Comparative Performance Analysis of Bayesian Hierarchical Models Versus Classical Statistical Approaches in Predicting Breast Cancer Treatment Outcomes: Evidence from Kenyan Healthcare Settings

MUHATI NELSON LWOYELO[1], VINCENT MARANI[2], RICHARD SIMWA[3]
[1,2] *Department of Mathematics, Kibabii University, Kenya*
[3] *School of Science and Engineering, Daystar University, Kenya*

Abstract- ***Current breast cancer treatment prediction models inadequately quantify uncertainty and fail to account for institutional clustering effects, particularly in resource-constrained healthcare settings. This study compared the performance of Bayesian hierarchical models against classical frequentist approaches for predicting pathological complete response (pCR) in breast cancer patients. We analyzed data from 5,400 breast cancer patients across 12 Kenyan treatment centers. Three progressively complex models were developed: single-level logistic regression (M0), Bayesian empty hierarchical model (M1), and Bayesian hierarchical model with clinical covariates (M2). Performance comparison utilized multiple metrics including Area Under the Curve (AUC), Brier Score, calibration measures, and information criteria. The Bayesian hierarchical model demonstrated superior performance with AUC = 0.837 compared to classical approaches (AUC = 0.752). Bayesian methods showed consistent 2-8 unit improvements in information criteria across all model complexity levels. The hierarchical structure captured 26.5% of outcome variation attributable to institutional clustering (ICC = 0.265), which classical models failed to address. Uncertainty quantification through credible intervals provided clinically meaningful prediction confidence assessment. Bayesian hierarchical approaches significantly outperform classical statistical methods in breast cancer treatment outcome prediction, particularly in settings with institutional clustering. The explicit uncertainty quantification and superior discrimination make Bayesian methods more suitable for clinical decision-making in resource-constrained environments.***

Indexed Terms- ***Bayesian Statistics, Breast Cancer, Treatment Outcomes, Model Comparison***

I. INTRODUCTION

Breast cancer represents the most prevalent malignancy among women globally, with approximately 2.3 million new cases diagnosed annually [1]. In sub-Saharan Africa, the disease disproportionately affects younger women and presents unique challenges due to resource constraints and healthcare infrastructure limitations [2]. Kenya, with 6,799 new breast cancer cases recorded in 2020, faces significant challenges in optimizing treatment outcomes through evidence-based clinical decision-making [3].

Current prediction models for breast cancer treatment outcomes exhibit critical limitations that compromise their clinical utility. Most existing approaches fail to adequately quantify uncertainty in treatment response predictions, providing point estimates without confidence bounds essential for informed clinical decision-making [4]. Additionally, these models typically ignore the hierarchical nature of healthcare data, where patients are clustered within treatment centers, leading to underestimated standard errors and poor generalizability across different healthcare settings [5].

The emergence of Bayesian statistical methods offers a paradigm shift in clinical prediction modeling. Unlike frequentist approaches, Bayesian methods naturally incorporate prior knowledge, handle uncertainty through probability distributions, and accommodate hierarchical data structures through partial pooling mechanisms [6]. These advantages are particularly relevant in resource-constrained settings where treatment decisions carry high stakes due to limited alternative options.

The Kenyan healthcare context presents unique challenges that traditional prediction models, predominantly developed using high-income country datasets, fail to address. Over 85% of existing breast cancer prediction models have been developed using Western populations, limiting their applicability to African settings [7]. The younger age distribution of breast cancer patients in Kenya (median age 45-50 years versus 60-65 years in Western countries) and the higher prevalence of aggressive molecular subtypes necessitate population-specific prediction approaches [8].

Institutional clustering effects are particularly pronounced in the Kenyan healthcare system, where variations in resource availability, specialist expertise, and treatment protocols across urban tertiary centers versus regional facilities create systematic differences in treatment outcomes [9]. These unmeasured institutional factors likely encompass clinical expertise levels, treatment protocol adherence, multidisciplinary team coordination, and healthcare infrastructure capabilities that systematically influence patient outcomes.

The objective of this study was to conduct a comprehensive comparison of Bayesian hierarchical modeling approaches against classical frequentist methods for predicting pathological complete response to neoadjuvant chemotherapy in Kenyan breast cancer patients. This comparison addresses a critical gap in clinical prediction methodology by demonstrating the superiority of Bayesian approaches in handling uncertainty quantification, institutional clustering, and clinical decision support in resource-constrained healthcare environments.

II. LITERATURE REVIEW

2.1 Evolution of Statistical Methods in Clinical Prediction

The field of clinical prediction modeling has undergone significant methodological evolution over the past two decades. Traditional frequentist approaches, including logistic regression and survival analysis, have dominated medical research due to their computational simplicity and established interpretation frameworks [10]. However, these methods exhibit fundamental limitations in clinical applications, particularly regarding uncertainty quantification and handling of complex data structures.

Recent systematic reviews have highlighted critical deficiencies in current prediction models. Collins et al. [11] demonstrated that fewer than 15% of published clinical prediction models provide meaningful uncertainty estimates for individual patient predictions. This limitation is particularly problematic in oncology, where treatment decisions carry significant consequences and uncertainty directly impacts risk-benefit assessments. Furthermore, validation practices remain methodologically flawed, with only 23% of models reporting calibration statistics and fewer than 10% performing comprehensive calibration assessment across patient subgroups [12].

2.2 Bayesian Methods in Medical Research

Bayesian statistical methods have gained increasing recognition in medical research due to their natural ability to incorporate prior knowledge and quantify uncertainty. The Bayesian framework provides a coherent probabilistic approach to statistical inference, enabling systematic integration of external evidence with observed data [13]. This capability is particularly valuable in resource-constrained settings where local research infrastructure may be limited, allowing leveraging of international evidence to inform locally relevant models.

Spiegelhalter [14] demonstrated that Bayesian approaches offer distinct advantages in clinical applications through their explicit quantification of uncertainty and natural incorporation of prior knowledge. The ability to provide credible intervals

for individual patient predictions enables clinicians to assess the reliability of model outputs and incorporate uncertainty into clinical decision-making processes. This contrasts sharply with frequentist confidence intervals, which provide information about the estimation procedure rather than the parameter of interest.

Recent advances in computational methods, particularly Markov Chain Monte Carlo (MCMC) techniques, have made sophisticated Bayesian models increasingly accessible to clinical researchers [15]. Modern software implementations, including Stan and PyMC, provide user-friendly interfaces for complex hierarchical models while automatically handling convergence diagnostics and uncertainty quantification.

2.3 Hierarchical Modeling in Healthcare Data

Healthcare data naturally exhibit hierarchical structures, with patients clustered within hospitals, repeated measurements within patients, and treatment protocols within institutions [16]. Failure to account for these structures leads to statistical and clinical problems, including underestimated standard errors, inflated Type I error rates, and poor generalizability across different healthcare settings.

Austin et al. [17] provided comprehensive guidance on when to use multilevel modeling in health services research, establishing that intraclass correlation coefficients exceeding 5% justify hierarchical approaches. In cancer care settings, institutional clustering effects are often substantial, reflecting variations in clinical expertise, treatment protocols, and healthcare infrastructure that systematically influence patient outcomes.

Bayesian hierarchical models offer optimal solutions for clustered clinical data through partial pooling mechanisms that automatically adjust information borrowing based on between-group similarity [18]. This approach enables improved estimation efficiency while maintaining appropriate uncertainty quantification, particularly valuable when some healthcare institutions have limited patient volumes.

2.4 Breast Cancer Prediction Models: Current State and Limitations

Existing breast cancer prediction models exhibit significant limitations that compromise their clinical utility in diverse healthcare settings. A systematic review by Wessels and van de Vijver [19] identified major deficiencies in current approaches, including inadequate external validation, poor calibration performance, and limited applicability to non-Western populations.

Population representativeness presents a critical limitation, with over 85% of existing models developed using high-income country datasets [20]. These models often perform poorly when applied to different populations due to variations in demographic characteristics, disease presentations, and healthcare systems. The gap is particularly acute for sub-Saharan African populations, who remain severely underrepresented in model development datasets despite unique disease characteristics and healthcare challenges [21].

Most current models fail to account for institutional clustering effects, treating all observations as independent despite clear hierarchical structures in healthcare delivery [22]. This oversight leads to biased parameter estimates and poor generalizability across different healthcare settings, limiting practical clinical implementation.

2.5 Comparative Studies: Bayesian versus Frequentist Approaches

Limited research has directly compared Bayesian and frequentist approaches in clinical prediction contexts. The available evidence consistently demonstrates Bayesian superiority across multiple dimensions, including predictive accuracy, uncertainty quantification, and clinical interpretability [23].

Berger [24] provided theoretical justification for Bayesian approaches in medical decision-making, emphasizing their natural incorporation of prior knowledge and coherent uncertainty quantification. Practical applications in oncology have demonstrated improved prediction accuracy and enhanced clinical decision support through explicit uncertainty estimation [25].

However, most comparative studies have been conducted in high-income country settings with well-established healthcare infrastructure. Limited evidence exists regarding Bayesian method performance in resource-constrained environments, where institutional clustering effects may be more pronounced and uncertainty quantification becomes even more critical for optimal resource allocation.

III. RESEARCH METHODOLOGY

3.1 Study Design and Setting

This retrospective cohort study was conducted across 12 major cancer treatment centers in Kenya, representing both academic medical centers and regional referral hospitals. The study design employed a systematic comparison of three progressively complex statistical modeling approaches to evaluate the performance advantages of Bayesian hierarchical methods over classical frequentist techniques.

3.2 Data Collection and Study Population

Electronic health records from 5,400 breast cancer patients receiving neoadjuvant chemotherapy between 2018 and 2023 were analyzed. Inclusion criteria encompassed female patients aged 18-80 years with histologically confirmed invasive breast carcinoma who completed at least four cycles of neoadjuvant chemotherapy. Exclusion criteria included patients with metastatic disease at diagnosis, incomplete treatment records, or missing pathological response assessments.

The primary outcome variable was pathological complete response (pCR), defined as the absence of residual invasive carcinoma in both breast and axillary lymph nodes following neoadjuvant chemotherapy, assessed according to standardized pathological protocols.

3.3 Model Development and Comparison Framework

Three statistical modeling approaches were systematically developed and compared:

Model M0 (Classical Single-level): Traditional logistic regression without hierarchical structure, treating all observations as independent and employing maximum likelihood estimation with asymptotic confidence intervals.

Model M1 (Bayesian Empty Hierarchical): Two-level Bayesian logistic regression with random intercepts for treatment centers but without patient-level covariates, enabling direct assessment of institutional clustering effects.

Model M2 (Bayesian Hierarchical with Covariates): Complete Bayesian hierarchical model incorporating both patient-level clinical covariates and center-specific random effects.

3.4 Bayesian Model Specification

The Bayesian hierarchical framework employed the following specification:

Level 1 (Patient): $y_{ij} \sim \text{Bernoulli}(\theta_{ij})$ Level 2 (Center): $\text{logit}(\theta_{ij}) = \beta_0 + \Sigma\beta_k x_{kij} + u_{0j}$

Where $u_{0j} \sim N(0, \sigma^2_{u0})$ represents center-specific random intercepts.

Prior distributions were specified based on established clinical knowledge and meta-analytic evidence: $\beta_0 \sim N(-0.85, 0.5^2)$, reflecting approximately 30% baseline pCR rates; established prognostic factors employed informative priors based on international literature; exploratory biomarkers received weakly informative priors $\beta \sim N(0, 0.5^2)$.

3.5 Computational Implementation

Bayesian inference was conducted using the No-U-Turn Sampler (NUTS) with four parallel chains, each running 2,000 warmup iterations followed by 2,000 sampling iterations. Convergence assessment employed multiple diagnostics including the Gelman-Rubin statistic ($\hat{R} < 1.01$), effective sample size calculations (ESS > 400), and visual inspection of trace plots.

Classical models were fitted using maximum likelihood estimation with robust standard errors to account for potential clustering, though these approaches cannot fully address hierarchical data structures.

3.6 Performance Evaluation Metrics

Modelcomparison employed multiple complementary evaluation approaches:

Discrimination Assessment: Area under the receiver operating characteristic curve (AUC) with 95% confidence intervals, sensitivity and specificity at

optimal thresholds, and positive/negative predictive values.

Calibration Assessment: Brier score decomposition, calibration slope and intercept, and Hosmer-Lemeshow goodness-of-fit tests with visualization through calibration plots.

Information Criteria: Akaike Information Criterion (AIC) for classical models, Deviance Information Criterion (DIC) and Widely Applicable Information Criterion (WAIC) for Bayesian approaches, and Leave-One-Out Cross-Validation (LOO-CV) for robust model comparison.

Clinical Utility: Decision curve analysis to quantify clinical benefit across threshold probabilities, net reclassification improvement, and integrated discrimination improvement.

3.7 Institutional Clustering Assessment

The extent of institutional clustering was quantified through the intraclass correlation coefficient (ICC):

$ICC = \sigma^2_{u0} / (\sigma^2_{u0} + \pi^2/3)$

Where σ^2_{u0} represents between-center variance and $\pi^2/3 \approx 3.29$ approximates within-center variance for logistic regression.

3.8 Statistical Analysis and Software

All Bayesian analyses were conducted using Stan via the R interface. Classical analyses employed standard R packages including glm for logistic regression and lme4 for mixed-effects models. Model validation utilized stratified 10-fold cross-validation with careful attention to maintaining hierarchical structure integrity.

IV. RESULTS

4.1 Study Population Characteristics

The study population comprised 5,400 breast cancer patients across 12 treatment centers, with a median age of 52 years (IQR: 45-61). Pathological complete response was achieved in 2,052 patients (38.0%), consistent with published international rates for neoadjuvant chemotherapy. The distribution showed 83.1% presenting with Stage II-III disease, reflecting typical patterns in resource-constrained settings with limited screening programs.

4.2 Institutional Clustering Assessment

The empty Bayesian hierarchical model (M1) revealed substantial institutional clustering, with an intraclass correlation coefficient of 26.5% (95% CrI: 4.3%-72.4%). This finding significantly exceeded conventional thresholds for multilevel modeling (ICC > 5%), providing compelling evidence for the necessity of hierarchical approaches in breast cancer outcome prediction.

The between-center variance ($\sigma^2_{u0} = 1.186$, 95% CrI: 0.148-8.650) demonstrated statistically significant variation across treatment centers, indicating systematic institutional factors influencing treatment outcomes beyond patient-level characteristics.

4.3 Model Performance Comparison

Table 1: Comprehensive Model Performance Comparison

Model	AUC	95% CI	Brier Score	Calibration Slope	AIC/DIC	LOO-CV
M0 (Classical)	0.752	0.731-0.773	0.198	0.89	368.4	367.1
M1 (Bayesian Empty)	0.683	0.661-0.705	0.247	0.92	339.9	341.2
M2 (Bayesian Hierarchical)	0.837	0.821-0.853	0.167	0.97	294.0	295.6

The Bayesian hierarchical model with covariates (M2) demonstrated superior performance across all evaluation metrics. The AUC improvement from 0.752 (classical) to 0.837 (Bayesian) represents a clinically meaningful enhancement in discrimination

ability, reaching the "excellent" classification threshold (AUC > 0.8).

4.4 Information Criteria Analysis

Bayesian approaches consistently outperformed classical methods across all information criteria. The progression from classical single-level (AIC = 368.4) to Bayesian hierarchical (DIC = 294.0) showed a dramatic 74.4-unit improvement, representing a 20.2% reduction in information criteria values.

Table 2: Model Improvement Metrics

Comparison	Absolute Improvement	Percentage Improvement
Classical to Bayesian Empty	28.5 units	7.7%
Empty to Full Hierarchical	45.9 units	13.5%
Classical to Full Bayesian	74.4 units	20.2%

4.5 Uncertainty Quantification Assessment

The Bayesian framework provided natural uncertainty quantification through credible intervals for all predictions. Unlike classical confidence intervals that describe estimation uncertainty, Bayesian credible intervals directly quantify parameter uncertainty, enabling more informed clinical decision-making.

For individual patient predictions, the Bayesian model generated prediction intervals that appropriately reflected underlying uncertainty. High-risk patients (predicted pCR probability > 60%) showed narrower credible intervals (typical width: 0.12-0.18), while intermediate-risk patients (30-60% predicted probability) exhibited wider intervals (typical width: 0.20-0.35), appropriately reflecting greater prediction uncertainty.

4.6 Clinical Covariate Effects

The Bayesian hierarchical model revealed several clinically significant associations:

- Tumor Stage: Progressive deterioration across stages (Stage IV OR: 24.38, 95% CrI: 6.60-93.11)
- Hormone Receptor Status: Reduced response probability (OR: 0.31, 95% CrI: 0.15-0.66)
- HER2 Status: Improved response with HER2 positivity (OR: 2.33, 95% CrI: 1.04-5.32)

4.7 Institutional Random Effects

Analysis of center-specific random effects revealed substantial heterogeneity across treatment centers. Three centers demonstrated significantly positive effects (credible intervals excluding zero), indicating above-average performance after controlling for patient characteristics. Conversely, two centers showed significantly negative effects, suggesting opportunities for quality improvement interventions.

The reduction in ICC from 26.5% (empty model) to 12.5% (full model) indicated that clinical covariates explained approximately 53% of between-center variation, while substantial clustering remained, supporting the continued necessity of hierarchical modeling.

4.8 Cross-Validation Performance

Stratified 10-fold cross-validation confirmed the robustness of performance differences. The Bayesian hierarchical model maintained superior discrimination (cross-validated AUC = 0.821) compared to classical approaches (cross-validated AUC = 0.741), with minimal optimism (0.016), indicating excellent generalization to new patients.

4.9 Calibration Assessment

Calibration analysis revealed superior performance of Bayesian approaches across the entire risk spectrum. The calibration slope of 0.97 for the Bayesian hierarchical model approached perfect calibration (slope = 1.0), while classical methods showed systematic deviation (slope = 0.89), indicating overfitting and poor calibration.

Figure Description: Calibration plots demonstrated excellent agreement between predicted probabilities and observed response rates for the Bayesian model across all risk deciles, while classical approaches showed systematic deviation, particularly in intermediate-risk ranges.

CONCLUSION

This comprehensive comparison provides compelling evidence for the superiority of Bayesian hierarchical modeling approaches over classical statistical methods in breast cancer treatment outcome prediction. The 8.5-unit AUC improvement (0.752 to 0.837) represents a clinically meaningful enhancement that translates directly into improved patient care through more accurate risk stratification and treatment planning.

The demonstration of substantial institutional clustering (ICC = 26.5%) addresses a critical methodological gap in current prediction modeling approaches. Classical single-level models fail to capture this hierarchical structure, leading to biased parameter estimates and poor generalizability across healthcare settings. The Bayesian framework's natural accommodation of clustering through partial pooling provides optimal solutions for healthcare data analysis.

Uncertainty quantification emerges as a fundamental advantage of Bayesian approaches, providing clinically essential information that classical methods cannot deliver. The ability to generate credible intervals for individual patient predictions enables more informed clinical decision-making and supports patient counseling regarding treatment options and expected outcomes.

The superior information criteria performance (20.2% improvement) combined with excellent cross-validation results demonstrates both the statistical and practical advantages of Bayesian methods. These findings are particularly relevant for resource-constrained healthcare settings where accurate prediction tools can optimize treatment allocation and improve outcomes within existing infrastructure limitations.

REFERENCES

[1] Bray, F., et al. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3), 229-263.

[2] Newman, L. A., et al. (2019). Breast cancer disparities in African women living in sub-Saharan Africa or the United States. *Cancer*, 125(24), 4456-4463.

[3] GLOBOCAN. (2020). Global Cancer Observatory: Cancer Today. International Agency for Research on Cancer, Lyon, France.

[4] Collins, G. S., et al. (2021). Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI). *BMJ Open*, 11(7), e048008.

[5] Austin, P. C., et al. (2018). To multilevel or not to multilevel: when to use multilevel modeling in health services research. *Health Services Research*, 53(4), 2531-2550.

[6] Gelman, A., et al. (2013). *Bayesian data analysis* (3rd ed.). CRC Press.

[7] Sidey-Gibbons, J. A. M., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology*, 19, 64.

[8] Cubasch, H., et al. (2017). Breast cancer characteristics and HIV among 1,092 women in Soweto, South Africa. *Breast Cancer Research and Treatment*, 166(3), 743-754.

[9] Mansell, J., et al. (2021). Improving quality and access to cancer care in sub-Saharan Africa. *Journal of Global Oncology*, 7, 1-4.

[10] Harrell, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

[11] Collins, G. S., et al. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Annals of Internal Medicine*, 162(1), W1-W73.

[12] Wessels, L. F., & van de Vijver, M. J. (2022). External validation of breast cancer prediction models: a systematic review. *Breast Cancer Research*, 24, 83.

[13] Spiegelhalter, D. J. (2019). *The art of statistics: How to learn from data*. Basic Books.

[14] Spiegelhalter, D. J. (2019). *The art of statistics: How to learn from data*. Basic Books.

[15] Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593-1623.

[16] Goldstein, H. (2011). *Multilevel statistical models*. John Wiley and Sons.

[17] Austin, P. C., et al. (2018). To multilevel or not to multilevel: when to use multilevel modeling in health services research. *Health Services Research*, 53(4), 2531-2550.

[18] Gelman, A., et al. (2013). *Bayesian data analysis* (3rd ed.). CRC Press.

[19] Wessels, L. F., & van de Vijver, M. J. (2022). External validation of breast cancer prediction models: a systematic review. *Breast Cancer Research*, 24, 83.

[20] Jedy-Agba, E., et al. (2022). Determinants of stage at diagnosis of breast cancer in Nigerian women. *Cancer Causes and Control*, 31(3), 233-241.

[21] Azubuike, S. O., et al. (2018). Rising global burden of breast cancer: the case of sub-Saharan Africa. *World Journal of Surgical Oncology*, 16, 63.

[22] Debray, T. P., et al. (2017). A framework for developing, implementing, and evaluating clinical prediction models in an era of big data. *BMC Medicine*, 15, 155.

[23] Berger, J. O. (2020). The case for objective Bayesian analysis. *Bayesian Analysis*, 15(3), 859-899.

[24] Berger, J. O. (2020). The case for objective Bayesian analysis. *Bayesian Analysis*, 15(3), 859-899.

[25] Lee, P. M. (2019). *Bayesian statistics: an introduction*. John Wiley & Sons.