American Journal of Theoretical and Applied Statistics
2026, Vol. 15, No. 5, pp. 213–230
https://doi.org/10.11648/j.ajtas.20261505.13

Research Article

Modelling Mortality Risk in Malaria Patients Using Logistic Regression Model: A Case Study of Nakuru Level 6 Hospital

Waruino John Mwangi*, Joseph Eyang'an Esekon, Harun Mwangi Gitonga

Department of Pure and Applied Sciences, Kirinyaga University, Kutus, Kenya

Abstract

Malaria remains a major cause of preventable illness and death in sub-Saharan Africa, yet routinely collected hospital data are not consistently integrated into objective tools for identifying patients at greatest risk of death. This study developed and internally evaluated a five-predictor logistic regression model for in-hospital mortality among malaria patients admitted to Nakuru Level 6 Hospital, Kenya. A retrospective observational design was used, drawing on electronic medical records from 1,500 patients admitted between 2020 and 2026. In-hospital death was the binary outcome, with age, sex, parasite density, haemoglobin level, and platelet count included as predictors. Fifty-four patients died, corresponding to a mortality rate of 3.6%. The original prespecified linear model showed strong discrimination (AUC = 0.922), but diagnostic testing identified significant nonlinearity for age, haemoglobin level, and platelet count. These three predictors were therefore refitted using restricted cubic splines while retaining the same five clinical predictors. The corrected nonlinear model significantly improved fit and discrimination, with an apparent AUC of 0.962. Bootstrap validation produced an optimism-corrected AUC of 0.955 and a Brier score of 0.0275, while repeated stratified cross-validation produced an AUC of 0.954 and a Brier score of 0.0273. Calibration intercepts were close to zero, although slopes below one indicated some residual optimism. Parasite density remained positively associated with mortality, sex remained non-significant, and nonlinear relationships were observed for age, haemoglobin, and platelet count. The conventional 0.5 threshold showed poor sensitivity, while lower thresholds improved detection but were not considered clinically established cut-offs. The findings support use of routine hospital data for mortality risk stratification, but external validation, possible recalibration, and prospective evaluation of clinical thresholds and utility are required before implementation in practice.

Keywords

Malaria Mortality, Logistic Regression, Clinical Prediction Model, Parasite Density, Haemoglobin, Platelet Count, Internal Validation, Kenya

1. Introduction

Malaria remains one of the most consequential infectious diseases worldwide, with morbidity and mortality concentrated in sub-Saharan Africa. The World Health Organization continues to identify the African region as the principal contributor to

*Correspondence: Waruino John Mwangi (johnmwangi3801@gmail.com)

Received: 7 August 2026; **Accepted:** 19 August 2026; **Published:** 11 September 2026

global malaria cases and deaths [1, 2]. Although prevention, diagnostic, and treatment strategies have improved, fatal outcomes persist among patients who present with severe disease, substantial parasite burden, haematological compromise, or delayed access to effective care. The continued occurrence of preventable deaths creates a need for methods that identify high-risk patients early enough to guide triage and clinical monitoring.

Kenya is malaria-endemic, but transmission intensity varies with altitude, climate, ecology, and population characteristics. National malaria policy therefore emphasizes prevention, effective case management, surveillance, and targeted use of limited resources [3]. The heterogeneity of malaria risk across Kenyan counties also means that patient profiles and clinical outcomes observed in one setting may not be directly transferable to another [4]. Referral hospitals, such as Nakuru Level 6 Hospital, receive patients with diverse disease severities and provide a useful setting for locally grounded prediction research based on information already available in electronic medical records.

Mortality in malaria is multifactorial. Age captures variation in acquired immunity, physiological reserve, and comorbidity. Parasite density reflects infection burden and is associated with severe anaemia, metabolic disturbance, organ dysfunction, and other complications [5, 6]. Haemoglobin concentration provides a direct measure of anaemia-related physiological compromise, while thrombocytopenia is frequently observed in malaria and may indicate systemic inflammation and severe infection [5, 7]. Sex may also influence exposure, biological vulnerability, pregnancy-related risk, and health-seeking behaviour, although its independent effect can diminish after adjustment for clinical severity. These predictors are attractive for a hospital model because they are routinely measured, objective, and available early in the admission pathway.

Logistic regression is widely used for binary clinical outcomes because it estimates event probabilities while providing interpretable regression coefficients, odds ratios, and confidence intervals [8, 9]. Its interpretability is particularly valuable in settings where clinicians need to understand why a patient has been classified as high risk. Nevertheless, the usefulness of a prediction model depends on more than statistical significance. Calibration, discrimination, threshold-dependent performance, internal validation, and model assumptions must be evaluated transparently. This is especially important when mortality is rare, because high overall accuracy can conceal poor detection of patients who die.

Previous malaria studies support the relevance of demographic, parasitological, and haematological predictors, but evidence remains context-dependent. Delayed treatment has been linked to severe malaria and anaemia [10], while logistic models of malaria severity have demonstrated that apparent accuracy may coexist with low recall for the severe-outcome group [11]. Effects can also vary by transmission intensity, patient group, and outcome definition, supporting the development of locally calibrated models rather than the direct adoption of estimates from other settings.

This study therefore modelled in-hospital mortality among malaria patients admitted to Nakuru Level 6 Hospital using age, sex, parasite density, haemoglobin level, and platelet count. The objectives were to fit the original prespecified five-predictor logistic regression model, estimate the associations of the predictors with 95% confidence intervals, assess model assumptions and functional form, compare the original linear specification with a corrected nonlinear specification where required, and evaluate model performance using discrimination, calibration, classification measures, and internal validation. Model complexity, shrinkage, threshold stability, and potential clinical utility were also examined.

2. Materials and Methods

2.1. Study Design and Setting

A retrospective observational study was conducted using routinely collected electronic medical records from Nakuru Level 6 Hospital in Nakuru County, Kenya. The hospital is a referral facility that manages patients requiring advanced assessment and treatment, including patients admitted with malaria. The design enabled evaluation of real-world demographic and laboratory measurements without introducing an intervention. Records from 2020 to 2026 were considered.

2.2. Study Population, Eligibility, and Data Source

The study population consisted of admitted patients with a documented diagnosis of malaria. A non-probability purposive approach was used to include eligible records with complete information for mortality status and the selected predictors. Records were excluded when the mortality outcome was missing, essential laboratory measurements were unavailable, or patient information was inconsistent. Where a patient had more than one eligible admission, only the first admission was retained to preserve independence of observations. The final analytical sample contained 1,500 patient records.

The outcome was in-hospital mortality, coded 1 when a patient died during admission and 0 when the patient survived to discharge. Predictor variables were age in completed years, sex coded female = 0 and male = 1, parasite density measured as parasites per microlitre, haemoglobin concentration in grams per decilitre, and platelet count in $\times 10^9$/L. Definitions and coding are summarized in Table 1.

Table 1. *Operational definitions of study variables.*

Variable	Operational definition	Coding/unit
Mortality	Death during malaria admission	1 = death; 0 = survival
Age	Age at admission	Years
Sex	Biological sex in record	0 = female; 1 = male
Parasite density	Parasites detected in blood	Parasites/μL
Haemoglobin	Blood haemoglobin concentration	g/dL
Platelet count	Circulating platelet concentration	$\times 10^9$/L

Variables were extracted from de-identified electronic medical records.

2.3. Data Preparation and Management

Direct identifiers, including patient names, contact details, and medical record numbers, were removed before analysis. Records were assessed for completeness, duplicate entries, internal consistency, and clinically implausible values. Only one record per patient was retained. The resulting analytical dataset contained the outcome and five prespecified predictors. Statistical analysis was performed in Python.

2.4. Statistical Analysis

Continuous variables were summarized using means, standard deviations, medians, quartiles, minima, and maxima. Categorical variables were summarized using frequencies and percentages. Mortality prevalence was reported for the complete sample and selected descriptive subgroups.

The statistical modelling proceeded sequentially from the original five-predictor linear logistic regression model to diagnostic assessment, correction of identified nonlinear effects, model comparison, and internal validation. In-hospital mortality was coded as 1 for death and 0 for survival. The same five prespecified predictors were retained throughout the analysis: age, sex, parasite density, haemoglobin level, and platelet count. Sex was coded as 0 for female and 1 for male. Age was measured in completed years, haemoglobin level in g/dL, and platelet count in $\times 10^9$/L. Parasite density was divided by 1,000 for modelling and reporting so that its coefficient represented an increase of 1,000 parasites/μL.

The original analysis used a multivariable binary logistic regression model in which the continuous predictors were initially entered as linear terms [8, 9]. Let $p_i = P(Y_i = 1)$ denote the probability that patient (i) died during admission. The original five-predictor linear model was specified as

$$\eta_{1i} = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 A_i + \beta_2 S_i + \beta_3 D_i + \beta_4 H_i + \beta_5 P_i \quad (1)$$

where (A_i) denotes age in years, (S_i) denotes sex coded 1 for male and 0 for female, (D_i) denotes parasite density divided by 1,000, (H_i) denotes haemoglobin level in g/dL, and (P_i) denotes platelet count in $\times 10^9$/L. Predicted mortality probability was obtained using the inverse-logit transformation:

$$\hat{p}_i = \frac{1}{1+\exp(-\eta_i)} \quad (2)$$

For the original linear model, regression coefficients were exponentiated to obtain adjusted odds ratios:

$$\mathrm{AOR}_j = e^{\beta_j} \quad (3)$$

The corresponding 95% confidence interval for each adjusted odds ratio was calculated as

$$95\%\ CI(AOR_j) = \left[exp\{\beta_j - 1.96\ SE(\beta_j)\}, exp\{\beta_j + 1.96\ SE(\beta_j)\}\right] \quad (4)$$

where SE (β_j) denotes the standard error of the estimated coefficient.

Model assumptions were evaluated before final interpretation. The binary-outcome assumption was addressed by coding mortality as death or survival, and the independence of observations was ensured by retaining only one eligible admission per patient. Multicollinearity was assessed using variance inflation factors, with values below 5 considered evidence against serious multicollinearity. Linearity between each continuous predictor and the logit of mortality was assessed using empirical-logit plots and the Box-Tidwell test.

The diagnostic assessment identified departures from line-

arity for age, haemoglobin level, and platelet count, while parasite density remained compatible with a linear specification. The original five-predictor linear model was therefore retained as the prespecified comparator, while a corrected nonlinear logistic regression model using the same five clinical predictors was fitted to address the identified violations of functional form. Age, haemoglobin level, and platelet count were represented using restricted cubic splines with three knots [12]. Three knots were selected to achieve clinically relevant curvature while limiting model complexity, given the relatively small number of mortality events. Knots were positioned at the 10th, 50th, and 90th percentiles of each variable. In the analytical dataset, these locations corresponded to 9, 44, and 78 years for age; 6.0, 10.2, and 14.4 g/dL for haemoglobin level; and 68, 198.5, and 316.1 $\times 10^9$/L for platelet count. Parasite density remained a linear term and sex remained binary.

For a continuous predictor (x), with three ordered knots ($k_1 < k_2 < k_3$), the nonlinear restricted cubic spline basis was defined as

$$h(x) = \frac{(x-k_1)_+^3 - \frac{k_3-k_1}{k_3-k_2}(x-k_2)_+^3 + \frac{k_2-k_1}{k_3-k_2}(x-k_3)_+^3}{(k_3-k_1)^2} \quad (5)$$

where

$$(z)_+ = \max(z, 0)$$

The corrected nonlinear logistic regression model was specified as

$$\eta_{si} = \alpha_0 + \alpha_1 A_i + \alpha_2 h_a(A_i) + \alpha_3 S_i + \alpha_4 D_i + \alpha_5 H_i + \alpha_6 h_H(H_i) + \alpha_7 P_i + \alpha_8 h_p(P_i) \quad (6)$$

Where (h_a) (A_i), $(h_H\ (H_i))$, and $(h_p\ (P_i))$ denote the nonlinear restricted cubic spline basis terms for age, haemoglobin level, and platelet count, respectively. The corrected specification therefore retained the same five clinical predictors as the original model. However, it required eight predictor parameters because age, haemoglobin level, and platelet count each contributed one linear and one nonlinear spline term, while sex and parasite density each contributed one parameter. Individual spline-basis coefficients were not interpreted as constant clinical odds ratios because the effect of a nonlinear predictor depends on the combination of its linear and nonlinear terms. The nonlinear predictors were therefore interpreted using their complete fitted functions, clinically meaningful contrasts, and graphical displays [12].

The original linear model and the corrected restricted cubic spline model were formally compared using log-likelihood, Akaike information criterion (AIC), likelihood ratio testing, discrimination, and overall prediction error. The contribution of nonlinearity for each spline predictor was assessed by comparing the complete spline model with a reduced model in which the corresponding nonlinear basis term was omitted. Statistical significance was evaluated at the 5% level.

Because the spline specification increased the number of predictor parameters without changing the five clinical predictors, model complexity and potential overfitting were assessed explicitly. The number of mortality events relative to the number of predictor parameters was reported descriptively, but a fixed events-per-variable rule was not used as the sole criterion of model adequacy. Formal sample-size adequacy was assessed using prediction-model methods that consider the number of candidate predictor parameters, outcome prevalence, anticipated model fit, and the desired degree of shrinkage [13, 14]. The assessment used eight predictor parameters, the observed mortality prevalence, a conservative anticipated Cox-Snell (R^2) of 0.05, and a target global shrinkage factor of 0.90. The numerical result of this assessment is reported in the Results section.

Ridge-penalized logistic regression using the same five predictors and the same restricted cubic spline representation was additionally fitted as a sensitivity analysis for potential overfitting [16]. For this analysis, all columns of the model design matrix were standardized within the training data before fitting. L2 penalization was applied using candidate inverse-regularization values (C) ranging from (10^{-3}) to (10^3) across 25 logarithmically spaced values. The value of (C) was selected using stratified cross-validation with logarithmic loss as the optimization criterion. For nested validation of the ridge model, penalty selection was performed using five-fold stratified cross-validation within each outer training fold so that information from held-out patients did not influence model tuning.

Bootstrap internal validation with 1,000 resamples was used to estimate optimism conditional on the corrected restricted cubic spline specification [15]. Each bootstrap sample contained the same number of observations as the original analytical dataset and was generated by sampling patients with replacement. Within each bootstrap sample, spline-knot locations were recalculated using the 10th, 50th, and 90th percentile rule, and the corrected five-predictor spline model was refitted. Performance in each bootstrap sample was compared with performance when the same fitted model was applied to the original dataset. Mean optimism was subtracted from apparent full-sample performance to obtain optimism-corrected estimates. Because this bootstrap analysis evaluated the corrected spline specification as a fixed model structure, validation of the complete model-building process, including functional-form selection, was conducted separately using repeated stratified cross-validation.

Discrimination was assessed using the area under the receiver operating characteristic curve. Overall probabilistic prediction error was assessed using the Brier score:

$$Brier = \left(\frac{1}{n}\right)\sum_{i=1}^{n}(Y_i - \hat{p}_i)^2 \quad (7)$$

Bootstrap distributions were used to derive 95% intervals for the optimism-corrected AUC and Brier score. Model calibration was assessed using the calibration intercept, calibration slope, and a smooth calibration curve. The calibration model was expressed as

$$\log\left[\frac{P(Y_i=1)}{1-P(Y_i=1)}\right] = a + b\log\left[\frac{\hat{p}_i}{1-\hat{p}_i}\right] \quad (8)$$

where (a) denotes the calibration intercept and (b) denotes the calibration slope. Ideal calibration corresponds to (a=0) and (b=1). Calibration-in-the-large was estimated by fixing the slope at 1 and estimating the intercept, while the calibration slope was estimated from the model in Equation (8). The Hosmer-Lemeshow test was retained only as a supplementary assessment of gross lack of fit and was not interpreted alone as evidence of good calibration.

Repeated stratified 10-fold cross-validation with 20 repetitions was used to validate the complete modelling process. Within each training fold, the Box-Tidwell functional-form assessment was repeated for all continuous predictors. A continuous predictor showing evidence of nonlinearity at $p < 0.05$ was represented using a three-knot restricted cubic spline, with knot locations recalculated from the 10th, 50th, and 90th percentiles of that training sample. Predictors satisfying the linearity assumption were retained as linear terms. The resulting model was fitted using the training observations only and then applied to the corresponding held-out observations. Youden threshold selection was also repeated using training data only before evaluation in the held-out fold. This procedure ensured that functional-form selection, spline-knot estimation, model fitting, and threshold selection were all performed without information from the validation observations. Cross-validated AUC, Brier score, calibration intercept, calibration slope, and threshold-dependent classification measures were calculated from held-out predictions.

Classification analyses were evaluated separately from threshold-independent model-performance measures. The original analysis considered the conventional probability threshold of 0.5 and a threshold selected using Youden's index. Youden's index for threshold (t) was defined as

$$J(t) = \text{Sensitivity}(t) + \text{Specificity}(t) - 1 \quad (9)$$

Because selecting and evaluating a threshold using the same development dataset may produce optimistic performance estimates, threshold selection was repeated within the training data of each cross-validation fold. The threshold that maximized (J(t)) in the training data was then applied to the corresponding held-out observations. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and confusion-matrix counts were reported. Ninety-five percent confidence intervals for sensitivity, specificity, PPV, and NPV were calculated using Wilson score intervals.

No externally established mortality-risk threshold was available for this hospital setting. A predicted mortality threshold of 5% was therefore examined only as an illustrative screening operating point to quantify the trade-off between identifying patients at increased mortality risk and the number of patients who would require additional clinical review. The 5% threshold was not treated as an established or recommended clinical cut-off.

Decision-curve analysis was used to evaluate the potential clinical utility of model-guided risk classification across candidate threshold probabilities from 1% to 20% [17]. Net benefit at threshold probability (t) was calculated as

$$\text{NB}(t) = \frac{TP(t)}{n} - \frac{FP(t)}{n}\left(\frac{t}{1-t}\right) \quad (10)$$

where (TP(t)) and (FP(t)) denote the numbers of true-positive and false-positive classifications at threshold (t), respectively. The model was compared with the default strategies of classifying all patients as high risk and classifying no patients as high risk.

All statistical analyses were performed using Python 3.13.5. The principal packages were NumPy 2.3.5, pandas 2.2.3, SciPy 1.17.0, statsmodels 0.14.6, scikit-learn 1.8.0, and Matplotlib 3.10.8. The original linear model and the unpenalized restricted cubic spline model were fitted by maximum likelihood using statsmodels. Logit with the Newton-Raphson algorithm. Complete-data models were allowed a maximum of 500 iterations, while unpenalized models fitted within bootstrap and cross-validation procedures were allowed a maximum of 200 iterations. No centering or standardization was applied to the original linear model or the unpenalized spline model apart from expressing parasite density per 1,000 parasites/μL. The ridge sensitivity analysis used StandardScaler and LogisticRegression with L2 penalization, the lbfgs solver, and a maximum of 5,000 iterations. Nested ridge validation used 10 outer stratified folds and five inner stratified folds for penalty selection. A random seed of 20260818 was used for reproducible resampling procedures. The reproducible Python analysis code and a data dictionary were prepared as supplementary materials.

2.5. Ethical Considerations

Approval and institutional authorization were obtained before access to hospital records. The analysis used secondary, de-identified data and involved no direct patient contact. Data were stored in password-protected files accessible only to the research team.

3. Results

3.1. Patient Characteristics and Mortality

The analysis included 1,500 malaria admissions. Fifty-four patients died in hospital, producing an event rate of 3.60%,

while 1,446 patients survived to discharge. The marked imbalance between deaths and survivals was relevant to interpretation of threshold-based classification performance.

The mean age was 43.24 years (standard deviation 24.94; range 0-85 years). Mean haemoglobin concentration was 10.19 g/dL, mean parasite density was 105,357.73 parasites/µL, and mean platelet count was 194.36 $\times 10^9$/L. Females represented 50.73% of the sample. Patients who died had a mean haemoglobin level of 7.42 g/dL compared with 10.30 g/dL among survivors. Mortality was 5.16% among patients classified as anaemic using the 11 g/dL threshold and 1.43% among non-anaemic patients. Selected descriptive characteristics are presented in Table 2.

Table 2. *Selected demographic, laboratory, and outcome characteristics.*

Characteristic	Value	Additional detail
Sample size	1,500	54 deaths; 3.60%
Age, years	43.24 ± 24.94	Median 44; range 0-85
Haemoglobin, g/dL	10.19 ± 3.02	Median 10.20; range 5.0-15.5
Parasite density, /µL	105,357.73 ± 57,357.08	Median 107,975
Platelet count, ×109/L	194.36 ± 89.78	Median 198.50; range 40-350
Female sex	761 (50.73%)	30 deaths; 3.94%
Male sex	739 (49.27%)	24 deaths; 3.25%
Anaemic (<11 g/dL)	872 (58.13%)	45 deaths; 5.16%
Non-anaemic (≥11 g/dL)	628 (41.87%)	9 deaths; 1.43%

Values are mean ± standard deviation or n (%), unless otherwise stated.

3.2. Model Assumptions and Overall Fit

The outcome was strictly binary, and no patient contributed more than one observation. Variance inflation factors ranged from 1.001 to 1.003, indicating negligible multicollinearity among the five prespecified predictors. The original linear logistic regression model contained five predictor parameters and 54 mortality events, corresponding to 10.8 events per predictor.

The Box-Tidwell assessment showed that parasite density was compatible with a linear relationship with the logit of mortality ($p = 0.519$), whereas age, haemoglobin level, and platelet count showed significant departures from linearity (all $p < 0.001$). These findings indicated that the original linear specification did not adequately represent the functional relationships of three continuous predictors. The original five-predictor linear model was therefore retained as the prespecified comparator, while age, haemoglobin level, and platelet count were represented using restricted cubic splines in the corrected model. Sex remained binary and parasite density remained linear. Importantly, the corrected model retained the same five clinical predictors but required eight predictor parameters because each of the three spline-modelled variables contributed one linear and one nonlinear term.

For the original prespecified linear model, the log-likelihood was -157.739 compared with -232.52 for the intercept-only model, and McFadden's pseudo-R2 was 0.3216. The Akaike information criterion was 327.479. The model was statistically significant compared with the intercept-only model (likelihood-ratio $p < 0.001$).

After correcting the identified nonlinear effects, the overall model fit improved substantially. The restricted cubic spline model had a log-likelihood of -131.410, an Akaike information criterion of 280.820, and McFadden's pseudo-R2 of 0.435. A nested likelihood-ratio comparison between the original linear specification and the corrected nonlinear specification showed that inclusion of the three nonlinear components significantly improved model fit, $\chi 2(3)=52.659$, $p<0.001$.

Because the corrected model contained eight predictor parameters, the 54 mortality events corresponded to 6.75 events per predictor parameter. The formal sample-size assessment specified in Section 2.4 indicated that approximately 1,400 participants and 51 mortality events would be required under the conservative assumptions of an outcome prevalence of 0.036, anticipated Cox-Snell R2 of 0.05, and target shrinkage factor of 0.90. The observed sample of 1,500 patients and 54 deaths therefore narrowly exceeded this estimated requirement, although the relatively small number of outcome events continued to justify cautious interpretation and additional penalized sensitivity analysis.

For the corrected restricted cubic spline model, the Hosmer-Lemeshow statistic was 6.524 with 8 degrees of freedom ($p=0.589$). This result indicated that no gross lack of fit was detected by the grouped goodness-of-fit test. It was not interpreted as evidence of good calibration by itself. Calibration intercept, calibration slope, Brier score, smooth calibration assessment, and resampling-based internal validation are reported separately in the model-validation results.

3.3. Predictors of In-Hospital Mortality

Under the original prespecified five-predictor linear logistic regression model, statistically significant coefficients were estimated for age, parasite density, haemoglobin level, and platelet count, while sex was not statistically significant. Because the diagnostic assessment reported in Section 3.2 identified significant departures from linearity for age, haemoglobin level, and platelet count, the adjusted odds ratios for these

three predictors are reported here as estimates from the original linear specification and should not be interpreted as constant effects across their complete observed ranges. Their corrected nonlinear relationships are presented in Section 3.4.

Under the original linear specification, each additional year of age was associated with an adjusted odds ratio of 1.036 (95% CI 1.022–1.051). Parasite density satisfied the linearity-in-the-logit assumption and remained positively associated with mortality. Each increase of 1,000 parasites/μL was associated with an adjusted odds ratio of 1.019 (95% CI 1.012–1.025).

The original linear model estimated an adjusted odds ratio of 0.652 (95% CI 0.569–0.748) for each 1 g/dL increase in haemoglobin level and 0.990 (95% CI 0.986–0.994) for each increase of 1 $\times 10^9$/L in platelet count. These haemoglobin and platelet estimates summarize the original linear specification only because subsequent diagnostic assessment demonstrated that their relationships with mortality were nonlinear. Male sex was not associated with a statistically detectable difference in mortality after adjustment for the remaining predictors (AOR = 0.930, 95% CI 0.505–1.714; p = 0.816). Complete coefficient estimates for the original prespecified model are presented in Table 3 and Figure 1.

Table 3. *Complete estimates from the original prespecified linear logistic regression model for in-hospital mortality.*

Model term	β	SE	95% CI for β	AOR (95% CI)	p value
Intercept	-2.1454	0.8622	-3.8353 to -0.4556	—	0.013
Age, per year	0.0354	0.0072	0.0213 to 0.0494	1.036 (1.022–1.051)	<0.001
Male vs female	-0.0724	0.3118	-0.6835 to 0.5387	0.930 (0.505–1.714)	0.816
Parasite density, per 1,000/μL	0.0185	0.0034	0.0120 to 0.0251	1.019 (1.012–1.025)	<0.001
Haemoglobin, per g/dL	-0.4271	0.0699	-0.5640 to -0.2901	0.652 (0.569–0.748)	<0.001
Platelet count, per $\times 10^9$/L	-0.0102	0.0020	-0.0142 to -0.0062	0.990 (0.986–0.994)	<0.001

AOR = adjusted odds ratio; CI = confidence interval; SE = standard error. Mortality was coded 1 for death and 0 for survival. Sex was coded 1 for male and 0 for female. Parasite density was divided by 1,000 before model fitting so that the coefficient and AOR represent an increase of 1,000 parasites/μL.

Substitution of the estimated coefficients into the prespecified linear model produced the following fitted linear predictor:

$$\hat{\eta}_L = -2.145425 + 0.035363A - 0.072404S + 0.018548D - 0.427057H - 0.010170P \quad (11)$$

where A denotes age in years, S=1 for male and S=0 for female, D denotes parasite density divided by 1,000, H denotes haemoglobin level in g/dL, and P denotes platelet count in $\times 10^9$/L.

Accordingly, an individual patient's predicted probability of in-hospital mortality under the original linear model was calculated as

$$\hat{p}_L = \frac{1}{1+\exp[-(-2.145425+0.035363A-0.072404S+0.018548D-0.427057H-0.010170P)]} \quad (12)$$

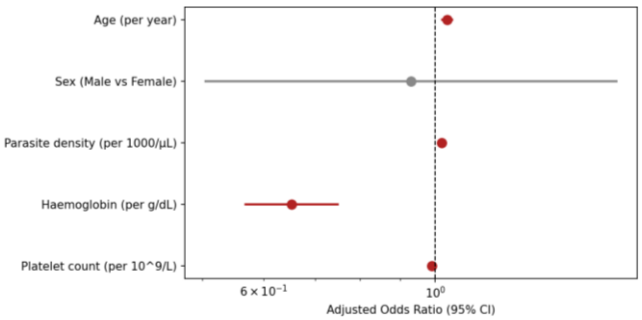

Figure 1. *Adjusted odds ratios and 95% confidence intervals from the original prespecified linear logistic regression model for in-hospital malaria mortality.*

The age, haemoglobin, and platelet-count estimates in Figure 1 represent the original linear specification; their corrected nonlinear relationships are presented in Section 3.4.

3.4. Nonlinear Model Refinement

To address the significant departures from linearity identified in Section 3.2, the same five prespecified clinical predictors were refitted using a restricted cubic spline logistic regression specification. Age, haemoglobin level, and platelet count were represented using three-knot restricted cubic splines, while parasite density remained a linear term and sex remained binary. The knot locations were 9, 44, and 78 years for age; 6.0, 10.2, and 14.4 g/dL for haemoglobin level; and

68, 198.5, and 316.1 ×109/L for platelet count. The corrected model therefore retained the original five clinical predictors but represented them using eight predictor parameters. The complete fitted coefficients are presented in Table 4

Table 4. *Complete restricted cubic spline logistic regression model for in-hospital mortality.*

Model term	β	SE	95% CI for β	p value
Intercept	4.3998	1.3583	1.7376 to 7.0621	0.001
Age, linear spline term	-0.0282	0.0216	-0.0705 to 0.0140	0.190
Age, nonlinear spline term	0.0794	0.0243	0.0318 to 0.1270	0.001
Male vs female	0.0730	0.3410	-0.5953 to 0.7414	0.830
Parasite density, per 1,000/μL	0.0215	0.0038	0.0141 to 0.0289	<0.001
Haemoglobin, linear spline term	-1.0016	0.1590	-1.3131 to -0.6901	<0.001
Haemoglobin, nonlinear spline term	0.8775	0.2146	0.4569 to 1.2982	<0.001
Platelet count, linear spline term	-0.0334	0.0053	-0.0439 to -0.0229	<0.001
Platelet count, nonlinear spline term	0.0345	0.0069	0.0211 to 0.0480	<0.001

CI = confidence interval; SE = standard error. Mortality was coded 1 for death and 0 for survival. Sex was coded 1 for male and 0 for female. Parasite density was divided by 1,000 before model fitting. Age, haemoglobin level, and platelet count were each represented by one linear and one nonlinear restricted cubic spline basis term. The spline-basis coefficients should not be interpreted separately as constant clinical odds ratios

Substitution of the estimated coefficients into the corrected restricted cubic spline model produced the following fitted linear predictor:

$$\hat{\eta}_s = 4.399833 - 0.028226\mathrm{A} + 0.079426\mathrm{h}_A(\mathrm{A}) + 0.073034\mathrm{S} + 0.021525\mathrm{D} - 1.001597\mathrm{H} + 0.877539\mathrm{h}_H(\mathrm{H}) - 0.033396\mathrm{P} + 0.034536\mathrm{h}_P(\mathrm{P}), \quad (13)$$

where A denotes age in years, S=1 for male and S=0 for female, D denotes parasite density divided by 1,000, H denotes haemoglobin level in g/dL, and P denotes platelet count in $\times 10^9$/L. The functions $\mathrm{h}_A(\mathrm{A})$, $\mathrm{h}_H(\mathrm{H})$, and $\mathrm{h}_P(\mathrm{P})$are the nonlinear restricted cubic spline basis functions defined in Equation (5), using the knot locations reported in Section 2.4.

Individual predicted mortality probability under the corrected model was therefore calculated as

$$\hat{p}_s = \frac{1}{1+\exp(-\hat{\eta}_s)} \quad (14)$$

For the two predictors that remained in conventional linear or binary form, the corrected model produced directly interpretable adjusted odds ratios. Parasite density remained positively associated with mortality, with an AOR of 1.022 per increase of 1,000 parasites/μL (95% CI 1.014–1.029; p < 0.001). Male sex remained non-significant after adjustment for the other predictors (AOR = 1.076, 95% CI 0.551–2.099; p = 0.830).

For age, haemoglobin level, and platelet count, a single constant odds ratio was not appropriate because their associations with mortality were nonlinear. Adjusted odds-ratio curves were therefore calculated from the complete fitted spline functions using the median value of each predictor as the reference. The resulting relationships are shown in Figures 2-4.

3.4.1. Age Interpretation

Age showed a nonlinear association with mortality. Using 44 years as the reference value, the adjusted odds ratio was 1.29 at age 50 years (95% CI 1.19–1.41), 2.48 at age 60 years (95% CI 1.87–3.29), 5.71 at age 70 years (95% CI 3.22–10.12), and 14.30 at age 80 years (95% CI 5.74–35.63). The fitted relationship remained relatively low through much of early and middle adulthood and then increased progressively at older ages. Uncertainty increased toward the extremes of the age distribution, as indicated by the wider confidence interval.

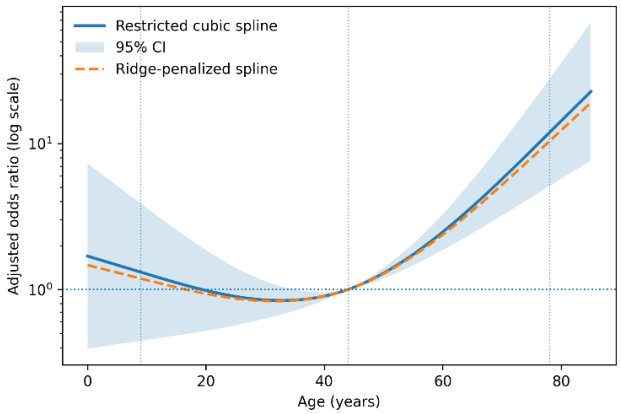

Figure 2. *Adjusted nonlinear relationship between age and in-hospital malaria mortality. Odds ratios are expressed relative to age 44 years. The solid curve represents the restricted cubic spline estimate, the shaded region represents the 95% confidence interval, and the dashed curve represents the ridge-penalized sensitivity estimate. Vertical dotted lines indicate spline knots at 9, 44, and 78 years. The odds-ratio axis is displayed on a logarithmic scale.*

3.4.2. Haemoglobin Interpretation

Haemoglobin level demonstrated a pronounced nonlinear inverse relationship with mortality at lower concentrations. Using 10.2 g/dL as the reference value, the adjusted odds ratio was 9.94 at 7 g/dL (95% CI 5.35–18.46), 3.98 at 8 g/dL (95% CI 2.78–5.69), and 1.85 at 9 g/dL (95% CI 1.58–2.17). At 5 g/dL, the estimated odds ratio increased to 72.74 (95% CI 21.35–247.79), although the wide confidence interval indicates substantially greater uncertainty at this extreme. Mortality odds declined sharply as haemoglobin approached approximately 10–11 g/dL, after which the fitted relationship became considerably flatter. Estimates at the upper extreme were less precise.

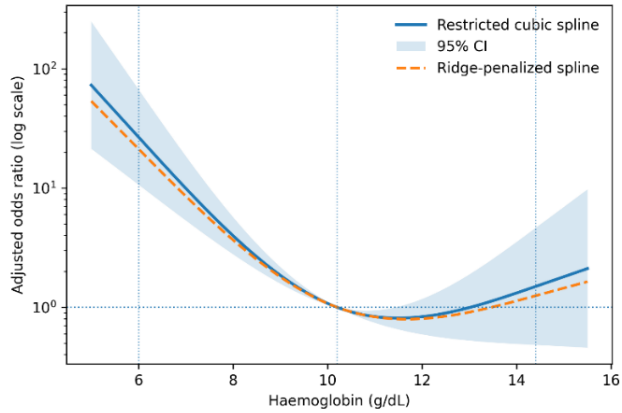

Figure 3. *Adjusted nonlinear relationship between haemoglobin level and in-hospital malaria mortality. Odds ratios are expressed relative to 10.2 g/dL. The solid curve represents the restricted cubic spline estimate, the shaded region represents the 95% confidence interval, and the dashed curve represents the ridge-penalized sensitivity estimate. Vertical dotted lines indicate spline knots at 6.0, 10.2, and 14.4 g/dL. The odds-ratio axis is displayed on a logarithmic scale.*

3.4.3. Platelet Interpretation

Platelet count also showed a marked nonlinear relationship with mortality. Using 198.5 $\times 10^9$/L as the reference value, the adjusted odds ratio was 17.77 at 75 $\times 10^9$/L (95% CI 7.62–41.47), 7.85 at 100 $\times 10^9$/L (95% CI 4.32–14.28), and 1.98 at 150 $\times 10^9$/L (95% CI 1.62–2.42). Estimated mortality odds were lowest around approximately 200 $\times 10^9$/L. At higher platelet counts, the fitted curve increased again; for example, the adjusted odds ratio at 300 $\times 10^9$/L was 3.10 (95% CI 1.36–7.08) relative to 198.5 $\times 10^9$/L. Confidence intervals widened toward both extremes of the platelet distribution, indicating greater uncertainty where observations were less concentrated.

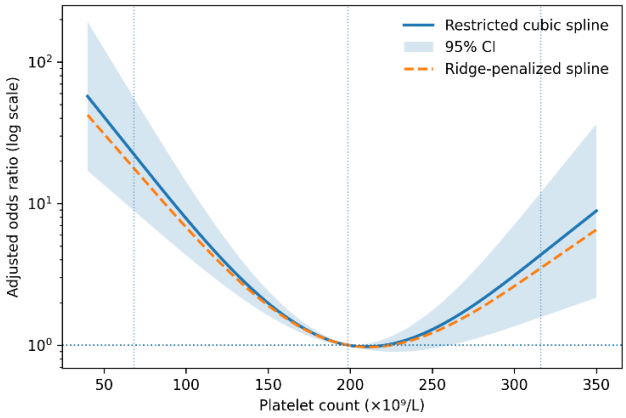

Figure 4. *Adjusted nonlinear relationship between platelet count and in-hospital malaria mortality. Odds ratios are expressed relative to 198.5 $\times 10^9$/L. The solid curve represents the restricted cubic spline estimate, the shaded region represents the 95% confidence interval, and the dashed curve represents the ridge-penalized sensitivity estimate. Vertical dotted lines indicate spline knots at 68, 198.5, and 316.1 $\times 10^9$/L. The odds-ratio axis is displayed on a logarithmic scale.*

The nonlinear refinement showed that the constant per-unit effects reported for age, haemoglobin level, and platelet count under the original linear specification did not adequately describe their relationships with mortality across the complete observed ranges. The corrected spline specification retained the direction of the major clinical associations while revealing substantial variation in effect magnitude across predictor values. Formal comparison of the original linear model and the corrected nonlinear model is presented in Section 3.5.

3.5. Comparison of Linear and Nonlinear Models

The original prespecified linear logistic regression model and the corrected restricted cubic spline model were compared to determine whether accounting for the nonlinear relationships identified during diagnostic assessment improved model fit and predictive performance. Both specifications retained the same five clinical predictors: age, sex, parasite density,

haemoglobin level, and platelet count. The original model used five predictor parameters, whereas the corrected model used eight predictor parameters because age, haemoglobin level, and platelet count each contributed one linear and one nonlinear spline term.

The corrected nonlinear specification provided substantially better overall fit than the original linear specification. The log-likelihood improved from -157.739 in the original model to -131.410 in the restricted cubic spline model. The Akaike information criterion decreased from 327.479 to 280.820, a reduction of 46.659 points, indicating considerably better fit after accounting for nonlinearity. McFadden's pseudo-R2 also increased from 0.322 to 0.435.

A formal nested likelihood-ratio test confirmed that addition of the three nonlinear spline components significantly improved model fit compared with the original linear model, $\chi^2(3)=52.659$, $p<0.001$. Separate tests of the nonlinear components showed significant evidence of nonlinearity for age, $\chi^2(1)=10.884$, $p=0.001$; haemoglobin level, $\chi^2(1)=16.544$, $p<0.001$; and platelet count, $\chi^2(1)=27.614$, $p<0.001$. These findings confirmed that the improvement in the corrected model was attributable to allowing the relationships of these three predictors with mortality to depart from a constant linear form.

Apparent discrimination also improved after correction of the functional form. The area under the receiver operating characteristic curve increased from 0.922 for the original linear model to 0.962 for the restricted cubic spline model. The apparent Brier score decreased from 0.0296 to 0.0259, indicating lower overall prediction error. These apparent performance estimates demonstrate improvement in the development sample but do not by themselves establish internal validity.

Table 5. *Comparison of the original prespecified linear logistic regression model and the corrected restricted cubic spline model.*

Model characteristic or performance measure	Original linear model	Corrected RCS model
Clinical predictors	5	5
Predictor parameters	5	8
Log-likelihood	-157.739	-131.410
AIC	327.479	280.820
McFadden pseudo-R2	0.322	0.435
Apparent AUC	0.922	0.962
Apparent Brier score	0.0296	0.0259

AIC = Akaike information criterion; AUC = area under the receiver operating characteristic curve; RCS = restricted cubic spline. Lower AIC and Brier score values indicate better performance, whereas higher log-likelihood, pseudo-R2, and AUC values indicate improvement. Both models contained the same five clinical predictors.

The model-comparison results showed that correcting the nonlinear relationships of age, haemoglobin level, and platelet count significantly improved model fit, apparent discrimination, and overall prediction error without introducing additional clinical predictors. The corrected restricted cubic spline model was therefore carried forward for the strengthened internal-validation and calibration analyses.

3.6. Discrimination, Classification, and Internal Validation

Following correction of the nonlinear functional relationships, the restricted cubic spline model was subjected to strengthened internal validation. Bootstrap validation with 1,000 resamples produced an optimism-corrected area under the receiver operating characteristic curve (AUC) of 0.955 (95% CI 0.940–0.970), compared with the apparent AUC of 0.962 reported in Section 3.5. The corresponding optimism-corrected Brier score was 0.0275 (95% CI 0.0208–0.0336). The bootstrap calibration intercept was 0.008 (95% CI -0.324 to 0.390), and the calibration slope was 0.927 (95% CI 0.722–1.153). The calibration intercept was close to the ideal value of zero, while the slope below one indicated a small degree of remaining model optimism, with uncertainty reflecting the limited number of mortality events.

Repeated stratified 10-fold cross-validation with 20 repetitions, in which functional-form assessment, spline-knot estimation, model fitting, and threshold selection where internal validation repeated within the training data, produced a cross-validated AUC of 0.954 and a Brier score of 0.0273. Calibration-in-the-large was 0.002, and the calibration slope was 0.915. Across the 200 training folds, age, haemoglobin level, and platelet count were selected for nonlinear modelling in 100% of folds, whereas parasite density was selected as nonlinear in only 0.5% of folds. These findings demonstrate that

the nonlinear specification identified in the complete dataset was highly stable during resampling and that the improved predictive performance was not dependent on functional-form decisions made using the full dataset.

The smooth calibration curve based on repeated cross-validated predictions is presented in Figure 5. Agreement between predicted and observed mortality was generally close across the lower and intermediate predicted-risk range, where most observations occurred. Greater divergence was observed toward the upper end of the predicted-risk distribution, where observations and mortality events were less numerous. This pattern is consistent with the calibration slope below one and indicates greater uncertainty for patients assigned very high predicted risks.

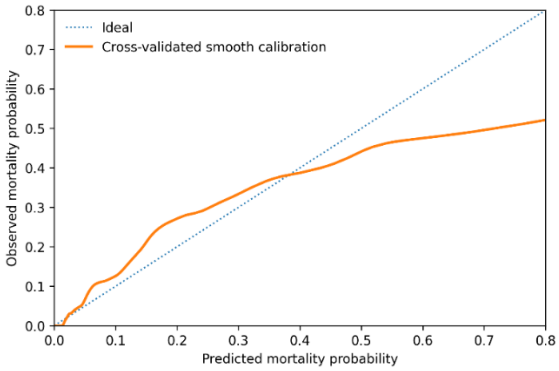

Figure 5. *Smooth calibration curve based on repeated cross-validated predictions from the complete model-building process, with functional-form assessment repeated within each training fold.*

Table 6. *Internal validation of the corrected restricted cubic spline mortality model.*

Validation approach	AUC	Brier score	Calibration intercept	Calibration slope
Bootstrap optimism-corrected RCS model	0.955 (0.940–0.970)	0.0275 (0.0208–0.0336)	0.008 (-0.324 to 0.390)	0.927 (0.722–1.153)
Repeated 20 × 10-fold CV of complete modelling process	0.954	0.0273	0.002	0.915
Nested 10-fold ridge RCS sensitivity model	0.952	0.0275	0.002	0.946

AUC = area under the receiver operating characteristic curve; CI = confidence interval; CV = cross-validation; RCS = restricted cubic spline. Values in parentheses are 95% bootstrap intervals where available. An ideal calibration intercept is 0 and an ideal calibration slope is 1.

A ridge-penalized version of the same five-predictor restricted cubic spline model was additionally evaluated as a sensitivity analysis for potential overfitting. Nested 10-fold cross-validation, with five-fold internal tuning of the penalty parameter, produced an AUC of 0.952, a Brier score of 0.0275, a calibration intercept of 0.002, and a calibration slope of 0.946. The median selected inverse-regularization parameter was C=7.81. The ridge model therefore produced discrimination and overall prediction error similar to the unpenalized spline model, while its calibration slope was closer to the ideal value of one. This finding supported the stability of the corrected model while also indicating a modest benefit from shrinkage for calibration.

The receiver operating characteristic curves based on internally validated predictions are shown in Figure 6. The original linear model achieved a cross-validated AUC of approximately 0.911, while the corrected restricted cubic spline model achieved an AUC of 0.954. The nested ridge-penalized spline sensitivity model achieved an AUC of 0.952. The close performance of the unpenalized and ridge-penalized spline models indicates that the improved discrimination after correction of nonlinearity was not dependent on an unpenalized full-sample fit.

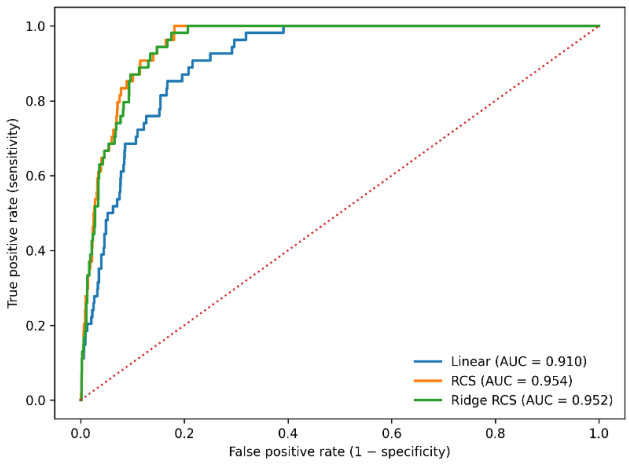

Figure 6. *Receiver operating characteristic curves comparing internally validated predictions from the original linear model, the corrected nonlinear model obtained through repeated functional-form assessment, and the ridge-penalized restricted cubic spline sensitivity model.*

Classification performance was considered separately from

threshold-independent measures of model quality. Under the original prespecified linear model, the conventional probability threshold of 0.5 classified only 6 of 54 deaths as positive. This produced sensitivity of 11.11%, specificity of 99.72%, positive predictive value of 60.00%, negative predictive value of 96.78%, and overall accuracy of 96.53%. The high accuracy therefore largely reflected correct classification of the much larger survival group and did not indicate adequate detection of mortality.

In the original development-sample analysis, Youden's index selected a threshold of 0.0363, reported previously as 0.036. At this threshold, 49 of the 54 deaths were identified, increasing sensitivity to 90.74%, while specificity decreased to 80.98%. Positive predictive value was 15.12%, negative predictive value was 99.57%, and accuracy was 81.33%. These estimates represent apparent classification performance because the threshold was selected and evaluated using the same development dataset. They should therefore not be interpreted as internally validated operating characteristics or as evidence that 0.036 is an established clinical threshold.

Table 7. *Development-sample classification performance of the original prespecified linear model.*

Metric	Threshold 0.50	Development Youden threshold 0.0363
True negatives	1,442	1,171
False positives	4	275
False negatives	48	5
True positives	6	49
Sensitivity	11.11% (5.19–22.19)	90.74% (80.09–95.98)
Specificity	99.72% (99.29–99.89)	80.98% (78.88–82.92)
Positive predictive value	60.00% (31.27–83.18)	15.12% (11.63–19.43)
Negative predictive value	96.78% (95.75–97.56)	99.57% (99.01–99.82)
Accuracy	96.53%	81.33%

Values in parentheses are 95% Wilson confidence intervals. The Youden threshold was selected and evaluated in the original development sample and therefore represents apparent rather than internally validated classification performance

To address optimism arising from development-sample threshold selection, Youden's index was recalculated separately within the training data of every fold during repeated stratified cross-validation and was then applied only to the corresponding held-out observations. Across the 200 training folds, the median selected threshold was 0.0343, with an interquartile range of 0.0313–0.0387. Mean held-out sensitivity was 90.97%, mean specificity was 85.26%, mean positive predictive value was 19.17%, mean negative predictive value was 99.62%, and mean accuracy was 85.47%. The variation in selected thresholds across training samples confirms that the development-data threshold of approximately 0.036 should not be interpreted as a fixed clinical cut-off.

The previously reported mean 10-fold cross-validation misclassification error of 0.036 was not retained as evidence of model validity. In the original analysis, that value was approximately equal to the observed mortality prevalence of 3.6%, making it possible for a model that classified nearly all patients as survivors to obtain an apparently low error. In the revised repeated cross-validation, use of the explicit 0.5 threshold resulted in a misclassification rate of approximately 0.0353, confirming that low overall error at this threshold primarily reflected the highly imbalanced outcome rather than adequate mortality detection. Discrimination, calibration, Brier score, and sensitivity were therefore given greater emphasis in the revised validation

Because no externally established mortality-risk threshold was available for Nakuru Level 6 Hospital, a 5% predicted-risk threshold was examined only as an illustrative operating point. It was selected to demonstrate the relationship between mortality detection and the number of patients who would require additional clinical review, not as a recommended clinical cut-off. Performance at 5% was compared with the conventional 50% threshold using averaged held-out predictions from repeated cross-validation.

At the illustrative 5% threshold, 47 of 54 deaths were classified as high risk, while 165 of the 1,446 survivors were also classified as high risk. Sensitivity was 87.04% (95% CI 75.58–93.58%), specificity was 88.59% (95% CI 86.85–90.13%), positive predictive value was 22.17% (95% CI 17.10–28.23%), and negative predictive value was 99.46% (95% CI 98.88–99.74%). Overall accuracy was 88.53%. A total of 212 of 1,500 admissions, or 14.13%, would have been

flagged for additional review at this illustrative threshold.

In contrast, the 50% threshold classified only 11 of the 54 deaths as high risk. Cross-validated sensitivity was 20.37% (95% CI 11.77–32.90%), specificity was 99.31% (95% CI 98.73–99.62%), positive predictive value was 52.38% (95% CI 32.37–71.66%), and negative predictive value was 97.09% (95% CI 96.11–97.83%). Although accuracy remained high at 96.47%, 43 of the 54 deaths were classified as low risk.

Table 8. *Cross-validated confusion counts and classification performance at illustrative 5% and conventional 50% mortality-risk thresholds.*

Metric	Threshold 0.05	Threshold 0.50
True negatives	1,281	1,436
False positives	165	10
False negatives	7	43
True positives	47	11
Sensitivity	87.04% (75.58–93.58)	20.37% (11.77–32.90)
Specificity	88.59% (86.85–90.13)	99.31% (98.73–99.62)
Positive predictive value	22.17% (17.10–28.23)	52.38% (32.37–71.66)
Negative predictive value	99.46% (98.88–99.74)	97.09% (96.11–97.83)
Accuracy	88.53%	96.47%

Values in parentheses are 95% Wilson confidence intervals. Classification was based on averaged held-out predictions from repeated stratified cross-validation. The 5% threshold is presented as an illustrative operating point and is not proposed as an established clinical cut-off.

Decision-curve analysis was used to examine whether model-guided risk classification provided greater net benefit than default strategies across a range of candidate threshold probabilities. The corrected restricted cubic spline model showed positive net benefit and exceeded both the classify-all and classify-none strategies across the evaluated threshold range from approximately 1% to 20%. At a 5% threshold, the model's net benefit was 0.0255, compared with -0.0147 for classifying all patients as high risk and 0 for classifying no patients as high risk. The decision curve therefore supports potential usefulness of model-guided risk stratification across a range of low mortality-risk thresholds, but it does not establish a single optimal clinical cut-off. Selection of an operational threshold would require prospective consideration of hospital monitoring capacity, the clinical consequences of missed high-risk patients, and the acceptable burden of false-positive alerts.

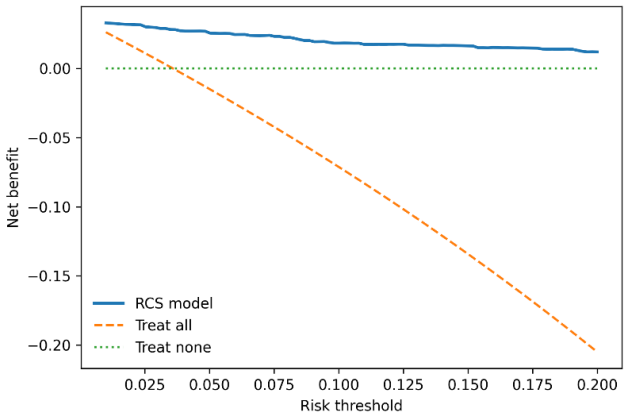

Figure 7. *Decision-curve analysis of the corrected restricted cubic spline mortality model across candidate risk thresholds from 1% to 20%. Net benefit is compared with strategies of classifying all patients as high risk and classifying no patients as high risk.*

Bootstrap validation, repeated cross-validation, calibration assessment, and penalized sensitivity analysis supported the improved performance of the corrected nonlinear model. The results also confirmed that threshold-dependent accuracy alone is unsuitable for evaluating mortality prediction in this highly imbalanced dataset. Although lower thresholds substantially improved mortality detection, no single threshold was identified as a definitive clinical operating point. External validation and prospective evaluation of clinical consequences and available monitoring capacity remain necessary before threshold-based implementation.

4. Discussion

4.1. Principal Findings

This study evaluated in-hospital malaria mortality using five routinely available clinical variables: age, sex, parasite density, haemoglobin level, and platelet count. The original prespecified linear logistic regression model identified statistically significant associations for age, parasite density, haemoglobin level, and platelet count, while sex was not independently associated with mortality. The original model demonstrated strong apparent discrimination, with an AUC of 0.922. Diagnostic assessment, however, showed that the assumption of linearity in the logit was not appropriate for age, haemoglobin level, and platelet count. The original linear model was therefore retained as the prespecified comparator, while these three predictors were re-expressed using restricted cubic splines in a corrected model containing the same five clinical predictors. Flexible modelling of continuous predictors is appropriate when their relationships with an outcome cannot be adequately represented by a single constant linear effect [12].

The nonlinear refinement materially improved the statisti-

cal representation of mortality risk. Compared with the original linear model, the corrected restricted cubic spline model improved log-likelihood, reduced AIC, increased McFadden's pseudo-R2, and increased apparent AUC from 0.922 to 0.962. The likelihood-ratio comparison confirmed that the additional nonlinear components significantly improved model fit. More importantly, the refinement altered the interpretation of age, haemoglobin level, and platelet count. Their effects were not constant across their observed ranges. Mortality risk increased more sharply at older ages, the association with haemoglobin was strongest at lower concentrations and became flatter as haemoglobin increased, and platelet count showed a marked nonlinear relationship with particularly high estimated risk at low values. Parasite density remained appropriately represented as a linear predictor, while sex remained non-significant.

The strengthened internal-validation analyses indicated that the improvement was not limited to the complete development dataset. Bootstrap validation produced an optimism-corrected AUC of 0.955, while repeated stratified cross-validation produced an AUC of 0.954 and a Brier score of 0.0273. Calibration intercepts were close to zero, although calibration slopes below one indicated some remaining optimism. The ridge-penalized sensitivity analysis produced similar discrimination and prediction error and a calibration slope closer to one, supporting the stability of the main findings while also showing the value of shrinkage when model complexity increases relative to the number of outcome events [15, 16]. Importantly, repeated cross-validation reproduced the functional-form assessment within every training fold; age, haemoglobin level, and platelet count were selected as nonlinear in all 200 training folds, providing additional evidence that the corrected specification was not an artefact of the complete dataset.

The classification findings also reinforce the distinction between discrimination and threshold-dependent clinical classification. In the original analysis, the conventional 0.5 threshold produced high overall accuracy but detected only 11.1% of deaths. The original development-data Youden threshold of approximately 0.036 substantially increased sensitivity, but resampling demonstrated that the selected threshold varied across training samples and should not be interpreted as a fixed clinical cut-off. This is important because only 3.6% of patients died, allowing apparently high accuracy to occur even when most deaths are missed. A similar problem has been reported in malaria severity modelling, where overall test accuracy remained relatively high despite low recall for the severe-outcome group [11]. The present findings therefore support evaluation of discrimination, calibration, prediction error, sensitivity, and clinical consequences together instead of relying on accuracy or a single statistically optimized threshold.

4.2. Interpretation of Clinical Predictors

Age remained an important predictor of in-hospital mortality, but the nonlinear refinement showed that its association with mortality was not adequately described by a constant increase in odds for every additional year of age. Under the original linear specification, the adjusted odds ratio was 1.036 per year; however, the restricted cubic spline analysis demonstrated that the magnitude of the age effect varied substantially across the observed age range. Mortality odds remained comparatively lower through much of early and middle adulthood and increased progressively at older ages. The World Health Organization identifies young children, particularly those under five years, as a population at substantial risk of severe malaria and death because of limited acquired immunity [1]. The present findings additionally suggest important vulnerability among older admitted patients. This may reflect declining physiological reserve, accumulated comorbidity, differences in previous malaria exposure, or reduced capacity to tolerate severe infection. The widening confidence intervals at the extremes of the fitted age curve indicate that estimates in the youngest and oldest patients should be interpreted cautiously. Overall, the spline analysis provides a more appropriate description of the age-mortality relationship than the constant per-year effect from the original linear model [12].

Parasite density remained positively associated with mortality and was the only continuous predictor for which the tested linearity assumption was not violated. In the corrected nonlinear model, each increase of 1,000 parasites/μL was associated with an adjusted odds ratio of approximately 1.022 for mortality. Although this change appears small when expressed per 1,000 parasites/μL, parasite densities in the study varied over a very wide range, so differences in total parasite burden can translate into substantial differences in estimated risk. This association is biologically plausible because high parasitaemia may contribute to red-cell destruction, microvascular sequestration, metabolic disturbance, tissue hypoxia, and organ dysfunction [5, 6]. The persistence of the association after correction of the nonlinear effects of age, haemoglobin, and platelet count supports parasite density as an independent indicator of infection burden in this hospital population.

Haemoglobin level showed a pronounced nonlinear inverse relationship with mortality. The original linear model estimated an AOR of 0.652 for every 1 g/dL increase in haemoglobin, but the spline analysis demonstrated that the protective association was not constant across the full haemoglobin range. Mortality odds declined most sharply at low haemoglobin concentrations, with the greatest risk observed among patients with severe anaemia, and the relationship became substantially flatter as haemoglobin approached approximately 10–11 g/dL. This pattern is clinically plausible because malaria-associated anaemia may result from destruction of infected and uninfected red blood cells, impaired erythropoiesis, and other mechanisms that reduce oxygen-carrying capacity during severe disease [5]. Mousa et al. also reported that delayed malaria treatment was associated with declining haemoglobin and substantially increased odds of severe malarial

anaemia, supporting the importance of haemoglobin as an indicator of disease severity [10]. The present spline findings therefore refine the original result by showing that the mortality gradient is particularly steep at lower haemoglobin levels, rather than implying the same percentage reduction in mortality odds for every 1 g/dL increase across the entire observed range.

Platelet count also showed a marked nonlinear relationship with mortality. The original linear specification suggested a constant protective association, with an AOR of 0.990 for each increase of 1 $\times 10^9$/L. The corrected spline model showed instead that mortality odds were particularly elevated at low platelet counts, declined toward the middle of the observed platelet distribution, and increased again at the upper end of the fitted curve. Thrombocytopenia is frequently observed in malaria and may result from immune-mediated platelet destruction, splenic sequestration, peripheral consumption, or systemic inflammatory activation [7]. The high estimated mortality risk at low platelet counts is therefore consistent with platelet depletion acting as a marker of more severe systemic disease. However, the apparent increase in mortality odds at high platelet values should be interpreted cautiously because confidence intervals widened toward the upper end of the distribution. The revised findings therefore support interpretation of platelet count through its complete nonlinear function rather than as a constant protective effect per unit increase.

Sex was not independently associated with mortality in either model specification. In the original linear model, the AOR for male versus female sex was 0.930, while the corrected spline model produced an AOR of approximately 1.076 with a confidence interval that crossed unity. The consistency of the non-significant sex effect after correction of the other predictors indicates that sex alone contributed little independent predictive information in this dataset. This finding does not imply that sex-related clinical circumstances are unimportant. Malaria risk may differ according to pregnancy, exposure patterns, acquired immunity, access to care, and health-seeking behaviour, factors that cannot be represented adequately by a single binary sex variable [18, 19]. The pregnancy literature, for example, demonstrates that physiological and immunological changes during pregnancy can alter malaria susceptibility and clinical outcomes [18]. The present result should therefore be interpreted specifically as an absence of an independent effect of recorded sex after adjustment for the other predictors included in this model.

4.3. Model Performance and Potential Clinical Use

The corrected restricted cubic spline model demonstrated strong discrimination while retaining the interpretability of a logistic regression framework. Its apparent AUC of 0.962 decreased only modestly to an optimism-corrected AUC of 0.955 after bootstrap validation and to 0.954 during repeated stratified cross-validation. The cross-validated Brier score of 0.0273 also indicated relatively low overall probabilistic prediction error. These findings suggest that routinely collected demographic and laboratory variables can meaningfully distinguish patients at different levels of mortality risk when their functional relationships with the outcome are represented appropriately. Logistic regression remains attractive for clinical prediction because its mathematical structure and predictor contributions can be examined directly, although nonlinear terms require interpretation through their complete fitted functions rather than individual spline coefficients [8, 20].

Calibration findings provide an important qualification to the strong discrimination. The bootstrap calibration intercept was close to zero, indicating little evidence of systematic overall underprediction or overprediction, while the calibration slope of 0.927 suggested some residual optimism. Repeated cross-validation produced a similar calibration pattern, with a slope of 0.915. The ridge-penalized sensitivity analysis yielded comparable discrimination and prediction error but a calibration slope closer to one, supporting the use of shrinkage when model complexity increases relative to the number of outcome events. Internal validation is important for estimating optimism in model performance, but it does not establish whether predictions will remain accurate in patients from different hospitals or clinical settings [15]. The present model should therefore be regarded as internally validated within the source population rather than ready for unrestricted clinical deployment.

Classification performance depended strongly on the probability threshold selected. The original development analysis showed that the conventional 0.5 threshold produced high overall accuracy while identifying only a small proportion of deaths. Although the original Youden threshold of approximately 0.036 increased sensitivity substantially, its selection and evaluation in the same dataset made its apparent performance optimistic. When threshold selection was repeated within the training samples during cross-validation, the selected thresholds varied, with a median of approximately 0.0343. This variability demonstrates that a statistically optimized threshold is not necessarily a stable or clinically appropriate operating point.

No mortality-risk threshold has been externally established for use at Nakuru Level 6 Hospital. Consequently, the 5% threshold examined in the revised analysis should be interpreted only as an illustrative operating scenario and not as a recommended clinical cut-off. At this threshold, the internally cross-validated model identified 87.0% of deaths and had a specificity of 88.6%, while approximately 14.1% of all admissions would have been flagged for additional review. The positive predictive value remained relatively low at 22.2%, whereas the negative predictive value was 99.46%. The low positive predictive value is expected in a population in which mortality prevalence is only 3.6% and means that most patients classified as high risk would ultimately survive. In a clinical setting, such a threshold would therefore need to be

judged against available monitoring capacity, the consequences of missed high-risk patients, and the workload generated by false-positive alerts.

Decision-curve analysis provided an additional assessment of potential clinical usefulness by evaluating net benefit across a range of threshold probabilities instead of selecting one threshold solely from sensitivity and specificity [17]. The corrected model showed greater net benefit than the classify-all and classify-none strategies across the examined low-risk threshold range. This supports the potential value of model-guided risk stratification but does not establish a universally optimal cut-off. The clinical consequences assigned to false-negative and false-positive classifications may differ substantially between hospitals, so threshold selection should ultimately be informed by prospective workflow assessment, available resources, and clinician-defined intervention consequences.

The findings are consistent with previous evidence showing that strong overall classification measures may conceal poor identification of uncommon severe outcomes. Ansong et al. [11] developed a logistic regression model for malaria severity among 417 children in Ghana and reported test accuracy of 0.83, while recall for the severe-malaria class was only 0.36. A related problem was observed in the original linear analysis of the present study, in which the conventional 0.5 threshold produced high accuracy but very low sensitivity. The strengthened analysis therefore gives greater emphasis to discrimination, calibration, Brier score, sensitivity, and clinical net benefit instead of overall accuracy alone.

The importance of haemoglobin as a component of mortality-risk assessment is also consistent with findings from Mousa et al. [10]. Their pooled multicentre analysis showed that treatment delay was associated with declining haemoglobin and substantially increased odds of severe malarial anaemia. Although their outcome was progression to severe malaria rather than in-hospital mortality, the findings support the biological and clinical importance of anaemia in malaria severity. The present nonlinear analysis extends this interpretation by showing that the association between haemoglobin and mortality was strongest at low concentrations and became substantially flatter as haemoglobin approached approximately 10–11 g/dL.

Finally, despite the favourable internal-validation results, the model has not been evaluated in an independent hospital, a later temporal cohort, or a prospectively enrolled population. Differences in malaria transmission, referral patterns, patient characteristics, laboratory measurement, treatment availability, and baseline mortality may alter both predictor distributions and calibration. External validation should therefore precede clinical implementation. If systematic differences in predicted and observed risk are identified in a new population, recalibration may also be required before prospective evaluation of whether model-assisted risk stratification improves clinical decision-making without producing an unacceptable alert burden.

4.4. Strengths and Limitations

A major strength of this study was the use of 1,500 routinely collected hospital records and five clinically available predictors. The revised analysis assessed several aspects of model quality, including functional form, discrimination, calibration, prediction error, classification performance, and clinical net benefit. Significant nonlinear relationships for age, haemoglobin level, and platelet count were addressed using restricted cubic splines while retaining the same five prespecified predictors [12]. Model complexity was also examined formally, and bootstrap validation, repeated cross-validation, ridge penalization, and decision-curve analysis were used to assess optimism, overfitting, and potential clinical utility [13-17].

Several limitations remain. First, the study was conducted at a single referral hospital, and the model has not been externally validated, so transportability to other settings is uncertain. Second, only 54 deaths occurred. Although the formal sample-size assessment indicated that the available sample narrowly met the estimated requirement, the limited number of events reduced precision, particularly at the extremes of the spline curves. Third, complete-case analysis may have introduced selection bias, and retrospective electronic records may contain measurement or documentation errors.

The model was also limited to five predictors and did not include potentially important variables such as vital signs, organ dysfunction, comorbidities, pregnancy status, treatment delay, treatment received, or referral pathway. Although the nonlinear specification improved model fit, the additional spline parameters increased model complexity, and calibration slopes below one suggested some residual optimism. Ridge penalization reduced this concern but could not eliminate it completely.

Finally, no clinically established mortality-risk threshold was available for Nakuru Level 6 Hospital. The original Youden threshold was development-derived, while the 5% threshold was used only as an illustrative operating point. Decision-curve analysis suggested potential net benefit across low-risk thresholds [17], but any operational threshold should be selected prospectively according to clinical consequences, monitoring capacity, staffing, and the burden of false-positive alerts.

The most important next step is external validation in geographically and clinically distinct hospitals, followed by recalibration where necessary and prospective assessment of whether model-assisted risk stratification improves clinical decision-making without creating an excessive alert burden [15].

5. Conclusions

Age, parasite density, haemoglobin level, and platelet count were associated with in-hospital mortality among malaria patients at Nakuru Level 6 Hospital, while sex was not independently predictive. The original five-predictor linear lo-

gistic regression model showed strong discrimination, but diagnostic assessment identified important nonlinear relationships for age, haemoglobin level, and platelet count. Refitting these three variables using restricted cubic splines, while retaining the same five clinical predictors, significantly improved model fit and discrimination.

The corrected model showed strong internally validated performance, with an optimism-corrected AUC of 0.955 and repeated cross-validated AUC of 0.954. Calibration intercepts were close to zero, although calibration slopes below one indicated some residual optimism and uncertainty remained at higher predicted risk. Threshold analyses also confirmed that the conventional 0.5 cut-off was unsuitable for this rare mortality outcome. Lower thresholds improved sensitivity, but no single threshold was established as clinically optimal.

The model should therefore be considered a promising internally validated risk-stratification tool rather than a deployment-ready clinical rule. External validation, possible recalibration, and prospective assessment of clinical utility and threshold selection are required before routine implementation.

Abbreviations

AOR	Adjusted Odds Ratio
AUC	Area Under the Receiver Operating Characteristic Curve
CI	Confidence Interval
CV	Cross-Validation
EMR	Electronic Medical Records
NPV	Negative Predictive Value
PPV	Positive Predictive Value
ROC	Receiver Operating Characteristic
VIF	Variance Inflation Factor
WHO	World Health Organization
AIC	Akaike Information Criterion
DCA	Decision Curve Analysis
RCS	Restricted Cubic Spline

Supplementary Material

The supplementary material can be accessed at https://doi.org/10.11648/j.ajtas.20261505.13

Acknowledgments

Sincere gratitude is expressed to the Almighty God for granting the strength, good health, and opportunity necessary for the successful and timely completion of this project. I am deeply grateful to my supervisors, Dr Eyang'an Esekon and Dr Harun Gitonga, for their continuous guidance, support, and constructive direction at every stage of this project. I would also like to extend my heartfelt appreciation to my late mother. Although she passed away before witnessing the completion of this work, her early lessons in perseverance and her unwavering belief in my potential laid a strong foundation for my academic journey. Finally, I am sincerely thankful to my brother, Haron Thuo, for his emotional support and encouragement, which helped me overcome numerous challenges throughout the course of this project.

Author Contributions

Waruino John Mwangi: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – original draft

Joseph Eyang'an Esekon: Methodology, Supervision, Validation, Writing – review & editing

Harun Mwangi Gitonga: Methodology, Supervision, Validation, Writing – review & editing

Data Availability Statement

The de-identified data supporting the findings may be available from the corresponding author upon reasonable request and subject to institutional and ethical approval.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] World Health Organization. *World Malaria Report 2023*. Geneva: World Health Organization; 2023.

[2] Venkatesan, P. WHO World Malaria Report 2024. *The Lancet Microbe*. 2025, 6(4), e101073. https://doi.org/10.1016/j.lanmic.2025.101073

[3] Ministry of Health, National Malaria Control Programme. *Kenya Malaria Strategy 2023–2027*. Nairobi, Kenya: Ministry of Health; 2023.

[4] Macharia, P. M., Giorgi, E., Noor, A. M., Waqo, E., Kiptui, R., Okiro, E. A., et al. Spatio-temporal analysis of *Plasmodium falciparum* prevalence to understand the past and chart the future of malaria control in Kenya. *Malaria Journal*. 2018, 17, 340. https://doi.org/10.1186/s12936-018-2489-9

[5] White, N. J. Severe malaria. *Malaria Journal*. 2022, 21(1), 284. https://doi.org/10.1186/s12936-022-04301-8

[6] Riley, E. M., Stewart, V. A., Ockenhouse, C. F. Parasite density and mortality in severe malaria. *Malaria Journal*. 2017, 16(1), 210.

[7] Krishna, P., Chalamalasetty, M. K. Thrombocytopenia in malaria and its diagnostic significance: a prospective study. *Journal of Clinical and Scientific Research*. 2023, 12(Suppl 1): S1–S4. https://doi.org/10.4103/jcsr.jcsr_9_23

[8] Schober, P., Vetter, T. R. Logistic regression in medical research. *Anesthesia & Analgesia*. 2021, 132(2), 365–366. https://doi.org/10.1213/ANE.0000000000005334

[9] Hosmer, D. W., Lemeshow, S., Sturdivant, R. X. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley; 2013.

[10] Mousa, A., Al-Taiar, A., Anstey, N. M., Badaut, C., Barber, B. E., Bassat, Q., et al. The impact of delayed treatment of uncomplicated *Plasmodium falciparum* malaria on progression to severe malaria: a systematic review and pooled multicentre individual-patient meta-analysis. *PLOS Medicine*. 2020, 17(10), e1003359. https://doi.org/10.1371/journal.pmed.1003359

[11] Ansong, M. O., Obeng, A. Y., Opoku, S. K. A logistic regression model to predict malaria severity in children. *European Journal of Electrical Engineering and Computer Science*. 2024, 8(2). https://doi.org/10.24018/ejece.2024.8.2.614

[12] Gauthier, J., Wu, Q. V., Gooley, T. A. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. *Bone Marrow Transplantation*. 2020, 55(4), 675–680. https://doi.org/10.1038/s41409-019-0679-x

[13] Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E. Jr., Moons, K. G. M., Collins, G. S. Minimum sample size for developing a multivariable prediction model: Part II—binary and time-to-event outcomes. *Statistics in Medicine*. 2019, 38(7), 1276–1296. https://doi.org/10.1002/sim.7992

[14] van Smeden, M., Moons, K. G. M., de Groot, J. A. H., Collins, G. S., Altman, D. G., Eijkemans, M. J. C., Reitsma, J. B. Sample size for binary logistic prediction models: beyond events per variable criteria. *Statistical Methods in Medical Research*. 2019, 28(8), 2455–2474. https://doi.org/10.1177/0962280218784726

[15] Steyerberg, E. W., Harrell, F. E. Jr., Borsboom, G. J. J. M., Eijkemans, M. J. C., Vergouwe, Y., Habbema, J. D. F. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*. 2001, 54(8), 774–781. https://doi.org/10.1016/S0895-4356(01)00341-9

[16] Van Calster, B., van Smeden, M., De Cock, B., Steyerberg, E. W. Regression shrinkage methods for clinical prediction models do not guarantee improved performance: simulation study. *Statistical Methods in Medical Research*. 2020, 29(11), 3166–3178. https://doi.org/10.1177/0962280220921415

[17] Vickers, A. J., Elkin, E. B. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*. 2006, 26(6), 565–574. https://doi.org/10.1177/0272989X06295361

[18] Rogerson, S. J., Hviid, L., Duffy, P. E., Leke, R. F. G., Taylor, D. W. Malaria in pregnancy: pathogenesis and immunity. *The Lancet Infectious Diseases*. 2007, 7(2), 105–117. https://doi.org/10.1016/S1473-3099(07)70022-1

[19] Atobrah, D., Kwansa, B. K., Okyere-Asante, P. G., Kyere, A., Badasu, D. M., Kretchy, I. A. Conceptualization of gender in published malaria and gender research: a systematic descriptive review. *International Journal for Equity in Health*. 2025, 24, 211. https://doi.org/10.1186/s12939-025-02545-9

[20] Norton, E. C., Dowd, B. E., Maciejewski, M. L. Odds ratios—current best practice and use. *JAMA*. 2018, 320(1), 84–85. https://doi.org/10.1001/jama.2018.7619

Biography

Waruino John Mwangi is a statistics researcher affiliated with the Department of Pure and Applied Sciences at Kirinyaga University, Kenya. His research interests include biostatistics, logistic regression, clinical prediction modelling, epidemiological data analysis, and the application of statistical methods to public-health decision-making.

Research Field

Waruino John Mwangi: Biostatistics, Clinical prediction modelling, Logistic regression, Epidemiological data analysis, Health statistics

Joseph Eyang'an Esekon: Applied statistics, Statistical modelling, Biostatistics, Research methodology, Predictive analytics

Harun Mwangi Gitonga: Applied statistics, Statistical modelling, Data analysis, Research methodology, Quantitative methods