Original Paper

Preoperative Mortality Risk Prediction and Equity Using Preoperative Score to Predict Postoperative Mortality in a Diverse US Surgical Population: Retrospective Cohort Study

Jackson Hall*, BS; Haley Warzecha*, BA; Chase Penhallurick, BA; Venkatsai Bellala, BS; Shyamal Asher, MD, MBA

Department of Anesthesiology, Brown University Health, Brown University, Providence, RI, United States
*these authors contributed equally

Corresponding Author:

Shyamal Asher, MD, MBA
Department of Anesthesiology
Brown University Health, Brown University
593 Eddy Street, Davol 129
Providence, RI 02903
United States
Phone: 1 401-444-5172
Email: Shyamal_asher@brown.edu

Abstract

Background: The Preoperative Score to Predict Postoperative Mortality (POSPOM), a point-based score using only preoperative variables, was developed and validated to predict in-hospital postoperative mortality in European surgical cohorts. Other widely used clinical risk tools perform differently across racial and socioeconomic groups. Although postoperative mortality has been studied in diverse US surgical populations, POSPOM performance across racial and ethnic subgroups remains insufficiently characterized.

Objective: This study aimed to evaluate POSPOM performance for 30-day postoperative mortality in a diverse US cohort and explore its performance for 1-year mortality among high-risk surgical patients and across racial and ethnic subgroups.

Methods: We performed a retrospective cohort study using the All of Us Research Program. Adults aged 18 to 91 years who underwent 1 of 21 POSPOM-eligible surgical procedures were included, and only each participant's most recent eligible procedure was analyzed. POSPOM scores were calculated from age, procedure type, and 15 weighted comorbidities. The primary outcome was 30-day postoperative mortality, as incomplete discharge data precluded reliable assessment of in-hospital mortality; 1-year mortality was examined as an exploratory secondary outcome. Logistic regression estimated odds ratios (ORs) per 1-point increase in the POSPOM score, and discrimination was evaluated using the area under the receiver operating characteristic curve (AUC). A high-risk subgroup (procedure points ≥14) was analyzed separately, and outcomes across racial and ethnic subgroups were explored descriptively. Given the small subgroup sizes and few events, subgroup analyses were interpreted cautiously.

Results: Among 17,262 patients, 12 (0.1%) died within 30 days and 45 (0.3%) within 1 year. For 30-day mortality, POSPOM yielded an AUC of 0.764, with higher odds of death per point (OR 1.112, 95% CI 1.040-1.190; *P*=.002). Discrimination was lower for 1-year mortality (AUC 0.638; OR 1.054, 95% CI 1.017-1.092; *P*=.004). In the high-risk subgroup (n=536; 30-day mortality 1.1%, 1-year mortality 2.1%), the 30-day AUC fell to 0.583, with no significant association (OR 0.942, 95% CI 0.812-1.092; *P*=.43). For 1-year mortality, the AUC was 0.692, with an inverse association (OR 0.882, 95% CI 0.781-0.998; *P*=.046). Descriptively, high-risk Black patients had higher 30-day mortality than White patients (2/61, 3.3% vs 4/348, 1.1%); no deaths were recorded among Asian, other-race, or Hispanic high-risk patients.

Conclusions: POSPOM demonstrated an AUC of 0.764 for 30-day mortality overall but provided limited risk stratification among high-risk surgical patients. Higher observed mortality among high-risk Black patients suggests the score may not fully capture baseline risk across racial groups, possibly reflecting structural and social factors outside its scope. As the 1-year and subgroup analyses were exploratory with few events, larger, adequately powered US studies are needed to evaluate POSPOM at its intended in-hospital end point, assess calibration across diverse populations, and inform recalibration before broader clinical use.

JMIR Perioper Med 2026;9:e99623; doi: 10.2196/99623

Keywords: preoperative care; postoperative mortality; risk assessment; health care disparities; socioeconomic factors; comorbidity

Introduction

The Preoperative Score to Predict Postoperative Mortality (POSPOM) is a validated, point-based tool that estimates postoperative mortality using only preoperative variables, including age, comorbidities, and surgical procedure type [1]. Developed to support early risk stratification, POSPOM intentionally excludes subjective clinician judgment and intraoperative data, making it relatively simple to implement across different hospitals and practice settings [1]. POSPOM was originally developed and validated to predict in-hospital postoperative mortality, with "good calibration and excellent discrimination" reported in large European surgical cohorts [1].

Postoperative mortality has previously been examined in large, demographically diverse US surgical cohorts. Smith et al [2] evaluated postoperative mortality risk estimates and 1-year survival among Veterans Health Administration surgical patients, while Kaiser et al [3] examined 30-day mortality in a large multicenter cohort of noncardiac surgical patients. However, these studies did not specifically evaluate POSPOM performance across racial and ethnic subgroups. Prior studies have shown that widely used clinical prediction tools, including the Sequential Organ Failure Assessment (SOFA) score, the American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator, and cardiovascular risk equations, can exhibit differential performance across racial and socioeconomic groups [1,4-8]. These findings have raised concerns that risk models may not perform equally across different patient populations, treatment decisions, and outcomes [1,4-8].

Although prior studies have examined postoperative mortality prediction in diverse US surgical populations, POSPOM performance across racial and ethnic subgroups has not been specifically characterized [2,3]. Evaluating POSPOM across these subgroups may help determine whether its predictive performance remains consistent across heterogeneous surgical populations.

Hospital discharge dates were not available for all participants in the All of Us dataset, preventing reliable assessment of in-hospital mortality across the full cohort [9]. We therefore selected 30-day mortality as the primary outcome because it provided a standardized postoperative interval that could be applied consistently across participants. We additionally examined 1-year mortality as an exploratory secondary outcome to assess the longer-term association between preoperative POSPOM scores and postoperative mortality.

The primary objective of this study was to evaluate POSPOM performance for 30-day postoperative mortality in a large, demographically diverse US cohort from the All of Us Research Program [9]. Secondary exploratory objectives were to evaluate performance for 1-year mortality, performance among high-risk surgical patients, and performance across racial and ethnic subgroups. We hypothesized that, similar to other clinical risk prediction tools, POSPOM may not perform consistently across demographic subgroups, particularly among patients undergoing high-risk surgical procedures [1,4-6].

Methods

Study Design and Data Source

This retrospective cohort study evaluated POSPOM performance for 30-day postoperative mortality in a large-scale, diverse, real-world US population, with secondary exploratory analyses of 1-year mortality and demographic subgroups. Data were derived from the National Institutes of Health (NIH) All of Us Research Program's Registered Tier Dataset version 8, which includes electronic health record (EHR) data from more than 1 million participants [9]. Clinical information was extracted from the Observational Medical Outcomes Partnership (OMOP) Common Data Model version 5.3.1, using tables for demographics, procedures, condition occurrences, surveys, and death records [10]. This retrospective study was not prospectively registered, and no formal study protocol was published.

Study Population

The initial cohort was identified from participants with recorded condition and procedure domains. To ensure independent observations among participants with multiple eligible surgeries, only the most recent procedure was included for the analysis. Participants were included in the final analytic cohort if they were aged 18 to 91 years at the time of data collection and underwent 1 of the 21 POSPOM-eligible surgical procedures.

Participants were excluded from the analysis if they fell outside the specified age range, had missing demographic or surgical data required for cohort eligibility or POSPOM score calculation, or underwent surgical procedures not covered by the POSPOM classification system.

Following these criteria, the final analytic cohort consisted of 17,262 patients.

Data Extraction and Processing

Demographic variables, including sex at birth, race, and ethnicity, were cleaned and standardized for analysis. Race was categorized as White, Black, Asian, or other, and ethnicity was coded as a binary variable (Hispanic vs non-Hispanic). Sex at birth was recoded as male or female.

POSPOM Score Calculation

Each participant's total POSPOM score was computed as the sum of 3 components—age, procedure type, and comorbidities—with total possible scores ranging from 0 to 54 points. The complete scoring criteria, including all variables and point allocations, are detailed in Table 1.

Table 1. Components and point allocations for the Preoperative Score to Predict Postoperative Mortality (POSPOM)[a].

Category and characteristic	POSPOM points
Patient age (y)	
<45	0
45-49	1
50-54	2
55-59	3
60-64	4
65-69	5
70-74	6
75-79	7
80-84	8
85-89	9
≥90	10
Comorbidities	
Ischemic heart disease	1
Cardiac arrhythmia or heart blocks	1
Chronic heart failure or cardiomyopathy	4
Peripheral vascular disease	1
Dementia	2
Cerebrovascular disease	1
Hemiplegia	4
Chronic obstructive pulmonary disease	1
Chronic respiratory failure	3
Chronic alcohol abuse	4
Cancer	4
Diabetes	1
Transplanted organs	2
Preoperative chronic dialysis	1
Chronic renal failure	2
Surgical category	
Endoscopic digestive	0
Ophthalmologic	0
Gynecologic	6
Other orthopedic	6
Interventional cardiorhythmology	8
Arthroplasty and spine	9
Ear, nose, and throat	9
Minor urologic	9
Plastic	9
Major urologic	12
Other surgery	12
Minor hepatic	12
Minor gastrointestinal	13
Renal transplant	13
Minor vascular	13
Orthopedic trauma	14

Category and characteristic	POSPOM points
Major hepatic	15
Thoracic	15
Neurosurgery	15
Major vascular	16
Major gastrointestinal	16
Interventional neuroradiology	17
Cardiac	17
Transplant	22
Multiple trauma related	22

[a]Adapted from Le Manach et al [1].

Points were assigned based on 5-year age brackets at the time of surgery, ranging from 0 points for patients younger than 45 years to 10 points for patients aged 90 years or older, as detailed in Table 1.

Procedures were mapped to POSPOM categories using Current Procedural Terminology and Systematized Nomenclature of Medicine Clinical Terms codes. Points associated with these procedures ranged from 0 for low-risk procedures (eg, colonoscopy and cataract surgery) to 22 for high-risk procedures (eg, liver transplantation).

Fifteen distinct comorbidities were identified using OMOP concept identifiers and mapped to diagnostic codes. To ensure comprehensive capture, structured query language queries used OMOP concept identifiers to include all descendant codes. Weights were applied to each condition, ranging from 1 point for conditions such as diabetes or chronic obstructive pulmonary disease to 4 points for chronic heart failure or malignant neoplastic disease.

Outcome Measures

The primary outcome was 30-day postoperative mortality, defined as death occurring within 30 days of the index surgical procedure. POSPOM was originally developed to predict in-hospital mortality; however, hospital discharge dates were not available for all participants, preventing reliable assessment of this outcome across the full cohort [1]. Thirty-day mortality was therefore selected as a standardized postoperative outcome that could be consistently assessed across participants. The secondary outcome was 1-year postoperative mortality, defined as death occurring within 365 days of the index surgical procedure. Mortality status and dates were derived from the OMOP death domain, and time from surgery to death was calculated. Participants without a recorded death date in the EHR were assumed to have survived through the relevant follow-up interval.

Statistical Analysis

A logistic regression model was used for the primary analysis, with the continuous POSPOM score as the predictor and 30-day mortality as the binary outcome. Model performance was evaluated through discrimination and calibration. Discrimination was measured by the concordance statistic (C-statistic or area under the receiver operating characteristic curve [AUC]). In accordance with recent consensus, qualitative labels for model discrimination were avoided, and continuous AUC values were reported directly [11]. Calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test and calibration plots comparing observed vs predicted mortality rates. Effect estimation used logistic regression to calculate odds ratios (ORs) for mortality per 1-point increase in the POSPOM score, reported with 95% CIs.

To evaluate model fairness and calibration across different demographic strata, a stratified analysis was conducted specifically for high-risk procedures, defined as a POSPOM procedure score of 14 or greater [1]. Exploratory subgroup analyses were conducted by race and ethnicity.

Data Quality and Validation

Data quality checks included assessment of missingness for key variables, validation of date ranges (restricting surgery dates to 1980-2023), verification of age calculations (≥18 years), and evaluation of mortality data completeness. The low number of mortality events limited statistical precision, particularly in subgroup analyses; these findings were therefore interpreted cautiously.

Software and Tools

Data extraction and processing were performed using Python (version 3.10; Python Software Foundation) with the pandas and NumPy libraries. Statistical analyses were conducted using *statsmodels* and scikit-learn, while visualizations were generated with Matplotlib and Seaborn. All queries and data management were conducted through the BigQuery environment within the All of Us researcher workbench.

Ethical Considerations

All data were deidentified in accordance with All of Us privacy protocols. No direct participant contact or intervention occurred. Analyses were performed under an approved data use agreement, and all findings are presented in aggregate form to protect participant confidentiality. The Brown University Institutional Review Board (IRB) formally reviewed this study and determined that it did not constitute human subjects research under federal regulations because it exclusively used deidentified registered tier data from the NIH All of Us Research Program. The overarching All of Us data collection protocol is independently reviewed and approved by the program's central IRB. Patients and

members of the public were not directly involved in the design, conduct, analysis, or reporting of this study.

Results

Cohort Characteristics and Overall Mortality

The study analyzed an analytic cohort of 17,262 participants who underwent POSPOM-eligible surgical procedures. Within this overall population, the absolute mortality event rate was low: 12 (0.1%) participants died within 30 days of their index surgery and 45 (0.3%) participants died within 1 year. A subset of 536 (3.1%) participants were categorized as high-risk based on a POSPOM surgical procedure score of 14 or greater. Mortality rates were notably higher in this high-risk subset, with 6 (1.1%) deaths occurring at 30 days and 11 (2.1%) deaths occurring at 1 year. The detailed predictive performance of the POSPOM score across these groups is summarized in Table 2.

Table 2. Predictive performance of the Preoperative Score to Predict Postoperative Mortality (POSPOM) score across the overall cohort and high-risk subgroups.

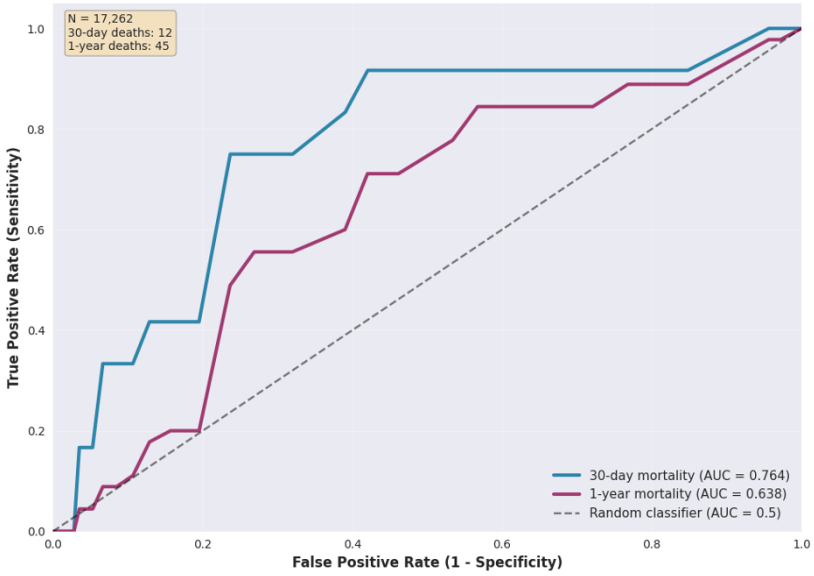
Group	Outcome type	Participants, n	Deaths, n (%)	AUC[a]	OR[b] (95% CI)	*P* value
All patients	30-d	17,262	12 (0.1)	0.764	1.112 (1.040-1.190)	.002
All patients	1-y	17,262	45 (0.3)	0.638	1.054 (1.017-1.092)	.004
High-risk (POSPOM score ≥14)	30-d	536	6 (1.1)	0.583	0.942 (0.812-1.092)	.43
High-risk (POSPOM score ≥14)	1-y	536	11 (2.1)	0.692	0.882 (0.781-0.998)	.046
Asian (High-risk)	30-d	9	0 (0)	—[c]	—	—
Black (High-risk)	30-d	61	2 (3.3)	—	—	—
Other (High-risk)	30-d	118	0 (0)	—	—	—
White (High-risk)	30-d	348	4 (1.2)	—	—	—

[a]AUC: area under the receiver operating characteristic curve.
[b]OR: odds ratio.
[c]Not applicable.

Overall POSPOM Score Performance

In the primary analysis of the overall cohort, the POSPOM model demonstrated an AUC of 0.764 for 30-day mortality (Figure 1). Logistic regression indicated that each 1-point increase in the POSPOM score was associated with a statistically significant increase in the odds of 30-day mortality (OR 1.112, 95% CI 1.040-1.190; *P*=.002). When evaluating longer-term outcomes, the model's discriminative ability decreased; for 1-year mortality, the model produced an AUC of 0.638 (Figure 1), and each 1-point score increase was associated with a smaller increase in mortality odds (OR 1.054, 95% CI 1.017-1.092; *P*=.004).

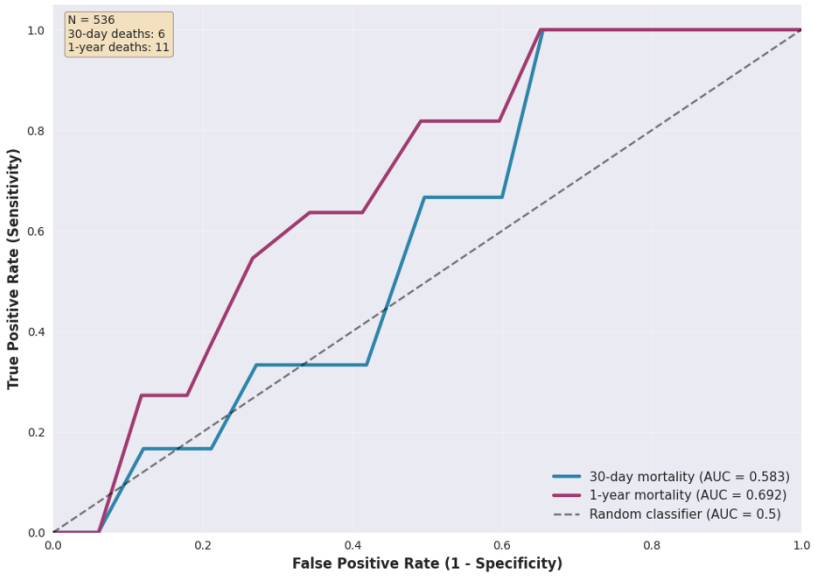
Figure 1. Receiver operating characteristic curves showing Preoperative Score to Predict Postoperative Mortality score performance for 30-day vs 1-year mortality in the overall cohort. AUC: area under the receiver operating characteristic curve.



Performance in High-Risk and Demographic Subgroups

When the analysis was restricted to the high-risk surgical cohort (n=536), model discrimination dropped considerably. For 30-day mortality in this subgroup, the POSPOM score yielded an AUC of 0.583 (Figure 2), with the score no longer showing a positive association with mortality (OR 0.942, 95% CI 0.812-1.092; *P*=.43). For 1-year mortality in the high-risk cohort, the AUC was 0.692 (Figure 2), and a 1-point increase in the POSPOM score was inversely associated with mortality odds (OR 0.882, 95% CI 0.781-0.998; *P*=.046).

Figure 2. Receiver operating characteristic curves showing Preoperative Score to Predict Postoperative Mortality score performance for 30-day vs 1-year mortality in the high-risk surgical cohort. AUC: area under the receiver operating characteristic curve.



Race-stratified analyses within the high-risk cohort were evaluated descriptively due to small sample sizes. Among Black patients (n=61), there were 2 (3.3%) recorded 30-day deaths. Among White patients (n=348), there were 4 (1.2%) recorded 30-day deaths. No mortality events were recorded among Asian (n=9) or other (n=118) racial groups within the high-risk cohort. In ethnicity-stratified analyses, the Hispanic subgroup (n=72) recorded no mortality events, while the non-Hispanic group (n=464) recorded 11 deaths at 1 year.

Discussion

Principal Findings

This retrospective cohort study evaluated POSPOM performance in a diverse US surgical population, with an exploratory assessment of outcomes across racial and ethnic subgroups. For 30-day mortality in the overall surgical cohort, POSPOM demonstrated an AUC of 0.764, but its performance decreased among high-risk patients (AUC 0.583) and for the 1-year outcome (AUC 0.638). In exploratory subgroup analyses, discrimination appeared to be higher among Black participants; however, event counts were small and should be interpreted cautiously. Although a higher absolute mortality rate was observed among Black patients compared with White patients within this high-risk subset, our data are insufficient to conclude whether the model differentially accounts for underlying baseline risk. These findings underscore the need for larger, adequately powered US validation studies to assess model calibration across minoritized populations.

General Performance and High-Risk Limitations

The initial validation of POSPOM in large European cohorts suggested that it was a robust and objective alternative to more subjective risk assessments [1,4]. The AUC of 0.764 observed for 30-day mortality supports the potential utility of POSPOM for standardized short-term postoperative risk assessment in diverse US populations. However, the decline in discrimination among high-risk surgical patients (AUC 0.583 for 30-day mortality in those with surgical points ≥14) indicates limited predictive performance in this subgroup. In these patients, the score did not provide meaningful risk stratification. For the 1-year outcome, an inverse association was observed within the high-risk subgroup. Given the small number of events and the fact that POSPOM was not developed for long-term mortality prediction, this finding should be interpreted cautiously and warrants evaluation in a larger cohort.

The outcomes evaluated in this study differ from the in-hospital mortality end point for which POSPOM was originally developed [1]. Because hospital discharge dates were unavailable for some participants, in-hospital mortality could not be reliably assessed across the full cohort. We therefore used 30-day mortality as a standardized primary outcome and examined 1-year mortality as an exploratory secondary outcome. Future studies with complete admission and discharge data should evaluate POSPOM performance at its intended in-hospital end point in diverse US populations.

Given the low number of events, all of the estimates, especially those from subgroup analyses, are subject to statistical uncertainty.

Equity and Racial Disparities

A central objective of this study was to assess whether POSPOM exhibits performance disparities across demographic lines, similar to other clinical tools [4,6,8]. Due to small sample sizes in the high-risk subgroup, outcomes across racial groups were evaluated descriptively. Notably, despite similar preoperative risk categorization, Black patients had a 30-day mortality rate 3 times higher than that of White patients (2/61, 3.3% vs 4/348, 1.2%) and a 1-year mortality rate more than triple that of their White counterparts (6.6% vs 2%).

These descriptive findings suggest that the point-based POSPOM system may underrepresent the absolute risk faced by Black patients. This disparity likely reflects broader structural inequities in health care access, the impact of social determinants of health, or differences in hospital quality that the POSPOM score is not designed to capture [1]. Although we did not analyze insurance status, hospital volume, or nonsurgical participants, prior literature indicates these systemic factors may heavily influence US surgical outcomes [12-14]. As a purely clinical tool, POSPOM does not account for these external health care system variances [1,4].

Comparison With Existing Risk Models

The performance variations and racial mortality gaps observed in this study align with emerging literature regarding algorithmic bias in clinical risk calculators [5,8,14-17]. For instance, studies evaluating the SOFA score have demonstrated that this model may overpredict mortality for certain racial groups while underpredicting mortality for others, potentially due to the inclusion of variables that do not account for baseline physiological differences or chronic exposure to systemic stressors [8,14,17]. Similarly, investigations into the National Surgical Quality Improvement Program surgical risk calculator have found that while the model provides strong aggregate discrimination, it often underestimates complications and mortality for minority patients and those from lower socioeconomic backgrounds [5,15,16]. This study suggests that POSPOM exhibits a similar pattern where the model fails to calibrate for the higher baseline incidence of adverse outcomes in minoritized cohorts. This consistent trend across various clinical tools suggests that preoperative risk models may require the inclusion of social and structural variables to achieve true equity in predictions.

Limitations in Subgroup Analyses

The absence of mortality events among several racial and ethnic subgroups highlights the small sample size challenge associated with these exploratory analyses [18]. Similarly, the absence of events in the Hispanic subgroup limited the ability to assess POSPOM performance in this population. Larger sample sizes are needed to evaluate fairness and calibration across ethnic groups. In addition, the extremely low number of mortality events in the high-risk surgical subgroup limits the statistical power of the model's discrimination in this specific cohort; therefore, our high-risk analyses should be interpreted as exploratory and signal-generating rather than definitive. Furthermore, while the All of Us database provided a more diverse sample than many traditional datasets, the overall low event rate for 30-day mortality (0.1%) limits the precision of our estimates [9]. These exploratory findings primarily underscore the need for larger, fully powered datasets to accurately evaluate subgroup performance.

Clinical and Policy Implications

These results exemplify the potential danger of assuming that a clinically validated tool is neutral simply because it avoids subjective physician input. If POSPOM is used to guide surgical decision-making or resource allocation without recalibration for US populations, it may inadvertently exacerbate existing health disparities [6,19,20]. For example, if a high-risk Black patient receives a risk score that discriminates well but overlooks a systematically higher baseline mortality, they may be given less aggressive preoperative optimization or postoperative surveillance than their true risk warrants. Ultimately, our findings suggest that the clinical utility of POSPOM in diverse US health care settings would be significantly enhanced by regional and demographic recalibration. Prioritized next steps should include using larger, multicenter US registries to reweight the tool's clinical variables and explicitly investigating whether the integration of specific social determinants of health can optimize baseline calibration across heterogeneous patient populations.

Generative AI Use

The authors used Cursor (an AI-powered editor) to help draft, debug, and refactor Python code for data processing and statistical analysis. All queries, analytic decisions, and outputs were reviewed and validated by the authors, who retain full responsibility for the manuscript.

Funding

No funding was received for this study.

Data Availability

The analytical code used for data processing and statistical analysis is not publicly available.
The data used in this study are available to registered, approved researchers through the National Institutes of Health All of Us Research Program's cloud-based platform. Detailed information on how to apply for data access can be found on the All of Us Research Program website [21].

Authors' Contributions

Conceptualization: JH (lead), HW (equal)
Formal analysis: HW (lead)
Methodology: HW, JH, CP
Project administration: JH (lead), HW (equal)
Supervision: SA
Visualization: HW, CP
Writing—original draft: JH (lead), CP (supporting), VB (supporting)
Writing—review and editing: JH, HW, CP, VB, SA

Conflicts of Interest

None declared.

References

1. Le Manach Y, Collins G, Rodseth R, et al. Preoperative Score to Predict Postoperative Mortality (POSPOM): derivation and validation. Anesthesiology. Mar 2016;124(3):570-579. [doi: 10.1097/ALN.0000000000000972] [Medline: 26655494]
2. Smith T, Li X, Nylander W, Gunnar W. Thirty-day postoperative mortality risk estimates and 1-year survival in Veterans Health Administration surgery patients. JAMA Surg. May 1, 2016;151(5):417-422. [doi: 10.1001/jamasurg.2015.4882] [Medline: 26747331]
3. Kaiser HA, Saied NN, Kokoefer AS, Saffour L, Zoller JK, Helwani MA. Incidence and prediction of intraoperative and postoperative cardiac arrest requiring cardiopulmonary resuscitation and 30-day mortality in non-cardiac surgical patients. PLoS One. Jan 22, 2020;15(1):e0225939. [doi: 10.1371/journal.pone.0225939] [Medline: 31967987]
4. Layer YC, Menzenbach J, Layer YL, et al. Validation of the Preoperative Score to Predict Postoperative Mortality (POSPOM) in Germany. PLoS One. Jan 27, 2021;16(1):e0245841. [doi: 10.1371/journal.pone.0245841] [Medline: 33503043]
5. Vigna C, Ore AS, Fabrizio A, Messaris E. Evaluation of racial/ethnic disparities in surgical outcomes after rectal cancer resection: an ACS-NSQIP analysis. Surg Pract Sci. Apr 17, 2024;17:100248. [doi: 10.1016/j.sipas.2024.100248] [Medline: 39845641]
6. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. Oct 25, 2019;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
7. Brindle P, May M, Gill P, et al. Primary prevention of cardiovascular disease: a web-based risk score for seven British Black and minority ethnic groups. Heart. Nov 2006;92(11):1595-1602. [doi: 10.1136/hrt.2006.092346] [Medline: 16762981]
8. Ashana DC, Anesi GL, Liu VX, et al. Equitably allocating resources during crises: racial differences in mortality prediction models. Am J Respir Crit Care Med. Mar 30, 2021;204(2):178-186. [doi: 10.1164/rccm.202012-4383OC] [Medline: 33751910]
9. All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The "All of Us" Research Program. N Engl J Med. Aug 15, 2019;381(7):668-676. [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
10. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19(1):54-60. [doi: 10.1136/amiajnl-2011-000376] [Medline: 22037893]
11. de Hond AA, Steyerberg EW, van Calster B. Interpreting area under the receiver operating characteristic curve. Lancet Digit Health. Dec 2022;4(12):e853-e855. [doi: 10.1016/S2589-7500(22)00188-1] [Medline: 36270955]
12. Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA Surg. May 2014;149(5):475-481. [doi: 10.1001/jamasurg.2013.5076] [Medline: 24623106]
13. Wakeam E, Hevelone ND, Maine R, et al. Failure to rescue in safety-net hospitals: availability of hospital resources and differences in performance. JAMA Surg. Mar 2014;149(3):229-235. [doi: 10.1001/jamasurg.2013.3566] [Medline: 24430015]
14. Tolchin B, Oladele C, Galusha D, et al. Racial disparities in the SOFA score among patients hospitalized with COVID-19. PLoS One. Sep 17, 2021;16(9):e0257608. [doi: 10.1371/journal.pone.0257608] [Medline: 34535009]

15. Matar DY, Knoedler S, Matar AY, et al. Surgical outcomes and sociodemographic disparities across all races: an ACS-NSQIP and NHIS multi-institutional analysis of over 7.5 million patients. Ann Surg Open. Jul 16, 2024;5(3):e467. [doi: 10.1097/AS9.0000000000000467] [Medline: 39310358]
16. Betancourt-Garcia MM, Vatcheva K, Gupta PK, et al. The effect of Hispanic ethnicity on surgical outcomes: an analysis of the NSQIP database. Am J Surg. Apr 2019;217(4):618-633. [doi: 10.1016/j.amjsurg.2018.10.004] [Medline: 30466953]
17. Sarkar R, Martin C, Mattie H, Gichoya JW, Stone DJ, Celi LA. Performance of intensive care unit severity scoring systems across different ethnicities in the USA: a retrospective observational study. Lancet Digit Health. Apr 2021;3(4):e241-e249. [doi: 10.1016/S2589-7500(21)00022-4] [Medline: 33766288]
18. Walker R, Behman R, Bubis L, Karanicolas P. Prospective and retrospective cohort studies. In: Thoma A, Sprague S, Voineskos SH, Serrano PE, Goldsmith CH, editors. Evidence-Based Surgery: A Guide to Understanding and Interpreting the Surgical Literature. Springer; 2025:191-204. [doi: 10.1007/978-3-031-87083-5_17]
19. Chin MH, Afsar-Manesh N, Bierman AS, et al. Guiding principles to address the impact of algorithm bias on racial and ethnic disparities in health and health care. JAMA Netw Open. Dec 1, 2023;6(12):e2345050. [doi: 10.1001/jamanetworkopen.2023.45050] [Medline: 38100101]
20. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight - reconsidering the use of race correction in clinical algorithms. N Engl J Med. Aug 27, 2020;383(9):874-882. [doi: 10.1056/NEJMms2004740] [Medline: 32853499]
21. All of Us Research Program. URL: https://www.nih.gov/allofus [Accessed 2026-11-13]

Abbreviations

AUC: area under the receiver operating characteristic curve
EHR: electronic health record
IRB: institutional review board
NIH: National Institutes of Health
OMOP: Observational Medical Outcomes Partnership
OR: odds ratio
POSPOM: Preoperative Score to Predict Postoperative Mortality
SOFA: Sequential Organ Failure Assessment

Edited by Ivan Steenstra, Kenneth Cummings; peer-reviewed by Daniel Riddle, Kate Fultz Hollis; submitted 28.Apr.2026; final revised version received 27.Oct.2026; accepted 27.Oct.2026; published 18.Nov.2026

Please cite as:
Hall J, Warzecha H, Penhallurick C, Bellala V, Asher S
Preoperative Mortality Risk Prediction and Equity Using Preoperative Score to Predict Postoperative Mortality in a Diverse US Surgical Population: Retrospective Cohort Study
JMIR Perioper Med 2026;9:e99623
URL: https://periop.jmir.org/2026/1/e99623
doi: 10.2196/99623

© Jackson Hall, Haley Warzecha, Chase Penhallurick, Venkatsai Bellala, Shyamal Asher. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 18.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.