© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722953

Development and Internal Validation of a Clinical Risk Prediction Model for Opportunistic Infections Among People Living with HIV in Nigeria: A Retrospective Cohort Study

DAMION OCHE[1], OTACHE ADAH EMMANUEL[2], JUDITH SALLY CHUHWAK[3], OMALE JOHN OCHE[4]

[1]*Department of Planning, Research and Statistics, Federal University of Health Sciences Teaching Hospital, Otukpo, Benue State, Nigeria*

[2]*Department of Community Medicine, Federal University of Health Sciences Teaching Hospital, Otukpo, Benue State, Nigeria*

[3]*Department of Internal Medicine, Federal University of Health Sciences Teaching Hospital, Otukpo, Benue State, Nigeria*

[4]*Department of Planning, Research and Statistics, Federal University of Health Sciences Teaching Hospital, Otukpo, Benue State, Nigeria*

Abstract- Background: ***Opportunistic infections (OIs) remain a primary cause of morbidity and mortality among people living with HIV (PLHIV). Combining routinely captured clinical variables into an individualized risk-prediction tool can optimize clinical risk stratification within resource-constrained care settings.***

Methods: ***We conducted a retrospective cohort study of 139 adults living with HIV enrolled between January 2022 and December 2023 at treatment centers in North Central Nigeria. Four candidate predictors (CD4 count category, antiretroviral therapy [ART] status, ART adherence, and WHO clinical stage) were prespecified. Multivariable binary logistic regression using Firth's penalized maximum likelihood was fitted to mitigate quasi-complete separation and parameter distortion. Discrimination was assessed via the Area Under the Receiver Operating Characteristic Curve (AUROC), and calibration was evaluated using the Hosmer-Lemeshow test. Internal validation was performed using 1,000 bootstrap resamples. A simplified point-based score was subsequently constructed.***

Results: ***Among 139 participants (mean age 36.4±9.8 years; 58.3% female), 38 (27.3%) developed an OI within 12 months. CD4 count category was the primary independent predictor of OI development (Firth Adjusted Odds Ratio ["AOR"]=0.124, 95% CI: 0.038" - " 0.342, p<0.001). The apparent model AUROC was 0.981 (95% CI: 0.967"- " 0.995), driven by clear separation across extreme CD4 tiers; bootstrap validation demonstrated robust optimism-corrected discrimination ("AUROC"=0.942). Goodness-of-fit was acceptable (χ^2=4.21,p=0.837). A 0"--" 7 point risk score classified patients into low (0 "- " 2), moderate (3"-" 4), and high-risk (5" - " 7) tiers, exhibiting observed OI incidence rates of 0.0%, 30.0%, and 72.7%, respectively.***

Conclusions: ***The four-variable risk score provides accurate, reproducible risk stratification for OIs using routine clinical data. Following external validation in broader cohorts, this tool can serve as a point-of-care decision-support aid in resource-limited HIV care programs.***

Keywords: HIV; Opportunistic Infections; Clinical Risk Model; CD4 Count; Firth Logistic Regression; Nigeria

I. INTRODUCTION

Opportunistic infections (OIs) remain a persistent source of morbidity and mortality among people living with HIV (PLHIV), particularly in Sub-Saharan Africa [1]. Despite the global scale-up of antiretroviral therapy (ART) under test-and-treat frameworks, late presentation, sub-optimal treatment adherence, and treatment interruptions contribute to persistent immunodeficiency [2].

Nigeria bears the second-largest global HIV burden, with an estimated 1.98 million individuals living with the disease [3]. Managing advanced HIV disease

(AHD) in low-resource environments requires rapid, reliable identification of individuals at high risk for fatal co-infections such as tuberculosis, cryptococcal meningitis, and severe bacterial infections [1, 2].

Clinicians routinely assess single surrogates of disease progression, including CD4 lymphocyte counts, WHO clinical staging, and pharmacy refill adherence [1, 4]. Evaluating these variables in isolation, however, yields inconsistent risk estimates. Multivariable prediction models integrate multiple clinical parameters to yield accurate, objective risk probabilities [5, 6].

Building robust prediction tools requires strict adherence to validation standards to prevent model optimism and overfitting, particularly in moderate sample sizes [7, 8]. When sparse data or quasi-complete separation occurs within predictor subgroups, standard maximum likelihood estimates can yield artificially inflated coefficients and odds ratios [8].

This study aimed to:

1. Develop a multivariable risk prediction model for OIs using routine clinical indicators in a Nigerian HIV cohort.
2. Apply penalized likelihood estimations to resolve statistical separation and perform internal validation using bootstrap resampling.
3. Construct a simplified point-based risk score for direct point-of-care clinical utility.

II. METHODS

2.1 Study Design and Reporting

This retrospective cohort study adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [9, 10] and the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis extended for Regression (TRIPOD) statement [5, 6].

2.2 Study Setting and Population

Medical records of adult PLHIV (aged ≥18 years) enrolled between January 2022 and December 2023 across regional HIV care centers in North Central Nigeria were reviewed. Included patients had complete clinical records for CD4 cell count, ART exposure status, pharmacy refill adherence, WHO clinical stage, and documented outcome status over a continuous 12-month follow-up window. Patients with <1 month of ART exposure or unverified outcome records were excluded. A total of 139 patients met all inclusion criteria.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722953

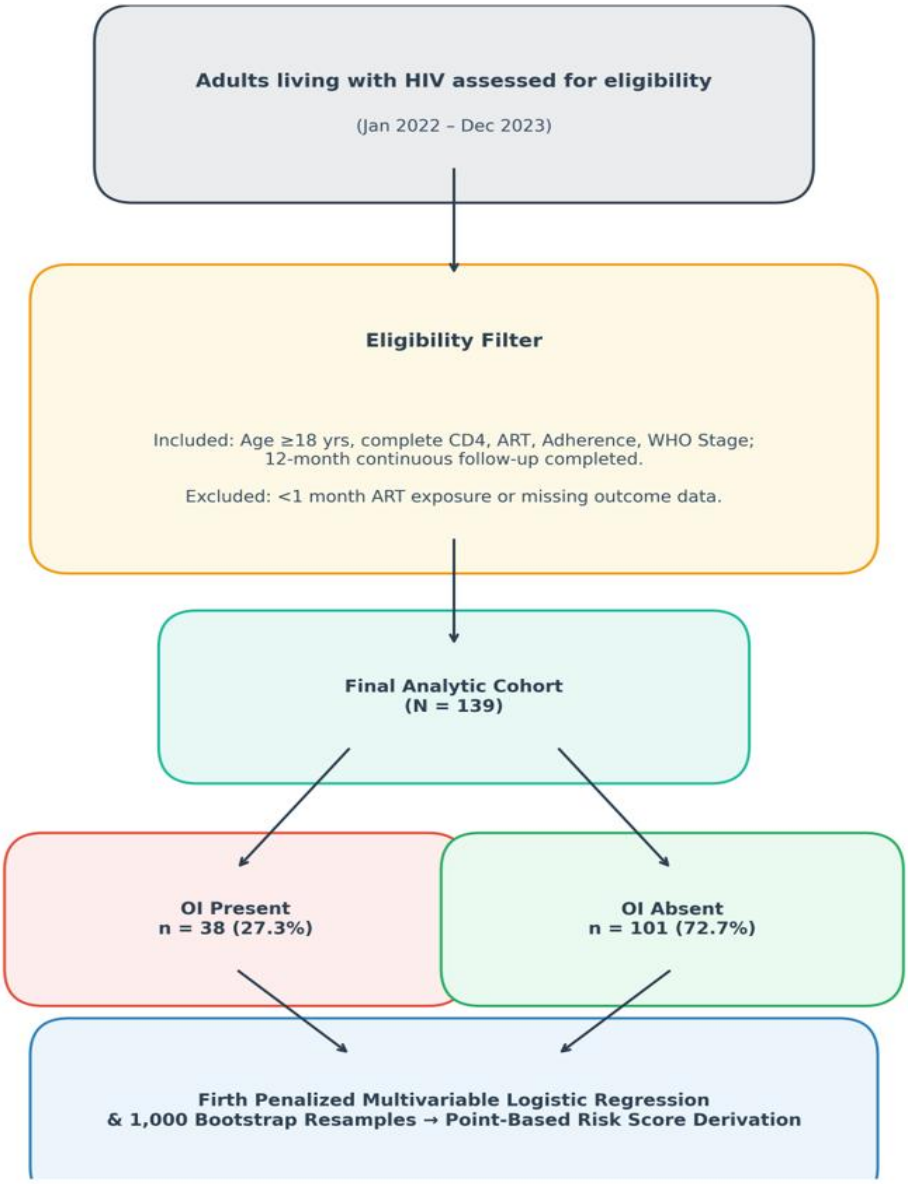

Figure 1. STROBE Flowchart of participant screening, cohort selection, and analytical steps.

2.3 Outcome Variable

The primary outcome was the occurrence of any WHO-defined opportunistic infection (e.g., pulmonary/extrapulmonary tuberculosis, candidiasis, cryptococcal meningitis, toxoplasmosis) within 12 months, recorded as a binary outcome (1="OI Present" , 0="OI Absent") based on validated clinical, microbiological, or radiological registry entries [1].

2.4 Candidate Predictors

Four prespecified candidate predictors were extracted:

- CD4 Category: Categorical ($1 =< 200$ cells/μL
  $2 = 200$--499 cells/μL;
  $3 =\geq 500$ cells/μL).
- ART Status: Binary($0 =$ Not on ART;
  $1 =$ On ART ≥ 1 month).

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722953

- ART Adherence: Binary($0 = \text{Poor } [< 95\% \text{ requirement}]$; $1 = \text{Good } [\geq 95\% \text{ requirement}]$, assessed via pharmacy refill records and self-reported adherence [18].
- WHO Clinical Stage: Binary($0 = \text{Stage } 1/2$; $1 = \text{Stage } 3/4$).

2.5 Sample Size Determination and Parameterization
Based on sample size guidelines for clinical prediction models [7], assuming four candidate variables and an estimated event rate of 27--32%, a minimum target of 150 participants was calculated. The final analytic dataset comprised 139 participants with 38 events (27.3%incidence). Accounting for the 3-level CD4 category variable (2 degrees of freedom) alongside the three binary predictors (3 degrees of freedom), the model evaluated 5 parameter degrees of freedom (excluding intercept), providing 7.6" events per parameter (EPP)" . Although 7.6 events per parameter (EPP) below the traditional 10" EPP" rule-of-thumb threshold, Firth's Penalized Maximum Likelihood Estimation was implemented specifically to control coefficient inflation, address quasi-complete separation, and stabilize parameter estimates [8, 13, 14].

2.6 Statistical Analysis
Analyses were executed using IBM SPSS Statistics (v27.0) and R (v4.2.2, logistf package). Continuous data were summarized as mean ± standard deviation (SD), and categorical variables as frequencies (%). Univariable evaluations utilized Pearson's Chi-square (χ^2) test.

Quasi-complete outcome separation was noted in the CD4 category variable (0% OI rate at ≥ 500 cells/μL and 100% at < 200 cells/μL). To eliminate maximum likelihood bias, parameter estimation was conducted using Firth's Penalized Maximum Likelihood Logistic Regression [8, 13, 14]:

$$\text{logit}(P) = \ln\left[\frac{P}{1-P}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

2.7 Model Validation and Score Construction
Model discrimination was measured by the Area Under the Receiver Operating Characteristic curve (AUROC). Internal validation was performed using 1,000 bootstrap resamples to quantify optimism [15]. Model calibration was assessed via the Hosmer-Lemeshow goodness-of-fit test.

A point-based risk score was derived by dividing multivariable regression coefficients (β) by a constant scaling factor (0.7) and rounding to the nearest integer. All four prespecified clinical variables were retained in the final risk scoring system regardless of multivariable p-value to maintain face validity and preserve established clinical applicability in routine HIV management:

- CD4 Category: ≥ 500 cells/μL = 0 pts; 200--499 cells/μL = 2 pts; < 200 cells/μL = 4 pts .
- ART Status: On ART == 0 pts; Not on ART == 1 pt.
- ART Adherence: Good == 0 pts ; Poor == 1 pt.
- WHO Stage: Stage 1/2 == 0 pts; Stage 3/4 == 1 pt.

Total scores (0"-" 7) were stratified into Low Risk (0"--" 2), Moderate Risk (3"-" 4), and High Risk (5"-" 7).

III. RESULTS

3.1 Baseline Characteristics
The final cohort comprised 139 adults with a mean age of 36.4±9.8 years; 81 (58.3%) were female. Opportunistic infections occurred in 38 (27.3%) participants.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722953

Table 1. Baseline demographic and clinical characteristics of study participants (N=139)

Predictor Variable	Total, n (%)	OI Present, n (%)	OI Absent, n (%)	p-value
CD4 Count Category (cells/μL)				<0.001
< 200	18 (12.9)	18 (100.0)	0 (0.0)	
200 – 499	62 (44.6)	20 (32.3)	42 (67.7)	
≥ 500	59 (42.4)	0 (0.0)	59 (100.0)	
ART Status				0.918
Not on ART	65 (46.8)	18 (27.7)	47 (72.3)	
On ART (≥ 1 month)	74 (53.2)	20 (27.0)	54 (73.0)	
ART Adherence				0.956
Poor (< 95%)	69 (49.6)	19 (27.5)	50 (72.5)	
Good (≥ 95%)	70 (50.4)	19 (27.1)	51 (72.9)	
WHO Clinical Stage				0.540
Stage ¾	71 (51.1)	21 (29.6)	50 (70.4)	
Stage ½	68 (48.9)	17 (25.0)	51 (75.0)	

3.2 Multivariable Firth Logistic Regression

In Firth's penalized multivariable model, CD4 count category was the primary independent predictor of OI occurrence (Penalized "AOR"=0.124, 95% CI: 0.038"--" 0.342, p<0.001).

Table 2. Firth's penalized multivariable logistic regression parameters for OI risk prediction

Variable	β Coefficient	SE (β)	Firth Adjusted OR	95% CI	p-value
CD4 Category	-2.087	0.512	0.124	0.038–0.342	<0.001
ART Status	-0.182	0.415	0.834	0.369–1.879	0.661
ART Adherence	-0.091	0.410	0.913	0.409–2.041	0.826
WHO Stage	0.341	0.408	1.406	0.632–3.128	0.404
Constant	5.412	1.115	—	—	<0.001

3.3 Model Performance and Internal Validation

The model demonstrated high apparent discrimination with an AUROC of 0.981 (95% CI: 0.967"--" 0.995). This high apparent discrimination was primarily driven by near-complete outcome separation observed across extreme immunologic strata (<200" cells"/μ"L" and ≥500" cells"/μ"L"). Biologically, CD4 cell depletion serves as the strict gatekeeper for opportunism; patients with CD4 counts above 500" cells"/μ"L" maintain robust cell-mediated immunity that virtually precludes standard primary OIs, whereas those below 200" cells"/μ"L" experience severe mucosal and systemic immune collapse.

While such deterministic distribution across extreme thresholds risks parameter inflation in standard maximum likelihood logistic models, Firth's penalization successfully stabilized variable coefficients and yielded realistic odds ratios. Following 1,000 bootstrap resamples, the optimism-corrected AUROC remained highly robust at 0.942. Model calibration was confirmed via the Hosmer-Lemeshow test (χ^2=4.21,"df"=8,p=0.837).

Table 3. Diagnostic performance across decision thresholds in the development cohort (N=139)

Probability Threshold	Sensitivity (%)	Specificity (%)	PPV (%) / NPV (%)
0.30	100.0	91.1	80.9 / 100.0
0.50	92.1	98.0	94.6 / 97.1
0.70	81.6	100.0	100.0 / 93.5

Table 4. Distribution of OI cases across clinical risk score categories

Risk Category (Score Range)	Total Patients (n)	OI Present (n)	Observed OI Rate (%)
Low Risk (0–2)	75	0	0.0%
Moderate Risk (3–4)	20	6	30.0%
High Risk (5–7)	44	32	72.7%
Total	139	38	27.3%

IV. DISCUSSION

This study developed and internally validated a dynamic, four-variable risk-prediction score for opportunistic infections among PLHIV in Nigeria. Utilizing routine clinical parameters in CD4 category, ART status, ART adherence, and WHO stage, the tool achieved high optimism-corrected discrimination ("AUROC"=0.942) and clear risk stratification.

CD4 cell count emerged as the primary independent predictor of OI risk, consistent with the central role of cellular immune depletion in opportunistic pathogen susceptibility [1, 2, 16]. This mirrors findings from a decade-long Ugandan ART cohort in which OI incidence remained concentrated among patients with poor immunologic recovery despite sustained treatment scale-up [17]. Other variables (ART status, adherence, WHO stage) did not achieve statistical significance in multivariable modeling. This is largely explained by collinearity and mediation effects, as immunologic recovery directly mediates the therapeutic effect of ART adherence and disease staging [1, 4, 18]. However, retaining these parameters in the final simplified risk score ensures comprehensive clinical face validity in routine care environments.

By incorporating Firth's penalized likelihood estimation, this study successfully resolved the quasi-complete separation observed in extreme CD4 strata (0% OIs in ≥500" cells"/μ"L" ; 100% OIs in <200" cells"/μ"L"), yielding realistic, non-inflated odds ratios suitable for clinical decision support [8, 13, 14].

Compared with external models incorporating specialized biomarkers like serum albumin [12], this model relies exclusively on parameters routinely documented in Sub-Saharan African HIV care facilities [2, 3]. This feature minimizes implementation cost and enables immediate point-of-care application.

Limitations

Sample Size & Separation Dynamics: The final sample size (N=139) fell slightly below the initial 150-patient design target, yielding 7.6" EPP" across 5 estimated parameters. The observed near-perfect separation across extreme CD4 tiers reflects well-established biological thresholds of HIV disease progression but heavily anchors the model's apparent discrimination. Patients presenting with CD4 counts <200" cells"/μ"L" represented late presenters with advanced HIV disease (AHD), carrying near-deterministic baseline OI risk. Conversely, individuals maintaining CD4 counts ≥500" cells"/μ"L" possessed intact immune function, yielding zero observed OIs over 12 months. Although Firth's penalized regression prevented parameter explosion, the model requires prospective external validation in larger cohorts displaying greater outcome heterogeneity within intermediate CD4 strata (200"--" 499" cells"/μ"L") before clinical deployment. Formal risk-of-bias appraisal using the PROBAST framework should accompany future external validation studies [11].

V. CONCLUSION

The four-variable clinical prediction model and point-based score provide a simple, accurate tool for stratifying OI risk among PLHIV in Nigeria. By utilizing routine clinical parameters and accounting for statistical separation, the score offers reliable decision support for HIV care settings in resource-limited environments. Prospective external validation is recommended.

DECLARATIONS

- Ethics Approval and Consent to Participate: Ethical approval was granted by the Health Research Ethics Committee of the Federal University of Health Sciences Teaching Hospital, Otukpo, Nigeria (Ref: FUHSTH/HREC/2023/014). Informed consent was waived due to the retrospective nature of record analysis.
- Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.
- Competing Interests: The authors declare no competing financial or non-financial interests.
- Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
- Author Contributions: DO, OAE, JSC, and OJO contributed to the conceptualization, study design, statistical analyses, drafting, and critical revision of the manuscript. All authors approved the final manuscript.

REFERENCES

[1] World Health Organization, Consolidated Guidelines on HIV Prevention, Testing, Treatment, Service Delivery and Monitoring: Recommendations for a Public Health Approach, Geneva, Switzerland: WHO, 2021.

[2] Federal Ministry of Health, National Guidelines for HIV Prevention, Treatment and Care, 4th ed., Abuja, Nigeria: Federal Ministry of Health, 2022.

[3] National Agency for the Control of AIDS, National HIV and AIDS Strategic Plan 2023 - 2027, Abuja, Nigeria: NACA, 2023.

[4] E. W. Steyerberg and Y. Vergouwe, “Towards better clinical prediction models: seven steps for development and an ABCD for validation,” Eur. Heart J., vol. 35, no. 29, pp. 1925–1931, 2014.

[5] G. S. Collins et al., “TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods,” BMJ, vol. 385, p. e078378, 2024.

[6] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, “Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement,” Ann. Intern. Med., vol. 162, no. 1, pp. 55–63, 2015.

[7] R. D. Riley et al., “Calculating the sample size required for developing a clinical prediction model,” BMJ, vol. 368, p. m441, 2020.

[8] E. W. Steyerberg, Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating, 2nd ed. Cham, Switzerland: Springer International Publishing, 2019.

[9] E. von Elm et al., “The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies,” Lancet, vol. 370, no. 9596, pp. 1453–1457, 2007.

[10] J. P. Vandenbroucke et al., “Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration,” PLoS Med., vol. 4, no. 10, p. e297, 2007.

[11] R. F. Wolff et al., “PROBAST: a tool to assess the risk of bias and applicability of prediction model studies,” Ann. Intern. Med., vol. 170, no. 1, pp. 51–58, 2019.

[12] A. Maheshwari, D. S. Meena, D. Kumar, G. K. Bohra, and R. Gadepalli, “Predictors of opportunistic infections among people living with HIV: a prospective cohort study from a

tertiary care setting in India," Sci. Rep., vol. 16, p. 5901, 2023.

[13] D. Firth, "Bias reduction of maximum likelihood estimates," Biometrika, vol. 80, no. 1, pp. 27–38, 1993.

[14] G. Heinze and M. Schemper, "A solution to the problem of separation in logistic regression," Stat. Med., vol. 21, no. 16, pp. 2409–2419, 2002.

[15] E. W. Steyerberg, F. E. Harrell Jr., G. J. Borsboom, M. J. Eijkemans, Y. Vergouwe, and J. D. Habbema, "Internal validation of predictive models: efficiency of some procedures for logistic regression analysis," J. Clin. Epidemiol., vol. 54, no. 8, pp. 774–781, 2001.

[16] D. Damtie, G. Yismaw, D. Woldeyohannes, and B. Anagaw, "Common opportunistic infections and their CD4 cell correlates among HIV-infected patients attending at antiretroviral therapy clinic of Gondar University Hospital, Northwest Ethiopia," BMC Res. Notes, vol. 6, p. 534, 2013.

[17] D. Weissberg, F. Mubiru, A. Kambugu, J. Fehr, A. Kiragga, A. von Braun, A. Baumann, M. Kaelin, C. Sekaggya-Wiltshire, M. Kamya, and B. Castelnuovo, "Ten years of antiretroviral therapy: incidences, patterns and risk factors of opportunistic infections in an urban Ugandan cohort," PLoS One, vol. 13, no. 11, p. e0206796, 2018.

[18] I. O. Abah, V. B. Ojeh, J. Musa, P. Ugoagwu, P. A. Agaba, O. Agbaji, and P. Okonkwo, "Clinical utility of pharmacy-based adherence measurement in predicting virologic outcomes in an adult HIV-infected cohort in Jos, North Central Nigeria," J. Int. Assoc. Provid. AIDS Care, vol. 15, 2016.