

External validation of a paediatric SMART triage model for use in resource limited facilities.

Joyce Kigo^{1*}, Stephen Kamau¹, Alishah Mawji³, Paul Mwaniki¹, Dustin Dunsmuir^{2,3}, Yashodani Pillay³, Cherri Zhang³, Katija Pallot³, Morris Ogero¹, David Kimutai⁵, Mary Ouma⁵, Ismael Mohamed⁵, Mary Chege⁶, Lydia Thurania⁶, Niranjana Kissoon⁴, J. Mark Ansermino^{2,3}, Samuel Akech¹

¹ Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme, Nairobi, Kenya

²Department of Anesthesiology, Pharmacology & Therapeutics, University of British Columbia, Vancouver, BC, Canada

³Centre for International Child Health, BC Children's Hospital Research Institute, Vancouver, BC, Canada

⁴Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada

⁵Mbagathi County Hospital

⁶Kiambu County Referral Hospital

***Correspondence**

jkigo@kemri-wellcome.org (JK)

Abstract

Introduction: Models for digital triage of sick children at emergency departments of hospitals in resource poor settings have been developed. However, prior to their adoption, external validation should be performed to ensure their generalizability.

Methods: We externally validated a previously published nine-predictor paediatric triage model (SMART Triage) developed in Uganda using data from two hospitals in Kenya. Both discrimination and calibration were assessed, and recalibration was performed by optimizing the intercept for classifying patients into emergency, priority, or non-urgent categories based on low-risk and high-risk thresholds.

Results: A total of 2539 patients were eligible at Hospital 1 and 2464 at Hospital 2, and 5003 for both hospitals combined; admission rates were 8.9%, 4.5%, and 6.8%, respectively. The model showed good discrimination, with area under the receiver-operator curve (AUC) of 0.826, 0.784 and 0.821, respectively. The pre-calibrated model at a low-risk threshold of 8% achieved a sensitivity of 93% (95% confidence interval, (CI):89%-96%), 81% (CI:74%-88%), and 89% (CI:85%–92%), respectively, and at a high-risk threshold of 40%, the model achieved a specificity of 86% (CI:84%–87%), 96% (CI:95%-97%), and 91% (CI:90%-92%), respectively. Recalibration improved the graphical fit, but new risk thresholds were required to optimize sensitivity and specificity.

Conclusion: The Smart Triage model showed good discrimination on external validation but required recalibration to improve the graphical fit of the calibration plot. There was no change in the order of prioritization of patients following recalibration in the respective triage categories. Recalibration required new site-specific risk thresholds that may not be needed if prioritization based on rank is all that is required. The Smart Triage model shows promise for wider application for use in triage for sick children in different settings.

Funder: Wellcome Trust (UK)

Key words: SMART, Digital triage, children, sub-Saharan Africa, LMICs

Introduction

The global burden of child mortality remains high in low and middle-income countries (LMICs). Despite significant progress globally, Sub-Saharan Africa continues to record mortality rates of 74 (95% confidence interval (CI), 68-86) deaths per 1000 live births, which is approximately 14 times higher than the mortality rate of children in Europe and North America (1, 2). These numbers, accounting for paediatric deaths outside the neonatal period, are largely attributed to infectious diseases including malaria, pneumonia, and diarrhoeal diseases which can be prevented or treated through simple interventions and training of healthcare workers (3).

Early recognition of critically ill children upon arrival to hospitals supports the attainment of the third Sustainable Development Goal (SDG) on ending preventable deaths of children by 2030, especially with increased availability and use of health facilities (4). Although the World Health Organization (WHO) has recommended the use of Emergency Triage and Treatment (ETAT) guidelines for the triage of sick children in resource-limited settings, its implementation in clinical practice has been challenging for many reasons including difficulties of scaling up the required training (5). Shortages of adequately trained frontline health workers in emergency areas has also been cited as impairing triage (6, 7). Without triage, children are frequently seen in order of arrival instead of according to the priority of illness.

Clinical prediction models that classify children on arrival to the hospital according to risk (that is, emergency, priority or non-urgent cases), can help frontline health workers identify critically ill children and prioritize care to reduce mortality and morbidity in resource-limited settings (8). Several predictive triage models have been developed but few models have been validated externally (8, 9).

Prediction models tend to perform well during internal validation, but this may not be replicated externally, therefore, it is recommended that model performance be examined in a different context (10-12). Ideally, clinical prediction models should be reproducible and generalizable to diverse patient groups and geographical locations and models that are not generalizable are a waste of limited research resources.

In this paper, we externally validate a previously developed nine-predictor pediatric triage model, referred to as the Smart Triage Model (13), used for the triaging of severely ill children. The original model was developed using data from Jinja, Uganda and this study performs external validation using data from two hospitals in Kenya. We also present results of updating the prediction model using recalibration-in-the-large, a method which adjusts the average predicted probability to be equal to the observed event rate.

Methods

Model external validation adhered to Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines on developing, validating, or updating a multivariable clinical prediction model (14) (Supplementary file S1).

Study registration

The study was registered in Clinical Trials.gov, identifier: NCT04304235, on 11th March 2020.

Study population and design

The Smart Triage model was developed in a study conducted at the pediatric emergency department (ED) in Jinja Regional Referral Hospital (JRRH), a public hospital within the Uganda Ministry of Health, between April 2020 and March 2021 (9). The Smart Triage model equation is a multiple logistic regression model that includes nine predictor variables which were selected using bootstrap stepwise regression and clinical expertise. The model

can be used for rapid identification of critically ill children at triage and can be integrated into any digital platform.

The validation of the Smart Triage model was performed using a dataset from two sites in Kenya, the Mbagathi County Hospital (Hospital 1) and Kiambu County Referral Hospital (Hospital 2), independently and in combination. Both hospitals are located in resource-limited urban settings and the pediatric outpatient departments (OPD) in each of the two hospitals receive approximately 20,000 patients per year with an admission rate of 10% and 7% respectively.

The study was approved by institutional review boards at Kenya Medical Research Institute Scientific Ethics Review Unit KEMRI (SERU#3958) in Kenya and the University of British Columbia in Canada (ID: H19-02398; H20-00484).

Sampling and Eligibility

Children seeking medical care at the OPD between 8:00 am and 5:00 pm on weekdays were screened for eligibility at JRRH, Uganda, and the two Kenya hospitals. In Kenya, assent was required for children above 13 years of age in addition to caregiver consent. Both sites only enrolled children presenting with medical illness and written informed consent was provided by a parent or guardian prior to enrollment. Children presenting for elective surgical procedures, scheduled clinic appointments, or those coming to be reviewed at the hospital for treatment of chronic illnesses were not eligible for enrollment. Emergency cases who were eligible but required immediate emergency treatment (within 15 minutes of arrival) were not enrolled to the study since care of the participant was prioritized. At both Kenya hospitals, three study nurses and two timekeepers were recruited and trained to conduct study-specific procedures. The timekeepers recorded arrival time for all patients arriving with an acute illness to the OPD and used a systematic sampling method based on 30-minute time cut-offs

to determine the next participant to approach for recruitment to the study. The study nurses verified patient eligibility, and if the first participant in a given cut off was not eligible or did not consent, they would examine the next participant in the same cut off depending on order of arrival. Study information was collected from those eligible after obtaining informed consent (6).

Data collection and management

The data collection method used to develop the initial model was repeated in hospitals 1 and 2(6). Data was collected using a custom-built Android mobile application installed on a password secured tablet. A Masimo iSpO2® Pulse Oximeter was connected to the tablet to measure the pulse oximetry and heart rate. Data on the tablet was stored in an encrypted format. Each day, data collected using the tablets was uploaded to a REDCap (Research Electronic Data Capture)(15) database in a central server housed at KEMRI Wellcome Trust Research Programme (KWTRP) office, Nairobi. Standard operating procedures were developed and used for data collection and these are available on the [Paediatric Sepsis CoLab Dataverse \(16\)](#).

Primary outcome

The primary outcome was defined as a composite of any one or more of the following: hospital admission for more than 24 hours as determined independently by the hospital clinician on duty (who was not part of the study team), mortality within 24 hours of admission or readmission, or referral within 48 hours to any other hospital after being seen by a hospital clinician/s, determined through follow up calls made 7 days after the initial hospital visit.

Predictors

The nine predictors included in the JRRH final model equation were square root of age in months, heart rate, temperature, mid-upper arm circumference, transformed oxygen saturation (using the concept of virtual shunt(17), parental concern, difficulty breathing, oedema, and pallor. The model was developed by multivariable logistic regression and the model equation which was referred to as the Smart Triage model equation is:

$$\text{logit}(p) = -32.888 + (0.252, \text{square root of age}) + (0.016, \text{heart rate}) + (0.819, \text{temperature}) + (-0.022, \text{mid-upper arm circumference}) + (0.048, \text{transformed oxygen saturation}) + (1.793, \text{parent concern}) + (1.012, \text{difficulty breathing}) + (1.814, \text{oedema}) + (1.506, \text{pallor}) \quad (13).$$

Sample size

We used a four-step procedure implemented in the `pmsampsize` R package (18) to determine the minimum required sample to perform validation of the model. Assuming an input C-statistic of 0.8, a prevalence of admission of 0.05, a Cox-Snell R-squared of 0.0697 based on 0.05 acceptable difference in apparent and adjusted R-squared, 0.05 margin of error in estimation of intercept, and a minimum number of events per predictor parameter(EPP) of 7, the minimal sample size required was 1117 participants with 64 events (19-22).

Observations with missing outcomes or missing 25% of the predictor variables were excluded from this analysis. All other missing values were imputed using K-Nearest neighbors (23).

Model validation and calibration

The Smart Triage model equation was applied to data from the two Kenya hospitals separately and then to the two data sets combined. The discriminative ability of the model was assessed using the area under the receiver-operator curve (AUC) and interpretation was based on the following criteria: non-informative ($AUC \leq 0.5$), poor discrimination ($0.5 <$

AUC < 0.7), and good discrimination (AUC ≥ 0.7) (24). Calibration, which is a measure of agreement between observed and predicted values, was assessed graphically by estimating the slope on a calibration plot of predicted versus observed outcome rates in each decile of predicted probability. A calibration slope of 1 and intercept of 0 is considered ideal (25), a slope value greater than 1 shows that the predictions are too narrow to distinguish a positive and negative outcome (26). We interpreted the calibration slope using the following criteria: non-informative (slope ≤ 0.5), poor calibration (0.5 < slope < 0.7), and good calibration (slope ≥ 0.7). We assessed calibration using the Hosmer- Lemeshow statistic (27) and a p-value <0.05 was considered significant.

To improve the calibration plot of predicted against observed values we performed recalibration-in-the-large (re-estimation of the model intercept) (28). Recalibration-in-the-large adjusts the intercept of the original model on a correction factor in equation 1.

$$\text{correction factor} = \ln\left(\frac{\frac{\text{observed outcome frequency}}{1 - \text{observed outcome frequency}}}{\frac{\text{mean predicted risk}}{1 - \text{mean predicted risk}}}\right) \dots\dots\dots \text{equation 1}$$

Risk stratification

A risk classification table was used to examine the accuracy of the recalibrated model in classifying patients into triage categories. The low-risk threshold was selected to maximize sensitivity to limit misclassification of emergency and priority cases as non-urgent (avoiding false negatives) while the high-risk threshold was selected to maximize specificity to limit misclassification of non-urgent or priority cases as emergency cases (avoiding false positives) (8). We computed the true and false positive rates, negative predictive values (NPVs), and positive predictive values (PPVs) for each triage group.

Results

Demographic characteristics

At Hospital 1, 2680 patients were screened for eligibility between 24th February 2021 and 6th November 2022; 2539 patients (94.7%) met the inclusion criteria and were included in analysis. Of those that were analyzed, 226 (8.9%) had a positive primary outcome and 79.8% of all participants were aged 5 years or younger. No participants had 25% of the predictor variables missing, but 13 participants were missing the admission outcome and were thus excluded from analysis (Fig 1). The most common reason for admission was pneumonia, diagnosed using clinical signs criteria, which accounted for 55.8% of the admissions; 56.2% of all admission were male (Table 1).

Figure 1: Sample population flow chart for dataset used in analysis at Mbagathi County Hospital (Hospital 1).

Table 1: Patient's characteristics stratified across those with and without outcome(admission)

	Mbagathi County Referral Hospital			Kiambu Teaching and Referral Hospital		
	Participant(N)	Admitted	Not admitted	Participant(N)	Admitted	Not admitted
Total Participants (%)	2539	226(8.9%)	2313(91.1%)	2464	112(4.5%)	2352(95.5%)
Sex						
Female (%)	1112(43.8%)	98(43.4%)	1014(91.2%)	1211(49.1%)	57(50.9%)	1154(49.1%)

Male (%)	1427(56.2%)	128(56.6%)	1299(91%)	1253(50.9%)	55(49.1%)	1198(50.9%)
Age						
Less than 30 days	78(3.1%)	23(10.2%)	55(2.4%)	90(3.7%)	24(21.4%)	66(2.8%)
30 days - <=1 Year	674(26.5%)	112(49.6%)	562(24.3%)	809(32.8%)	46(41.1%)	763(32.4%)
1 - <=2 years	511(20.1%)	49(21.7%)	462(20%)	628(25.5%)	26(23.2%)	602(25.6%)
2 - <=3 years	312(12.3%)	14(6.2%)	298 (12.9%)	421(17.1%)	7(6.3%)	414 (17.6%)
3 - <=4 years	250(9.8%)	12(5.3%)	238(10.3%)	276(11.2%)	8(7.1%)	268(11.4%)
4 - <=5 years	202(8%)	3(1.3%)	199(8.6%)	199(8.1%)	1(0.9%)	198(8.4%)
>5 - <=12 years	512(20.2%)	13(5.8%)	499(21.6%)	41(1.7%)	0	41(1.8%)
Admission diagnosis						
Malaria		7(3.1%)			0(0%)	
Septicemia		5(2.2%)			1(0.9%)	
Neonatal Sepsis		8(3.6%)			5(4.5%)	
Pneumonia		126(55.8%)			67(59.8%)	

Bronchiolitis		3 (1.3%)			1(0.9%)	
Gastroenteritis/Diarrhea		5(2.2%)			1(0.9%)	
Meningitis/encephalitis or other CNS infection		9(4%)			1(0.9%)	
Any skin or soft tissue infection		3(1.3%)			0(0%)	
Malnutrition		8(3.6%)			0(0%)	
Reactive Airway Disease/Asthma		0(0%)			1(0%)	
Dehydration		16(7.1%)			5(4.5%)	
Other (e.g., fever, convulsions, Jaundice, Neonatal Jaundice etc.)		36(15.9%)			30(26.8%)	

At Hospital 2, 2671 participants were screened for eligibility between 24th February 2021 and 6th November 2022, of whom 2464 (92.3%) participants met the inclusion criteria and 112(4.5%) participants had a positive primary outcome. One participant had missing admission outcome and 3 withdrew from the study after enrolment and thus were excluded from this analysis. No patient had more than 25% of predictor variables missing (Fig 2). Among the 112 admitted participants, 50.9% were female and 98.3% of all the participants were aged 5 years or younger. The most common reason for admission was pneumonia which

accounted for 59.8% of the admissions (Table 1). For both hospitals combined 5003 participants were analyzed and 338(6.8%) participants had positive primary outcome.

Figure 1: Sample population flow chart for dataset used for analysis at Kiambu County Referral Hospital (Hospital 2).

Model performance

At Hospital 1 the model achieved good discrimination with an AUC value of 0.826 (Fig3A) and a calibration slope value of 0.72. Graphically the calibration plot improved on performing recalibration-in-the-large (Fig 4A, 4B) when the model intercept was adjusted from -32.888 to -34.393. The Hosmer-Lemeshow test pre-calibration had a p-value < 0.00001 and on recalibration the p-value was <0.01, both p-values were below 0.05 significance level.

Figure 2: Area Under the receiver-operator curve-AUC; This shows the discrimination ability of model (C-statistic) in respective hospitals.

Figure 3: Pre-calibration and post-calibration plot; On the left are the pre-calibration plots for respective sites and on the right are post-recalibration plots for respective site; The shift of the predicted values can be explained by the histograms

At Hospital 2 the model achieved good discrimination with an AUC value of 0.784 (Fig 3B) and a calibration slope value of 1.00. Graphically the calibration plot improved (Fig 4C,4D) when the new model intercept was adjusted from -32.888 to -34.201. The Hosmer-Lemeshow statistic pre-calibration had (p-value < 0.00001) and on recalibration had p-value =0.31.

For both sites combined the model achieved good discrimination with an AUC value of 0.821(Fig 3C) and calibration slope value of 0.796. On performing recalibration, the model calibration slope improved (Fig 4E, 4F) when the new model intercept was adjusted from -32.888 to -34.325. The Hosmer-Lemeshow test pre-calibration had a (p-value < 0.00001) and on recalibration had p-value <0.05.

Risk classification

At Hospital 1 before performing recalibration the low-risk threshold of 0.08 achieved a sensitivity (CI) of 93% (89% to 96%) and at a higher risk threshold of 0.4 the model achieved specificity (CI) of 86% (84% to 87%). On performing recalibration, the predicted values shifted toward zero and this required a new set of thresholds to maintain the numbers assigned to the various triage categories; at a low-risk threshold of 0.026 the model achieved sensitivity of 90% CI (86% to 93%) and at high-risk threshold of 0.13 the model achieved specificity of 86% CI (85% to 88%). Before calibration 60% of the admitted participants were categorized in the emergency category and this was unchanged with the use of the optimized thresholds (Table 2).

Table 2: Pre-calibration and post-recalibration risk classification table at selected thresholds with low-risk threshold selected to maximize sensitivity (limit misclassification of emergency and priority cases as non-urgent (avoiding false negatives)) while the high-risk threshold

selected to maximize specificity (increase correct classification of emergency cases (avoiding false positives)).

Precalibration table							
	Triage Category(Risk Threshold range)	Participant n (%)	Participant with outcome, n (%)	Sensitivity(95%CI)	Specificity(95%CI)	NPV	PPV
Hospital 1	Emergency(>=0.4)	463(18%)	136(60%)	0.60(0.54,0.66)	0.86(0.85,0.87)	0.96(0.95,0.96)	0.29(0.27,0.32)
	Priority(>0.08<0.4)	1027(40%)	73(32)%	0.75(0.69,0.81)	0.77(0.75,0.79)	0.97(0.96,0.98)	0.24(0.22,0.26)
	Non-urgent(<=0.08)	1049(41%)	17(8%)	0.93(0.89,0.96)	0.45(0.43,0.47)	0.98(0.98,0.99)	0.14(0.13,0.15)
Hospital 2	Emergency(>=0.4)	135(5%)	41(37%)	0.37(0.28,0.46)	0.96(0.95,0.97)	0.97(0.97,0.97)	0.30(0.24,0.37)
	Priority(>0.08<0.4)	992(40%)	50(45%)	0.64(0.55,0.73)	0.81(0.80,0.83)	0.98(0.97,0.99)	0.14(0.12,0.16)
	Non-urgent(<=0.08)	1337(54)	21(19%)	0.81(0.74,0.88)	0.56(0.54,0.58)	0.98(0.98,0.99)	0.08(0.07,0.09)
Combined	Emergency(>=0.4)	598(12%)	177(52%)	0.52(0.47,0.58)	0.91(0.90,0.92)	0.96(0.96,0.97)	0.23(0.27,0.33)
	Priority(>0.08<0.4)	2019(40%)	123(36%)	0.67(0.62,0.72)	0.83(0.82,0.84)	0.97(0.97,0.98)	0.23(0.21,0.24)
	Non-urgent(<=0.08)	2386(48%)	38(11%)	0.89(0.85,0.92)	0.50(0.49,0.52)	0.98(0.98,0.99)	0.12(0.11,0.12)
Recalibration table							
Hospital 1	Emergency (>=0.130)	458(18%)	135(60%)	0.60(0.53,0.66)	0.86(0.85,0.87)	0.96(0.95,0.96)	0.30(0.27,0.33)
	Priority(>0.026<0.130)	750(30%)	68(30%)	0.71(0.65,0.77)	0.79(0.78,0.81)	0.97(0.96,0.97)	0.25(0.23,0.27)
	Non-urgent(<=0.026)	1331(52%)	23(10%)	0.90(0.86,0.94)	0.57(0.55,0.59)	0.98(0.98,0.99)	0.17(0.16,0.18)
Hospital 2	Emergency (>=0.108)	210(9%)	51(46%)	0.46(0.36,0.55)	0.93(0.92,0.94)	0.97(0.97, 0.98)	0.24(0.20, 0.29)
	Priority(>0.020<0.0.108)	1088(44%)	43(38%)	0.57(0.48,0.66)	0.86(0.85,0.88)	0.98(0.97,0.98)	0.14 (0.17 ,0.19)
	Non-urgent(<=0.020)	1166(47%)	18(16%)	0.84(0.77,0.9)	0.49(0.47,0.51)	0.99(0.98,0.99)	0.07(0.06,0.08)
Combined	Emergency (>=0.106)	744(15%)	199(59%)	0.59(0.54,0.64)	0.88(0.87,0.89)	0.97(0.96,0.97)	0.27(0.25,0.29)
	Priority(<0.020<0.106)	1904(38%)	101(30%)	0.69(0.65,0.74)	0.82(0.80,0.83)	0.97(0.97,0.98)	0.21(0.20,0.23)
	Non-urgent(<=0.020)	2355(47%)	38(11%)	0.89(0.85,0.92)	0.50(0.48,0.51)	0.98(0.98,0.99)	0.11(0.11,0.12)

At Hospital 2 before performing recalibration, the model achieved sensitivity of 81 CI (74% to 88%) and specificity of 96% CI (95% to 97%). On performing recalibration with a low-risk threshold of 0.020 the model achieved sensitivity of 84% CI (77% to 90%) and at a high-risk threshold of 0.108 the model achieved specificity of 93% CI (92% to 94%). Before calibration, 37% of the admitted participants were categorized in the emergency category and on recalibration, 46% were categorized in the emergency category (Table 2).

For both hospitals combined before performing recalibration at low-risk threshold of 0.08 the model achieved a sensitivity of 89% CI (85% to 92%) and at the higher risk threshold of 0.4

the model achieved a specificity of 91.0% CI (90% to 92%). On performing recalibration at low-risk threshold of 0.020 the model achieved sensitivity of 89% CI (85% to 92%) and at high-risk threshold of 0.106 the model achieved a specificity of 88% (87% to 89%). Before calibration 53% of the admitted participants were categorized in the emergency category and on recalibration, 59% were categorized as emergency (Table 2).

Discussion

Key results

Performance of the Smart Triage model showed stable discrimination in all the three sets of data which suggests that for a pair of randomly selected children, the model would assign the higher risk score to the one with positive outcome compared to an individual with a negative outcome. On graphically assessing calibration of the predicted against observed outcomes, the graphical plot deteriorated in all the three sets of data shifting towards overprediction. Recalibration-in-the-large improved the calibration plot, but this required a change in risk thresholds to optimize sensitivity and specificity and organize patients into clinically manageable risk classification triage categories.

The stability of discrimination and increasing overprediction or underprediction has been observed in parallel prediction validation studies (29-32). The change in performance often occurs because of data variation in the patient population including changes in the outcome rate, disease incidence, and prevalence across different regions, patient case mix and clinical practice (33-36). In our case, this was attributed to fewer admissions in both sites compared to the hospital where the model was developed and the lower number of participants with pallor (anaemia), one of the predictor variables. The lower prevalence of pallor at the study hospitals compared to the primary hospital is expected because the primary hospital was located in a malaria endemic zone (anaemia is a common complication of malaria) while the

validation hospitals are in an area without local malaria transmission. When an algorithm is developed in a setting with high disease prevalence it may systematically overestimate risk when used in settings with lower disease incidence (37).

Appropriately implemented prediction models can be helpful in supporting decision making to improve patient outcomes and prioritize allocation of resources. Integrating predictive analytics into electronic health record systems enables the use of predictive models and can allow incorporation of probability-based tools into clinical decision support systems (29, 33, 38). In this study we performed external validation on an existing paediatric triage model (Smart Triage Model) which could be integrated into a health record systems in LMIC, after appropriate contextualization, to strengthen the triage system.

While recalibration improved the fit of the model, it also required an adjustment of risk thresholds. This would limit more generalized adoption of the model if recalibration was required at each site. Future work will investigate options to predict optimal calibration based on easily collected site specific information (such as admission rate or malaria prevalence) or the option of using the same model and thresholds, even if calibration is not optimal.

Limitations

A significant limitation of this study was the choice of hospital admission as primary outcome measure as it is not the most robust measure of illness severity. However, to exclude admitted cases that lacked severe illness we included only children admitted more than 24 hours. To capture children who were sent home but had severe illness we included children who were readmitted within 48 hours determined from a follow up call made 7 days after discharge. We also included mortality within 24 hours of admission or readmission to still capture the severely ill.

Patient characteristics tend to change over time and in different geographical settings. There may be arguments that the findings need replication in more geographical settings with varied event rate, case mix, and hospital contexts but we believe that findings of external validation of the Smart Triage model from these two hospitals is reassuring on its adaptability to varied contexts and patient populations.

Conclusion

The Smart Triage model showed good discrimination on external validation but required modest recalibration to improve the graphical fit of the calibration plot. On recalibration, new site-specific set of thresholds were required to maintain the same sensitivity and specificity across the triage categories. There was no significant change in the distribution of patients into the three triage categories, which alleviates concerns about model updating for prediction if prioritization based on rank is all that is required. Future research could examine whether the Smart Triage model can be applied to different populations if only the risk thresholds are adjusted without recalibration. The Smart Triage model shows promise for wider application for use in triage for sick children in different settings.

Declarations

Author contributions

SA, AM, NK DD, and JMA designed the study. SK, JK, YP, KP DD, DK, MO, IM, MC, LT SA and JMA were responsible for study implementation and monitoring. DD was responsible for software development. JK, DD, SK, MO and PM were responsible for data management. JK, CZ conducted the statistical analysis. JK wrote the original manuscript, which was critically reviewed and revised by the team. All authors approved the final manuscript.

Declaration of Interests

We declare no competing interests.

Data sharing

The data studied and data dictionary is available upon request and users who wish to reuse the source data can make a request through KEMRI-Wellcome Trust Research Program data governance committee. This committee can be contacted at dgc@kemri-wellcome.org.

Acknowledgements

We would like to thank the administration in both Mbagathi County Hospital and Kiambu County Referral Hospital and all participants and caregivers for their contribution in making this study a success. We would like to thank the research team in both hospitals i.e., Sidney Korir, Kevin Bosek, Verah Karasi, John Mboya, Brian Ochieng, Angela Wanjiru, Mercy Mutuku, Anastasia Gathigia, Esther Muthoni, Emmah Kinyanjui, Felix Kimani, Faith Wairimu for their hard work and dedication in data collection. This research was supported by Wellcome Trust.

Role of the Funding Source

The study's funders were not involved in the study's conception, data collecting, analysis, interpretation, or manuscript-writing processes. Each author had complete access to the study's data and agreed to submit it for publication.

References

1. UNICEF. UNICEF Data: Monitoring the situation of children and women 2020 [Available from: <https://data.unicef.org/topic/child-survival/under-five-mortality/>].
2. WHO. Child mortality and causes of death 2020 [Available from: <https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/child-mortality-and-causes-of-death>].
3. Kissoon N, Carapetis J. Pediatric sepsis in the developing world. *Journal of Infection*. 2015;71:S21-S6.
4. WHO. Children: improving survival and well-being 2020 [Available from: <https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality>].
5. Organization WH. Guideline: updates on paediatric emergency triage, assessment and treatment: care of critically-ill children. Guideline: updates on paediatric emergency triage, assessment and treatment: care of critically-ill children 2016.
6. Mawji A, Li E, Komugisha C, Akech S, Dunsmuir D, Wiens MO, et al. Smart triage: triage and management of sepsis in children using the point-of-care Pediatric Rapid Sepsis Trigger (PRST) tool. *BMC health services research*. 2020;20(1):1-13.
7. Baker T. Critical care in low-income countries. *Tropical Medicine & International Health*. 2009;14(2):143-8.
8. Mawji A, Akech S, Mwaniki P, Dunsmuir D, Bone J, Wiens MO, et al. Derivation and internal validation of a data-driven prediction model to guide frontline health workers in triaging children under-five in Nairobi, Kenya. *Wellcome Open Research*. 2019;4.
9. Mawji A. Smart Triage Jinja Model. 2021.
10. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *Bmj*. 2009;338.

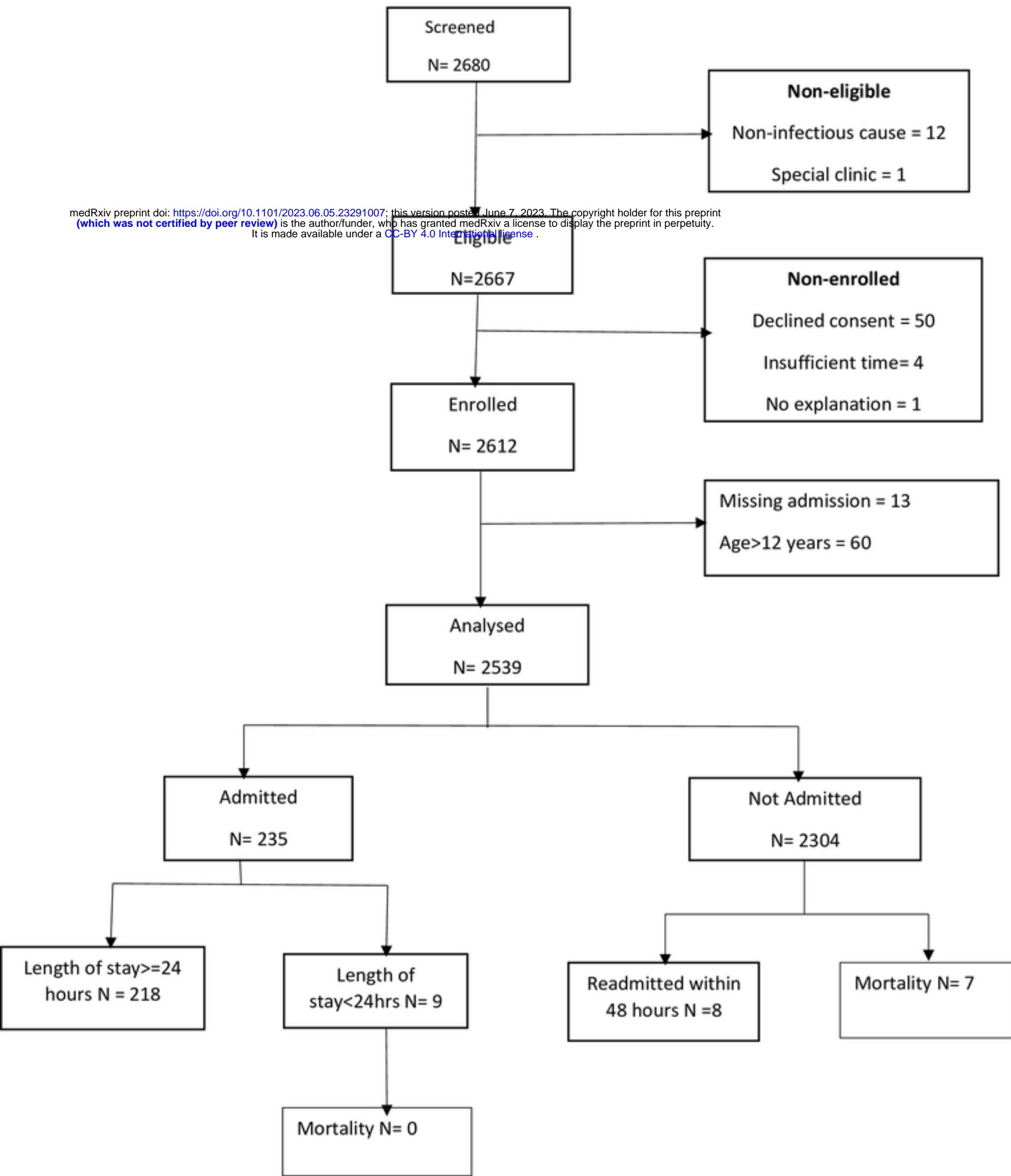
11. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *Bmj*. 2009;338.
12. Bleeker S, Moll H, Steyerberg Ea, Donders A, Derksen-Lubsen G, Grobbee D, et al. External validation is necessary in prediction research:: A clinical example. *Journal of clinical epidemiology*. 2003;56(9):826-32.
13. Mawji A, Li E, Dunsmuir D, Komugisha C, Novakowski SK, Wiens MO, et al. Smart triage: Development of a rapid pediatric triage algorithm for use in low-and-middle income countries. *Frontiers in Pediatrics*. 2022;10.
14. Moons KG, Altman DG, Reitsma JB, Collins GS. New guideline for the reporting of studies developing, validating, or updating a multivariable clinical prediction model: the TRIPOD statement. *Advances in anatomic pathology*. 2015;22(5):303-5.
15. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics*. 2009;42(2):377-81.
16. Mawji A. Smart Triage Jinja Standard Operating Protocols. V1 ed: Borealis; 2021.
17. Tüshaus L, Moreo M, Zhang J, Hartinger SM, Mäusezahl D, Karlen W. Physiologically driven, altitude-adaptive model for the interpretation of pediatric oxygen saturation at altitudes above 2,000 m asl. *Journal of Applied Physiology*. 2019;127(3):847-57.
18. Ensor J, Martin EC, Riley RD. Package ‘pmsampsize’. 2022.
19. Van Smeden M, de Groot JA, Moons KG, Collins GS, Altman DG, Eijkemans MJ, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC medical research methodology*. 2016;16(1):1-12.

20. van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. *Statistical methods in medical research*. 2019;28(8):2455-74.
21. Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II-binary and time-to-event outcomes. *Statistics in medicine*. 2019;38(7):1276-96.
22. Riley RD, Ensor J, Snell KI, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *Bmj*. 2020;368.
23. Ling W, Dong-Mei F, editors. Estimation of missing values using a weighted k-nearest neighbors algorithm. *2009 International Conference on Environmental Science and Information Application Technology*; 2009: IEEE.
24. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass)*. 2010;21(1):128.
25. Van Hoorde K, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW, Van Calster B. Assessing calibration of multinomial risk prediction models. *Statistics in Medicine*. 2014;33(15):2585-96.
26. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *bmj*. 2016;353.
27. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. John Wiley & Sons. New York. 2000.
28. Vergouwe Y, Nieboer D, Oostenbrink R, Debray TP, Murray GD, Kattan MW, et al. A closed testing procedure to select an appropriate method for updating prediction models. *Statistics in medicine*. 2017;36(28):4529-39.

29. Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. *Journal of the American Medical Informatics Association*. 2017;24(6):1052-61.
30. Harrison DA, Lone NI, Haddow C, MacGillivray M, Khan A, Cook B, et al. External validation of the Intensive Care National Audit & Research Centre (ICNARC) risk prediction model in critical care units in Scotland. *BMC anesthesiology*. 2014;14(1):1-9.
31. Madan P, Elayda MA, Lee V-V, Wilson JM. Risk-prediction models for mortality after coronary artery bypass surgery: application to individual patients. *International journal of cardiology*. 2011;149(2):227-31.
32. Hickey GL, Grant SW, Murphy GJ, Bhabra M, Pagano D, McAllister K, et al. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *European Journal of Cardio-Thoracic Surgery*. 2013;43(6):1146-52.
33. Parikh RB, Kakad M, Bates DW. Integrating predictive analytics into high-value care: the dawn of precision delivery. *Jama*. 2016;315(7):651-2.
34. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *Bmj*. 2009;338.
35. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS medicine*. 2013;10(2):e1001381.
36. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-8.

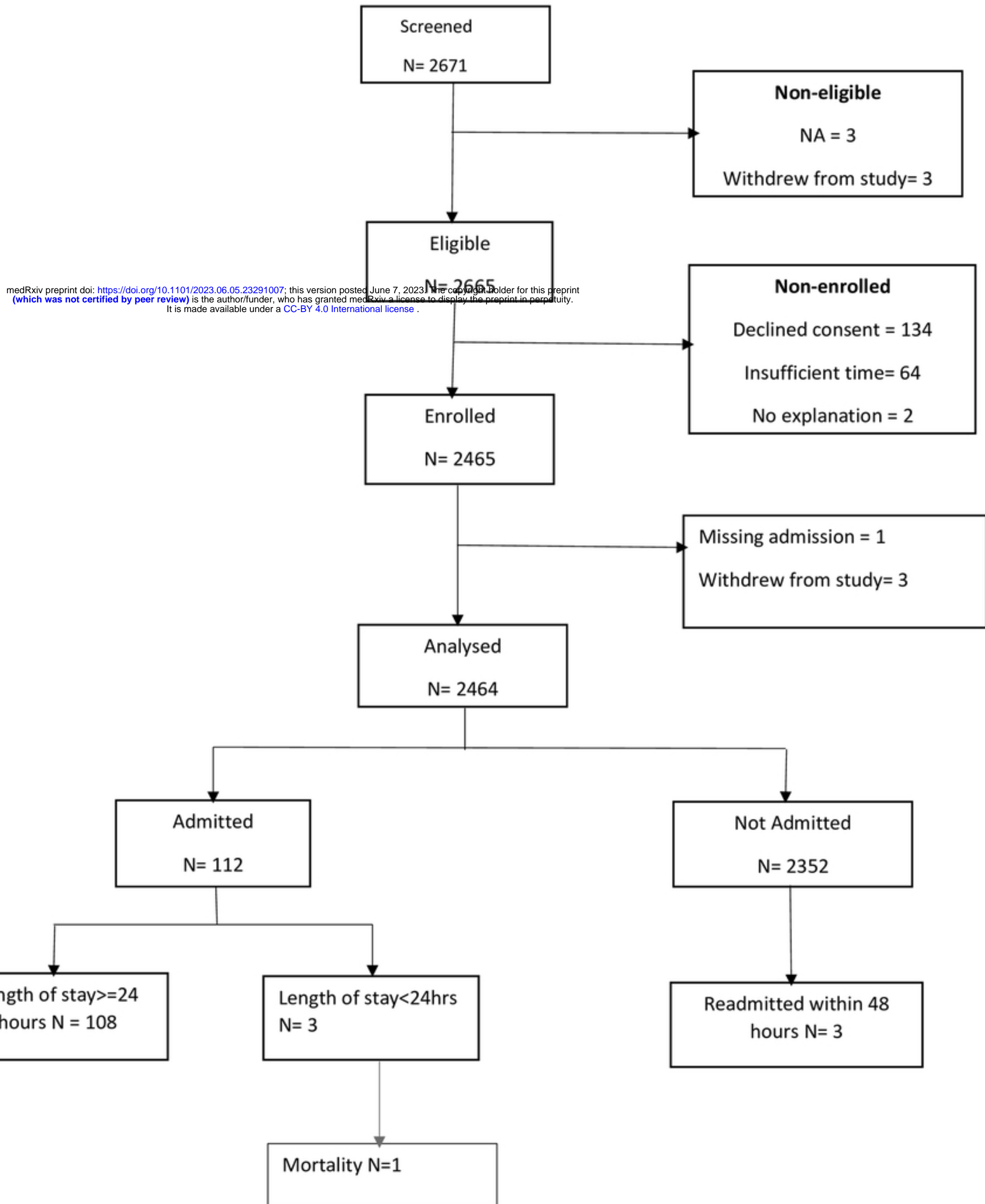
37. Steyerberg E, Roobol M, Kattan M, Van der Kwast T, De Koning H, Schröder F.
Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram.
The Journal of urology. 2007;177(1):107-12.
38. Steyerberg EW, van der Ploeg T, Van Calster B. Risk prediction with machine
learning and regression methods. Biometrical Journal. 2014;56(4):601-6.

Figure 1: Diagram of participant flow Mbagathi County Hospital



Figure

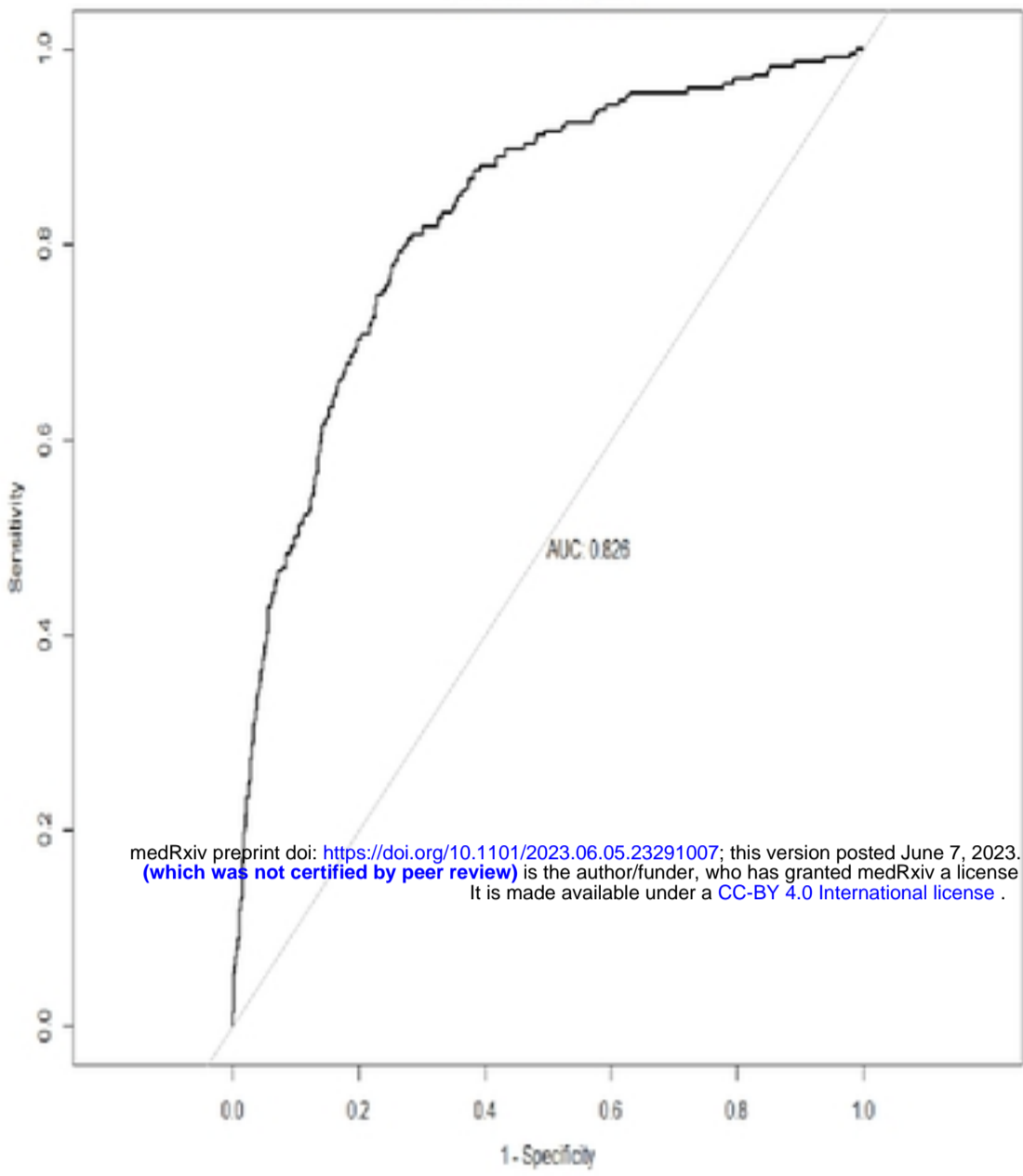
Figure 2: Diagram of participant flow- Kiambu teaching and referral hospital



Figure

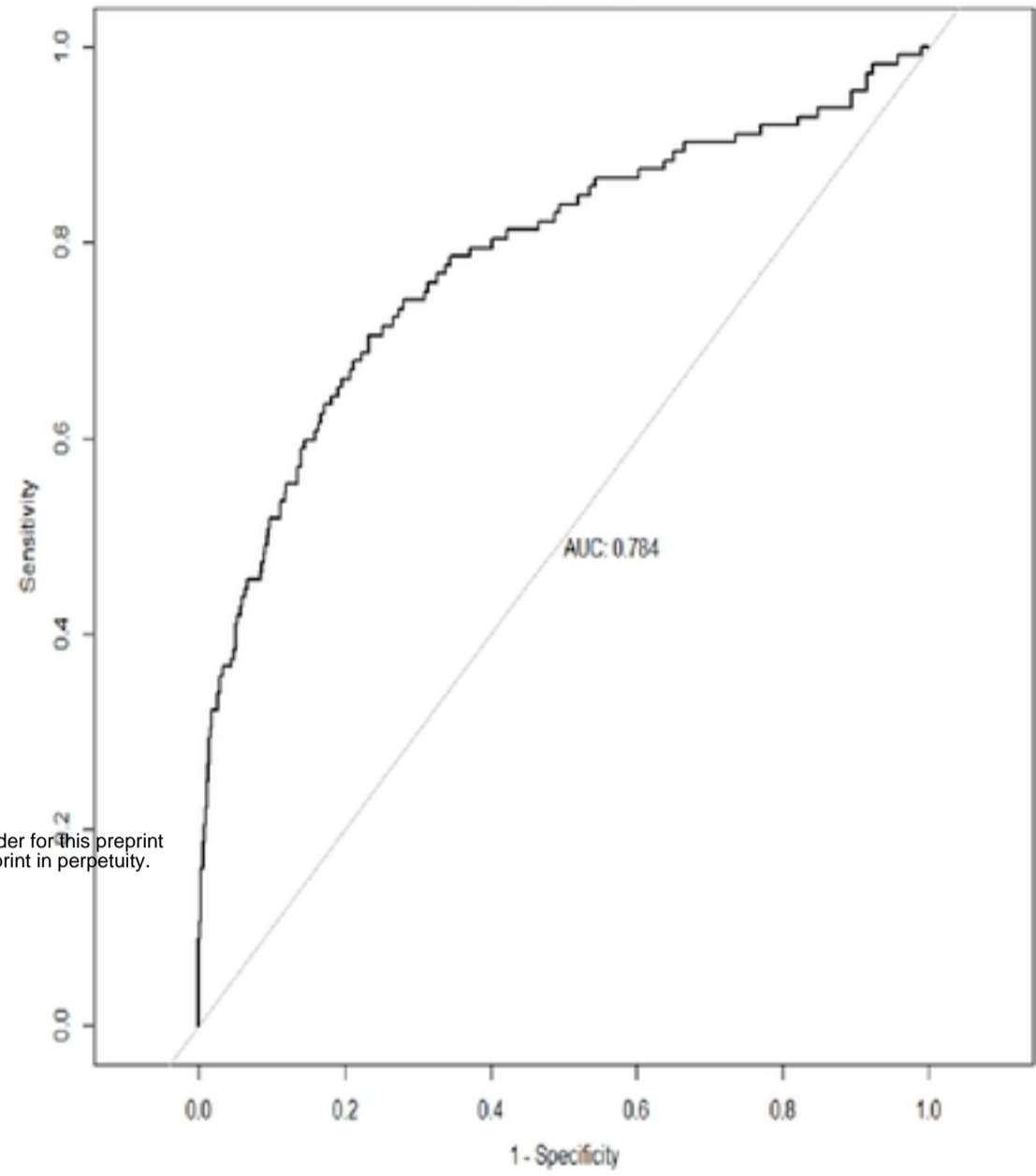
(a)

AUC Curve- Hospital 1



(b)

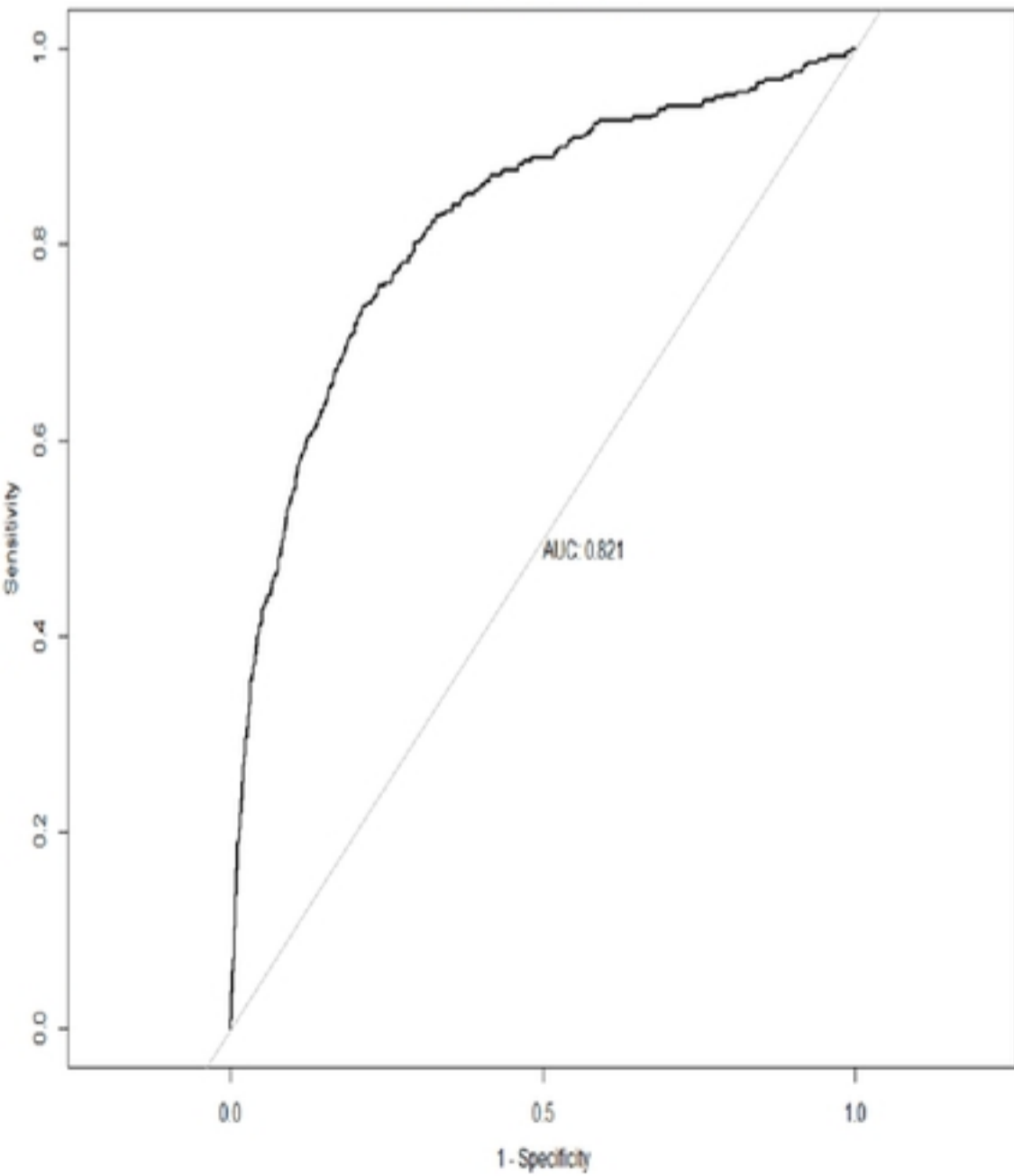
AUC Curve- Hospital 2



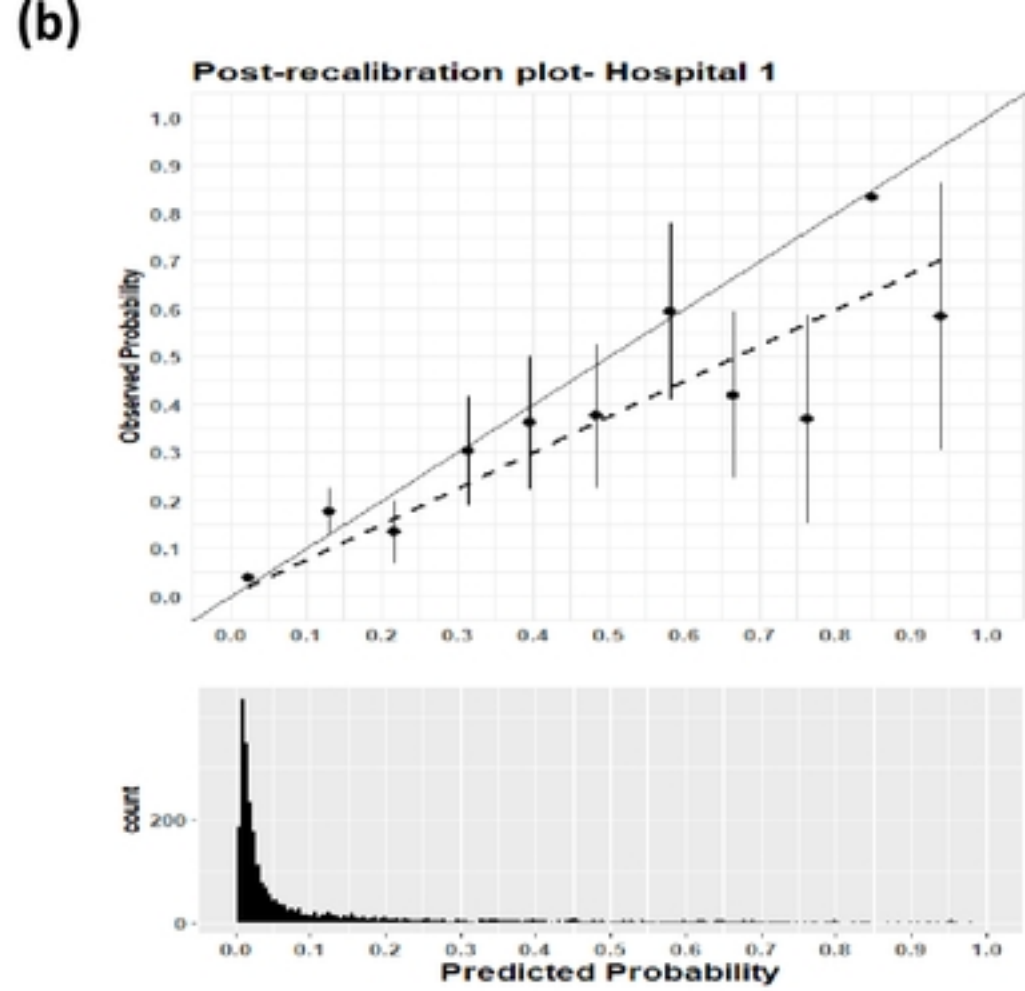
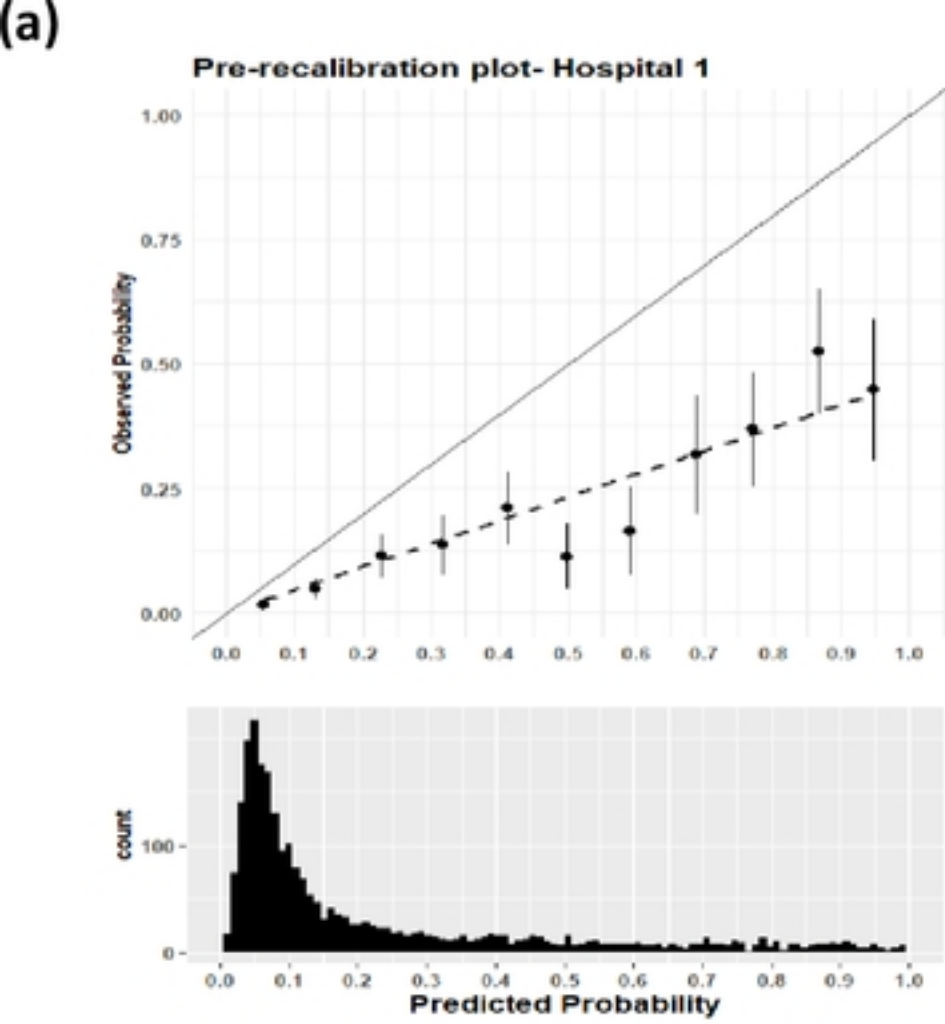
medRxiv preprint doi: <https://doi.org/10.1101/2023.06.05.23291007>; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

(c)

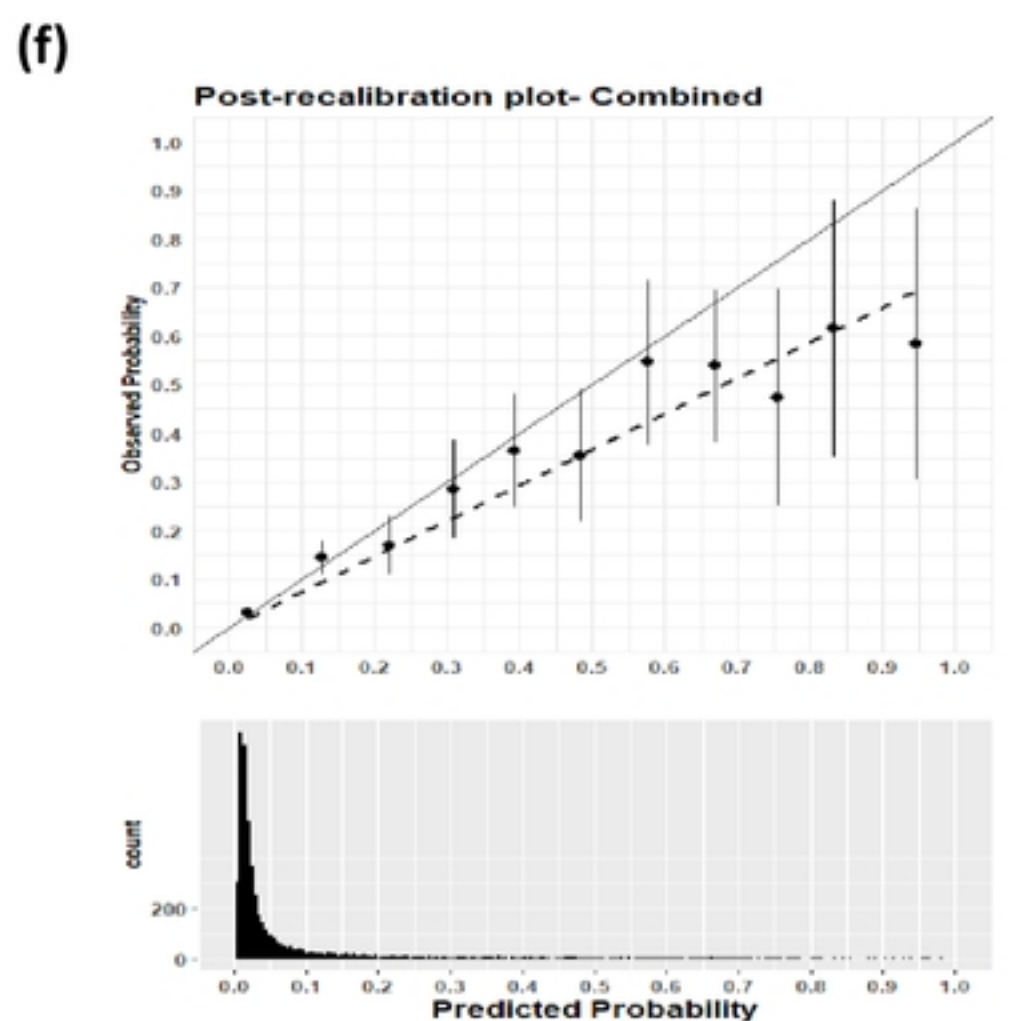
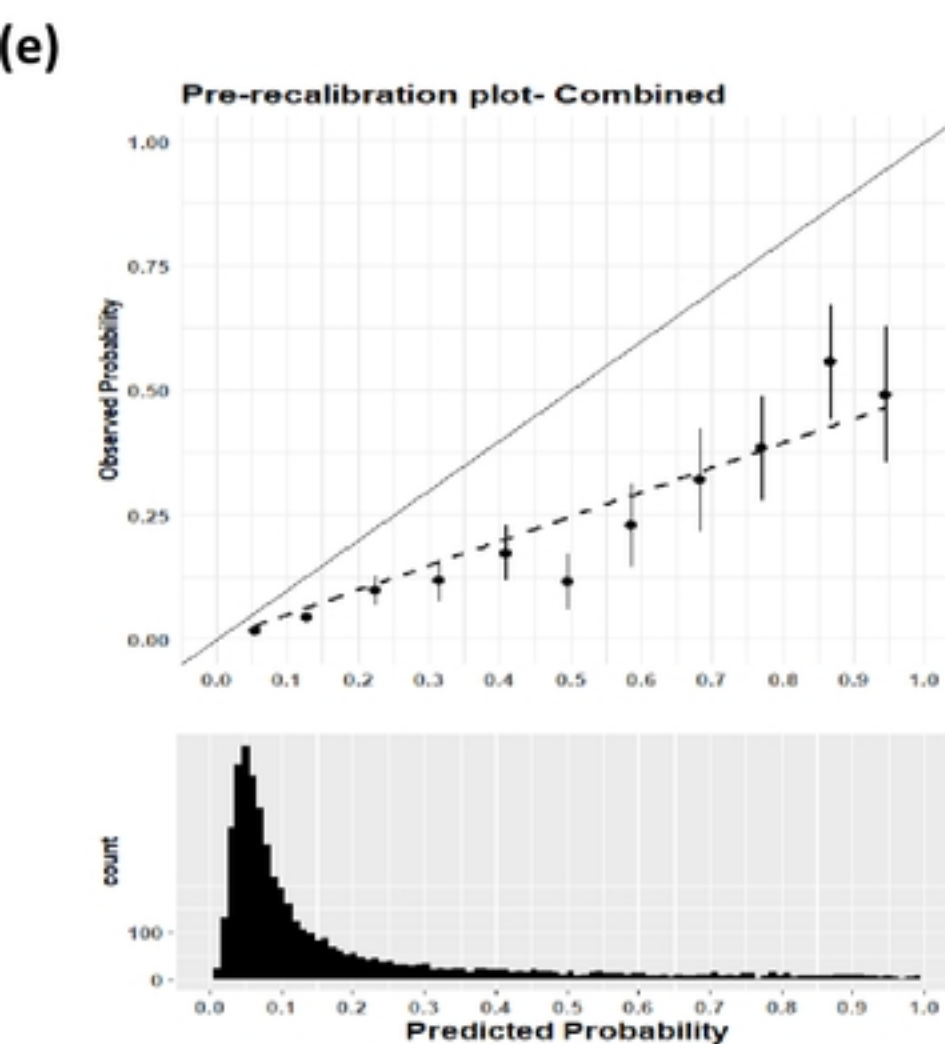
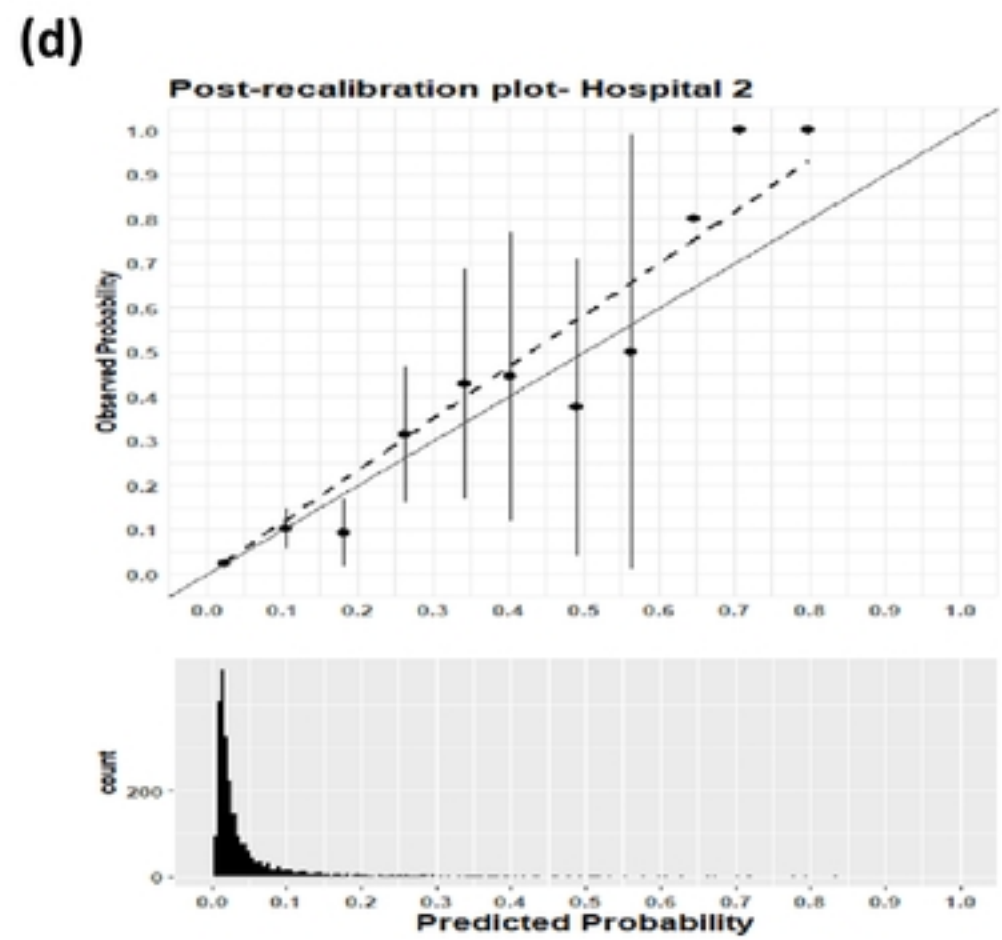
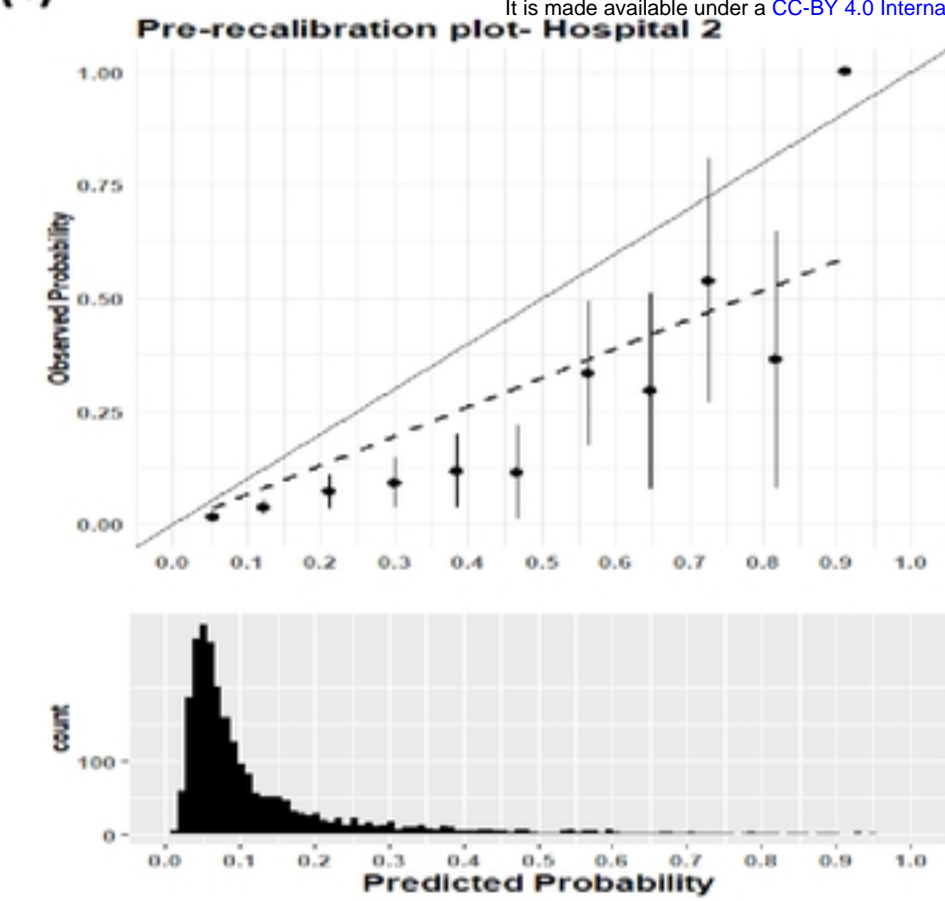
AUC Curve- Combined



Figure



(c) medRxiv preprint doi: <https://doi.org/10.1101/2023.06.05.23291007>; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



Figure