] Dr Tehreem Ahmed[4], Dr Rimsha Ashfaq[5], Dr Saeed Ahmed[6]

*[*1,2,3]MBBS, FCPS Training, Department of Surgery, Dow Medical College & Dr. Ruth K. M. Pfau Civil Hospital, Karachi, Pakistan. Dow University of Health Sciences.*

[4]MBBS, FCPS, Senior Registrar, Department of Surgery, Creek General Hospital, Karachi, Pakistan, United Medical and Dental College.

[5]MBBS, FCPS training, Department of Surgery, Dow International Medical College and Dow University Hospital, Karachi, Pakistan. Dow University of Health Sciences.

[6]MBBS, FCPS, FRCS, Professor of Surgery, Department of Surgery, Creek General Hospital, Karachi, Pakistan. United Medical and Dental College.

[*1]drmehreen.ansari@gmail.com, [2]anumsharif93@gmail.com, [3]barkhatejwani31@gmail.com
[4]tas1193@gmail.com, [5]rimsha.ashfaq@hotmail.com, [6]dr.saeedshaikh230@gmail.com

DOI: https://doi.org/10.5281/zenodo.16307759

Keywords
APACHE-II Score, POSSUM Scoring System, Postoperative Mortality, Risk Assessment

Article History
Received: 22 April, 2025
Accepted: 07 July, 2025
Published: 22 July, 2025

Corresponding Author: *
Mehreen Ansari

Abstract

Objeective *To compare the predictive ability of the APACHE-II and P-POSSUM scoring systems in predicting thirty day postoperative mortality among patients undergoing non-traumatic emergency laparotomy in a large public sector hospital in Pakistan.*

Methods *A prospective observational study was carried out at Dr. Ruth KM Pfau Civil Hospital Karachi among patients undergoing non-traumatic emergency laparotomy. APACHE-II and P-POSSUM scores were calculated for each patient using physiological and operative parameters. The mortality was assessed within 30 days of post-operation. Receiver Operating Characteristic (ROC) curve analysis was used to determine the discriminative ability of both scores. Optimal cutoff points were identified using Youden's Index, and diagnostic accuracy was assessed through sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).*

Results *Of 192 patients, the 30-day postoperative mortality rate was 24 (12.5%). Both APACHE-II and P-POSSUM showed excellent discriminatory power, with area under the AUC values exceeding 0.90. The optimal cutoff for APACHE-II mortality risk was 27.35, reporting sensitivity, specificity, PPV, NPV, and diagnostic accuracy of 95.8%, 22.6%, 15.0%, 97.4%, and 31.8% respectively. For P-POSSUM, the optimal cutoff was 49.75, with sensitivity, specificity, PPV, NPV, and diagnostic accuracy of 95.8%, 45.8%, 20.2%, 98.7%, and 52.1% respectively.*

Conclusion *Both APACHE II and P-POSSUM scores are effective in predicting 30-day postoperative mortality, with P-POSSUM offering superior diagnostic accuracy due to the inclusion ofintraoperative variables.*

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

INTRODUCTION

Emergency laparotomy remains one of the most frequently performed procedures in acute surgical care worldwide. Despite advances in surgical techniques, anesthetic practices, and postoperative critical care, the mortality rates associated with emergency laparotomies continue to be alarmingly high. A systematic review reported a one-year mortality rate of 24.6% among such patients[1], while outcomes among geriatric populations are even more dismal, with one-year mortality rates ranging from 30% to 47%.[2]

There is a notable paucity of data on emergency laparotomy outcomes from low- and middle-income countries (LMICs) such as Pakistan.[3] This gap underlines the need for context-specific studies to understand the burden of surgical mortality and the effectiveness of predictive tools in resource-limited settings. Early prognostic evaluation can help identify high-risk patients who may benefit from aggressive perioperative management.[4-6]

Risk stratification tools such as scoring systems are increasingly utilized to support surgical audits and improve clinical governance.[7-9] The Portsmouth Physiological and Operative Severity Score for the enumeration of Mortality and morbidity (P-POSSUM), which is a modification of the original POSSUM score, is now widely applying to evaluate outcomes in emergency laparotomy.[7] P-POSSUM incorporates both physiological and intraoperative parameters, requiring eighteen factors including estimates of blood loss and peritoneal contamination.[10] In contrast, the Acute Physiology and Chronic Health Evaluation II (APACHE-II) score which was first introduced in 1985 remains one of the most extensively validated illness severity scoring systems all around the world.[11] It relies on twelve preoperative physiological variables and is valued for its applicability in diverse clinical settings.[12,13]

Although both tools are employed in surgical practice, they differ significantly in their components and clinical applicability.[6,14] Nevertheless, no consensus exists regarding the superiority of one system over the other, and comparative performance remains underexplored in a developing country like Pakistan. Moreover, the late presentation of most of these cases in public sector hospitals of Pakistan and consequently, their severity necessitates pre-assessment that can predict the prognosis of such cases. Therefore, this study was designed to compare the predictive accuracy of APACHE-II and P-POSSUM scoring systems in predicting postoperative mortality among patients undergoing non-traumatic emergency laparotomy.

METHODS

This prospective observational study was conducted in the Department of General Surgery at Dr Ruth K.M. Pfau Civil Hospital Karachi, Pakistan. The study was conducted after getting approval from the Institutional Review Board (IRB) of Dow University of Health Sciences. Moreover, signed informed consent was obtained from all eligible study participants or their attendants.

All patients aged 18 years and above who presented with non-traumatic abdominal pathology requiring emergency open laparotomy were considered eligible for inclusion. All these patients were enrolled through non-probability consecutively sampling. However, patients undergoing re-exploration procedures or those unwilling to provide informed consent were excluded. The sample size was calculated using an expected postoperative mortality rate of 14.6%, margin of error 5%, and 95% CI. The estimated sample size came out to be 192.

After the clinical decision for surgery was made, demographic and physiological data were recorded preoperatively. The APACHE-II score ranged from 0 to 71, with higher scores indicating increased severity of illness and a correspondingly higher risk of postoperative mortality. The estimated mortality risk based on this score was obtained using validated scoring tables or online calculators.

The P-POSSUM score was computed using twelve physiological parameters and six operative severity factors. Physiological data was collected before surgery, while operative variables were documented intraoperatively and postoperatively. The total physiological and operative scores were entered into a logistic regression equation to calculate the predicted mortality risk, which was expressed as a percentage. While APACHE-II relies exclusively on preoperative data and allows for early triage, the P-POSSUM scoring system includes intraoperative

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

findings, making it more suitable for postoperative audit and institutional outcome evaluation.

Interpretation of these scoring systems was standardized for the study. The APACHE-II score was interpreted as a composite measure of acute physiological derangement and chronic health burden, with scores above 20 indicating a high likelihood of poor outcome. The mortality risk derived from this score represented the estimated probability of in-hospital death. The P-POSSUM physiological score reflected the baseline clinical status, while the operative severity score captured the complexity and intraoperative challenges of the surgical intervention. The predicted mortality risk calculated using P-POSSUM represented the combined impact of physiological and surgical risk on the likelihood of death.

All data were entered into a structured proforma and patients were followed until discharge and subsequently for 30 days post-discharge through telephone interviews or scheduled follow-up visits. The primary outcome of interest was 30-day postoperative mortality, defined as any death occurring during the index hospital admission or within 30 days following discharge. Secondary outcomes included duration of hospital stay, need for postoperative mechanical ventilation, requirement of inotropic support for hemodynamic instability (mean arterial pressure less than 65 mmHg despite adequate fluid resuscitation), development of acute kidney injury diagnosed according to the KDIGO criteria, need for re exploration after the initial surgery, and occurrence of cardiac morbidity such as myocardial infarction or arrhythmias requiring medical intervention.

Statistical analysis was performed using IBM SPSS version 24. Continuous variables were summarized using means and standard deviations or medians as appropriate, while categorical variables were reported as frequencies and percentages. The distribution of continuous data was assessed for normality using the Shapiro-Wilk test. Between-group comparisons for continuous variables were conducted using the independent sample t-test or the Mann–Whitney U test, depending on the data distribution. The Chi-square test was used for comparison of categorical variables. To assess the predictive accuracy of the APACHE-II and P-POSSUM scoring systems for 30-day mortality, receiver operating characteristic (ROC) curve analysis was performed. The area under the ROC curve (AUC) was calculated for each score to evaluate their discriminatory power, with values closer to 1.0 indicating superior predictive performance.

Results

Of 192 patients, the mean age of the patients was 49.81 ±18.58 years. There were 103 (53.6%) males and 89 (46.4%) females. The mean height, weight and BMI of the patients were 166.35 ±9.92 cm, 71.05 ±15.43 kg, and 25.94 ±6.45 kg/m^2 respectively. Diabetes mellitus was observed in 51 (26.6%), hypertension in 69 (35.9%) and chronic kidney disease in 25 (13.0%) patients. The mean length of hospital stay was 16.14 ±7.44 days.

Ventilatory support was provided in 46 (24.0%) patients whereas inotropic support was provided in 35 (18.2%) patients. The mean APACHE-II predictive mortality score was 49.07 ±24.52 whereas mean P-POSSUM predictive mortality score was 58.67 ±25.67. The 30-day postoperative mortality was observed in 24 (12.5%) patients.

Figure 1 illustrates the ROC curves comparing the predictive performance of APACHE-II mortality risk and P-POSSUM predicted mortality risk. Both scoring systems demonstrated excellent discriminative ability, with Area Under the Curve (AUC) values exceeding 0.90 (APACHE-II: AUC 0.92, p-value <0.001) and (P-POSSUM: AUC 0.95, p-value <0.001).

The optimal cutoff for APACHE-II mortality risk was identified as 27.35, yielding a sensitivity of 95.8% and specificity of 22.6%, with a corresponding Youden's Index of 0.184. For P-POSSUM, the optimal cutoff was 49.75, also achieving a sensitivity of 95.8% but with improved specificity of 45.8%, and a higher Youden's Index of 0.416. (Table 1)

Cross-tabulation analysis showed that both APACHE-II and P-POSSUM had high sensitivity (95.8%) for predicting 30-day mortality. However, P-POSSUM demonstrated superior specificity (45.8% vs. 22.6%) and overall accuracy (52.1% vs. 31.8%). (Table 2)

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Discussion

The prediction of postoperative mortality remains a cornerstone of surgical risk assessment, particularly in emergency settings where decision-making must be both timely and accurate. This study compared the prognostic utility of two widely used scoring systems—APACHE-II and P-POSSUM for predicting 30-day postoperative mortality among surgical patients in a tertiary care setting in Pakistan. Both scoring systems demonstrated high sensitivity; however, P-POSSUM outperformed APACHE-II in terms of specificity and overall diagnostic accuracy. These findings align with prior literature and support the clinical relevance of incorporating objective risk stratification tools into perioperative care.

In our cohort of 192 patients, the observed 30-day mortality rate was 12.5%. The ROC analysis showed that both APACHE-II and P-POSSUM mortality risk scores had strong discriminatory ability, with AUC values exceeding 0.90. These values reflect excellent predictive performance, consistent with those reported in international literature. The optimal cutoff value for APACHE-II yielded a high sensitivity, but relatively low specificity. In contrast, the optimal P-POSSUM cutoff maintained a similar high sensitivity but improved specificity and diagnostic accuracy. The higher Youden's Index for P-POSSUM further confirmed its superior performance in balancing sensitivity and specificity.

Several studies have evaluated the comparative performance of APACHE-II and P-POSSUM scoring systems in predicting mortality. The JLUMHS study by Ashfaq et al. (2024)[16] found a statistically significant association between increasing APACHE-II scores and mortality in patients undergoing exploratory laparotomy for peritonitis, reinforcing the predictive validity of this tool in high-risk abdominal surgery. Similarly, the Elsevier article by Yelamanchi et al. (2020) confirmed the strong discriminative ability of both APACHE-II and P-POSSUM, identifying them as superior predictors among various risk scoring models evaluated for emergency laparotomy patients. However, they prefer APACHE-II over P-POSSUM because of the easy calculation and reliance only on pre-operative parameters in APACHE-II scoring.[17] Similar findings were reported by by Nag et al. (2019) as well.[15] In the Journal of Surgery Pakistan article (2023), although POSSUM showed acceptable predictive capacity for mortality, it performed poorly for morbidity prediction, indicating the value of adopting the modified P-POSSUM variant in contemporary clinical settings.[18]

Our findings are consistent with these reports and extend their relevance to a Pakistani surgical population. While APACHE-II remains a robust and easy-to-use score based solely on physiological variables, its limitations stem from its inability to incorporate intraoperative factors, which are central to the P-POSSUM framework. The inclusion of operative severity in P-POSSUM likely accounts for its better specificity and diagnostic accuracy in our sample. Moreover, given that both scores were evaluated against actual mortality outcomes and not just intermediate endpoints, the results hold strong external validity.

The strengths of this study include its prospective design, use of validated risk scoring systems, and analysis based on clinically meaningful endpoints. Furthermore, the inclusion of patients from a tertiary care center provides a realistic reflection of surgical case complexity in resource-limited settings. However, several limitations must be acknowledged. First, the study was conducted at a single center, which may limit the generalizability of findings. Second, despite using validated cutoff values derived from ROC analysis, a relatively small number of events (i.e., deaths) may have impacted the precision of diagnostic metrics such as specificity and PPV. Third, the mortality data were limited to a 30-day window and may not capture long-term postoperative outcomes.

Future research should focus on validating these findings in larger multicenter cohorts, with stratification by type of surgery and comorbidities. Integration of these scores with electronic health records may also enhance their real-time applicability. Additionally, comparative evaluation with machine-learning models or combined scoring indices may further optimize predictive accuracy and personalize perioperative risk assessment. Despite the limitations, this study contributes valuable evidence supporting the application of APACHE-II and P-POSSUM scoring systems in surgical mortality

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

prediction and highlights P-POSSUM's potential as a more accurate tool in complex surgical populations.

Conclusion

The predictive ability of APACHE-II and P-POSSUM scores was found almost similar in predicting 30-day postoperative mortality among patients undergoing non-traumatic emergency laparotomy.

Figure 1: Receiver Operating Characteristic (ROC) curves comparing the predictive performance of APACHE-II Mortality Risk and P-POSSUM Predicted Mortality Risk for 30-day postoperative mortality.

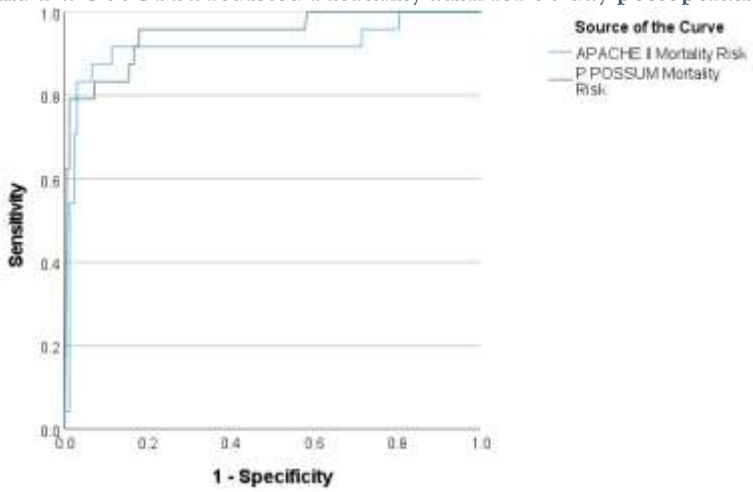

Table 1. Optimal cutoff, sensitivity, specificity, and Youden's Index for APACHE-II and P-POSSUM scores

Score System	Optimal Cutoff	Sensitivity (%)	Specificity (%)	Youden's Index
APACHE-II	27.35	27.35	95.8	22.59
P-POSSUM	49.75	49.75	95.8	45.8

Table 2: Cross-tabulation of observed 30-day mortality with predicted mortality risk using APACHE-II and P-POSSUM optimal cutoffs

Predicted Mortality	Mortality 30 days		p-value	Se	Sp	PPV	NPV	Overall Diagnostic Accuracy
	Yes	No						
	n (%)	n (%)						
APACHE-II ≥ 27.35								
Yes	23 (15.0)	130 (85.0)	0.032	95.83	22.62	15.03	97.44	31.77
No	1 (2.6)	38 (97.4)						
P-POSSUM ≥ 49.75								
Yes	23 (20.2)	91 (79.8)	<0.001	95.83	45.83	20.18	98.72	52.08
No	1 (1.3)	77 (98.7)						

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

References

Fagan G, Barazanchi A, Coulter G, Leeman M, Hill AG, Eglinton TW. New Zealand and Australia emergency laparotomy mortality rates compare favourably to international outcomes: a systematic review. ANZ J Surg. 2021 Dec;91(12):2583-2591. doi: 10.1111/ans.16563.

Ng ZQ, Weber D. One-Year Outcomes Following Emergency Laparotomy: A Systematic Review. World J Surg. 2022 Mar;46(3):512-523. doi: 10.1007/s00268-021-06385-w.

Hussain A, Mahmood F, Teng C, Jafferbhoy S, Luke D, Tsiamis A. Patient outcome of emergency laparotomy improved with increasing "number of surgeons on-call" in a university hospital: Audit loop. Ann Med Surg (Lond). 2017 Sep 28;23:21-24. doi: 10.1016/j.amsu.2017.09.013.

Neary WD, Heather BP, Earnshaw JJ. The Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (POSSUM). Br J Surg. 2003 Feb;90(2):157-65. doi: 10.1002/bjs.4041.

Mercer S, Guha A, Ramesh V. The P-POSSUM scoring systems for predicting the mortality of neurosurgical patients undergoing craniotomy: Further validation of usefulness and application across healthcare systems. Indian J Anaesth. 2013 Nov;57(6):587-91. doi: 10.4103/0019-5049.123332.

Nag DS. Assessing the risk: Scoring systems for outcome prediction in emergency laparotomies. Biomedicine (Taipei). 2015 Dec;5(4):20. doi: 10.7603/s40681-015-0020-y.

Huddart S, Peden CJ, Swart M, McCormick B, Dickinson M, Mohammed MA, Quiney N; ELPQuiC Collaborator Group; ELPQuiC Collaborator Group. Use of a pathway quality improvement care bundle to reduce mortality after emergency laparotomy. Br J Surg. 2015 Jan;102(1):57-66. doi: 10.1002/bjs.9658.

Rix TE, Bates T. Pre-operative risk scores for the prediction of outcome in elderly people who require emergency surgery. World J Emerg Surg. 2007 Jun 5;2:16. doi: 10.1186/1749-7922-2-16.

Chieng TH, Roslan AC, Chuah JA. Risk-adjusted analysis of patients undergoing laparotomy using POSSUM and P-POSSUM score in Queen Elizabeth Hospital, Sabah. Med J Malaysia. 2010 Dec;65(4):286-90.

Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. Br J Surg. 1998 Sep;85(9):1217-20. doi: 10.1046/j.1365-2168.1998.00840.x.

Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE-II: a severity of disease classification system. Crit Care Med. 1985 Oct;13(10):818-29.

Vincent JL, Moreno R. Clinical review: scoring systems in the critically ill. *Crit Care.* 2010;14(2):207. doi: 10.1186/cc8204.

Oliver CM, Walker E, Giannaris S, Grocott MP, Moonesinghe SR. Risk assessment tools validated for patients undergoing emergency laparotomy: a systematic review. *Br J Anaesth.* 2015;115:849–860.

Ahmed A, Azim A. Emergency Laparotomies: Causes, Pathophysiology, and Outcomes. Indian J Crit Care Med. 2020 Sep;24(Suppl 4):S183-S189. doi: 10.5005/jp-journals-10071-23612.

Nag DS, Dembla A, Mahanty PR, Kant S, Chatterjee A, Samaddar DP, Chugh P. Comparative analysis of APACHE-II and P-POSSUM scoring systems in predicting postoperative mortality in patients undergoing emergency laparotomy. World J Clin Cases. 2019 Aug 26;7(16):2227-2237. doi: 10.12998/wjcc.v7.i16.2227.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Ashfaq H, Ahmed R, Kumar D, Kazim E, Dastagir S, Gul A. The Prediction of Mortality Using APACHE-II Score in Patients Presenting with Peritonitis and Undergoing Emergency Exploratory Laparotomy. Journal of Liaquat University of Medical & Health Sciences. 2024 Apr 3;23(01):21-5.

Yelamanchi R, Gupta N, Durga CK, Korpal M. Comparative study between P- POSSUM and APACHE-II scores in predicting outcomes of perforation peritonitis: Prospective observational cohort study. Int J Surg. 2020 Nov;83:3-7. doi: 10.1016/j.ijsu.2020.09.006.

Rizvi SAF, Gul A, Khan MU, Rajput A, Anwer E, Ahmad R. POSSUM System In Predicting Morbidity and Mortality In Emergency Laparotomy Cases. Journal of Surgery Pakistan. 2023 Nov 21;28(3):79-83.