Author Manuscript

Development of a perioperative risk mortality calculator for humanitarian surgical care

Christopher W. Reynolds[1]*, Hannah Wild[2,3]*†, Yen Sia Low[4], Saurabh Gombar[4], Sherry M. Wren[5]

1. Department of Surgery, University of Michigan Medicine, Ann Arbor, MI
2. Department of Surgery, University of Washington, Seattle, WA
3. Explosive Weapons Trauma Care Collective, International Blast Injury Research Network, University of Southampton, Southampton, UK
4. Atropos Health, Palo Alto CA
5. Stanford University School of Medicine, Stanford, CA

* Joint first authors

† To whom correspondence should be addressed: hbwild@uw.edu, 1959 NE Pacific Street, Seattle, WA 98195

Ethical exemption for this study was determined by the University of Washington Human Subjects Division (STUDY00020362).

Keywords: Humanitarian surgical care, perioperative mortality, low-resource settings, surgical quality improvement

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/wjs.12485.

This article is protected by copyright. All rights reserved.

Abstract:

Background: Risk models to predict perioperative mortality rates (POMR) are critical to surgical quality improvement yet are not widely adapted for use in humanitarian and low-resource settings (LRS). We developed a POMR and corresponding nomogram and calculator for use in humanitarian surgical care.

Methods: Electronic health record data from a high-income academic medical center from 2015-2019 were retrospectively extracted, selecting variables and operations specific to LRS. This development dataset was used to create a logistic regression POMR model which was then prospectively validated using data from 2022-2023 from the same institution.

Results: EHR from a total of 49,277 patients were used. The model fitted eight variables feasibly obtainable in LRS: age >65 years (OR=4.05, 95% CI: 3.71-4.43), male sex (OR=1.32, 95% CI: 1.25-1.40), GCS < 13 (OR=5.20, 95% CI: 4.73-5.73), glucose > 200 mg/dL (OR=2.19, 95% CI: 2.01-2.38), Hgb ≤ 11 g/dL (OR=2.65, 95% CI: 2.43-2.89), HR > 120 bpm (OR=2.49, 95% CI: 2.35-2.64), T > 38 degrees Celsius (OR=1.32, 95% CI: 1.19-1.45), and SBP > 180 mmHg (OR=1.18, 95% CI: 1.02-1.37). The model demonstrated high area under the curve (0.847, 0.867, 0.925), sensitivity (0.739, 0.886, 0.844), specificity (0.807, 0.780, 0.864), and negative predictive value (0.750, 0.997, 0.999) on training, holdout and prospective validation sets

Conclusion: We validated a POMR model for use in LRS using eight variables that are readily available in the target environment. This model's predictors and accompanying clinical tools of an Excel calculator and nomogram make it simultaneously comprehensive and accessible in LRS.

This article is protected by copyright. All rights reserved.

Author Manuscript

Introduction:

Global disparities in access to safe surgical care are amplified in humanitarian settings where conflict and other crises disrupt already fragile infrastructure with limited material and human resources [1,2,3]. In such environments, novel strategies to optimize surgical care are critical. Perioperative risk stratification (PORS) is widely used in high-resource health systems as a tool to quantitatively benchmark patient outcomes, assess care quality, and evaluate the impact of quality improvement interventions [4]. Despite being a critical component to delivering and evaluating surgical care quality, use of POMR predictive models has not yet been widely adopted in humanitarian surgical care [5]. At the patient level, POMR helps to categorize patients by risk and optimize resource allocation in environments with significant resource constraints [6]. At the population level, PORS is a key component for developing population-specific risk prediction models and has been shown to significantly reduce mortality in high-income health systems [7].

Numerous POMR prediction tools have been developed for care delivery in high-income settings such as the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) Surgical Risk Calculator. Such tools cannot be directly translated to humanitarian surgical care in low-resource settings (LRS) as they depend on data that are rarely available in LRS including laboratory values, knowledge of existing comorbidities, imaging, and other diagnostics [8]. To be feasibly deployed in a humanitarian context, PORS must be performed in a manner that poses minimal burden on strained personnel and does not require extensive data collection infrastructure [9]. Numerous studies have evaluated the applicability and adaptability of PORS models created for high-resource settings in LRS [10,11]. However, a recent scoping review of POMR in humanitarian settings demonstrated that most models do not encompass all relevant patient populations (e.g., both traumatic and non-traumatic conditions; children and adults; obstetrics), require data that is unavailable or too time-intensive to collect, and do not consistently correlate with patient outcomes [4].

Several attempts have been made to develop and validate PORS tools for LRS, including the African Surgical Outcomes Study (ASOS) and Kampala Trauma Score (KTS) [12,13]. While helpful models for the contexts in which they were validated, these tools are insufficient to apply as a universal standard in humanitarian settings. Specifically, ASOS relies only on static demographic characteristics (e.g. age, sex) and urgency/type of surgery without taking into account metrics of patient physiology although validated models from high-income settings (e.g., the Revised Trauma Score) have demonstrated the importance of incorporating physiologic parameters such as vitals, Glasgow Coma Score (GCS), and basic laboratory values [14]. The KTS has only

This article is protected by copyright. All rights reserved.

Author Manuscript

been used for trauma patients, and both models exclude specific populations including pediatric and pregnant patients [15]. While a pediatric version (ASOS Peads) has been developed, these algorithms are designed to be population-specific and therefore lose the benefit of wide applicability independent of patient characteristics and presentation. There remains a need for a single model encompassing all relevant populations for feasible use in highly resource-constrained humanitarian settings [16].

To address the need for PORS models suitable for humanitarian settings, we developed a prototype POMR calculator designed to meet three objectives: 1) utilizes variables that can be feasibly obtained in LRS, 2) specific POMR value can be computed with minimal technological infrastructure, and 3) risk can be easily interpretable in a clinical setting to a broad patient population. We then developed an accompanying nomogram visual aid and Excel file that can be used to calculate patient risk with limited technological capabilities.

Methods

Data collection and screening

As there is no comprehensive patient data registry from humanitarian settings, to develop a POMR model generalizable to humanitarian surgical care, we used electronic health records (EHR) from a high-resource setting (Stanford Hospital) but considered only procedures that would likely be performed in the target context (LRS). In all, 3498 surgical *Current Procedural Terminology* (*CPT*) codes were grouped by the Healthcare Cost and Utilization Project Clinical Classifications Software for Services and Procedures [17]. As previously published, 11 major surgical procedure categories were combined to create major categories, defined as cardiovascular; cataract; ear, nose, and throat (ENT); general surgical; musculoskeletal; nervous system; obstetrics and gynecology; skin; thoracic; transplant; and urology procedures [18]. Further subdivision into 25 minor categories had been previously done for an analysis of World Bank Disease Control Priorities defined 44 essential surgical procedures in the United States during COVID [19,20]. Eligible surgical procedures (n=2135) were determined using independent screening techniques between two study personnel with experience in humanitarian surgical care (HW, SW; Supplementary file 1). Absolute exclusions included the transplant category, procedures requiring cardiopulmonary bypass, endovascular, laparoscopic, or robotic approaches.

Author Manuscript

This article is protected by copyright. All rights reserved.

To predict POMR, defined as mortality within 30 days of a surgical procedure, we used the EHR of eligible patients (N=19,686) to train a logistic regression model. Eligible patients were between 12-90 years old and underwent surgical procedures between 2015 and 2019.

From this cohort, predictors of mortality were identified as potential risk factors. The nine candidate predictors considered were age, sex, GCS, vitals including temperature, systolic blood pressure (SBP), diastolic blood pressure (DBP) and heart rate (HR), and labs including hemoglobin (Hgb) and blood glucose (BG) up to 24 hours immediately after admission. The highest SBP within this period was selected. As systolic and diastolic blood pressure were highly correlated, diastolic blood pressure was removed to avoid multicollinearity for improved model fit.

To assess the model's applicability beyond the training period (2015-2019), data were prospectively collected between 2022 and 2023 at the same institution for the same list of eligible surgical procedures. This patient cohort (N=29,591) was used to prospectively validate the model. The years 2020-2021 during the peak of the COVID-19 pandemic were excluded to avoid non-generalizable conditions which may have influenced the model.

Model development

Continuous predictors of mortality were binned into categorical levels according to expert-recommended clinical thresholds to improve ease of usability for the model. Predictors with >5% of variables missing were binned as "NA" to be handled separately for model and nomogram development; predictors with <5% missing were omitted case-wise according to the default regression settings in the R (v4.2) statistical software. Model development data (2015-2019) were split into 90% training and 10% holdout sets. The holdout set was reserved for later testing to ensure the model performed well using previously unseen data. A multivariate logistic regression model was trained on a balanced training (N=34,774) set whereby the minority class (328 dead) was repeatedly sampled to reach a size of 17,384 to balance the major class (17,390 living). Interaction terms crossing SBP with age levels were also modeled, as age and SBP influenced each other's association with mortality. The trained model was assessed for model fit and classification performance using area under the curve (AUC), accuracy, specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV) using both training and holdout datasets.

Prospective validation of model

This article is protected by copyright. All rights reserved.

Author Manuscript

We selected a second dataset of patients from 2022-2023 (N=29,591) as a prospective evaluation cohort and used the model to predict mortality. We compared the classification performance of this same model on this prospective validation cohort using AUC, accuracy, specificity, sensitivity, PPV, and NPV to those on the training and holdout datasets.

Model calibration for improved interpretability

Given the low prevalence of death cases within the dataset (1.8%), model calibration was performed to align the predicted risk of death with the observed fraction of deaths and therefore improve interpretability in a clinical context. For example, a 0.99 risk being very close to the maximum limit of 1 suggests that to be a classifier which is *highly confident* of mortality risk for this patient. To improve the interpretability of predicted risk, model calibration was performed to align the predicted risk of death with the observed fraction of deaths.

Model calibration is the process of rescaling a model's predicted probabilities to correspond with observed probabilities. The ultimate goal of model calibration is to provide an intuitive scale such that users can apply a real-world interpretation of the predicted probabilities. The quality of model calibration is demonstrated in a reliability diagram of calibrated predictions vs observed percentages. A well-calibrated model should have calibrated predictions closely track the observed percentages according to a 45-degree line.

Using the CalibratR [21] framework proposed for models built in R, three calibration modes were performed: histogram binning [22], Bayesian Binning in Quantiles (BBQ) [23] and GUESS [24]. The optimal calibration method was selected based on the highest AUC and lowest expected calibration error (ECE) and how well the calibrated predictions matched the observed percentages compared with a 45-degree line in a reliability diagram.

Conversion to nomogram and Excel calculator

To facilitate implementation in a low-resource setting, the logistic regression model was converted to a nomogram using the rms package (v6.8) in R as well as programmed into an Excel spreadsheet (Supplementary file 2). In the nomogram conversion, the weights of the logistic model were rescaled and shifted such that the smallest weights would be in integers and reference levels of categorical variables would be zero for easy summation into a nomogram mortality score. Mortality scores ranged from 0 (low mortality risk) to 260 (high mortality risk) and corresponded to defined mortality probabilities.

This article is protected by copyright. All rights reserved.

Author Manuscript

Author Manuscript

Results

Patient characteristics

The model development set for model training and holdout validation consisted of 19,686 patients (19,322 alive, 364 dead) who underwent surgical procedures applicable to LRS between 2015-2019 at a high-income country academic medical center (Table 1). The group of alive patients were more likely than those who died to be female (56% vs 40%), younger (mean age: 58 years vs 65 years), not be of Asian, Black or White race (60 vs 51%) and have fewer comorbidities according to mean Charlson Comorbidity Index (4.4 vs. 7.9), which is an objective measure of comorbidity burden across 17 chronic conditions [25]. The most common comorbidities across both groups were malignancy, chronic pulmonary disease, and diabetes. For prospective model validation, a second cohort of 29,591 patients was utilized. The prospective validation cohort demonstrated similar characteristics to the development cohort. The dead and alive groups differed largely along the same characteristics as observed in the model development cohort (Table 1). Participants underwent a diverse set of operations which reflect those feasible in LRS (Table 2).

Predictors of mortality

Using a 90% training set from the model development set to create a logistic regression model, multiple factors were found to be predictive of patient mortality (Table 3). These factors included age 51-65 years (OR=2.99, 95% CI: 2.73-3.27), age >65 years (OR=4.05, 95% CI: 3.71-4.43), male sex (OR=1.32, 95% CI: 1.25-1.40), GCS < 13 (OR=5.20, 95% CI: 4.73-5.73), blood glucose > 200 mg/dL (OR=2.19, 95% CI: 2.01-2.38), Hgb ≤ 11 g/dL (OR=2.65, 95% CI: 2.43-2.89), HR > 120 bpm (OR=2.49, 95% CI: 2.35-2.64), temperature > 38 degrees Celsius (OR=1.32, 95% CI: 1.19-1.45), and SBP > 180 mmHg (OR=1.18, 95% CI: 1.02-1.37).

Generally higher mortality risk correlated with higher age (age 51-65 years: OR=2.99, 95% CI: 2.73-3.27; age >65 years: OR=4.05, 95% CI: 3.71-4.43) and higher SBP (OR=1.18, 95% CI: 1.02-1.37) (Table 2). However, we observed significant interaction effects between age and SBP on mortality (age 51- 65 years and SBP > 180: OR=0.41, 95% CI: 0.34-0.50; age > 65 years and SBP > 180: OR=0.64, 95% CI: 0.54-0.75). When SBP exceeded 180 mmHg, the effect of age was not as pronounced as when SBP was 180 mmHg or less as shown by the respective nomogram scales of different ranges (Fig 1).

Model performance

This article is protected by copyright. All rights reserved.

Author Manuscript

The training, holdout, and prospective datasets were used to evaluate model performance using AUC, sensitivity, specificity, PPV and NPV. AUC, sensitivity, specificity and NPV performed robustly across all data sets, including an AUC of 0.847, 0.867, and 0.925 for the training, holdout, and prospective validation set, respectively (Table 4). PPV had less robust performance within the holdout and prospective validation sets, with calculated PPVs of 0.070 and 0.044 respectively. The model had good discriminatory power as shown in the histogram (Fig 2) where the living patients (pink) were likely to have low uncalibrated predicted risk and were well separated from the deceased patients (blue) who were likely to have high uncalibrated predicted risk.

Model calibration

Given the low observed mortality rate in the training cohort (1.8%), model calibration was performed to rescale the predicted risks and better align these results with the observed mortality prevalence. Of the three attempted model calibration processes: histogram binning, Bayesian binning, and GUESS calibration, histogram binning was selected as the final calibration approach, given its highest AUC (0.838), lowest ECE (0.00003) and how closely it tracked the 45-degree line (i.e. calibrated probabilities equal observed percentages) in the reliability diagram (Fig 3).

Nomogram

To create a more accessible version of the mortality model for clinical use, a nomogram visual aid was derived from the logistic regression of the mortality risk model. In the nomogram (Fig 1), mortality scores ranged from 0 to 260. The range is a consequence of the smallest predictor (GCS is NA: OR=0.25) given 0 points and the largest predictor (GCS<13: OR=1.65) given 100 points and all other predictors in between. This 0-to-100 scale is the default set in the *rms* package used to construct the nomogram. When all possible predictors in the nomogram were summed, we obtained a theoretical maximum total of 260 mortality score points. The nomogram's predictors were the same as the model's predictors except that the former were rescaled to integer mortality score points from the latter and reported in odds ratios.

To illustrate the use of the nomogram and Excel calculator, refer to Table 5 where a male person between 18 and 50 years old with SBP <= 180 mmHg and hemoglobin > 6.8 mmol/L but

This article is protected by copyright. All rights reserved.

missing GCS and blood glucose measurements has a total mortality score of 12. This mortality score corresponds to an uncalibrated predicted probability of 0.07 and a calibrated probability of 0.0006, all of which signify a very low mortality risk. The nomogram points per predictor also map to the individual predictor scales in the nomogram diagram (Fig 1). For example, being male is given 12 nomogram points in both Table 4 and Fig 1. The nomogram table (Table 4) and figure (Fig 1) are designed to complement each other: the table explains how the nomogram points are awarded and maps to the logistic model while the figure provides a quick visual for interpreting the scale in a LRS.

Discussion:

In this study, we created and validated a novel POMR model for use in resource-constrained humanitarian settings. Our model incorporates predictors that can be feasibly obtained, risk which can be computed with minimal hardware, and outputs which are easily interpretable for a broad range of patients in LRS. This model showed evidence for validation with a high sensitivity, specificity, NPV, and AUC across training, holdout, and prospective datasets. These findings from our perioperative risk mortality model demonstrate potential for use to benchmark quality of humanitarian surgical care and identify opportunities for quality improvement with minimal burden on constrained health systems in LRS.

Several novel aspects of this model make it highly applicable to humanitarian and LRS contexts. First, it utilized predictors that can be feasibly obtained in such settings and still be calculated even when certain of these variables are not available (e.g., age, sex, vital signs, and GCS). Point-of-care laboratory values which are available in most LRS and humanitarian contexts including blood glucose levels and hemoglobin also demonstrated impact within our model. Factors of GCS <13, age 65 and older paired with SBP > 180 mmHg, and hemoglobin ≤ 11g/dL demonstrated the highest contributions to mortality, with GCS < 13 as the most impactful with more than double the mortality score contribution of any other variable within our nomogram. In acknowledgment of the highly variable quality of data collection in LRS, our model also accounts for context discrepancy (i.e., where certain variables may not be a part of routine data collection) by having a categorical level to code for missing if the variable is unavailable without disrupting the final mortality risk. This adjustment allows the calculator to be used when the variable is either present or absent. The finding of SBP >180 mmHg as predictive of mortality is less compatible with clinical observations of low SBP correlating with mortality. We believe this is a result of using highest SBP within 24 hours after surgical admission for model development and acknowledge this as a limitation of the derivation strategy. Future validations of the

This article is protected by copyright. All rights reserved.

Author Manuscript

calculator will be needed to investigate the role of SBP in predicting POMR by using admission or preoperative SBP.

Second, model use and risk calculation could be used quickly and with minimal technological infrastructure. Our approach utilizes only eight variables, six of which can be ascertained upon clinician evaluation and binned into yes vs. no or binary numeric categories. Only the final two variables require laboratory capabilities, which can be performed as rapid point-of-care testing. The model does not require an understanding of a patient's past medical history, which can be difficult to obtain in humanitarian settings [26,27].

Third, the model outputs provide easy implementation and interpretability in a low-resource clinical setting given its Excel calculator and nomogram versions. The simple Excel calculator derived from the logistic regression version of the model allows clinicians to easily access and utilize the model's calculating power in real-time for patients being evaluated as surgical candidates. The nomogram provides an additional rapid tool to calculate risk without the need for computer or Excel access, while also conceptualizing the impact of individual variables on mortality risk. POMR risk models which are regularly implemented in humanitarian settings were often developed without inclusion of adolescent, obstetric, or non-trauma patients, or clinicians are required to use different versions of these tools for varied patient characteristics or disease states [16,28,29,30,31,32]. This model is the most inclusive to-date and is purposefully designed for application to surgical candidates in humanitarian settings regardless of age, clinical history, or presentation.

This study had several limitations. First, our model has not yet been tested on any resource-constrained humanitarian population, as our dataset was obtained from a single, tertiary level hospital in a high-income country. However, our options for model development were exceptionally limited, given that such comprehensive datasets are not available in LMICs. To overcome this limitation, we selected procedures likely to be performed in LRS to mitigate this challenge. A necessary future study to further evaluate the applicability of this model includes piloting within the intended low-resource and humanitarian context, ideally with a comprehensive and diverse dataset of surgical patients. Second, despite high AUC, sensitivity, specificity, and NPV value of the model across training, holdout, and prospective validation sets, PPV was low when evaluated for the holdout and prospective sets. This is expected given the low prevalence of death cases in the holdout and prospective sets (<2%) which puts an upper limit on PPV, a function of prevalence [33]. We hypothesize that applying this model to a patient cohort with a higher mortality prevalence, as would likely be the case when piloting the

This article is protected by copyright. All rights reserved.

Author Manuscript

model in a humanitarian context, would reveal a more favorable PPV. Third, our sample was mostly female and older adults above the age of 50, which may not accurately reflect the patient demographics of all humanitarian contexts. This inherent limitation to our dataset further supports additional testing with LMIC specific datasets, should they be feasible to attain. However, seven of the eight variables of this model correlated significantly with patient outcome independent of age, and the age demographics of our dataset are not drastically different from those of patients in other LRS [34]. Fourth, the strategy used to select highest SBP within a 24-hour window likely resulted in higher-than-expected SBP. In future validation strategies of this calculator, this methodology will be revised to use SBP on admission or preoperative SBP.

Conclusion

In this study, we developed and validated a POMR model designed for use in resource-constrained humanitarian surgical care. The model was validated with eight predictors that can be feasibly obtained, compute POMR with minimal technical infrastructure, and generate outputs that are easily interpretable in LRS and directly applicable to patient care encompassing diverse patient populations. We anticipate that this POMR calculator has significant utility to benchmark the quality of humanitarian surgical care in a streamlined and easy-to-implement manner. Next steps include conducting a pilot of this tool within a humanitarian environment to evaluate performance in the target context. With multiple simple implementation formats including a logistic model, Excel calculator, and nomogram, this model holds potential to be a powerful tool in benchmarking the quality of humanitarian surgical care and identifying targeted opportunities for performance improvement.

Author Manuscript

This article is protected by copyright. All rights reserved.

Author Manuscript

References:

1. Meara JG, Leather AJ, Hagander L, Alkire BC, Alonso N, Ameh EA, Bickler SW, Conteh L, Dare AJ, Davies J, Mérisier ED, El-Halabi S, Farmer PE, Gawande A, Gillies R, Greenberg SL, Grimes CE, Gruen RL, Ismail EA, Kamara TB, Lavy C, Lundeg G, Mkandawire NC, Raykar NP, Riesel JN, Rodas E, Rose J, Roy N, Shrime MG, Sullivan R, Verguet S, Watters D, Weiser TG, Wilson IH, Yamey G, Yip W. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. Lancet. 2015 Aug 8;386(9993):569-624. doi: 10.1016/S0140-6736(15)60160-X.

2. Colombo S, Pavignani E. Recurrent failings of medical humanitarianism: intractable, ignored, or just exaggerated? Lancet. 2017 Nov 18;390(10109):2314-2324. doi: 10.1016/S0140-6736(17)31277-1.

3. Wild H, Stewart BT, LeBoa C, Stave CD, Wren SM. Epidemiology of Injuries Sustained by Civilians and Local Combatants in Contemporary Armed Conflict: An Appeal for a Shared Trauma Registry Among Humanitarian Actors. World J Surg. 2020 Jun;44(6):1863-1873. doi: 10.1007/s00268-020-05428-y.

4. Kim M, Wall MM, Li G. Risk Stratification for Major Postoperative Complications in Patients Undergoing Intra-abdominal General Surgery Using Latent Class Analysis. Anesth Analg. 2018 Mar;126(3):848-857. doi: 10.1213/ANE.0000000000002345.

5. Wild H, Stewart BT, LeBoa C, Jewell T, Mehta K, Wren SM. Perioperative Risk Assessment in Humanitarian Settings: A Scoping Review. World J Surg. 2023 May;47(5):1092-1113. doi: 10.1007/s00268-023-06893-x.

6. Ng-Kamstra JS, Arya S, Greenberg SLM, Kotagal M, Arsenault C, Ljungman D, Yorlets RR, Agarwal A, Frankfurter C, Nikouline A, Lai FYX, Palmqvist CL, Fu T, Mahmood T, Raju S, Sharma S, Marks IH, Bowder A, Pi L, Meara JG, Shrime MG. Perioperative mortality rates in low-income and middle-income countries: a systematic review and meta-analysis. BMJ Glob Health. 2018 Jun 22;3(3):e000810. doi: 10.1136/bmjgh-2018-000810.

7. Vernooij JEM, Koning NJ, Geurts JW, Holewijn S, Preckel B, Kalkman CJ, Vernooij LM. Performance and usability of pre-operative prediction models for 30-day peri-operative mortality risk: a systematic review. Anaesthesia. 2023 May;78(5):607-619. doi: 10.1111/anae.15988.

8. Matthews L, Levett DZH, Grocott MPW. Perioperative Risk Stratification and Modification. Anesthesiol Clin. 2022 Mar;40(1S):e1-e23. doi: 10.1016/j.anclin.2022.03.001.

This article is protected by copyright. All rights reserved.

Author Manuscript

9. Grigorescu BL, Săplăcan I, Petrișor M, Bordea IR, Fodor R, Lazăr A. Perioperative Risk Stratification: A Need for an Improved Assessment in Surgery and Anesthesia-A Pilot Study. Medicina (Kaunas). 2021 Oct 19;57(10):1132. doi: 10.3390/medicina57101132.

10. Davies JF, Lenglet A, van Wijhe M, Ariti C. Perioperative mortality: Analysis of 3 years of operative data across 7 general surgical projects of Médecins Sans Frontières in Democratic Republic of Congo, Central African Republic, and South Sudan. Surgery. 2016 May;159(5):1269-78. doi: 10.1016/j.surg.2015.12.022.

11. Ndayizeye L, Ngarambe C, Smart B, Riviello R, Majyambere JP, Rickard J. Peritonitis in Rwanda: Epidemiology and risk factors for morbidity and mortality. Surgery. 2016 Dec;160(6):1645-1656. doi: 10.1016/j.surg.2016.08.036. Epub 2016 Oct 4.

12. Biccard BM, Madiba TE, Kluyts HL, Munlemvo DM, Madzimbamuto FD, Basenero A, Gordon CS, Youssouf C, Rakotoarison SR, Gobin V, Samateh AL, Sani CM, Omigbodun AO, Amanor-Boadu SD, Tumukunde JT, Esterhuizen TM, Manach YL, Forget P, Elkhogia AM, Mehyaoui RM, Zoumeno E, Ndayisaba G, Ndasi H, Ndonga AKN, Ngumi ZWW, Patel UP, Ashebir DZ, Antwi-Kusi AAK, Mbwele B, Sama HD, Elfiky M, Fawzy MA, Pearse RM; African Surgical Outcomes Study (ASOS) investigators. Perioperative patient outcomes in the African Surgical Outcomes Study: a 7-day prospective observational cohort study. Lancet. 2018 Apr 21;391(10130):1589-1598. doi: 10.1016/S0140-6736(18)30001-1.

13. Rosenkrantz L, Schuurman N, Hameed MS, Boniface R, Lett R. The Kampala Trauma Score: A 20-year track record. J Trauma Acute Care Surg. 2022 Jun 1;92(6):e132-e138. doi: 10.1097/TA.0000000000003567.

14. Yolcu S, Sener K, Tapsiz H, Ozer AI, Avci A. Revised Trauma Score and CRAMS better predicted mortality in high-energy-trauma patients than Early-Warning Score. Ir J Med Sci. 2023 Aug;192(4):1855-1860. doi: 10.1007/s11845-022-03208-2.

15. Weeks SR, Juillard CJ, Monono ME, Etoundi GA, Ngamby MK, Hyder AA, Stevens KA. Is the Kampala trauma score an effective predictor of mortality in low-resource settings? A comparison of multiple trauma severity scores. World J Surg. 2014 Aug;38(8):1905-11. doi: 10.1007/s00268-014-2496-0.

16. ASOS-Paeds Investigators. Outcomes after surgery for children in Africa (ASOS-Paeds): a 14-day prospective observational cohort study. Lancet. 2024 Apr 13;403(10435):1482-1492. doi: 10.1016/S0140-6736(24)00103-X.

17. Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project: Clinical Classifications Software for Services and Procedures, Version 2020.1. Accessed May 14, 2021. https://www.hcup-us.ahrq.gov/toolssoftware/ccs_svcsproc/ccssvcproc.jsp

This article is protected by copyright. All rights reserved.

Author Manuscript

18. Mattingly AS, Rose L, Eddington HS, et al. Trends in US Surgical Procedures and Health Care System Response to Policies Curtailing Elective Surgical Operations During the COVID-19 Pandemic. JAMA Netw Open. 2021;4(12):e2138038. doi:10.1001/jamanetworkopen.2021.38038

19. Debas HT, Donkor P, Gawande A, Jamison DT, Kruk ME, Mock CN, eds. Disease Control Priorities, Third Edition (Volume 1): Essential Surgery. World Bank; 2015.

20. Mattingly AS, Eddington HS, Rose L, et al. Defining Essential Surgery in the US During the COVID-19 Pandemic Response. JAMA Surg. 2023;158(1):99–100. doi:10.1001/jamasurg.2022.3944

21. Schwarz J and Heider D, Bioinformatics 2019, 35(14):2458-2465. Available from: https://cran.r-project.org/web/packages/CalibratR/index.html

22. Zadrozny B., Elkan C. (2001) Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. In International Conference on Machine Learning, pp. 609–616.

23. Naeini MP, Cooper GF, Hauskrecht M. Obtaining Well Calibrated Probabilities Using Bayesian Binning. Proc AAAI Conf Artif Intell. 2015 Jan;2015:2901-2907.

24. Schwarz J, Heider D. GUESS: projecting machine learning scores to well-calibrated probability estimates for clinical decision-making. Bioinformatics. 2019 Jul 15;35(14):2458-2465. doi: 10.1093/bioinformatics/bty984.

25. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005 Nov;43(11):1130-9. doi: 10.1097/01.mlr.0000182534.19832.83.

26. Copeland GP. The POSSUM system of surgical audit. Arch Surg. 2002 Jan;137(1):15-9. doi: 10.1001/archsurg.137.1.15.

27. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, Cohen ME. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg. 2013 Nov;217(5):833-42.e1-3. doi: 10.1016/j.jamcollsurg.2013.07.385.

28. Akay S, Ozturk AM, Akay H. Comparison of modified Kampala trauma score with trauma mortality prediction model and trauma-injury severity score: A National Trauma Data

This article is protected by copyright. All rights reserved.

Author Manuscript

Bank Study. Am J Emerg Med. 2017 Aug;35(8):1056-1059. doi: 10.1016/j.ajem.2017.02.035.

29. Jeong JH, Park YJ, Kim DH, Kim TY, Kang C, Lee SH, Lee SB, Kim SC, Lim D. The new trauma score (NTS): a modification of the revised trauma score for better trauma mortality prediction. BMC Surg. 2017 Jul 3;17(1):77. doi: 10.1186/s12893-017-0272-4.

30. Paterson A, Maswime S, Hardy A, Pearse RM, Biccard BM; African Surgical Outcomes (ASOS) group and International Surgical Outcomes (ISOS) group. Postoperative outcomes associated with surgical care for women in Africa: an international risk-adjusted analysis of prospective observational cohorts. BJA Open. 2022 Oct 21;4:100100. doi: 10.1016/j.bjao.2022.100100.

31. Manoochehry S, Vafabin M, Bitaraf S, Amiri A. A Comparison between the Ability of Revised Trauma Score and Kampala Trauma Score in Predicting Mortality; a Meta-Analysis. Arch Acad Emerg Med. 2019 Jan 15;7(1):e6.

32. Gaba F, Mohammadi SM, Krivonosov MI, Blyuss O, On Behalf Of The Go Soar Collaborators. Predicting Risk of Post-Operative Morbidity and Mortality following Gynaecological Oncology Surgery (PROMEGO): A Global Gynaecological Oncology Surgical Outcomes Collaborative Led Study. Cancers (Basel). 2024 May 26;16(11):2021. doi: 10.3390/cancers16112021.

33. Tenny S, Hoffman MR. Prevalence. [Updated 2023 May 22]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK430867/

34. Wong EG, Trelles M, Dominguez L, Mupenda Mwania J, Kasonga Tshibangu C, Haq Saqeb S, et al. Operative procedures in the elderly in low-resource settings: a review of Médecins Sans Frontières facilities. World J Surg. 2015 Mar;39(3):652-7. doi: 10.1007/s00268-014-2855-x.

This article is protected by copyright. All rights reserved.

Author Manuscript

Tables and Figures

Table 1: Characteristics of patients in model development data set (2015-2019) and in prospective validation set (2022-2023)

	2015 - 2019 (model development)			2022 - 2023 (prospective validation)		
	Alive	**Dead**	***P* value**	**Alive**	**Dead**	***P* value**
N (%)	19322 (98.2%)	364 (1.8%)		29383 (99.3%)	208 (0.7%)	
Female, n (%)	10,884 (56%)	145 (40%)	<0.001	16,587 (56%)	80 (38%)	<0.001
Age, years			<0.001			<0.001
Mean (SD)	58 (17)	65 (14)		55 (18)	67 (15)	
Range	13, 90	20, 90		12, 90	22, 90	
Age Group, n (%)			<0.001			<0.001
Less than 18 years	149 (0.8%)	0 (0%)		478 (1.6%)	0 (0%)	
18 to <30 years	1,367 (7.1%)	10 (2.7%)		2,621 (8.9%)	6 (2.9%)	
30 to <40 years	2,147 (11%)	13 (3.6%)		4,303 (15%)	5 (2.4%)	
40 to <50 years	2,306 (12%)	26 (7.1%)		3,921 (13%)	17 (8.2%)	
50 to <60 years	3,581 (19%)	66 (18%)		4,838 (16%)	24 (12%)	
60 to <70 years	4,595 (24%)	102 (28%)		5,969 (20%)	58 (28%)	
70 to <80 years	3,646 (19%)	93 (26%)		5,096 (17%)	57 (27%)	
80 to <90 years	1,531 (7.9%)	54 (15%)		2,157 (7.3%)	41 (20%)	
90 years and above	0 (0%)	0 (0%)		0 (0%)	0 (0%)	
Race, n (%)			0.003			0.038
Asian	2,899 (15%)	64 (18%)		5,551 (19%)	39 (19%)	
Black	704 (3.6%)	16 (4.4%)		994 (3.4%)	5 (2.4%)	
White	11,628 (60%)	184 (51%)		16,935 (58%)	106 (51%)	
Other	4,091 (21%)	100 (27%)		5,903 (20%)	58 (28%)	
Hispanic, n (%)	2,945 (15%)	62 (17%)	0.347	4,754 (16%)	45 (22%)	0.033
Observation days prior to index			<0.001			0.713
Mean (SD)	12,466 (8,947)	17,511 (7,707)		3,940 (2,512)	3,997 (2,242)	
Range	90 - 32,854	110 - 32,687		90 - 26,556	92 - 10,998	
Follow-up days			0.6			<0.001

This article is protected by copyright. All rights reserved.

Mean (SD)	2,992 (3,788)	3,110 (4,270)		441 (273)	251 (159)	
Range	0 - 26,852	2 - 22,554		0 - 7,150	3 - 688	
Charlson Comorbidity Index			<0.001			<0.001
Mean (SD)	4.4 (3.9)	7.9 (4.0)		3.5 (3.6)	8.4 (4.1)	
Range	0.0, 24.0	0.0, 21.0		0.0, 24.0	0.0, 23.0	
Comorbidities						
Malignancy, n (%)	6,810 (35%)	139 (38%)	0.245	7,536 (26%)	76 (37%)	<0.001
Metastatic Solid Tumor, n (%)	2,825 (15%)	70 (19%)	0.014	2,033 (6.9%)	35 (17%)	<0.001
Diabetes, n (%)	3,199 (17%)	110 (30%)	<0.001	4,331 (15%)	72 (35%)	<0.001
Diabetes with Complications, n (%)	1,211 (6.3%)	57 (16%)	<0.001	1,976 (6.7%)	50 (24%)	<0.001
Congestive Heart Failure, n (%)	1,557 (8.1%)	161 (44%)	<0.001	2,159 (7.3%)	92 (44%)	<0.001
Myocardial Infarction, n (%)	874 (4.5%)	89 (24%)	<0.001	1,071 (3.6%)	49 (24%)	<0.001
Peripheral Vascular Disease, n (%)	1,748 (9.0%)	128 (35%)	<0.001	2,264 (7.7%)	67 (32%)	<0.001
Chronic Pulmonary Disease, n (%)	3,477 (18%)	103 (28%)	<0.001	5,481 (19%)	68 (33%)	<0.001
Cerebrovascular Disease, n (%)	1,767 (9.1%)	135 (37%)	<0.001	2,264 (7.7%)	85 (41%)	<0.001
Dementia, n (%)	272 (1.4%)	13 (3.6%)	<0.001	399 (1.4%)	16 (7.7%)	<0.001
Hemiparaplegia, n (%)	582 (3.0%)	64 (18%)	<0.001	556 (1.9%)	23 (11%)	<0.001
Mild Liver Disease, n (%)	2,079 (11%)	106 (29%)	<0.001	2,958 (10%)	57 (27%)	<0.001
Severe Liver Disease, n (%)	292 (1.5%)	33 (9.1%)	<0.001	353 (1.2%)	18 (8.7%)	<0.001
Renal Disease, n (%)	2,481 (13%)	145 (40%)	<0.001	3,133 (11%)	106 (51%)	<0.001
Peptic Ulcer Disease, n (%)	500 (2.6%)	15 (4.1%)	0.069	744 (2.5%)	25 (12%)	<0.001
Rheumatic Disease, n (%)	735 (3.8%)	19 (5.2%)	0.163	1,156 (3.9%)	12 (5.8%)	0.176
HIV, n (%)	60 (0.3%)	0 (0%)	0.287	95 (0.3%)	2 (1.0%)	0.319

Author Manuscript

This article is protected by copyright. All rights reserved.

Author Manuscript

Table 2: Operation type according to major CPT code for the POMR model development and validation cohorts

	Total		Development		Prospective validation	
	N	%	N	%	N	%
Non-fracture musculoskeletal (e.g. – tendon repair, fasciotomy)	7295	13.71%	2585	13.13%	4710	15.92%
Excision of Skin and Soft Tissue (e.g. – repair, excise benign/malignant lesions)	7073	13.29%	1363	6.92%	5710	19.30%
Fractures (e.g. upper and lower extremity, face/skull)	4874	9.16%	2152	10.93%	2722	9.20%
Spine (e.g. arthrodesis, I/D spine)	4696	8.83%	2867	14.56%	1829	6.18%
Breast, Endocrine (e.g. – mastectomy, thyroidectomy)	4178	7.85%	1206	6.13%	2972	10.04%
Grafts, Flaps, and Other Reconstruction (e.g. – skin graft, local flap)	4016	7.55%	1967	9.99%	2049	6.92%
Ear, Face, Nose, Mouth, Pharynx (e.g.- I/D or excision tongue, repair mouth vestibule)	3554	6.68%	1754	8.91%	1800	6.08%
Incision, Drainage, and Debridement (e.g.- debride skin/soft tissue abdomen, breast)	2356	4.43%	1022	5.19%	1334	4.51%
Gynecology (e.g. – hysterectomy, vaginal repair)	2161	4.06%	845	4.29%	1316	4.45%
Vascular (e.g.- repair blood vessel, artery ligation)	1726	3.24%	1070	5.44%	656	2.22%
C-section	1719	3.23%	796	4.04%	923	3.12%
Solid organ, hepatobiliary, spleen (e.g. – splenectomy, suture liver hemorrhage)	1703	3.20%	1097	5.57%	606	2.05%
Thoracic (e.g. – thoracotomy, diaphragm repair)	1484	2.79%	996	5.06%	488	1.65%
Kidney, Bladder, Urethra (e.g.- nephrectomy, urethral stricture)	1290	2.42%	742	3.77%	548	1.85%
Abortion	1073	2.02%	260	1.32%	813	2.75%
Skull, Brain, Meninges, Nerves (e.g. – burr hole, craniotomy subdural)	1022	1.92%	610	3.10%	412	1.39%

This article is protected by copyright. All rights reserved.

Arthroplasty, Arthroscopy (e.g. – patella removal, knee ligament reconstruction)	998	1.88%	521	2.65%	477	1.61%
Hernia (e.g.- inguinal, umbilical)	664	1.25%	0	0.00%	664	2.24%
Amputation	607	1.14%	311	1.58%	296	1.00%
Male reproductive (e.g. – orchiectomy, hydrocele)	369	0.69%	31	0.16%	338	1.14%
Prostate (e.g. – perineal or suprapubic prostatectomy)	215	0.40%	102	0.52%	113	0.38%
Ocular (e.g. canthotomy, repair laceration)	78	0.15%	26	0.13%	52	0.18%
Cardiac (e.g. – repair cardiac wound w/o bypass, pericardiotomy)	50	0.09%	50	0.25%	0	0.00%

Author Manuscript

This article is protected by copyright. All rights reserved.

Table 3: Risk predictors in logistic regression of the POMR model

	Coefficient	OR	95% CI	*P* value
Age 12- 50 yr	ref	-	-	
Age 51- 65 yr	1.09	2.99	(2.73 – 3.27)	< 0.0001
Age > 65 yr	1.40	4.05	(3.71 – 4.43)	< 0.0001
Male	0.280	1.32	(1.24 – 1.40)	< 0.0001
GCS 13 -15	ref	-	-	-
GCS < 13	1.65	5.20	(4.73 – 5.73)	< 0.0001
GCS=NA	-1.39	0.25	(0.23 – 0.26)	< 0.0001
glucose <= 200 mg/dL	ref	-	-	-
glucose > 200 mg/dL	0.785	2.19	(2.01 – 2.38)	< 0.0001
glucose = NA	-0.0595	0.94	(0.80 – 1.11)	0.47
Hb > 11 g/dL	ref	-	-	-
Hb <= 11 g/dL	0.975	2.65	(2.43 – 2.89)	< 0.0001
Hb = NA	0.696	2.01	(1.71 – 2.36)	< 0.0001
HR > 120 bpm	0.912	2.49	(2.35 – 2.64)	< 0.0001
Temperature > 38 degC	0.274	1.32	(1.19 – 1.45)	< 0.0001
BP systolic >180 mmHg	0.166	1.18	(1.02 – 1.37)	0.027
age 51- 65 yr & BP systolic >180	-0.885	0.41	(0.34 - 0.50)	< 0.0001
age > 65 yr & BP systolic >180	-0.454	0.64	(0.54 – 0.75)	< 0.0001
Intercept	-1.40	0.25	(0.22 – 0.27)	< 0.0001

Author Manuscript

This article is protected by copyright. All rights reserved.

Table 4: Model performance in training, holdout and prospective validation data sets

	Training set	Holdout set	Prospective validation set
AUC	0.847	0.867	0.925
Sensitivity	0.739	0.886	0.844
Specificity	0.807	0.780	0.864
PPV	0.798	0.070	0.044
NPV	0.750	0.997	0.999

Author Manuscript

This article is protected by copyright. All rights reserved.

Table 5: Example calculation of mortality score using nomogram and logistic regression model, assuming a 40-year-old adult male with SBP <= 180 mmHg and hemoglobin > 6.8 mmol/L but missing GCS and blood glucose measurements. Table 5 shows the logistic model weights and nomogram points per predictor (if Y=1) of such a patient whose nomogram points sum up to 12 and correspond to a low calibrated probability of 0.0006.

	Y=1 N=0	Nomogram points	Logistic model coefficient
GCS < 13	0	100	1.65
GCS from 13 to 15	0	46	0
GCS = NA	1	0	-1.39
Male	1	12	0.28
Temp > 38 deg C	0	9	0.274
HR > 120 bpm	0	30	0.912
Glucose <= 11.1 mmol/L	0	2	0
Glucose >11.1 mmol/L	0	28	0.78
Glucose = NA	1	0	-0.0595
Hb <= 6.8 mmol/L	0	32	0.975
Hb > 6.8 mmol/L	1	0	0
Hb = NA	0	23	0.696
BPsys > 180 mmHg	0		0.166
Age from 12 to 50 yr	1	0	0
Age from 51 to 65 yr	0	0	1.094
Age > 65 yr	0	0	1.4
Age from 12 to 50 if BPsys <= 180 mmHg	1	0	0
Age from 51 to 65 if BPsys <= 180 mmHg	0	36	0
Age >65 if BPsys <= 180 mmHg	0	45	0
Age from 18 to 50 if BPsys > 180 mmHg	0	5	0
Age from 51 to 65 if BPsys > 180 mmHg	0	12	-0.885
Age >65 if BPsys > 180 mmHg	0	37	-0.45
Intercept constant	1	0	-1.4
Nomogram score*		12	-
logistic model score+			-2.57
logistic probability (uncalibrated)			0.07

This article is protected by copyright. All rights reserved.

Author Manuscript

logistic probability (calibrated)			6E-04

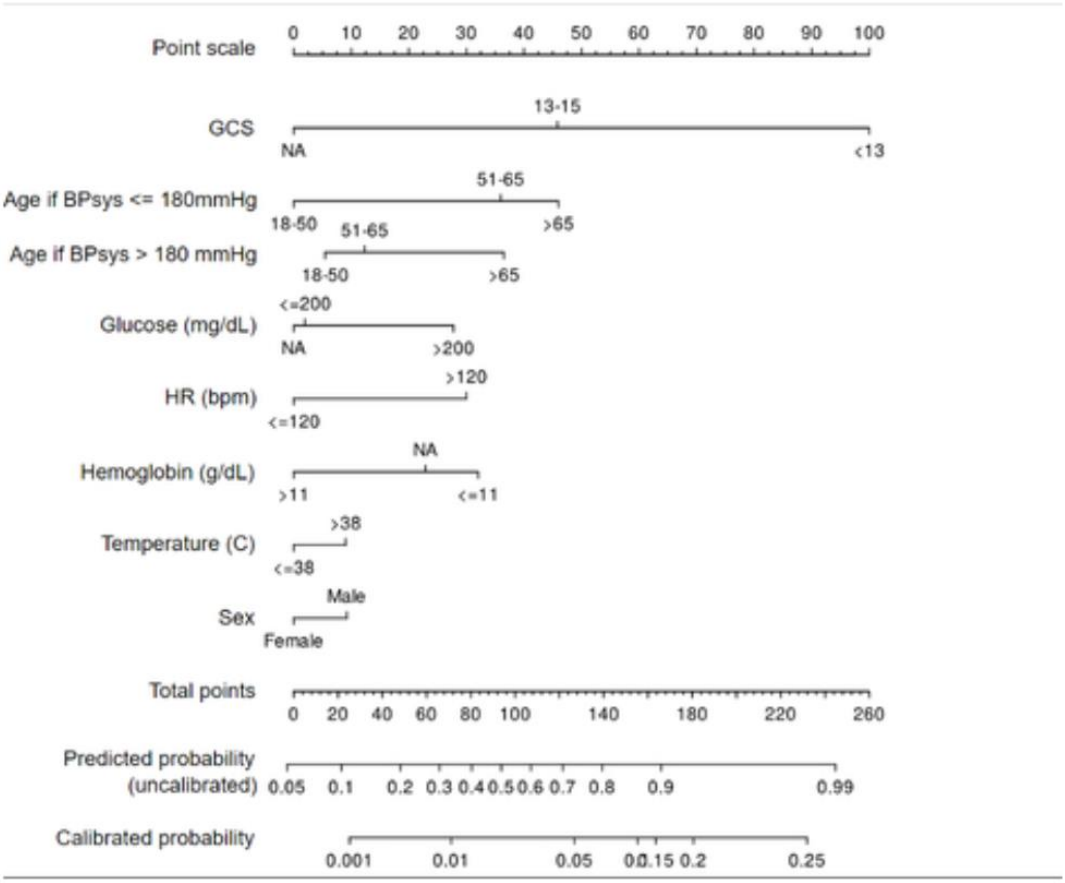
*The nomogram score is the sum of nomogram points for the rows marked Y=1

+The logistic model score is the sum of the logistic model coefficients for the rows marked Y=1

Author Manuscript

This article is protected by copyright. All rights reserved.

Author Manuscript



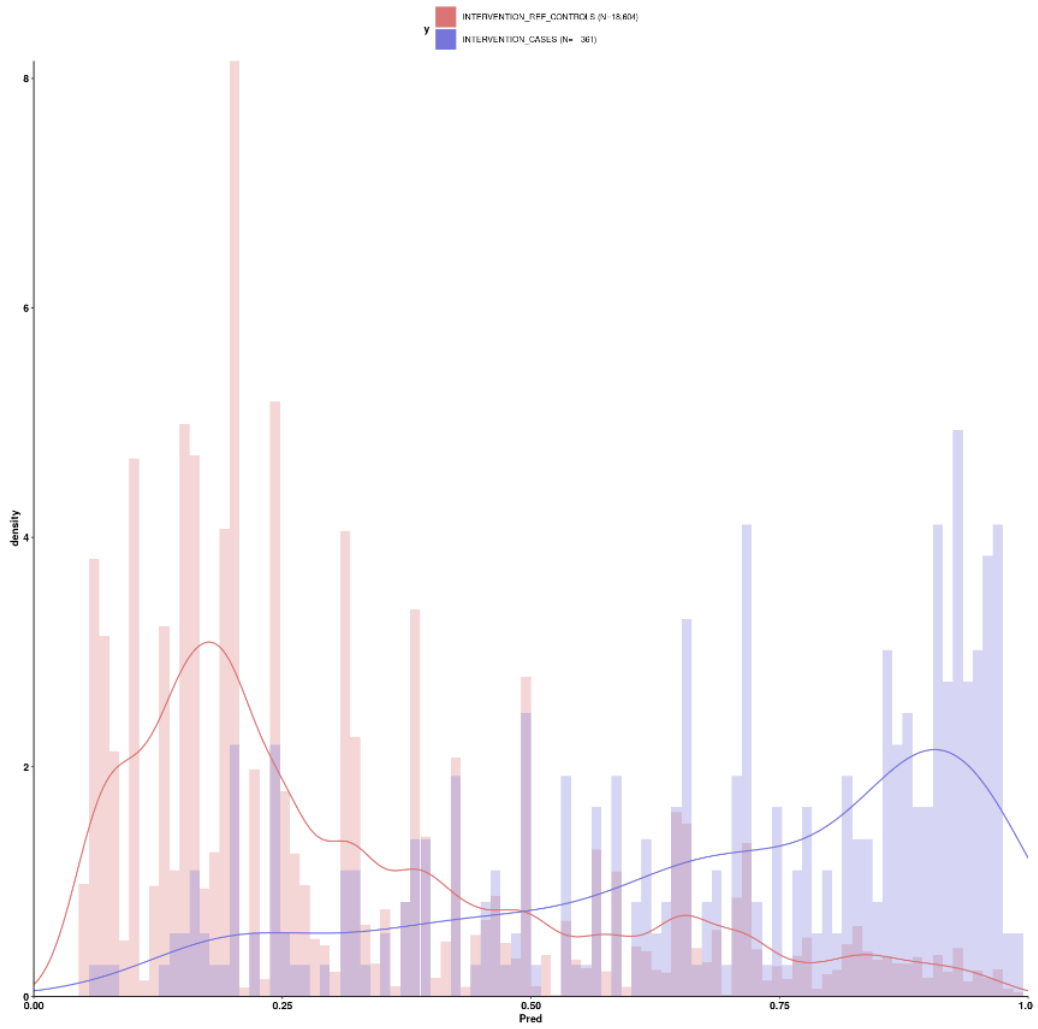
Nomogram calculating mortality risk of patients depending on their predictors

165x136mm (72 x 72 DPI)

Fig 1: Nomogram calculating mortality risk of patients depending on their predictors. This nomogram can be used by aligning the numeric or categorical variables for a given patient along the x-axis, corresponding to an assigned point value. Once all available variables have been determined, they can be summed for a total point score and plotted on the third horizontal line from the bottom. Total point scores correspond to an uncalibrated and calibrated predicted probability of mortality, according to how they align on the nomogram. For an example of a more specific application to a hypothetical patient case, see Table 5.

This article is protected by copyright. All rights reserved.

Author Manuscript



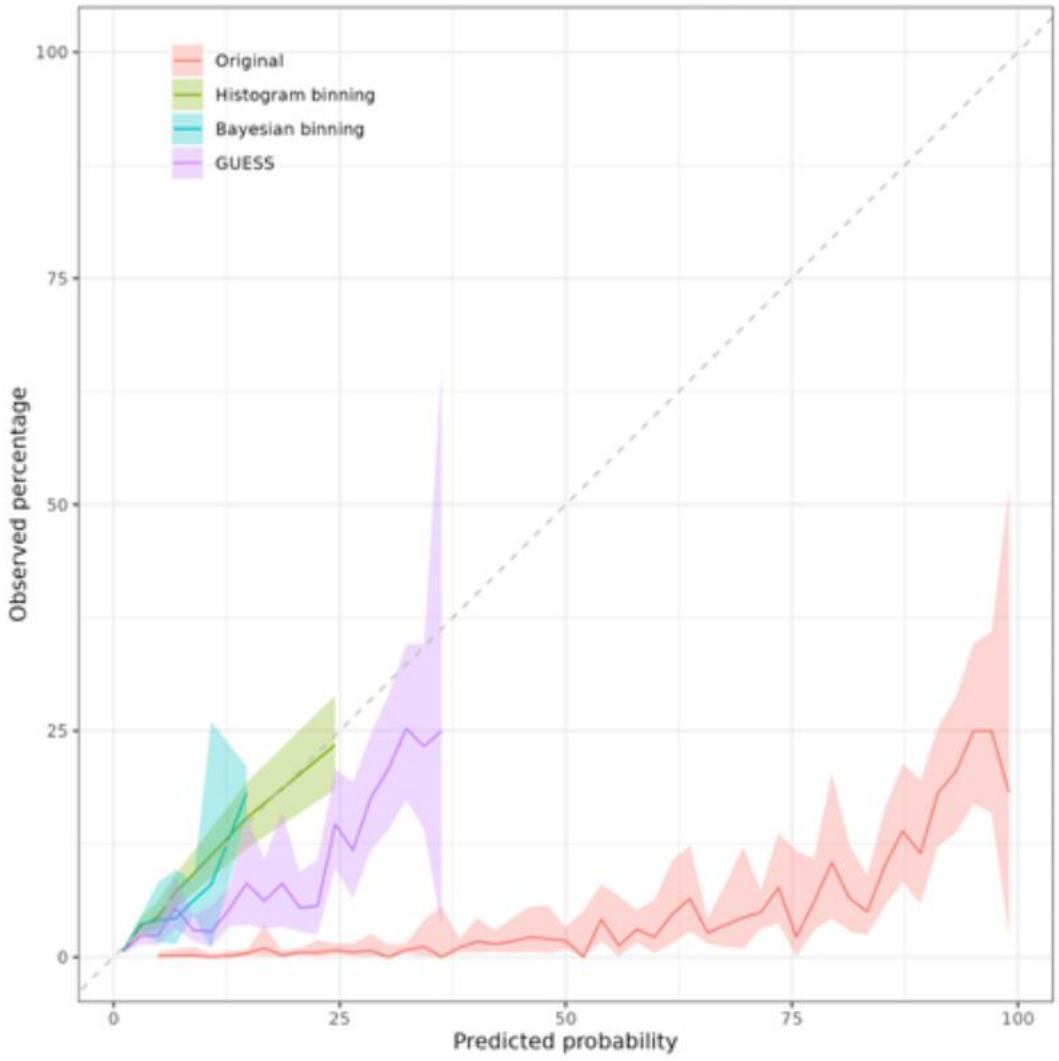
Histogram showing number of patients per predicted risk probability in model training data

710x710mm (59 x 59 DPI)

Fig 2: Histogram showing number of patients per predicted risk probability value in the model training data. Controls (living, pink) were often predicted to have low probabilities of mortality, while mortality cases (dead, blue) were predicted to have high probabilities of mortality. This visual representation suggests that the model can reasonably discriminate between patients with alive and dead perioperative outcomes.

This article is protected by copyright. All rights reserved.

Author Manuscript



Model calibration curves

165x165mm (72 x 72 DPI)

Fig 3: Reliability diagram showing quality of model calibration by histogram binning, Bayesian binning and GUESS. A well-calibrated model should closely track the 45-degree line where calibrated probability (predicted probability, x-axis) is most similar to the observed percentage of deaths (y-axis). Among the three calibration models tested, histogram binning was the most accurately calibrated model, as can be seen by its close accuracies between predicted and observed mortality.

This article is protected by copyright. All rights reserved.

Supplemental files

Supplemental file 1: Eligible procedures (CPT code) and their categories

Supplemental file 2: Excel spreadsheet of POMR calculation

Author Manuscript

This article is protected by copyright. All rights reserved.

Author Manuscript

Ethics statement: Ethical exemption for this study was determined by the University of Washington Human Subjects Division (STUDY00020362).

Conflict of interest: The authors have no competing interests to declare.

This article is protected by copyright. All rights reserved.