Nigerian Medical Journal
NIGERIA MEDICAL ASSOCIATION
Print ISSN 0300-1652, E-ISSN 2229-774X

Original Article

Predictors of Optimal Uptake of Intermittent Preventive Treatment in Pregnancy (Iptp3+) in Nigeria: Survey-Weighted Regression and Exploratory Machine-Learning Analysis of the 2021 Nigeria Malaria Indicator Survey

Happiness Opeyemi Agboola[1], Adewale Jamiu Lasisi[2], Olufisayo Elugbadebo[3], Salim Khan[4], Olatunde Aremu[5],

[1]Department of Health Promotion, Oyo State Primary Healthcare Board, Ibadan, Nigeria, [2]Department of Obstetrics and Gynaecology, Adeoyo Maternity Teaching Hospital, Ibadan, Nigeria, [3]Department of Psychiatry, College of Medicine, University of Ibadan, Ibadan, Nigeria, [4]Faculty of Health, Education and Life Sciences, Birmingham City University, Birmingham, United Kingdom, [5]Department of Public Health, Faculty of Health, Education and Life Sciences, Birmingham City University, Birmingham, United Kingdom

Abstract

Background: Malaria in pregnancy remains a public health challenge in Nigeria despite the effectiveness of intermittent preventive treatment with sulfadoxine-pyrimethamine (IPTp-SP). Uptake of the WHO-recommended minimum of three doses (IPTp3+) remains below national targets. This study examined factors associated with optimal uptake of IPTp and explored machine-learning-based prediction.

Methodology: We analysed 2021 Nigeria Malaria Indicator Survey data for 3,103 women aged 15–49 years with a live birth, guided by Andersen's Behavioural Model. Descriptive and regression analyses were conducted in R. Descriptive summaries were unweighted, whereas the primary regression incorporated sampling weights, strata, and primary sampling units. Following a survey-adjusted nonlinearity test, ANC visits were categorised as 0–3, 4–7, and ≥8. Random Forest and XGBoost were fitted in Python as unweighted exploratory models, with calibration, permutation importance, and SHAP assessed on held-out data.

Results: Overall, 52.4% of women in the unweighted analytic sample reported IPTp3+ uptake. The association with ANC was nonlinear ($F(2,84)=18.12$; $p<0.001$). Compared with 0–3 visits, adjusted odds ratios were 2.10 (95% CI:1.64–2.68) for 4–7 visits and 2.05 (95% CI:1.49–2.83) for ≥8 visits; both $p<0.001$. Geopolitical zone was associated with uptake. XGBoost and Random Forest showed modest discrimination (ROC-AUC:0.661 and 0.640) and Brier scores of 0.225 and 0.228.

Conclusion: Only about half of women achieved optimal IPTp uptake. ANC attendance was strongly associated with IPTp3+ uptake, while geographic disparities persisted after adjustment. The exploratory machine-learning models showed modest predictive performance. Strengthening ANC access and improving IPTp delivery remain important priorities.

Keywords: Intermittent preventive treatment in pregnancy; Malaria in pregnancy; Antenatal care; explainable machine learning; Nigeria

Correspondence: Happiness Opeyemi Agboola, Department of Health Promotion, Oyo State Primary Healthcare Board, Ibadan, Nigeria. Telephone: +2348168579398. Email: happyagboola7@gmail.com

How to Cite: Agboola HO, Lasisi AJ, Elugbadebo O, Khan S, Aremu O. Predictors of Optimal Uptake of Intermittent Preventive Treatment in Pregnancy (Iptp3+) in Nigeria: Survey-Weighted Regression and Exploratory Machine-Learning Analysis of the 2021 Nigeria Malaria Indicator Survey. Niger Med J 2026; 67 (5): 1801 – 1818. https://doi.org/10.71480/nmj.v67i5.1537

Quick Response Code:

Introduction

Malaria remains a major global public health challenge, with the greatest burden concentrated in sub-Saharan Africa [1]. Recent estimates indicate more than 280 million cases worldwide, with Africa accounting for approximately 94–95% of malaria cases and deaths [1,2]. Nigeria contributes substantially to this burden, and transmission remains intense because ecological and climatic conditions favour Anopheles breeding and survival [2,3]. Transmission often increases during the rainy season and varies across geographical settings, with rural communities and populations living near water bodies frequently experiencing greater exposure [2,3]. Plasmodium falciparum is the predominant parasite, while major vectors include members of the Anopheles gambiae complex and Anopheles funestus group [1,2]. Pregnant women are particularly vulnerable because pregnancy-related immunological changes increase susceptibility to infection and adverse outcomes, including maternal anaemia, stillbirth, preterm delivery, and low birth weight [3,4].

Intermittent preventive treatment in pregnancy with sulfadoxine–pyrimethamine (IPTp-SP) is therefore a key WHO-recommended intervention. It should be administered at scheduled antenatal care (ANC) visits from the second trimester, with at least three doses (IPTp3+) considered optimal [3]. Evidence shows that IPTp-SP reduces maternal morbidity and improves birth outcomes [5,6]. However, IPTp3+ uptake in Nigeria remains below the national target of 80%, with marked geographical and socioeconomic inequalities [7–9]. Previous studies have identified maternal education, household wealth, ANC attendance, malaria-related information, service availability, and facility readiness as important correlates of uptake [8,10,11].

This study was guided by Andersen's Behavioural Model of Health Services Utilisation [12]. Maternal age and education were treated as predisposing factors; household wealth, residence, and geopolitical zone as enabling factors; and ANC attendance as a measure of healthcare utilisation. Existing population health research has mainly used logistic regression, which may not fully capture nonlinear relationships and predictor interactions [7–9]. Machine-learning methods can complement regression by exploring nonlinear predictive patterns and out-of-sample performance [13,14]. However, predictive importance does not imply causality, and machine learning may add limited value when few predictors are available. Applications to IPTp uptake using recent nationally representative Nigerian data remain limited.

This study addressed this gap by combining survey-weighted logistic regression with explainable machine-learning approaches to examine IPTp3+ uptake among Nigerian women. Exploratory Random Forest and XGBoost models were interpreted using SHAP and permutation importance to explain model predictions rather than rank causal determinants or set policy priorities. The objectives were to identify factors associated with IPTp3+ uptake, assess out-of-sample discrimination, and determine whether the machine-learning models added information beyond regression.

Methods

Study Design and Data Source

This study was a secondary analysis of the 2021 Nigeria Malaria Indicator Survey (NMIS), a nationally representative cross-sectional survey conducted by the National Malaria Elimination Programme in collaboration with international partners [15]. The NMIS used a stratified two-stage cluster sampling design to ensure representativeness across Nigeria's six geopolitical zones. In the first stage, enumeration areas were selected with probability proportional to size, followed by systematic household sampling in the second stage. The survey collected data on household characteristics, malaria prevention practices, maternal health, and child health indicators. Eligible participants were women aged 15–49 years who were usual residents or visitors in selected households. For this analysis, the sample was restricted to women who had a live birth in the two years preceding the survey, consistent with prior DHS-based analyses [15].

Outcome Variable

The outcome was IPTp3+ uptake, defined as maternal report of receiving three or more doses of sulfadoxine-pyrimethamine during the most recent pregnancy, in accordance with the WHO-recommended minimum number of doses. The outcome was recoded as a binary variable, with 1 representing receipt of at least three doses and 0 representing receipt of fewer than three doses. Because IPTp receipt was based primarily on maternal report, the measure may have been affected by recall error, incorrect identification of sulfadoxine-pyrimethamine and incomplete information regarding the timing and spacing of doses.

Predictor Variables

Predictor variables were selected based on epidemiological relevance, data availability and consistency with previous studies of IPTp uptake. The variables included maternal age, place of residence, educational attainment, household wealth index, geopolitical zone and number of antenatal care visits. The analytic dataset did not include linked measures of gestational age at first ANC contact, facility readiness, sulfadoxine–pyrimethamine stock availability, provider adherence to IPTp guidelines, or the precise timing and frequency of IPTp administration. These potentially important health-system and service-delivery variables could therefore not be incorporated into the present analysis.

Statistical Analysis

Tables 1 and 2 present unweighted descriptive summaries of the analytic sample. Categorical variables are reported as counts and percentages, while the number of antenatal care (ANC) visits is reported as the median and interquartile range. Unweighted bivariate associations were evaluated using Pearson's chi-squared tests for categorical variables and the Wilcoxon rank-sum test for the number of ANC visits. Because sampling weights and survey-design variables were not applied to these descriptive and bivariate analyses, the findings describe the observed analytic sample and should not be interpreted as nationally representative estimates.

A multivariable survey-weighted logistic regression model was fitted to examine factors associated with IPTp3+ uptake. Sampling weights, primary sampling units and survey strata were incorporated to account for unequal selection probabilities, clustering and stratification.

The number of ANC visits was initially examined as a continuous variable. ANC visits were centred and standardised, and linear, quadratic and cubic terms were included in a full survey-adjusted model. Departure from linearity was evaluated by jointly testing the quadratic and cubic terms. The design-adjusted joint test indicated evidence of nonlinearity, $F(2, 84) = 18.12$, $p < 0.001$. ANC utilisation was therefore included in the primary regression model as a categorical variable comprising 0–3, 4–7 and ≥8 visits.

Variances were estimated using Taylor series linearisation based on primary-sampling-unit contributions within strata. Statistical inference used the residual survey-design degrees of freedom. Associations are reported as adjusted odds ratios with 95% confidence intervals.

Multicollinearity was assessed using variance inflation factors calculated from the regression design matrix. Values below 5 were considered to indicate no problematic multicollinearity. All inferential tests were two-sided, with statistical significance evaluated at $p < 0.05$. Exact p values are reported to three decimal places, while values below 0.001 are reported as $p < 0.001$.

Exploratory Machine-Learning Analysis

Random Forest and XGBoost models were developed as exploratory predictive comparators using maternal and household characteristics. Unweighted models excluded survey-design variables and were not intended for nationally representative or causal inference. Data were split 80:20 using stratified sampling, with preprocessing and five-fold cross-validation performed within the training set. Models were evaluated using ROC-AUC, accuracy, sensitivity, specificity, balanced accuracy, precision, recall, F1-score, and Brier scores. Calibration was also assessed. Sensitivity analyses incorporated sampling-design variables as predictors to determine whether predictive performance depended on sampling structure, without implementing survey-weighted estimation.

Machine-Learning Model Interpretation

Permutation importance and Shapley additive explanations (SHAP) were used to examine the fitted models. Permutation importance assessed the reduction in held-out ROC-AUC following the random permutation of each predictor. SHAP values described the direction and magnitude of each predictor's contribution to individual model predictions.

Permutation importance and SHAP findings were interpreted as model-specific associational summaries. They were not interpreted as causal effects, population-level effect estimates, an epidemiological hierarchy of determinants or rankings of intervention priorities.

Statistical Software

Descriptive and regression analyses were conducted using R (R Foundation for Statistical Computing, Vienna, Austria). Exploratory machine-learning analyses were conducted using Python within the Jupyter Notebook environment, with the scikit-learn, XGBoost and SHAP packages.

Ethical Considerations

Permission to access and use the 2021 Nigeria Malaria Indicator Survey (NMIS) dataset was obtained from the DHS Program/ICF through the standard online registration and data access request process. The dataset is available through the DHS Program website: (https://dhsprogram.com/data/dataset/Nigeria_MIS_2021.cfm?flag=1.) The NMIS was conducted in accordance with established ethical procedures, including informed consent from participants and measures to ensure confidentiality and anonymity. The dataset used for this secondary analysis was anonymised, with all identifying information removed before release. Therefore, no additional ethical approval was required for the present study.

Results

Study population and IPTp uptake

A total of 3,103 women from the NMIS 2021 dataset were included in the analysis. Of these, 1,625 (52.4%) reported optimal IPTp uptake, while 1,478 (47.6%) reported fewer than three doses. These figures describe the unweighted analytic sample and should not be interpreted as a survey-weighted national coverage estimate.

Table 1. Unweighted distribution of IPTp3+ uptake in the NMIS 2021 analytic sample.

IPTp uptake status	Frequency	Percentage (%)
Not optimal	1,478	47.6
Optimal	1,625	52.4
Total	3,103	100.0

Note: Counts and percentages are unweighted analytic-sample summaries.

Table 2. Unweighted bivariate comparison of factors by IPTp3+ uptake in the NMIS 2021 analytic sample.

Characteristic	Category	Not optimal n (%)	Optimal n (%)	p value
Age group	15–24	345 (23%)	427 (26%)	0.120
	25–34	792 (54%)	818 (50%)	
	35–49	341 (23%)	380 (23%)	
Residence	Urban	505 (34%)	596 (37%)	0.140
	Rural	973 (66%)	1,029 (63%)	
Geopolitical zone	North Central	206 (14%)	348 (21%)	<0.001
	North East	282 (19%)	291 (18%)	
	North West	434 (29%)	442 (27%)	
	South East	203 (14%)	203 (12%)	
	South South	178 (12%)	185 (11%)	
	South West	175 (12%)	156 (9.6%)	
Education	No education	496 (34%)	525 (32%)	0.300
	Primary	250 (17%)	260 (16%)	
	Secondary	518 (35%)	621 (38%)	
	Higher	214 (14%)	219 (13%)	
Wealth index	Poorest	238 (16%)	211 (13%)	0.130
	Poorer	245 (17%)	273 (17%)	
	Middle	303 (21%)	365 (22%)	
	Richer	343 (23%)	374 (23%)	
	Richest	349 (24%)	402 (25%)	
ANC visits	Median (Q1, Q3)	4.0 (3.0, 6.0)	5.0 (4.0, 6.0)	<0.001

Table 2 presents the unweighted bivariate association between selected maternal, household, and healthcare-related characteristics and optimal IPTp uptake. Geopolitical zone and number of antenatal care (ANC) visits differed significantly by IPTp3+ status. Women with optimal uptake attended more ANC visits than women with non-optimal uptake [median 5.0 (IQR: 4.0–6.0) versus 4.0 (IQR: 3.0–6.0); $p < 0.001$]. Significant regional variation was also observed ($p < 0.001$). No statistically significant differences were observed for age group, place of residence, educational attainment, or household wealth index (all $p > 0.05$).

Table 3. Survey-weighted adjusted odds ratios for factors associated with IPTp3+ uptake in NMIS 2021.

Variable	Category	AOR	95% CI	p value
Maternal age (years)	Per additional year	1.00	0.98–1.01	0.580
Residence	Urban	Reference	–	–
	Rural	0.92	0.72–1.18	0.506
Geopolitical zone	North Central	Reference	–	–
	North East	0.69	0.49–0.98	0.037
	North West	0.71	0.52–0.98	0.036
	South East	0.57	0.39–0.83	0.004

	South South	0.59	0.32–1.08	0.086
	South West	0.52	0.33–0.83	0.006
Education	No education	Reference	–	–
	Primary	0.93	0.69–1.26	0.641
	Secondary	0.95	0.72–1.25	0.689
	Higher	0.88	0.62–1.24	0.448
Wealth index	Poorest	Reference	–	–
	Poorer	1.04	0.74–1.47	0.810
	Middle	1.15	0.80–1.66	0.448
	Richer	1.07	0.75–1.53	0.686
	Richest	1.08	0.73–1.59	0.689
ANC visits	0–3 visits	Reference	–	–
	4–7 visits	2.10	1.64–2.68	<0.001
	≥8 visits	2.05	1.49–2.83	<0.001

Note: The primary model incorporates sampling weights, 17 strata, and 118 PSUs; Taylor-linearized standard errors and residual survey-design degrees of freedom were used for inference. ANC was modelled categorically because the survey-adjusted nonlinearity test was significant. AOR = adjusted odds ratio; CI = confidence interval. Associations should not be interpreted causally.

Table 3 presents adjusted odds ratios from the primary survey-adjusted multivariable logistic regression model. After controlling for maternal age, place of residence, educational attainment, household wealth index, geopolitical zone, and ANC category, geopolitical zone and ANC attendance remained associated with optimal IPTp uptake. Compared with women in the North Central zone, lower adjusted odds were observed in the North East (AOR = 0.69; 95% CI: 0.49–0.98; p = 0.037), North West (AOR = 0.71; 95% CI: 0.52–0.98; p = 0.036), South East (AOR = 0.57; 95% CI: 0.39–0.83; p = 0.004), and South West (AOR = 0.52; 95% CI: 0.33–0.83; p = 0.006). The estimate for South South was lower but not statistically significant (AOR = 0.59; 95% CI: 0.32–1.08; p = 0.086). The overall ANC-category test was significant ($F(2, 85) = 19.50$; $p < 0.001$). Compared with 0–3 ANC visits, 4–7 visits were associated with an AOR of 2.10 (95% CI: 1.64–2.68; $p < 0.001$), while ≥8 visits were associated with an AOR of 2.05 (95% CI: 1.49–2.83; $p < 0.001$). The survey-adjusted diagnostic supported the categorical ANC specification ($F(2, 84) = 18.12$; $p < 0.001$), and the maximum VIF was 3.97. Maternal age, residence, education, and household wealth were not independently associated with optimal IPTp uptake.

Exploratory Machine-Learning Results

The exploratory analysis included all 3,103 women. The primary models used the six substantive predictors listed in Table 4. State identifiers and survey-design fields were excluded from the primary predictor set. State, the original NMIS sampling weight, primary sampling units and survey strata were added only to the sensitivity models as ordinary predictors. The rescaled sampling-weight field was excluded because it duplicated the original sampling weight on a different scale. Both the primary and sensitivity models remained unweighted predictive models; therefore, their results describe prediction within the analytic sample and should not be interpreted as nationally representative estimates.

Table 4. Feature audit for the primary and sampling-structure sensitivity models

Variable	Decision	Rationale
Optimal IPTp uptake	Outcome only; not used as a predictor	Binary study outcome.

Variable	Decision	Rationale
Maternal age	Retained in the primary and sensitivity models	Substantive predictor.
Residence	Retained in the primary and sensitivity models	Substantive predictor.
State	Excluded from the primary model; included only in the sensitivity analysis	High-cardinality geographic field nested within geopolitical zone.
Education	Retained in the primary and sensitivity models	Substantive predictor.
Wealth index	Retained in the primary and sensitivity models	Substantive predictor.
ANC visits	Retained in the primary and sensitivity models	Substantive predictor.
Original NMIS sampling weight	Excluded from the primary model; included only as an ordinary predictor in the sensitivity analysis	Survey-design field representing unequal selection probabilities; not a substantive individual-level predictor.
Age group	Excluded from both models	Derived duplicate of maternal age; retained only for descriptive reporting.
Geopolitical zone	Retained in the primary and sensitivity models	Substantive geographic predictor.
Rescaled NMIS sampling weight	Excluded from both models	Duplicate representation of the original sampling weight, calculated as the original weight divided by 1,000,000.
Primary sampling unit	Excluded from the primary model; included only in the sensitivity analysis	Survey-design field representing the clustering structure.
Survey stratum	Excluded from the primary model; included only in the sensitivity analysis	Survey-design field representing the stratification structure.

Note: The sensitivity models added state, the original NMIS sampling weight, primary sampling units and survey strata to the primary predictor set. The sampling weight was entered as an ordinary predictor and was not used as an observation-level training weight. The sensitivity models therefore remained unweighted and did not implement design-based survey estimation. Survey-design fields were not interpreted as substantive predictors, causal determinants or intervention priorities.

Exploratory Model Performance on Held-Out Data

Table 5. Unweighted Random Forest and XGBoost performance on the held-out test set.

Model	Accuracy	ROC-AUC	Sensitivity	Specificity	Balanced accuracy	Precision	Recall	F1-score
Random Forest	0.607	0.640	0.652	0.557	0.605	0.618	0.652	0.635
XGBoost	0.599	0.661	0.723	0.463	0.593	0.596	0.723	0.654

Note: Sensitivity and recall refer to the IPTp3+ class. Metrics were calculated on the unweighted held-out 20% test set using the fixed 0.50 probability threshold. ROC-AUC is threshold-free; Brier scores and calibration curves assess probability calibration.

Both models showed poor-to-moderate discrimination on the held-out test data. XGBoost achieved a higher ROC-AUC than Random Forest (0.661 versus 0.640) and higher sensitivity (0.723 versus 0.652), whereas Random Forest was more specific (0.557 versus 0.463). Brier scores were 0.225 for XGBoost and 0.228 for Random Forest. Overall, these findings do not indicate robust individual prediction and do not support the use of either model for policy prioritisation.

Table 6. Unweighted Random Forest confusion matrix on the held-out test set.

Actual status	Predicted: non-optimal	Predicted: optimal
Actual: non-optimal	165	131
Actual: optimal	113	212

Table 7. Unweighted XGBoost confusion matrix on the held-out test set.

Actual status	Predicted: non-optimal	Predicted: optimal
Actual: non-optimal	137	159
Actual: optimal	90	235

At the fixed probability threshold of 0.50, Random Forest correctly classified 165 non-IPTp3+ and 212 IPTp3+ records, while XGBoost correctly classified 137 and 235, respectively. XGBoost produced more false positives but fewer false negatives, which is consistent with its higher sensitivity and lower specificity. This descriptive comparison does not make either model suitable for individual decision-making.

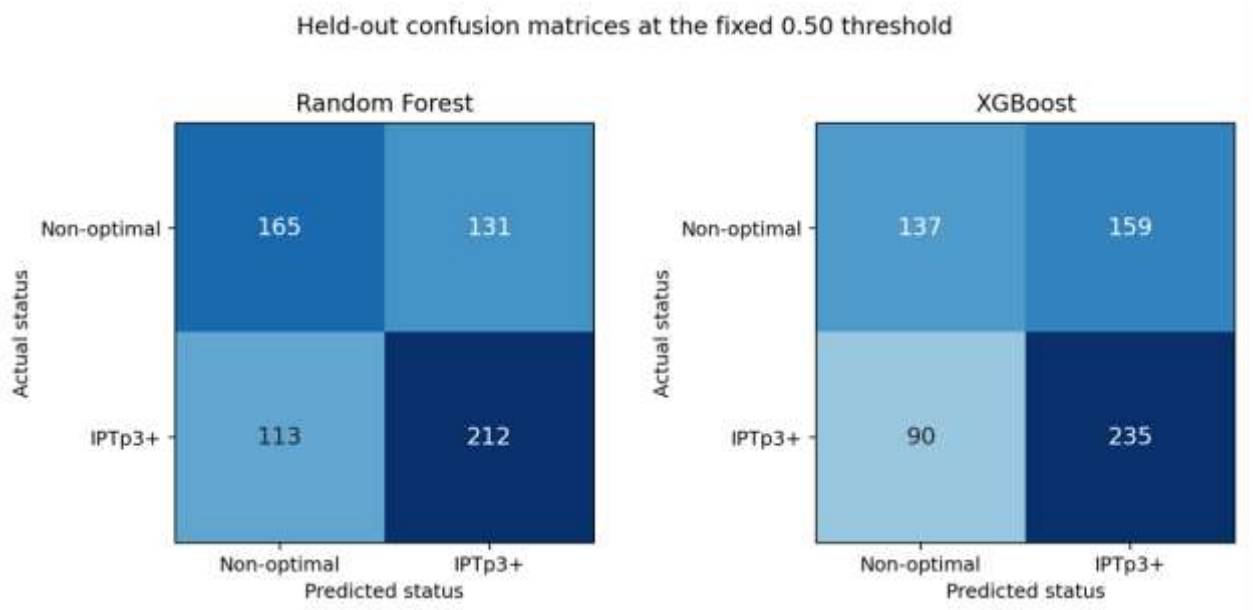

Figure 1. Unweighted held-out test confusion matrices for Random Forest and XGBoost.

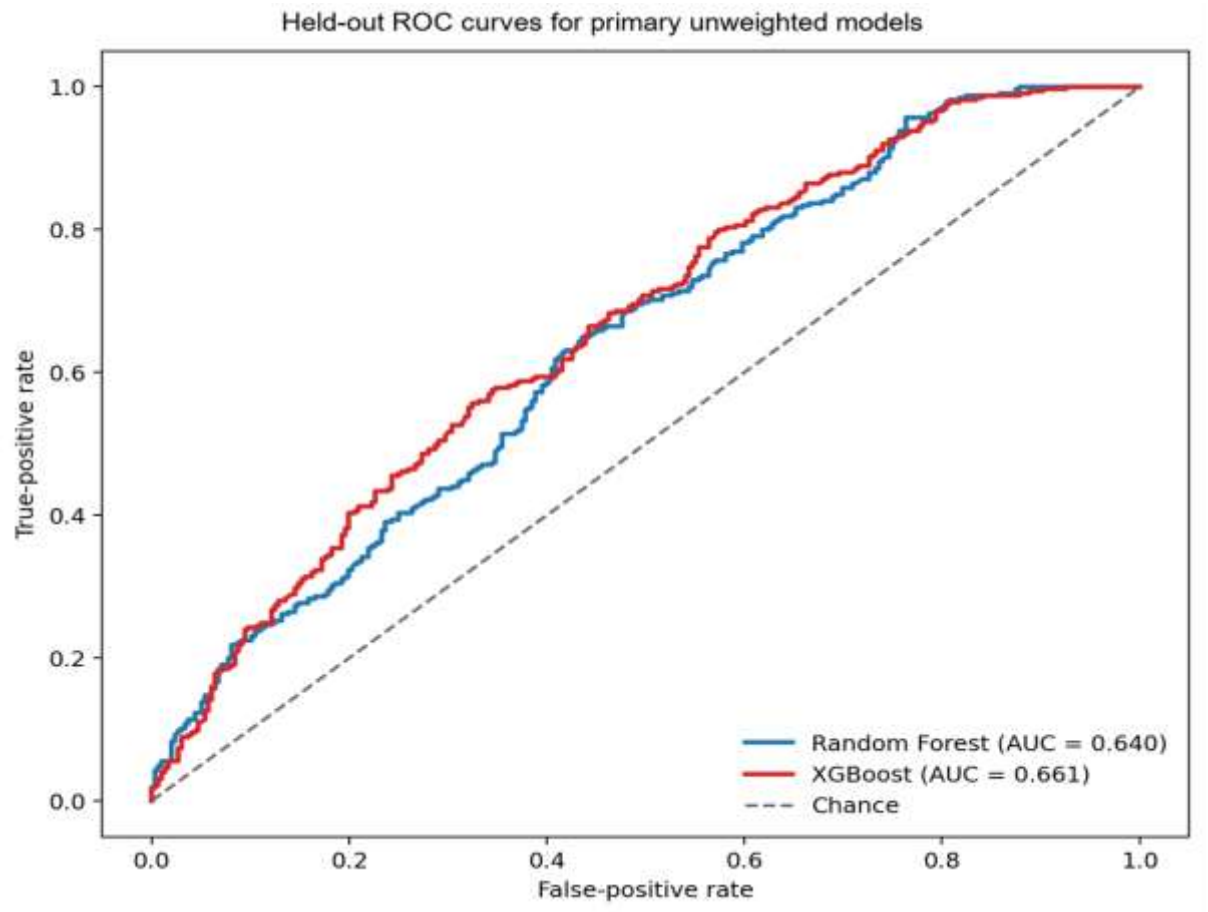

Figure 2. Held-out ROC curves for the primary unweighted Random Forest and XGBoost models.

The ROC curves lay above the chance diagonal but remained in the poor-to-moderate range. XGBoost achieved a numerically higher ROC-AUC than Random Forest (0.661 versus 0.640), but neither model demonstrated strong discrimination.

Sensitivity Analysis Including Survey-Design Fields

Table 8. Unweighted performance of the primary and sampling-structure sensitivity models.

Model	Accuracy	ROC-AUC	Sensitivity	Specificity	Balanced accuracy	F1-score
Random Forest (sampling-structure sensitivity model)	0.660	0.731	0.754	0.557	0.656	0.699
XGBoost (sampling-structure sensitivity model)	0.675	0.742	0.702	0.645	0.673	0.693
Random Forest (primary unweighted model)	0.607	0.640	0.652	0.557	0.605	0.635
XGBoost (primary unweighted model)	0.599	0.661	0.723	0.463	0.593	0.654

Note: The sampling-structure sensitivity models included the six primary predictors together with state, the original NMIS sampling weight, primary sampling unit and survey stratum. The duplicate rescaled sampling weight and derived age group were excluded. The primary models retained only the six substantive predictors. Both model sets used the same train–test split, preprocessing procedures, cross-validation strategy and fixed probability threshold of 0.50. Neither model set supports population inference. ROC-AUC was higher in the sampling-structure sensitivity models than in the primary models (Random Forest: 0.731 versus 0.640; XGBoost: 0.742 versus 0.661), suggesting that the algorithms could use sampling-structure information to improve held-out discrimination within the analytic sample. This finding does not establish that the sampling-structure fields are substantive determinants or that their exclusion eliminates all design-related bias.

Model-Based Variable Contributions

Table 9. Unweighted Random Forest permutation importance on held-out data.

Predictor	Mean decrease in ROC-AUC	SD
ANC visits	0.114	0.017
Geopolitical zone	0.048	0.016
Education	0.023	0.010
Wealth index	0.018	0.015
Residence	-0.004	0.006
Maternal age	-0.015	0.008

Table 10. Unweighted XGBoost permutation importance on held-out data.

Predictor	Mean decrease in ROC-AUC	SD
ANC visits	0.126	0.017
Geopolitical zone	0.064	0.016
Education	0.018	0.008
Maternal age	0.002	0.009
Wealth index	0.017	0.010
Residence	0.002	0.005

ANC visits produced the largest model-specific decrease in held-out ROC-AUC when permuted in both models, followed by geopolitical zone. These results describe contributions to prediction within the fitted models and should not be interpreted as causal rankings.

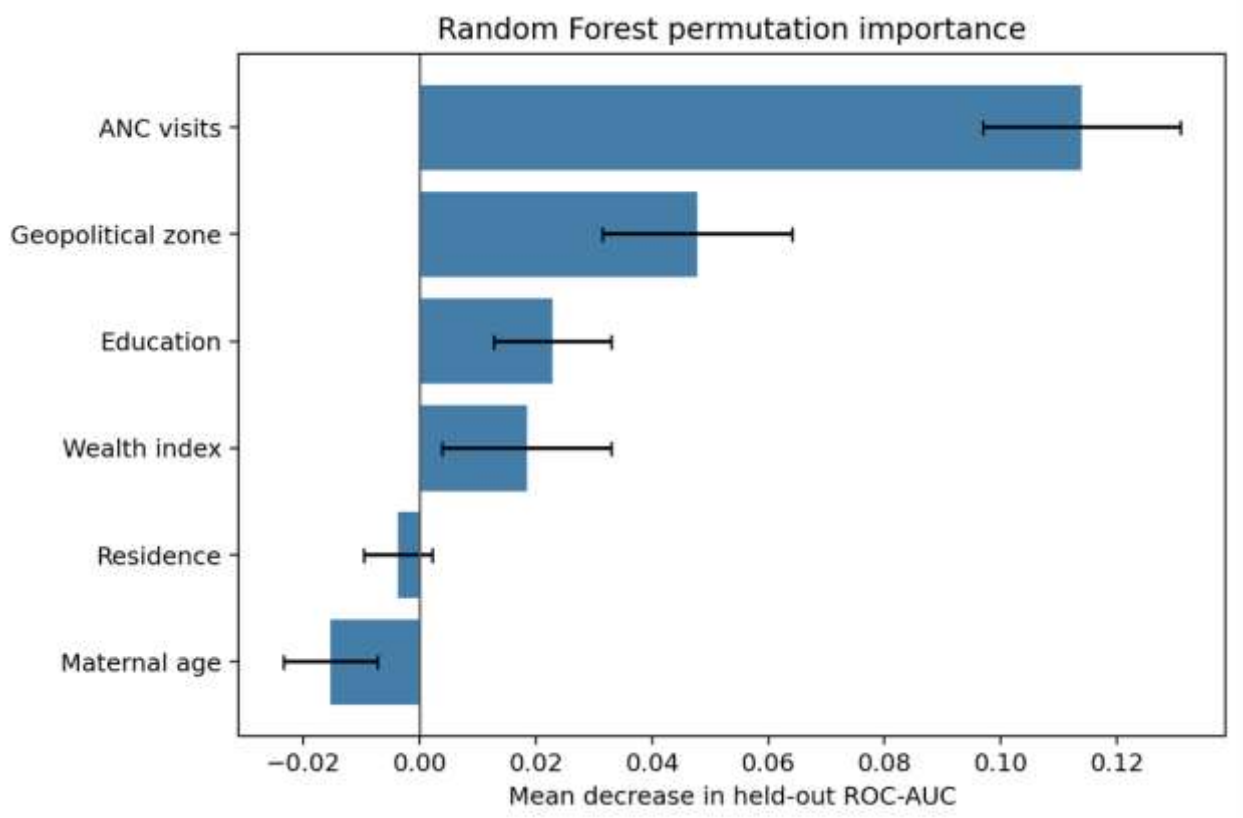

Figure 3. Random Forest permutation importance for the primary predictors.
Permutation importance showed that ANC visits produced the largest decrease in Random Forest test ROC-AUC, followed by geopolitical zone. Predictors with small or negative permutation values made little stable contribution to discrimination in the fitted model.

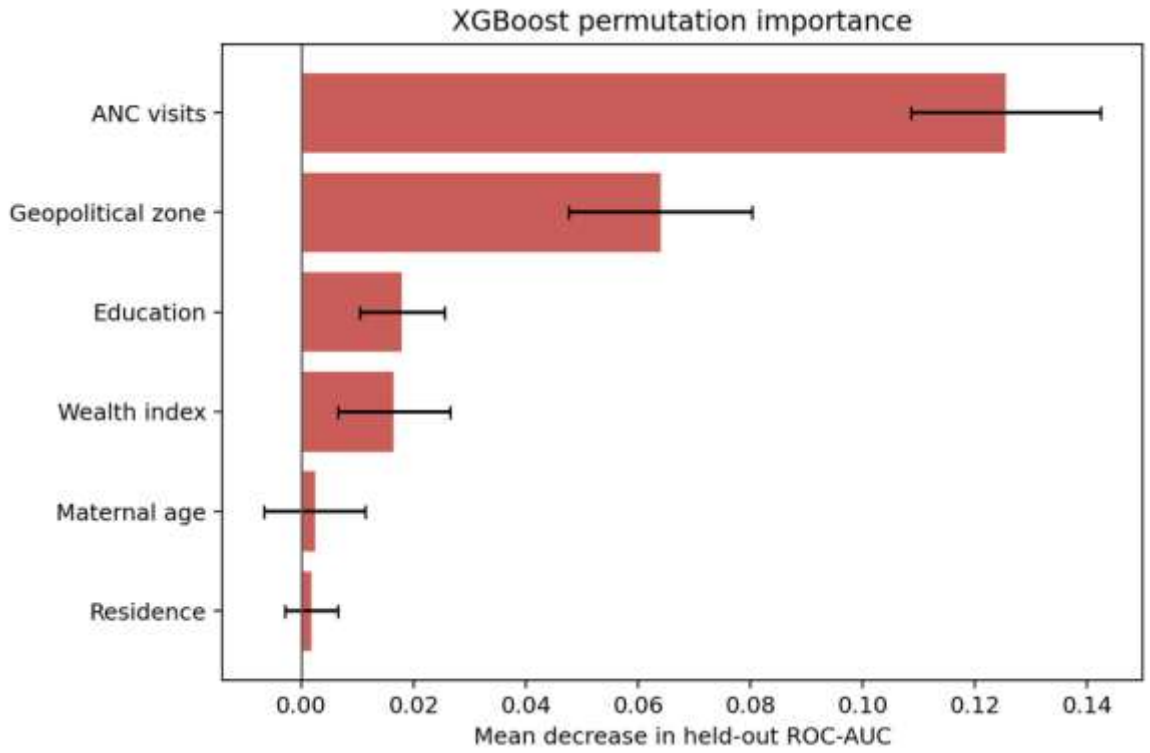

Figure 4. XGBoost permutation importance for the primary predictors.
The XGBoost importance profile also showed the largest model-specific contributions for ANC visits and geopolitical zone. Education and wealth contributed modestly, while residence and maternal age added little to discrimination in the fitted sample.

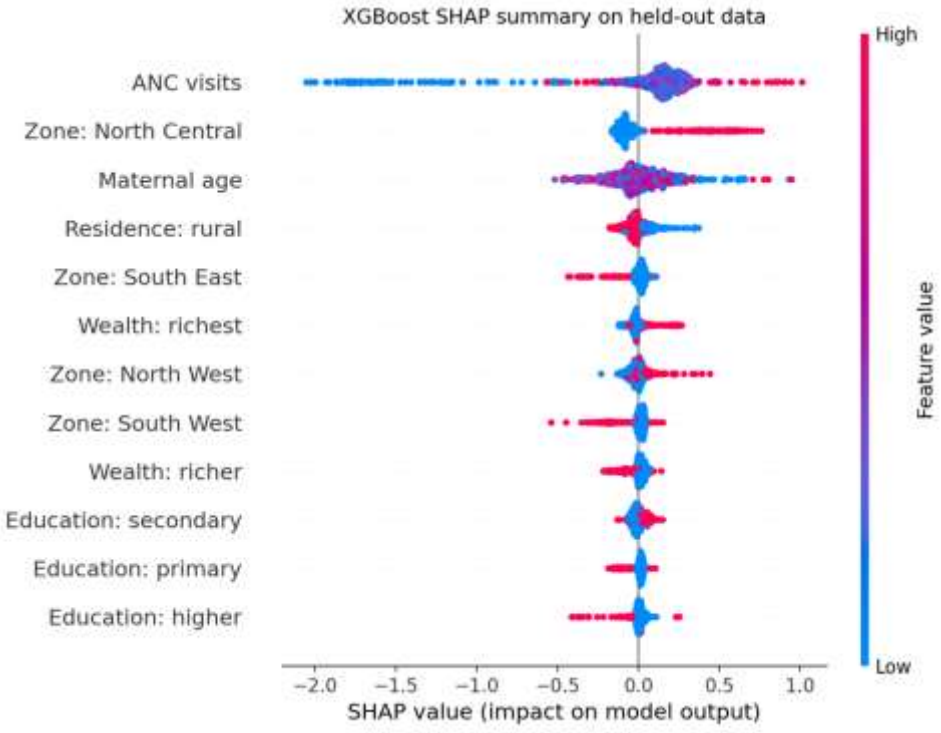

Figure 5. XGBoost SHAP summary plot for the primary predictors.
The SHAP plot provides directionality for the fitted model. Higher ANC-visit values generally shifted predictions towards optimal IPTp uptake, while lower values shifted predictions away from optimal uptake. This describes model behaviour and does not establish that increasing ANC visits would itself cause a change in uptake.

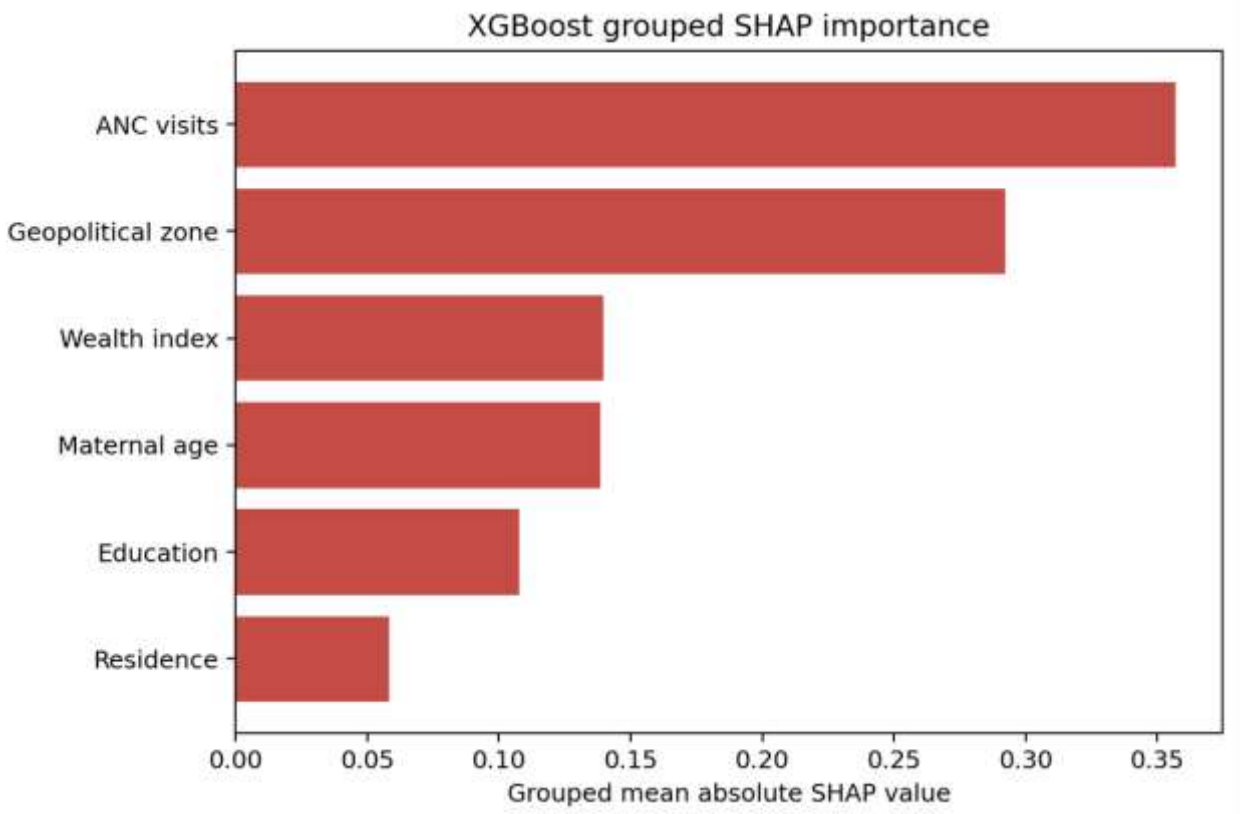

Figure 6. XGBoost mean absolute SHAP importance for the primary predictors.
Mean absolute SHAP values showed the largest model-specific contributions for ANC visits and geopolitical zone, followed by smaller contributions from sociodemographic variables. These values are associational summaries of the fitted unweighted model and should not be interpreted as population-level or causal importance.

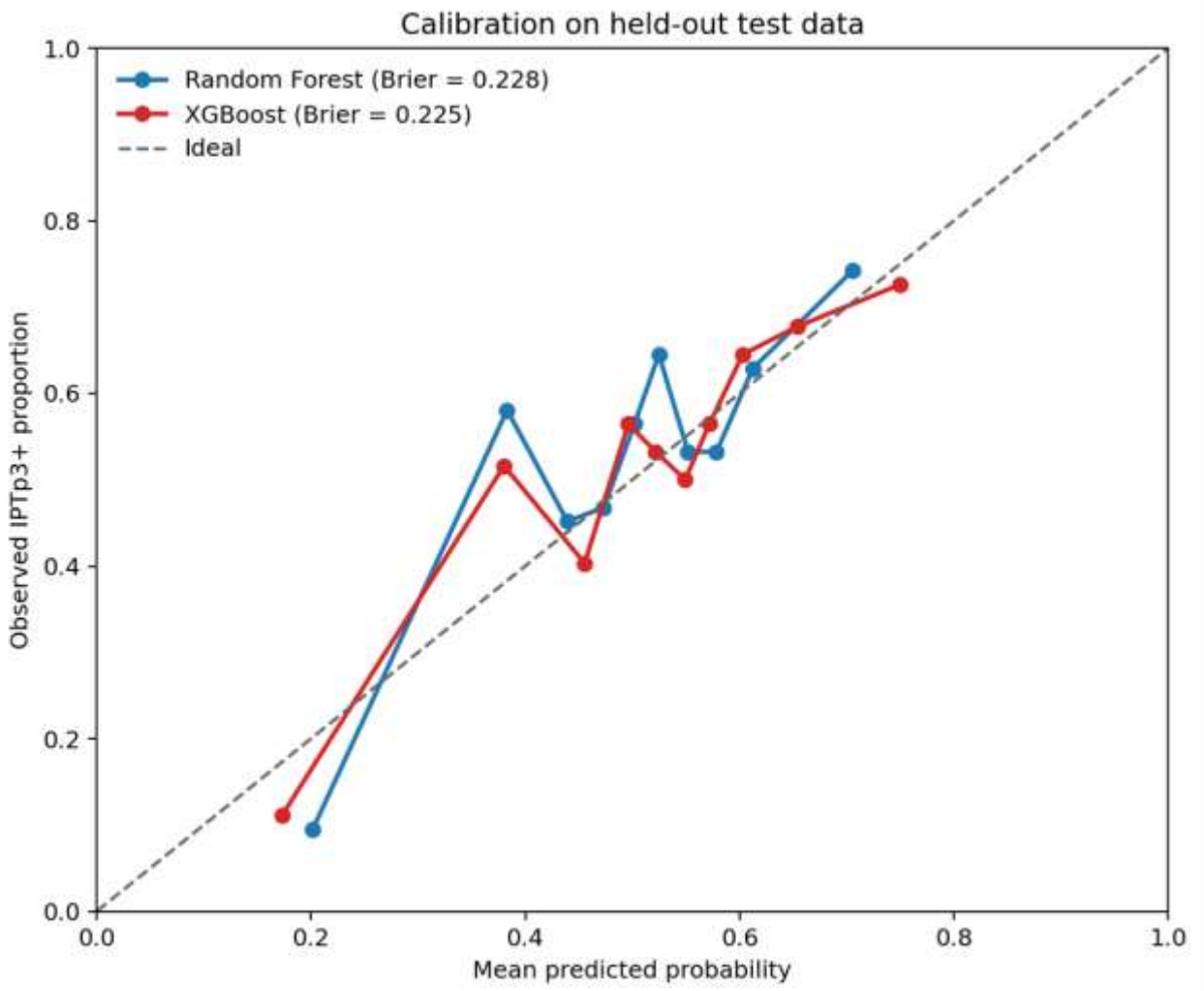

Figure 7. Calibration curves and Brier scores for the primary unweighted models on the held-out test set.

Brier scores were 0.228 for Random Forest and 0.225 for XGBoost. The quantile-binned calibration curves followed the ideal line only approximately and showed local deviations. These findings further indicate that the internally evaluated models are exploratory and are not ready for individual decision support.

Discussion

This study analysed 2021 Nigeria Malaria Indicator Survey data from 3,103 women to examine IPTp3+ uptake. Just over half (52.4%) of the unweighted sample reported IPTp3+ uptake; this is not a survey-weighted national estimate. In the survey-adjusted model, ANC category and geopolitical zone were associated with uptake, whereas maternal age, residence, education, and household wealth were not independently associated. The ANC association was nonlinear, with similar adjusted odds for 4–7 and ≥8 visits. The unweighted machine-learning models showed a comparable pattern but only poor-to-moderate discrimination.

Women reporting 4–7 or ≥8 ANC visits had about twice the adjusted odds of IPTp3+ uptake compared with those reporting 0–3 visits. Routine ANC provides opportunities for counselling, malaria-prevention education, directly observed therapy, and IPTp-SP administration [18–20]. However, ANC attendance may also reflect access to, continuity of, and engagement with health services. The cross-sectional design does

not establish that additional visits alone cause higher uptake; service content and readiness remain important [11].

Maternal education, household wealth, and residence were not significant after adjustment, possibly because ANC attendance and geopolitical zone partly captured their effects. This does not imply absent socioeconomic inequalities. Nigerian studies document socioeconomic, geographic, and service-related differences in IPTp uptake [8,21,22]. Poorer women may face delayed ANC initiation, inconsistent attendance, distance and cost barriers, or limited access to quality services [10,22,23,24]. Socioeconomic disadvantage may therefore operate indirectly through care access, making ANC attendance an access-related correlate rather than solely individual behaviour.

Geopolitical zone remained associated with uptake, with lower adjusted odds in the North East, North West, South East, and South West than North Central. The South South estimate was lower, although its confidence interval included the null. These differences may reflect facility readiness, sulfadoxine–pyrimethamine distribution, staffing, provider adherence, ANC organisation, security constraints, or cultural patterns in ANC timing and use [11,15,21,22,25]. Because NMIS could not test these explanations, the pattern should not be attributed to one factor. Region-specific linked household–facility studies are needed.

Machine learning added limited epidemiological information beyond regression and is best viewed as a methodological comparison. Random Forest and XGBoost achieved ROC-AUCs of 0.640–0.661, Brier scores of 0.225–0.228, and imperfect calibration. ANC visits and zone made the largest model-specific contributions, but SHAP and permutation importance do not estimate causal effects or identify policy priorities. Modest performance suggests that facility readiness, drug availability, provider adherence, dose timing, and ANC quality may explain additional variation [11]. Survey-aware external validation is required before predictive use. These findings have implications for malaria control and maternal health programmes in Nigeria. Policy decisions should prioritise survey-weighted associations, WHO guidance, and broader epidemiological and health-systems evidence, while treating the exploratory, unweighted machine-learning results as supplementary. Improving IPTp3+ coverage requires timely ANC access, IPTp-SP delivery at every eligible contact, reliable drug supplies, provider adherence, quality ANC services, and investigation of regional implementation gaps [11,17,25,26]. Local assessments of supply chains, service readiness, security, and community barriers should guide intervention priorities.

The strengths include the large national sample, survey-adjusted regression, assessment of ANC nonlinearity and multicollinearity, and comparison with exploratory machine learning. Limitations include the cross-sectional design, possible recall and dose misclassification, and unavailable data on ANC timing, drug stocks, provider adherence, facility readiness, security, and community norms. ANC categories should not be interpreted as causal thresholds. The unweighted machine-learning models had modest discrimination, single-split internal validation, and no external validation; therefore, their sample-specific outputs are unsuitable for population inference, causal ranking, clinical use, or policy prioritisation. Zone-level estimates may also conceal local variation.

Conclusion

Only about half of women achieved optimal IPTp uptake. ANC attendance and geographic differences were strongly associated with uptake, while age, education, wealth, and residence were not independently associated. Machine-learning models added limited epidemiological value. Strengthening timely ANC, IPTp delivery, drug supply, and provider adherence is recommended. Future studies should incorporate facility data and survey-aware predictive methods.

Acknowledgements
We gratefully acknowledge the National Malaria Elimination Programme, the National Population Commission, ICF, and the DHS Program for conducting the 2021 Nigeria Malaria Indicator Survey and for granting authorised access to the dataset used in this study.

Funding
This study received no specific funding.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
The dataset used in this study is available from the DHS Program following registration and approval through the DHS Program data access platform. The data are anonymised and were accessed with permission from the DHS Program/ICF. The 2021 Nigeria Malaria Indicator Survey dataset can be accessed at: https://dhsprogram.com/data/dataset/Nigeria_MIS_2021.cfm?flag=1

References

1. World Health Organization. *World malaria report 2025*. Geneva: World Health Organization; 2025. Available from: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2025
2. Mukhtar A, Ismail M, Usman A, Odetokun IA, Ibrahim MA, Salman AA, Bawa-Sani HG, Nass SS, Goni BW, Shehu MN, Nuhu A, Zhang X, Isah MB. Prevalence and associated risk factors of asymptomatic malaria in Nigeria: a systematic review and meta-analysis. *Malar J*. 2025;25(1):71. doi:10.1186/s12936-025-05671-5. PMID: 41476218; PMCID: PMC12866038. https://pubmed.ncbi.nlm.nih.gov/41476218/
3. Dambazau ZB, Cyril-Egware K, Charles O, Eshikhena G, Adegbite G, Akoma D, et al. Malaria transmission dynamics and the impact of treatment interventions in Kano State, Nigeria: a compartmental model approach. *PAMJ-One Health*. 2025;18(11). doi:10.11604/pamj-oh.2025.18.11.48342. Available from: https://www.one-health.panafrican-med-journal.com/content/article/18/11/ful
4. World Health Organization. Intermittent preventative treatment to reduce the risk of malaria during pregnancy [Internet]. Geneva: World Health Organization; 2023 [cited 2026 Jun 3]. Available from: https://www.who.int/tools/elena/interventions/iptp-pregnancy
5. Liu E, Partap U, Shinde S, Wang D, Costa JC, Cliffer IR, et al. Effect of intermittent preventive treatment during pregnancy with sulfadoxine-pyrimethamine on maternal gestational weight gain in low-income and middle-income countries: a systematic review and individual participant data meta-analysis of randomised clinical trials. EClinicalMedicine. 2025;84:103279. https://doi.org/10.1016/j.eclinm.2025.103279
6. Chukwuanukwu RC, Eleje GU, Menkiti FE, Ikpeze GC, Mbachu CN, Obiegbu NP, et al. Intermittent preventive treatment, malaria, HIV, hepatitis and pregnancy outcomes in Nigerian women: a cross-sectional study in two healthcare facilities. BMC Pregnancy Childbirth. 2025;25(1):809. https://doi.org/10.1186/s12884-025-07943-z
7. Adegbola AJ, Ogboye RM, Ijarotimi OA, Ubom AE, Adesoji BA, Bolaji OO. Assessment of uptake of sulphadoxine-pyrimethamine for intermittent preventive treatment among pregnant women in Osun State, Nigeria. *Trans R Soc Trop Med Hyg*. 2025;119(1):77–84. doi:10.1093/trstmh/trae076. PMID: 39397751
8. Kalu GO, Francis JM, Ibisomi L, Chirwa T, Kagura J. Factors associated with the uptake of intermittent preventive treatment (IPTp-SP) for malaria in pregnancy: Further analysis of the 2018 Nigeria Demographic and Health Survey. PLoS Glob Public Health. 2023;3(2):e0000771. doi:10.1371/journal.pgph.0000771.

9. Adeyemo MOA, Adeniran GO, Adeniyi VA, Olabisi EO, Oyekale RA, Akinwale O, et al. Assessment of uptake of intermittent preventive treatment of malaria among pregnant women attending antenatal clinic in public health facilities in Osogbo metropolis, Osun State, Nigeria. Int J Afr Nurs Sci. 2024;20:100742. https://doi.org/10.1016/j.ijans.2024.100742
10. Alain DC, Cédric Y, Nadia NAC, Tabah EN, Ezéchiel D, Hu H. Obstacles to the use of intermittent preventive treatment and insecticide-treated nets for malaria prevention among pregnant women in the N'Djamena South Health District. *Discov Public Health*. 2026;23:23. doi:10.1186/s12982-025-01326-7.
11. Xu X, Ameyaw EK, Ahinkorah BO, Seidu AA, Yaya S, Bishwajit G. The readiness of malaria services and uptake of intermittent preventive treatment of malaria during pregnancy in sub-Saharan Africa. J Glob Health. 2024;14:04112. https://doi.org/10.7189/jogh.14.04112
12. Andersen RM. Revisiting the behavioral model and access to medical care: Does it matter? *J Health Soc Behav*. 1995;36(1):1–10. https://doi.org/10.2307/2137284
13. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems*. 2017;30. Available from: https://arxiv.org/abs/1705.07874
14. Sadeghi Z, Alizadehsani R, Cifci MA, Kausar S, Rehman R, Mahanta P, et al. A review of explainable artificial intelligence in healthcare. *Comput Comput Electr Eng*. 2024;118:109370. https://doi.org/10.1016/j.compeleceng.2024.109370
15. National Malaria Elimination Programme (NMEP) [Nigeria], National Population Commission (NPC) [Nigeria], and ICF. Nigeria Malaria Indicator Survey 2021 Final Report. Abuja, Nigeria, and Rockville, Maryland, USA: NMEP, NPC, and ICF; 2022.
16. Toth D, Eltinge JL. Building consistent regression trees from complex sample data. J Am Stat Assoc. 2011;106(496):1626–36. https://doi.org/10.1198/jasa.2011.tm10383
17. Nalenz M, Rodemann J, Augustin T. Learning de-biased regression trees and forests from complex samples. Mach Learn. 2024;113:3379–3398. https://doi.org/10.1007/s10994-023-06439-1
18. Bayuo FB, Antwi EO, Adu-Boahen N, Effah EA, Appiah SK, Opoku HA, Efiong MB, Acheampomaa G, Senyah MM, Bofa A, Afoakwah G, Haile ZT. Association between adequacy of antenatal care and uptake of intermittent preventive treatment of malaria in pregnancy among women in Ghana. *Malar J*. 2025;24(1):385. doi:10.1186/s12936-025-05537-w. PMID: 41204273; PMCID: PMC12595622.
19. Hill J, Hoyt J, van Eijk AM, D'Mello-Guyett L, ter Kuile FO, Steketee R, Smith H, Webster J. Factors affecting the delivery, access, and use of interventions to prevent malaria in pregnancy in Sub-Saharan Africa: A systematic review and meta-analysis. PLoS Med. 2013;10(7):e1001488. https://doi.org/10.1371/journal.pmed.1001488
20. Agbi DM, Mwebesa E, Jimmy AI, Valère SK, Kerich GK. Impact of eight or more antenatal care visits on intermittent preventive treatment of malaria uptake during pregnancy and facility-based delivery in Ghana: a propensity score-matched analysis. BMC Pregnancy Childbirth. 2026;26:351.
21. Adeojo OS, Fayehun O. Determinants of the optimal uptake of sulfadoxine-pyrimethamine doses for intermittent treatment of malaria in pregnancy in urban Nigeria. Sci Rep. 2024;14:19097. doi:10.1038/s41598-024-52927-2.
22. Ladan SM, Lima NP, Orimadegun AE, Wado YD, Akinyemi JO, Jatau AA, et al. Prevalence and inequalities of missed opportunities for intermittent preventive therapy (IPTp) of malaria among pregnant women: an analysis of 2018 Nigeria DHS. Malar J. 2026;25:201. https://doi.org/10.1186/s12936-026-05840-0
23. Owembabazi M, Mpamize H, Naggulu IP, Nalubega JF. Socio-economic factors and health facility-related factors affecting the utilization of ANC services among women aged 18–40 attending ANC at Kyankwanzi Health Center III: A cross-sectional study. Student J Health Res Afr. 2025;6(3):10. doi:10.51168/sjhrafrica.v6i3.1549.
24. Bachan EG, Abu J, Adokiya MN, Tobin-West C. Effectiveness of malaria control interventions in resource-limited countries: A systematic review. *J Interv Epidemiol Public Health*. 2025;8(3):74.

25. Salihu T, Boakye K, Ayerakwah PA, Dorgbetor CI, Owusu-Ansah E, Boateng G, et al. What has recommended antenatal care utilisation got to do with optimal intake of IPTp-SP among rural women aged 15–49 in Nigeria? A population-based survey. Research Square. 2022. https://doi.org/10.21203/rs.3.rs-1965954/v1
26. World Health Organization. WHO policy brief for the implementation of intermittent preventive treatment of malaria in pregnancy using sulfadoxine-pyrimethamine (IPTp-SP). Geneva: World Health Organization; 2014. Available from: https://www.who.int/publications/i/item/WHO-HTM-GMP-2014.4