ARTICLE

Flexible decay metrics for early prediction of gestational trophoblastic neoplasia: A comparative modeling approach

Alvin Duke R. Sy*[1,2], Fernando B. Garcia, Jr.[3], Clarissa L. Velayo[4], Michael Daniel C. Lucagbo[5], Abubakar S. Asaad[1], and Maria Stephanie Fay S. Cagayan[2,6]

[1]Department of Epidemiology and Biostatistics, College of Public Health, University of the Philippines Manila, Manila, Philippines 1000
[2]Department of Pharmacology and Toxicology, College of Medicine, University of the Philippines Manila, Manila, Philippines 1000
[3]Department of Health Policy and Administration, College of Public Health, University of the Philippines Manila, Manila, Philippines 1000
[4]Department of Physiology, College of Medicine, University of the Philippines Manila, Manila, Philippines 1000
[5]School of Statistics, University of the Philippines Diliman, Quezon City, Philippines 1101
[6]Section of Trophoblastic Diseases, Department of Obstetrics and Gynecology, University of the Philippines Philippine General Hospital, Taft Avenue, Manila, Philippines 1000

ABSTRACT

Background: Molar pregnancy is a rare condition carrying a 15-22% risk of progression to gestational trophoblastic neoplasia (GTN). Although serial β-hCG monitoring is standard, its ability to predict early malignant change early remains uncertain.

This study examined whether flexible β-hCG decay metrics could improve early GTN prediction compared with conventional thresholds, and whether machine-learning classifiers provide meaningful gains beyond interpretable statistical models.

Methods: This retrospective cohort analyzed 413 post-molar patients with longitudinal β-hCG data. Seven decay metrics were derived from early follow-up measurements and evaluated using logistic and gradient boosting machine (GBM) models. Model performance was assessed through cross-validated discrimination (area under the curve (AUC)), calibration, and decision-curve analysis (DCA).

Results: The GBM model achieved higher apparent discrimination (AUC: 0.96) but negligible net clinical benefit (NCB) across thresholds, indicating probable overfitting and limited bedside utility. A parsimonious logistic model (AUC: 0.77; calibration slope: 0.85) showed stable calibration and consistent net benefit for identifying low-risk patients. Among the decay metrics, time-to-75% β-hCG decline (~25 days) emerged as the most robust and interpretable predictor (sensitivity 92%, negative predictive value 88%), offering a simple signal of malignant persistence.

Conclusion: Early β-hCG decay dynamics, particularly time-to-75% decline, can guide risk-adaptive follow-up after molar evacuation. Complex machine-learning models contributed little beyond traditional approaches in this moderate-sized cohort. These findings support prospective validation and exploration of

*Corresponding author
Email Address: arsy3@outlook.up.edu.ph
Date received: 02 September 2025
Dates revised: 16 November 2025
Date accepted: 24 November 2025
DOI: https://doi.org/10.54645/202518SupIFI-95

KEYWORDS

gestational trophoblastic disease, prognosis, retrospective studies, area under curve, chorionic gonadotropin, logistic models

cost-effective surveillance models tailored to resource-limited settings.

INTRODUCTION

Gestational trophoblastic diseases (GTD) progress to GTN in roughly 14.9-21.5% of cases, yet early progression is often clinically silent (Shana Rahman and Sudhamani 2023; Riahi et al. 2020). Routine surveillance relies on serial β-hCG monitoring, but static thresholds and fixed cut-offs often fail to capture the heterogeneous regression patterns seen after molar evacuation (Galingan and Cagayan 2021; Bakhtiyari et al. 2015; Seckl et al. 2010). Timely identification of patients who require intensified surveillance or intervention, therefore, remains a key clinical challenge.

Multiple analytic strategies have been proposed to capture the dynamics of β-hCG regression given its potential as a basis for early risk stratification. Slope-based decay metrics using weekly β-hCG assessments can identify GTN up to two weeks earlier than conventional criteria (Zhao et al. 2017), while early ratio-based markers at fixed intervals (e.g., one- or two-week percentage declines) reach accuracies approaching 90% in some studies (Galingan and Cagayan 2021). More advanced approaches, like joint models and latent class trajectory analyses, showed strong associations between non-linear β-hCG trends and post-molar GTN risk, with hazards increasing roughly three-fold per log-unit rise (Riahi et al. 2020).

Despite these advances, translation into clinical practice remains limited. The siloed application of predictive measures like regression slopes (Bakhtiyari et al. 2015) or class-based models (Burny et al. 2016) hinders synthesis into a broader decision framework. Previously proposed cut-offs (e.g., 508 mIU/mL at 3 weeks, 185 mIU/mL at 5 weeks) lack generalizability across populations (Sy and Cagayan 2023), while probabilistic models balancing accuracy with interpretability remain rarely used. This challenge is compounded in rare diseases like GTD in resource-limited settings, where small sample sizes, incomplete surveillance, and high-dimensional data create additional constraints.

This study evaluates a set of early, flexible β-hCG decay metrics within interpretable and adaptable predictive frameworks (i.e., logistic regression, machine learning or ML) to determine whether these measures yield complementary, clinically actionable prognostic information before GTN becomes apparent. The researchers hypothesize that these decay metrics can provide additional, usable prognostic signals beyond conventional thresholds when incorporated into parsimonious, well-calibrated models, thereby enhancing early, risk-adapted follow-up care without presupposing categorical superiority over existing rules.

MATERIALS AND METHODS

Study Design

A retrospective cohort study was conducted using pooled data from the Philippine Society for the Study of Trophoblastic Diseases, Inc. (PSSTD). Formal sample size and power calculations indicated that approximately 1,049 women would ideally be required to detect a small effect size (OR ~1.75) at 95% confidence and 80% power, assuming a 10–15% prevalence of malignant transformation or chemoresistance (Finch et al 2023; Altieri et al. 2003). However, given the rarity of GTD and registry constraints, all available observations were included, yielding a final analytic cohort of 413 patients who underwent molar evacuation and serial β-hCG monitoring until remission or progression to GTN. This dataset combined an earlier cohort of 258 women with both partial (PHM) and complete (CHM) moles and a later cohort of 155 women with CHM only.

The achievable sample (compared to the target) limits statistical power for detecting small effects and constrains model complexity. A parsimonious model specification was prioritized to mitigate potential overfitting and instability, with results validated through nested cross-validation and bootstrap resampling. Findings are to be interpreted as exploratory and hypothesis-generating pending prospective validation.

Variables and Metrics

The dataset was anonymized and de-identified, with ethical approval obtained from the University of the Philippines Manila Research Ethics Board (UPMREB 2024-0368-01) prior to data processing. A complete-case analysis was used, as β-hCG metrics require at least two valid follow-up values per patient, thus making single-value imputation mathematically inconsistent. Furthermore, missingness was concentrated in early follow-up points and largely due to loss to follow-up rather than data omission.

The primary outcome was progression to GTN, defined using the International Federation of Gynecology and Obstetrics (FIGO) criteria. GTD patients were monitored serially from molar evacuation until biochemical remission or progression to GTN.

The registry originally contained clinico-demographic variables such as maternal age, gravidity, parity, mode of evacuation, histologic type, and chemoprophylaxis status, along with all available β-hCG values during follow-up. Seven flexible decay metrics were derived from the initial post-evacuation β-hCG values to quantify early hormone decline. These were selected for their clinical interpretability, simplicity, and feasibility in low-resource settings. The operational definitions, derivation formulae, and clinical interpretations are summarized in Table 1. These derived variables were computed post hoc from serial β-hCG data and were not part of the original registry dataset. Segmental slopes refer to piecewise log-linear slopes computed within defined early follow-up windows using simple regressions, while the parametric slope corresponds to the fixed-effect coefficient from a mixed-effects model, representing the average early decline rate per patient.

Table 1: Summary of Flexible β-hCG Decay Metrics for Early GTN Prediction

Metric	Formula	Interpretation
Percent Drop	$[(\beta_0 - \beta_7) / \beta_0] \times 100$	Approximates proportional decline, between pre-evacuation and seventh follow-up β-hCG
Log β-hCG Ratio	Log (Pre-evacuation β-hCG / First post-evacuation β-hCG)	Reflects early treatment response, a ratio >2 to 5 suggests sharp drop (protective)
Log Decline Rate	$[\log(\beta_0) - \log(\beta_7)] / (t_7 - t_0)$	Models log-linear decay rate, flatter slopes imply slower decline

Time to Threshold	Earliest day β-hCG < 100 / 10 / 5	Time to reach clinical reference β-hCG levels (<100, <10, and <5 mIU/mL)
Early Doubling	$\beta_7 > 2 \times \beta_0$	Binary indicator of rapid rise, potentially suggestive of GTN
Parametric Slope	Estimated from first 2–5 points or model-based	Continuous rate of decline from mixed-effects or linear regression
Segmental Slopes	Slopes T1 (first 2 visits), T2 (visits 3-4), T3 (5th visit) from linear fits over follow-up bins	Captures phase-specific decay patterns in early GTD surveillance

Statistical Analysis

Bivariate comparisons between GTN and non-GTN cases used χ^2 and Wilcoxon rank-sum tests. Logistic regression, summarized using odds ratios (OR), was performed hierarchically: starting with a core model of decay metrics, extending to additional predictors, and subsequently adding interaction terms. Interaction terms (i.e., slope × baseline β-hCG) were explored because the prognostic meaning of an early decline rate may depend on initial tumor burden. A slow slope from a very high baseline can imply different risk than the same slope from a low baseline. Clinically important covariates were incorporated into the best-fitting model and used for an adjusted model.

A gradient boosting model (GBM) was chosen as the ML comparator because it accommodates non-linear relationships and variable interactions efficiently without the high computational demand or sample size required by neural networks or other ML frameworks (Li et al. 2022). Alternative algorithms (random forests, support vector machines, neural networks) were initially explored in pilot runs but offered no clear performance advantage and required larger samples to tune reliably. GBM therefore provided a pragmatic balance between flexibility, interpretability, and computational feasibility for this modest-sized dataset.

The GBM model was trained on the same set of predictors following a 70/30 train-test split, with feature importance summarized by gain, cover, and frequency. Hyperparameters followed default boosting settings (i.e., learning rate: 0.1, maximum tree depth: 6, boosting rounds: 100) with internal five-fold cross-validation to minimize overfitting. No further optimization was done to avoid inflating model performance given the limited event count. As such, its results reflect baseline GBM behavior rather than tuned performance.

Model discrimination was assessed using the area under the receiver operating characteristic (ROC) curve (AUC, 95% confidence intervals (CI), DeLong method). Optimal binary cutoffs were determined via Youden's index, and diagnostic accuracy was summarized as sensitivity, specificity, predictive values (PV), and correct classification proportion (CCP). Model fit was compared using Akaike (AIC) and Bayesian information criteria (BIC), sample size reduction, and variance inflation factors (VIF) for multicollinearity.

Net clinical benefit (NCB) was evaluated with decision curve analysis (DCA) across thresholds of 10%, 20%, and 30%, representing clinically plausible intervention points (Vickers and Elkin 2006). Five-fold cross-validation was used to compute standardized net clinical benefit (sNCB) with confidence intervals. Calibration was assessed with 200 bootstrapped resamples, reporting slope, intercept, and mean absolute error (MAE), and generating bias-corrected calibration plots.

The robustness of early β-hCG decay metrics was examined via a series of sensitivity and subgroup analyses. Stratified analyses across relevant subgroups (i.e., histological type, mode of evacuation, chemoprophylaxis status) were summarized using relative risks (RR) estimated from Poisson regression with robust errors. Similarly, subgroup-specific AUCs were computed from logistic regression models, while cross-validation procedures were repeated for both logistic and GBM models to evaluate generalizability.

All analyses were performed in R version 4.5.0 (R Core Team 2025) using the following packages: broom (Robinson et al. 2024), caret (Kuhn 2008), ggplot2 (Wickham 2016), lme4 (Bates et al. 2015), lmtest (Zeileis and Hothorn 2002), pROC (Robin et al. 2011), purrr (Wickham and Henry 2025), randomForest (Liaw and Wiener 2002), rmda (Brown 2018), sandwich (Zeileis et al. 2024), xgboost (Chen and Guestrin 2016), and zoo (Zeileis and Grothendieck 2005).

RESULTS

Sample Characteristics

Table 2 compares the outcome groups for context and was meant to summarize distributions prior to multivariable modeling. Of the total, the majority of the women experienced spontaneous remission, while a small yet significant proportion (18.6%, 95% CI: 15 to 22.74%) experienced GTN.

Table 2: Baseline Characteristics of Study Participants

Characteristics	*Total* (*n = 413*)	*Non-GTN* (*n = 336*)	*GTN* (*n = 77*)	*p*
Maternal Age				
<40 years	334 (80.87%)	271 (80.65%)	63 (81.82%)	0.82
≥40 years	79 (19.13%)	65 (19.35%)	14 (18.18%)	
Gravidity				
G1	115 (27.85%)	93 (27.68%)	22 (28.57%)	
G2 – G3	148 (35.84%)	113 (33.63%)	35 (45.45%)	0.07
G4 and above	150 (36.32%)	130 (38.69%)	20 (25.97%)	
Parity				
Nulliparous	125 (30.27%)	101 (30.06%)	24 (31.17%)	0.85
Parous	288 (69.73%)	235 (69.94%)	53 (68.83%)	
Pre-evacuation β-hCG				
Median	197,369	192,585	222,815	0.16
Interquartile range	10,000, 413,216	10,000, 407,949	14,911, 594,406	
Post-evacuation β-hCG				
1st Follow-up	1,250 (300, 4,520)	955.11 (164.95, 3,062.50)	5,046 (1,371, 18,400)	0.01**

Characteristics	*Total* (*n = 413*)	*Non-GTN* (*n = 336*)	*GTN* (*n = 77*)	*p*
2nd Follow-up	131 (23.07, 508.24)	80.45 (14.68, 235.18)	1,883.75 (518.39, 9,028.59)	0.01**
3rd Follow-up	23.10 (4.10, 134.60)	14.30 (3.30, 44.75)	1,779 (525, 3,914.26)	0.01**
Histological Type				
PHM	43 (10.41%)	37 (11.01%)	6 (7.79%)	0.40
CHM	370 (89.59%)	299 (88.99%)	71 (92.21%)	
Mode of Evacuation				
Suction curettage	323 (78.21%)	253 (75.30%)	70 (90.91%)	0.01**
Surgical evacuation	90 (21.79%)	83 (24.70%)	7 (9.09%)	
Chemoprophylaxis				
Not given	87 (21.07%)	59 (17.56%)	28 (36.36%)	0.01**
Given	326 (78.93%)	277 (82.44%)	49 (63.64%)	
Follow-up (in weeks)	33 (19, 39)	34 (29, 44)	9 (7, 12)	0.01**

*p<0.05, **p<0.01

Maternal age and parity were comparable across groups, with most women being multigravida. However, the GTN group had relatively more women with two to three pregnancies and fewer patients with at least four pregnancies. Pre-evacuation β-hCG levels were predominantly high without significant differences between outcomes. Histology was overwhelmingly CHM with both groups, and slightly more common in GTN, though not statistically significant.

GTN was more frequent among women who underwent suction curettage and who did not receive chemoprophylaxis. Median follow-up was substantially longer in remission than in GTN, indicating earlier recognition and treatment initiation in the latter.

Furthermore, median β-hCG values fell sharply over the first three visits, but GTN cases showed markedly slower declines ($p < 0.01$ at each follow-up). Figure 1 illustrates these divergent trajectories, with non-GTN cases exhibiting rapid decreases, and GTN cases showing slower, irregular declines or early plateaus. Longitudinal distributions across 22 visits (Supp Table 1) demonstrated convergence toward very low values (<2 mIU/mL) after week 6, but wide early interquartile ranges highlighted heterogeneous decline rates. These findings underscore the prognostic relevance of early follow-up and justify modeling focused on initial decay patterns rather than absolute thresholds.

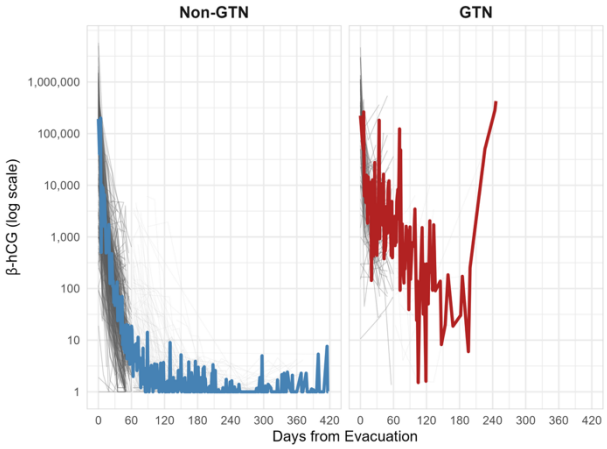

Figure 1: Individual β-hCG Trajectories over Time

Predictive Utility of Flexible β-hCG Decay Metrics

Univariable analyses first evaluated each decay metric individually before proceeding to multivariable modeling. Flexible decay metrics (Table 3) revealed marked differences between outcomes. GTN cases had apparent mean percent increase in β-hCG (−164.81% vs. 99.35%, p<0.01), where negative values indicate increases due to drop metric convention, more frequent early doubling (11.7% vs. 1.8%, p<0.01), and shallower or reversed slopes across early phases (T1–T3). Although overall parametric slopes and drop times were not significantly different, segmental analyses showed that GTN cases declined more slowly, plateaued earlier, and occasionally re-elevated, highlighting the diagnostic utility of granular slope metrics.

Table 3: Summary of Flexible β-hCG Decay Metrics across GTN Status

Metrics	*Total* (*n = 413*)	*Non-GTN* (*n = 336*)	*GTN* (*n = 77*)	*p*
Percent drop	50.1 ± 645.5	99.4 ± 10.4	-164.8 ± 1483.4	0.01**
Log β-hCG Ratio	-0.06 ± 0.04	-0.06 ± 0.03	-0.05 ± 0.05	0.07
Drop rate				
Time-to-50%	23.9 ± 38.6	25.6 ± 41.7	15.7 ± 13.7	0.40
Time-to-75%	26.6 ± 41.2	28.5 ± 44.3	16.5 ± 14.8	0.14
Time-to-90%	28.9 ± 41.5	31.1 ± 44.3	15.9 ± 12.9	0.03*
Log decline rate	-0.06 ± 0.04	-0.06 ± 0.03	-0.05 ± 0.05	0.07
Time-to-threshold				
Time to 100 mIU/mL	50.4 ± 41.7	50 ± 42.2	58.4 ± 33.2	0.01**
Time to 10 mIU/mL	73.8 ± 46.7	73.9 ± 47.3	70 ± 11.1	0.01**
Time to 5 mIU/mL	82.6 ± 50.6	82.8 ± 51.1	72.6 ± 13.5	0.01**
Early doubling	15 (3.63%)	6 (1.79%)	9 (11.69%)	0.01**
Parametric slope	-0.36 ± 0.12	-0.36 ± 0.12	-0.36 ± 0.11	0.55
Segmental slopes				
T1 (≤2 follow-ups)	-0.22 ± 0.14	-0.23 ± 0.14	-0.15 ± 0.13	0.01**
T2 (≤4 follow-ups)	-0.07 ± 0.10	-0.07 ± 0.09	-0.01 ± 0.09	0.01**

Metrics	Total (n = 413)	Non-GTN (n = 336)	GTN (n = 77)	p
T3 (5th follow-up)	-0.01 ± 0.05	-0.02 ± 0.03	0.01 ± 0.16	0.01**

*p<0.05, **p<0.01

Crude logistic models (Supp Table 2) identified percent drop from baseline as the strongest discriminator (OR: 0.94 per 1% increase; AUC: 0.94). Early β-hCG doubling strongly increased GTN risk (OR: 7.28, p<0.01), and segmental slopes were predictive of malignant progression. Extremely high or unstable odds for some metrics (e.g., log decline rate, T2 slope) reflected separation artifacts.

By contrast, time-to-threshold metrics and parametric slopes did not show a clear association in this sample, consistent with their computational structure.

Multivariable Models

In this section, all subsequent estimates refer to multivariable models adjusting for the variables specified in each model.

Core Model

The first multivariable model, which contained percent drop, log β-hCG ratio, and time-to-75%, produced stable and interpretable estimates (Table 4). All three predictors were statistically significant, and aligned with biological plausibility, such that a greater percent drop and higher log ratio of follow-up to baseline β-hCG were protective, while a slower decline to 75% of pre-evacuation levels was associated with higher odds of GTN adjusting for these decay metrics.

Table 4: Core Model

Predictors	OR	95% CI	p
Percent drop	0.95	0.91 – 0.98	0.01**
Log β-hCG Ratio	0.86	0.74 – 0.98	0.03*
Time-to-75%	0.98	0.96 – 0.99	0.03*

*p<0.05, **p<0.01

Extended and Interaction Models

Adding segmental slopes in the extended model (Supp Table 3) modestly improved apparent performance, but attenuated effects of other metrics. After adjustment for clinical covariates, only the segmental slope during the fifth follow-up (T3) remained a significant predictor, suggesting prognostic value of late-phase β-hCG trajectories. However, huge standard errors, wide CI, and reduction in statistical significance for other predictors indicate collinearity and instability.

The interaction model (Supp Table 4) suggested that a declining slope at the fifth follow-up was protective in women with low baseline β-hCG controlling for other covariates. This is consistent with a synergistic risk pattern as supported by an interaction plot (Supp Figure 1), with crossing lines between slope and baseline β-hCG levels. However, wide CI, large standard errors, and borderline p-value for the T3 slope (p: 0.06) highlight the fragility of this model.

Model Selection

Comparison of model performance across candidate specifications (Supp Table 5) showed that the core model achieved the most favorable balance between fit and parsimony. It had the lowest AIC and favorable BIC while retaining the largest number of observations. Likewise, it demonstrated good discrimination (AUC: 0.75) and low multicollinearity (all VIFs <1.20). Contrastingly, both extended and interaction models exhibited slightly higher discrimination, but these gains were offset by reduced sample size, wide CIs, and high multicollinearity.

Adjusted Model

Upon adding clinically relevant covariates (Table 5), surgical evacuation and chemoprophylactic administration were protective, aligned with expected patterns. Importantly, all three decay metrics (i.e., percent drop, log β-hCG ratio, and time-to-75%) remained independently associated with GTN after adjusting for maternal age, evacuation method, and chemoprophylaxis. This model balances parsimony with explanatory power and was chosen for subsequent analyses. It potentially offers a clinically relevant risk stratification tool that maintains statistical soundness and clinical interpretability.

Table 5: Adjusted Model for Prediction

Predictors	OR	95% CI	p
Maternal age	1.03	0.99 – 1.07	0.15
Mode of evacuation			
Suction curettage	1.00		
Surgical evacuation	0.31	0.10 – 0.82	0.03*
Chemoprophylaxis	0.49	0.26 – 0.98	0.04*
Percent drop	0.96	0.92 – 0.98	0.01*
Log β-hCG Ratio	0.86	0.74 – 0.99	0.04*
Time-to-75%	0.98	0.96 – 0.99	0.04*

*p<0.05, **p<0.01

Clinical Utility and Model Stability

To complement the adjusted logistic regression model, a GBM model was developed using the same predictors. While GBM models do not provide interpretable ORs, feature importance metrics like gain, cover, and frequency were extracted (Supp Table 6). Percent drop contributed most substantially (72.3% of total gain), followed by time-to-75% and log β-hCG ratio, whereas clinical variables such as maternal age, mode of evacuation, and chemoprophylaxis contributed minimally. This finding underscores the dominant predictive value of early β-hCG decay metrics in ML–based risk stratification.

Comparing discrimination and clinical utility, GBM showed excellent discrimination (AUC: 0.96) but zero NCB across clinically meaningful thresholds (Table 6; Supp Figure 2). Such discrepancy reflected poor calibration and overfitting, with the GBM model tending to predict too many patients as low risk and producing extreme probability estimates without corresponding true positives. Conversely, the logistic model demonstrated more modest discrimination with consistently positive NCB across decision thresholds.

Table 6: Comparative Model Performance

Metrics	Logistic Model		Gradient Boosting Model	
	NCB	*CV (sNCB)*	*NCB*	*CV (sNCB)*
Threshold				
10%	0.090	0.516 (0.338 to 0.703)	0.053	0.040 (-0.041 to 0.121)
20%	0.048	0.328 (0.149 to 0.519)	0	-0.015 (-0.160 to 0.044)
30%	0.034	0.243 (0.075 to 0.481)	0	-0.073 (-0.326 to 0.004)
40%	0.027	0.216 (0.069 to 0.461)	0	-0.141 (-0.547 to 0)
50%	0.028	0.198 (0.013 to 0.423)	0	-0.236 (-0.856 to 0)
AUC (95% CI)	0.77 (0.70 to 0.83)		0.96 (0.92 to 0.99)	

Both models underwent five-fold cross-validation using predictors in the adjusted model. Cross-validated sNCB for logistic regression peaked at 0.516 (95% CI: 0.338–0.703) at the 10% threshold and remained positive through 30 to 50% thresholds (Table 6). However, GBM yielded lower or negative sNCB values, with CIs often crossing zero, reflecting poor generalizability.

Figure 2 highlights such divergence with the logistic model maintaining stable and positive sNCB across thresholds, whereas GBM performance declined sharply with widening CIs at higher thresholds.

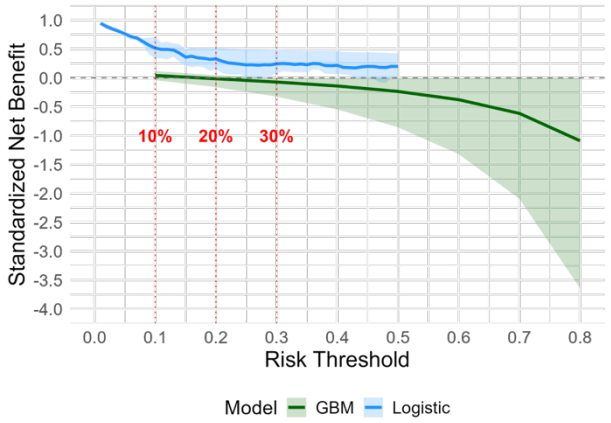

Figure 2: Logistic and GBM Cross-Validated Decision Curve Plots

Bootstrapped calibration of the logistic model yielded an intercept of –0.12 and a slope of 0.85, indicating slight overestimation at higher predicted risks but acceptable overall calibration (Figure 3). Predictions aligned closely with observed probabilities, with slight overestimation at higher predicted risks and minor underestimation at lower levels. The calibration curve demonstrated that predicted risks were directionally and proportionally consistent with actual outcomes, reinforcing the model's reliability in supporting patient-specific risk stratification (Supp Fig 3).

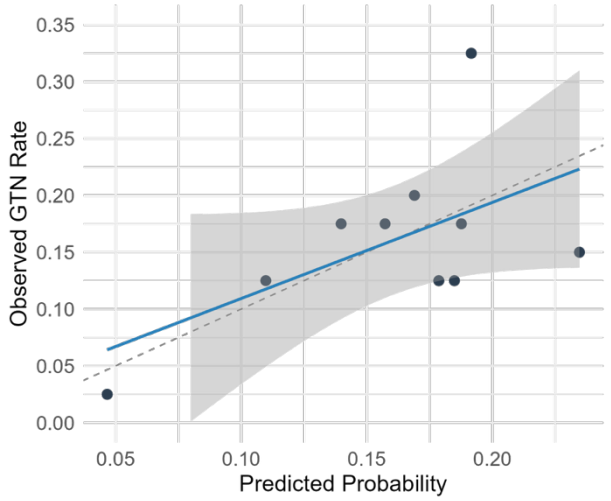

Figure 3: Calibration Plot of the Adjusted Logistic Model

Stratified and Subgroup Analysis

Stratified analyses evaluated performance of the adjusted logistic model across key subgroups (Supp Table 7). Subgroup differences indicate performance heterogeneity such as perfect prediction for PHM (AUC: 1) but only moderate for CHM (AUC: 0.75). Surgical evacuation cases (AUC: 0.90) showed better discrimination compared to suction curettage (AUC: 0.74), while patients who did not receive chemoprophylaxis exhibited notably higher accuracy (AUC: 0.97 vs. 0.74). These findings suggest that β-hCG decay patterns are more informative prior to chemotherapy influence.

Diagnostic thresholds optimized using the Youden Index (Supp Table 8) identified clinically interpretable cutoffs specifically, log β-hCG ratio of 4.12 (approximately 60-fold decline from baseline), a time-to-75% decline of 25.5 days, and an early clearance metric of 99.97% drop. Among these, time-to-75% exhibited the strongest diagnostic performance, with high sensitivity and negative predictive value for ruling out GTN progression (Table 7). Other metrics, while statistically significant, demonstrated limited specificity or impractically extreme cutoff values, limiting their standalone clinical utility.

A forest plot (Figure 4) summarizes stratified robust Poisson regression models (Supp Table 9). Slower time to 75% decline remained significantly associated with GTN across most subgroups, including complete moles, suction curettage, and chemoprophylaxis recipients. The effect appeared stronger among surgical evacuation cases, although small sample size limited precision. In this analysis, partial moles or patients who did not receive chemotherapy were unable to reach statistical significance.

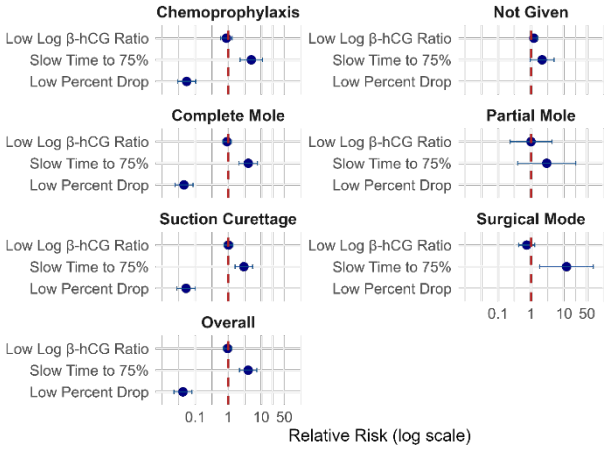

Figure 4: Forest Plot of Stratified Relative Risk across Covariates

Table 7: Diagnostic Performance of Decay Metrics

Metrics	*Sensitivity*	*Specificity*	*Positive PV*	*Negative PV*	*CCP*
Percent drop	69% (62-76%)	9% (6-14%)	37% (32-43%)	29% (19-40%)	36% (31-40%)
Time to 75%	92% (86-97%)	19% (15-24%)	29% (24-34%)	88% (77-94%)	38% (34-43%)
Log β-hCG Ratio	30% (21-40%)	3% (2-6%)	9% (6-12%)	13% (6-23%)	9% (7-13%)

Percent drop occasionally exhibited paradoxical protective associations, likely reflecting early measurement variability or non-monotonic β-hCG patterns in non-GTN cases. Similarly, the log β-hCG ratio did not add significant discriminatory value when time-to-decline metrics were already incorporated. Together, these findings position β-hCG decay metrics as clinically informative early prognostic markers.

DISCUSSION

This study examined the role of early β-hCG decay kinetics in predicting the development of post-molar GTN. Among several candidate metrics, the time-to-75% β-hCG decline consistently emerged as the most robust and clinically interpretable predictor. Patients who failed to achieve this decline within 25 days are likely to have a four-fold increased risk of GTN. This pattern held across subgroups, with this early, shape-sensitive threshold providing more actionable decision point than conventional FIGO definitions especially when used with other prognostic variables.

These findings build upon previous studies underscoring the importance of early β-hCG decline dynamics. Much of the earlier literature emphasized slope-based declines (Kader et al. 2024) or β-hCG measurements within the first three weeks post-evacuation (Sy and Cagayan 2023; Khosravirad et al. 2017). Other studies highlighted that reaching <5 mIU/mL by 56 days lowers risk (Albright et al. 2020; Braga et al. 2015), while ratios and absolute levels at two to four weeks reliably predicted persistence (Rakprasit et al. 2023; Wolfberg et al. 2005). Together, these studies support the premise that early biochemical trajectories contain clinically meaningful signals.

The biological mechanism underlying delayed β-hCG decay is well established. Persistent or malignant trophoblastic tissue continues producing β-hCG after evacuation, resulting in slower clearance (Braga et al. 2019). Instead of the rapid decline expected from physiologic involuting syncytiotrophoblasts, neoplastic cytotrophoblasts sustain β-hCG secretion, producing plateauing or secondary rises (Taylor et al. 2016). The time-to-75% metric therefore approximates the window in which normal trophoblastic regression should occur, making delayed clearance a plausible early marker of malignant persistence.

These observations align with existing FIGO and National Comprehensive Cancer Network (NCCN) guidelines, which define GTN using plateauing or rising β-hCG values across consecutive weekly measurements (Ngan et al. 2021; Abu-Rustum et al. 2019). Because the time-to-75% captures kinetic abnormalities earlier, it offers a complementary signal while remaining compatible with existing follow-up schedules. Furthermore, the finding that failure to reach a 75% drop within 25 days increases the risk of progression reinforces previous studies showing that two-week β-hCG ratios <30 predicted GTN (Kang et al. 2012). Translating it to clinical practice, if β-hCG levels do not drop by 75% within the first three to four weeks, clinicians may consider intensifying surveillance, ordering additional diagnostics, or anticipating earlier intervention.

From a model development perspective, the inclusion of demographic (e.g., maternal age) and treatment-related variables (e.g., suction curettage) did not enhance prediction (Savage et al. 2013). This aligns with reports suggesting that early biochemical behavior often supersedes baseline clinical characteristics in GTD prognosis (Ngan et al. 2018). In resource-constrained settings like the Philippines, which have a relatively high GTD burden, a shift toward β-hCG-driven predictive models may offer a practical and rapid approach to post-evacuation triage.

The percent drop metric, although statistically significant, had limited ability to rule out GTN. Similarly, log-ratio thresholds reflected the protective effect of sharper β-hCG declines but may be less intuitive for bedside decision-making. These findings support using multiple metrics in combination to reduce the risk of misclassification, while maintaining alignment with standard practice benchmarks such as β-hCG normalization (<5 mIU/mL) by eight weeks (Abu-Rustum et al. 2019).

The contrasting performance of the GBM and logistic models warrants attention. Although GBM achieved a high apparent AUC (0.96), it exhibited poor cross-validated calibration and minimal net clinical benefit, likely due to sample size limitations and sparse events contributing to overfitting. This emphasizes the importance of appropriately matching model complexity to dataset structure rather than suggesting inherent flaws in ML methods (Huber et al. 2023; Christodoulou et al. 2019). Traditional logistic regression retained advantages in interpretability and biological grounding, making it a more suitable option for early clinical translation (Topol 2019).

Additionally, this reinforces a broader point in prediction modeling: discrimination alone is insufficient. Calibration, validation, and out-of-sample performance are essential and frequently overlooked in prediction model studies (Parker et al. 2023; Van Calster et al. 2019). In this study, both regression models confirmed the salience of time-based and relative decay metrics, with the simpler model offering better calibration and decision-analytic utility. As a result, these metrics appear more practical for implementation.

A key strength of this study is its multi-method analytic design. By combining traditional regression with ML-based approaches and applying ROC, calibration, and decision-curve analyses, the findings were triangulated. Stratified AUCs and RR estimates provided sensitivity analyses across relevant subgroups, while Youden-based thresholds accompanied by cross-validated DCA improved the translational value of the models.

Its novelty lies in the use of flexible, trajectory-based markers rather than fixed β-hCG cutoffs. These metrics remain intuitive, actionable during early follow-up, and do not require additional cost. The simplicity of time-to-75% has the potential to improve patient–clinician communication and support shared decision-making, especially in the context of early follow-up uncertainty. However, the retrospective design limits control over β-hCG sampling intervals, contributing to variable values and occasional separation issues. Shorter follow-up among GTN cases may introduce survivor bias or truncation effects in slope-based metrics. The use of data from single-institution and inconsistent laboratory sources may also affect generalizability. Thresholds identified here may not extrapolate directly to other healthcare settings.

GBM performance remained inconsistent, with limited NCB likely due to scaling procedures, data limitations, or sparse events. Although hierarchical and parsimonious modeling approaches, along with diagnostic checks, helped address these constraints, caution remains essential when interpreting outlier estimates, especially in stratified analyses.

Based on the findings, percent drop and time-to-75% may be incorporated into post-evacuation protocols as preliminary screening tools during the first two to four weeks. For example, a delayed or prolonged time-to-75% (>25 days) may prompt early referral or intensified surveillance. Prospective validation in diverse clinical and assay contexts is needed to test generalizability. Future directions include developing point-of-care tools, Bayesian hybrid models, and cost-effectiveness analyses to support implementation.

While ML-based models offered some insights, traditional logistic regression remains the most appropriate starting point for clinical exploration. Further improvements may involve adding complementary diagnostic markers and standardizing β-hCG assays across laboratories. Low-cost mobile monitoring tools also warrant exploration, particularly in settings with constrained access to laboratory and diagnostics.

Finally, this study highlights the broader gap in the Philippines regarding locally validated, dynamic prediction tools. Given the rarity of GTD, establishing a harmonized GTD registry would enable larger-scale evaluation of β-hCG trajectory metrics and support more timely detection and management, particularly in resource-limited environments.

CONCLUSION

This study offers a paradigm shift in GTN surveillance, moving beyond fixed biomarker thresholds to maximize repeated measurements using flexible, trajectory-based β-hCG decay metrics. These metrics exhibited early, reliable, and clinically meaningful signals of malignant progression. Machine learning offered the potential of non-linear models yet traditional regression methods showed superior calibration and clinical utility, thus, reinforcing the importance of model interpretability.

Time-to-75% decline offered actionable insights within four weeks after molar evacuation. The use of these clinically interpretable measures empowers both clinicians and patients with timely, personalized insights which could prompt intensified surveillance or earlier intervention, compared with current practice, which delays action until fixed cutoffs are met, often several weeks later.

Furthermore, these simple, intuitive decay thresholds offer a feasible framework for tailoring GTN follow-up, especially in resource-constrained settings where follow-up is not always optimal. However, external validation across diverse populations, refinement of β-hCG decay patterns, and improved disease registries will be needed to ensure that these early detection strategies translate into improved outcomes for GTD patients.

ACKNOWLEDGMENT

The authors sincerely thank Dr. Maria Febi de Ramos, Dr. Maria del Carmen Castillo, Dr. Melissa de Quiros, and all trophoblastic disease specialists under PSSTD for allowing us to make use of their data for secondary analysis.

CONFLICT OF INTEREST

All authors have stated explicitly that there are no conflicts of interest in connection with this article.

CONTRIBUTIONS OF INDIVIDUAL AUTHORS

Conceptualization: ARS; Data curation: ARS, MSC; Formal analysis: ARS; Funding acquisition: ARS, MSC; Investigation: ARS, MSC; Methodology: ARS, ASA, MCL; Project administration: MSC; Resources: MSC; Software: ARS; Supervision: MSC, CLV, ASA, MCL, FBG; Validation: MSC, CLV, ASA, MCL, FBG; Visualization: ARS; Writing – original draft: ARS; Writing–review & editing: all authors.

REFERENCES

Abu-Rustum NR, Yashar CM, Bean S, Bradley K, Campos SM, Chon HS, Chu C, Cohn D, Crispens MA, Damast S, Dorigo O, Eifel PJ, Fisher CM, Frederick P, Gaffney DK, Han E, Huh WK, Lurain JR, Mariani A, Mutch D, Nagel C, Nekhlyudov L, Fader AN, Remmenga SW, Reynolds RK, Sisodia R, Tillmanns T, Ueda S, Wyse E, McMillian NR, Scavone J. Gestational Trophoblastic Neoplasia, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2019; 17(11):1374-91.

Albright B, Shorter J, Mastroyannis S, Ko E, Schreiber C. Gestational Trophoblastic Neoplasia After Human Chorionic Gonadotropin Normalization Following Molar Pregnancy: A Systematic Review and Meta-analysis. Obstet Gynecol 2020; 135(1):12-23.

Altieri A, Franceschi S, Ferlay J, Smith J, La Vecchia C. Epidemiology and aetiology of gestational trophoblastic diseases. Lancet Oncol 2003; 4:670-8.

Bakhtiyari M, Mirzamoradi M, Kimyaiee P, Aghaie A, Mansournia MA, Ashrafi-Vand S, Sarfjoo FS. Postmolar gestational trophoblastic neoplasia: beyond the traditional risk factors. Fertil Steril 2015; 104(3):649-54.

Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw 2015; 67(1):1-48.

Braga A, Maestá I, Matos M, Elias K, Rizzo J, Viggiano M. Gestational trophoblastic neoplasia after spontaneous human chorionic gonadotropin normalization following molar pregnancy evacuation. Gynecol Oncol 2015; 139(2):283-7.

Braga A, Mora P, de Melo A, et al. Challenges in the diagnosis and treatment of gestational trophoblastic neoplasia worldwide. World J Clin Oncol 2019; 10(2):28-37.

Brown M. rmda: Risk model decision analysis [Computer software]. Version 1.6. Published 2018. https://github.com/mdbrown/rmda.

Burny C, Rabilloud M, Golfier F, Massardier J, Hajri T, Schott AM, Subtil F. Early diagnosis of gestational trophoblastic neoplasia based on trajectory classification with compartment modeling. BMC Med Res Methodol 2016; 16:3.

Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Paper presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016; San Francisco.

Christodoulou E, Ma J, Collins G, Steyerberg E, Verbakel J, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019; 110:12-22.

Finch L, Hosterman T, Huang M. A comprehensive assessment of differences in gestational trophoblastic neoplasia (GTN) by race and ethnicity: A national program of cancer (NPCR) and surveillance, epidemiology, and end Results (SEER) database study (2117). Gynecol Oncol 2023; 176(S1):S207-8.

Galingan M, Cagayan M. Beta-human chorionic gonadotropin levels as early predictor for progression to Gestational Trophoblastic Neoplasia after molar pregnancy evacuation at a Philippine tertiary hospital. Philipp J Obstet Gynecol 2021; 45(4):153-9.

Huber M, Schober P, Petersen S, Luedi M. Decision curve analysis confirms higher clinical utility of multi-domain versus single-domain prediction models in patients with open abdomen treatment for peritonitis. BMC Med Inform Decis Mak 2023; 23(1):63.

Kader M, Elora A, Chowdhury M, Afrose T, Amin R, Sultana Jolly R. Evaluation of β-hCG regression after evacuation of molar pregnancy as a predictive factor for malignant GTN. Int J Reprod Contracept Obstet Gynecol 2024; 13(12):3569-74.

Kang W, Choi H, Kim S. Prediction of persistent gestational trophoblastic neoplasia: the role of hCG level and ratio in 2 weeks after evacuation of complete mole. Gynecol Oncol 2012; 124(2):250-3.

Khosravirad A, Zayeri F, Baghestani A, Yoosefi M, Bakhtiyari M. Predictive Power of Human Chorionic Gonadotropin in Post-Molar Gestational Trophoblastic Neoplasia: A Longitudinal ROC Analysis. Int J Cancer Manag 2017; 10(9):e9015.

Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008; 28(5):1-26.

Li K, Yao S, Zhang Z, Cao B, Wilson CM, Kalos D, Kuan PF, Zhu R, Wang X. Efficient gradient boosting for prognostic biomarker discovery. Bioinformatics 2022; 38(6):1631–8.

Liaw A, Wiener M. Classification and regression by randomForest. R News 2002; 2(3):18-22.

Ngan HYS, Seckl MJ, Berkowitz RS, Xiang Y, Golfier F, Sekharan PK, Lurain JR, Massuger L. Update on the diagnosis and management of gestational trophoblastic disease. Int J Gynaecol Obstet 2018; 143(Suppl 2):79-85.

Ngan HYS, Seckl MJ, Berkowitz RS, Xiang Y, Golfier F, Sekharan PK, Lurain JR, Massuger L. Diagnosis and management of gestational trophoblastic disease: 2021 update. Int J Gynaecol Obstet 2021; 155(Suppl 1):86-93.

Parker VL, Winter MC, Tidy JA, Hancock BW, Palmer JE, Sarwar N, Kaur B, McDonald K, Aguiar X, Singh K, Unsworth N, Jabbar I, Pacey AA, Harrison RF, Seckl MJ. PREDICT-GTN 1: Can we improve the FIGO scoring system in gestational trophoblastic neoplasia? Int J Cancer 2023; 152(5):986-97.

R Core Team. R: A Language and Environment for Statistical Computing. version 4.5.0. Vienna, Austria: R Foundation for Statistical Computing; 2025. Retrieved from https://www.R-project.org.

Rakprasit C, Ruengkhachorn I, Therasakvichya S, Inthasorn P, Achariyapota V, Kuljarasnont S, Khemworapong K, Jareemit N. Combined analysis of clinical features, human chorionic gonadotropin (hCG) value, and hCG ratios for early prediction of postmolar gestational trophoblastic neoplasia. Arch Gynecol Obstet 2023; 307(4):1145-54.

Riahi R, Rahimiforoushani A, Nourijelyani K, Akbari Sharak N, Bakhtiyari M. Early Detection of Gestational Trophoblastic Neoplasia Based on Serial Measurement of Human Chorionic Gonadotrophin Hormone in Women with Molar Pregnancy. Int J Prev Med 2020; 11:187.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011; 12(77):1-8.

Robinson D, Hayes A, Couch S. broom: Convert statistical objects into tidy tibbles [Computer software]. Version 1.0.6. Published 2024. https://broom.tidyverse.org.

Savage PM, Sita-Lumsden A, Dickson S, Iyer R, Everard J, Coleman R, Fisher RA, Short D, Casalboni S, Catalano K, Seckl MJ. The relationship of maternal age to molar pregnancy incidence, risks for chemotherapy and subsequent pregnancy outcome. J Obstet Gynaecol 2013; 33(4):406-11.

Seckl M, Sebire N, Berkowitz R. Gestational trophoblastic disease. Lancet 2010; 376(9742):717-29.

Shana Rahman K, Sudhamani C. Incidence and risk factors of post-molar gestational trophoblastic neoplasia: a prospective study. Int J Reprod Contracept Obstet Gynecol [Internet] 2023; 12(4):924-30.

Sy A, Cagayan M. Comparison of beta human chorionic gonadotropin based prognostic models on the clinical outcomes of gestational trophoblastic disease patients in the Philippines. Philipp J Obstet Gynecol 2023; 47(3):99-107.

Taylor F, Short D, Harvey R, Winter MC, Tidy J, Hancock BW, Savage PM, Sarwar N, Seckl MJ, Coleman RE. Late spontaneous resolution of persistent molar pregnancy. BJOG 2016; 123(7):1175-81.

Topol E. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019; 25(1):44-56.

Van Calster, B; McLernon, DJ; van Smeden, M; Wynants, L; Steyerberg, EW; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. BMC Med 2019; 17(1):230.

Vickers A, Elkin E. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006; 26(6):565-74.

Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.

Wickham H, Henry L. purrr: Functional programming tools [Computer software]. Version 1.2.0. Published 2025. https://purrr.tidyverse.org.

Wolfberg A, Berkowitz R, Goldstein D, Feltmate C, Lieberman E. Postevacuation hCG levels and risk of gestational trophoblastic neoplasia in women with complete molar pregnancy. Obstet Gynecol 2005; 106(3):548-52.

Zeileis A, Grothendieck G. zoo: S3 infrastructure for regular and irregular time series (Z's ordered observations). J Stat Softw 2005; 14(6):1-27.

Zeileis A, Hothorn T. Diagnostic checking in regression relationships. R News 2002; 2(3):7-10.

Zeileis A, Lumley T, Graham N, Koell S. sandwich: Robust covariance matrix estimators [Computer software]. Version 3.1-1. Published 2024. https://cran.r-project.org/package=sandwich.

Zhao P, Wang S, Zhang X, Lu W. A Novel Prediction Model for Postmolar Gestational Trophoblastic Neoplasia and Comparison with Existing Models. Int J Gynecol Cancer 2017; 27(5):1028-34.

SUPPLEMENTAL DATA

Appendix A: Supplementary Tables

Table 1: Distribution of Serum β-hCG across Follow-up Visits

Follow-up	Median (IQR)	Obs.	Follow-up	Median (IQR)	Obs.
1	1,250 (300, 4,520)	413	12	1.20 (0.10, 1.21)	108
2	131 (23.04, 508.24)	412	13	1.20 (0.10, 1.21)	57
3	23.10 (4.10, 134.60)	402	14	1.20 (0.69, 1.20)	33
4	5.25 (1.66, 38.29)	372	15	1.20 (0.52, 1.20)	20
5	2.05 (1.00, 8.60)	346	16	0.10 (0.10, 1.10)	12
6	1.20 (0.11, 3.40)	308	17	0.14 (0.10, 1.10)	8
7	1.20 (0.15, 2.00)	267	18	0.10 (0.10, 0.10)	6
8	1.20 (0.10, 1.21)	238	19	0.10 (0.10, 0.10)	5
9	1.20 (0.10, 1.21)	217	20	0.10 (0.10, 0.10)	3
10	1.20 (0.10, 1.21)	195	21	0.10	1
11	1.20 (0.10, 1.21)	155	22	0.10	1

Table 2: Crude Association of Flexible Decay Metrics with GTN

Metrics	*OR (95% CI)*	*p-value*	*AUC (95% CI)*
Percent drop	0.94 (0.91 – 0.97)	<0.01*	0.94 (0.92 – 0.97)
Log β-hCG Ratio	0.77 (0.68 – 0.86)	<0.01*	0.68 (0.62 – 0.74)
Drop rate			
Time-to-50%	0.99 (0.97 – <1.00)	0.05*	0.52 (0.45 – 0.58)
Time-to-75%	0.99 (0.97 – <1.00)	0.03*	0.54 (0.47 – 0.61)
Time-to-90%	0.98 (0.96 – 0.99)	0.01*	0.57 (0.51 – 0.64)
Log decline rate	4,397 (2.86 – 11,075,723)	0.03*	0.57 (0.48 – 0.66)
Time-to-threshold			
Time to 100 mIU/mL	1.00 (0.99 – 1.01)	0.39	0.64 (0.50 – 0.77)
Time to 10 mIU/mL	1.00 (0.98 – 1.01)	0.80	0.42 (0.33 – 0.51)
Time to 5 mIU/mL	0.99 (0.97 – 1.01)	0.56	0.49 (0.38 – 0.61)
Early doubling	7.28 (2.54 – 22.40)	<0.01*	0.55 (0.51 – 0.59)
Parametric slope	0.62 (0.07 – 5.18)	0.66	0.52 (0.46 – 0.59)
Segmental slopes			
T1 (≤2 follow-ups)	117.21 (17.20 – 888.52)	<0.01*	0.68 (0.62 – 0.75)
T2 (≤4 follow-ups)	22,4180.93 (1,979.05 – 42,828,994)	<0.01*	0.76 (0.68 – 0.85)
T3 (5th follow-up)	915.77 (0.75 – 10,538,688)	0.07	0.76 (0.61 – 0.91

Table 3: Extended Model

Predictors	*OR*	*95% CI*	*p-value*
Segmental slopes			
T1 (≤2 follow-ups)	2.09	0.13 – 25.90	0.61
T2 (≤4 follow-ups)	1.74	0.21 – 28.90	0.68
T3 (5th follow-up)	0.26	0.06 – 0.66	0.03*
Percent drop	0.99	0.87 – 0.99	0.69
Time-to-75%	0.77	0.44 – 1.03	0.35
Time-to-10%	1.00	0.94 – 1.05	0.99
Log β-hCG Ratio	1.04	0.32 – 3.33	0.95

Table 4: Interaction Model

Predictors	*OR*	*95% CI*	*p-value*
Segmental slopes			
T1 (≤2 follow-ups)	2.69	0.29 – 28.30	0.44
T2 (≤4 follow-ups)	0.81	0.44 – 2.43	0.56
T3 (5th follow-up)	0.22	0.04 – 1.12	0.06
Slope T3 * β-hCG	0.01	<0.01 – 0.43	0.02*
Baseline β-hCG	0.07	<0.01 – 0.96	0.32
Time-to-75%	0.96	0.83 – 1.04	0.53
Time-to-10%	1.00	0.96 – 1.02	0.82
Log β-hCG Ratio	1.17	0.42 – 2.94	0.74

Table 5: Summary of Model Comparison Metrics

Model	*n*	*AIC*	*BIC*	*LR x^2*	*p-value*	*AUC*	*VIF*
Core	400	313.43	329.39	46.3	<0.01*	0.75 (0.68 – 0.81)	All < 1.20
Extended	286	31.03	60.27	27.1	<0.01*	0.91 (0.74 – 1.00)	Max ≈ 1.59
Interaction	286	43.04	75.94	17.1	<0.01*	0.84 (0.54 – 1.00)	Max ≈ 4.76
Adjusted	400	309.19	337.13	56.5	<0.01*	0.77 (0.70 – 0.83)	All < 1.40

Table 6: GBM Feature Importance Metrics

Feature	Gain	Cover	Frequency
Percent drop	0.723	0.475	0.315
Time-to-75%	0.142	0.174	0.172
Log β-hCG ratio	0.066	0.182	0.295
Maternal age	0.061	0.161	0.207
Mode of evacuation	0.006	0.006	0.009
Chemoprophylaxis	0.001	0.001	0.002

Table 7: Stratified Discrimination Metrics

Subgroup	AUC	95% CI
Histology		
Partial mole	1.00	-
Complete mole	0.75	0.68 – 0.81
Mode of evacuation		
Suction curettage	0.74	0.67 – 0.82
Surgical evacuation	0.90	0.70 – 1.00
Chemoprophylaxis		
Not given	0.97	0.95 – 1.00
Given	0.74	0.67 – 0.82

Table 8: Summary Measures of Early β-hCG Decay Metrics

Metrics	Median	Range	Youden's
Percent drop	99.99	-12,720.51 to 100	99.97
Time-to-75%	13.00	2 to $_{+}\infty$	25.50
Log β-hCG ratio	4.50	0.03 to 15.61	4.12

Table 9: Stratified Risk Estimation Models of Early Decay Metrics and GTN

Subgroup	Predictor	Robust RR (95% CI)	p-value
Overall	Low Percent Drop	0.04 (0.02 – 0.08)	<0.01*
	Slow Time to 75%	4.02 (2.20 – 7.36)	<0.01*
	Low Log β-hCG Ratio	0.96 (0.72 – 1.27)	0.77
Histology			
Partial mole	Low Percent Drop	1.27^{-9}	0.009
	Slow Time to 75%	3.00 (0.39 – 22.87)	0.29
	Low Log β-hCG Ratio	1.00 (0.23 – 4.31)	0.99
Complete mole	Low Percent Drop	0.05 (0.02 – 0.09)	<0.01*
	Slow Time to 75%	4.06 (2.14 – 7.69)	<0.01*
	Low Log β-hCG Ratio	0.92 (0.69 – 1.24)	0.59
Mode of evacuation			
Suction curettage	Low Percent Drop	0.05 (0.03 – 0.10)	<0.01*
	Slow Time to 75%	2.98 (1.62 – 5.50)	<0.01*
	Low Log β-hCG Ratio	1.02 (0.75 – 1.38)	0.92
Surgical approach	Low Percent Drop	1.19^{-10} (8.59^{-11} – 3.24^{-10})	<0.01*
	Slow Time to 75%	11.94 (1.84 – 77.40)	0.01*
	Low Log β-hCG Ratio	0.74 (0.42 – 1.30)	0.30
Chemoprophylaxis			
Not given	Low Percent Drop	2.36^{-10} (1.71^{-10} – 3.24^{-10})	<0.01*
	Slow Time to 75%	2.17 (0.94 – 5.01)	0.07
	Low Log β-hCG Ratio	1.22 (0.95 – 1.57)	0.12
Given	Low Percent Drop	0.05 (0.03 – 0.10)	<0.01*
	Slow Time to 75%	5.02 (2.29 – 11.00)	<0.01*
	Low Log β-hCG Ratio	0.88 (0.60 – 1.30)	0.53

Appendix B: Supplementary Figures

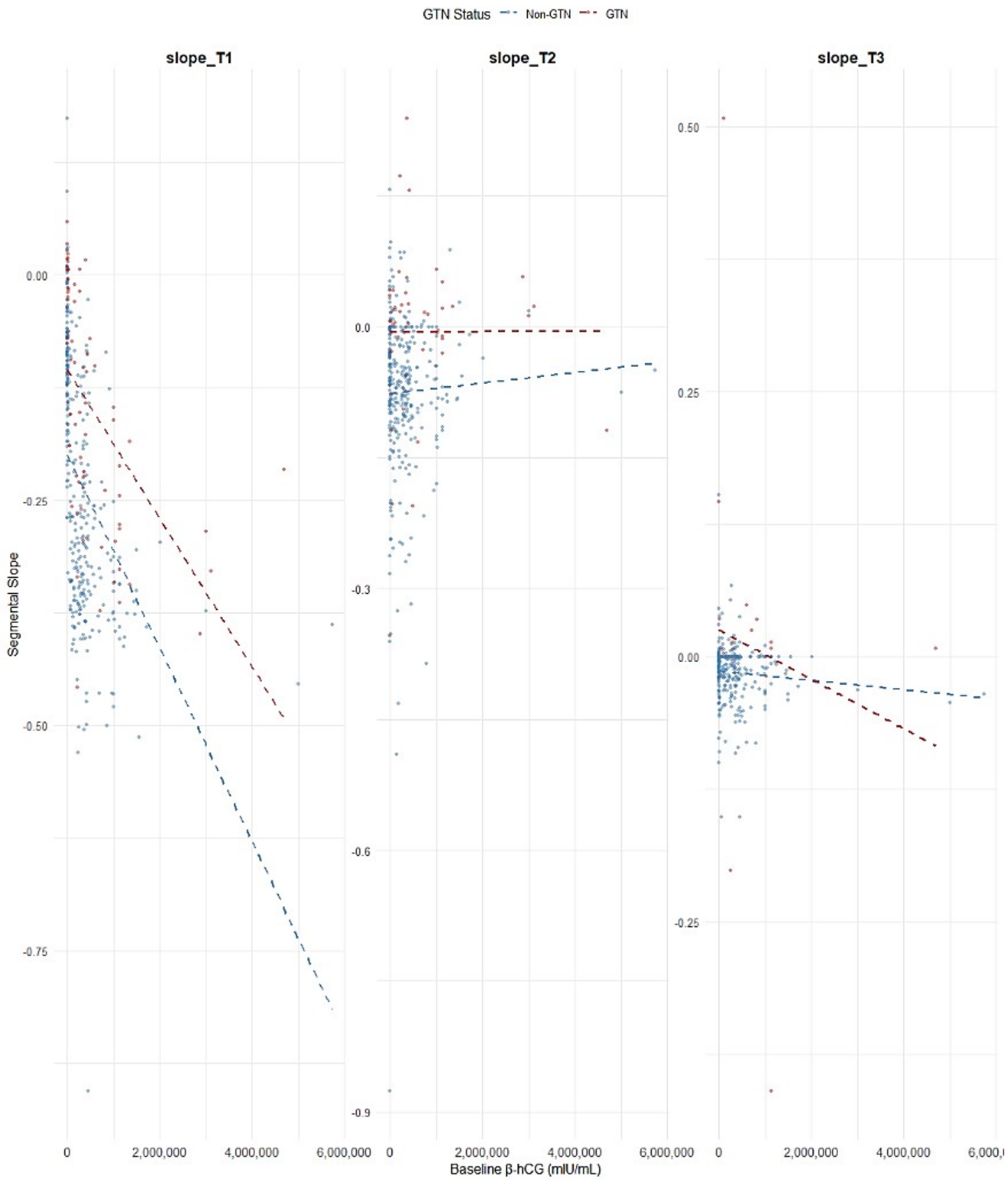

Figure 1: Interaction Plot between Segmental Slopes and Baseline β-hCG

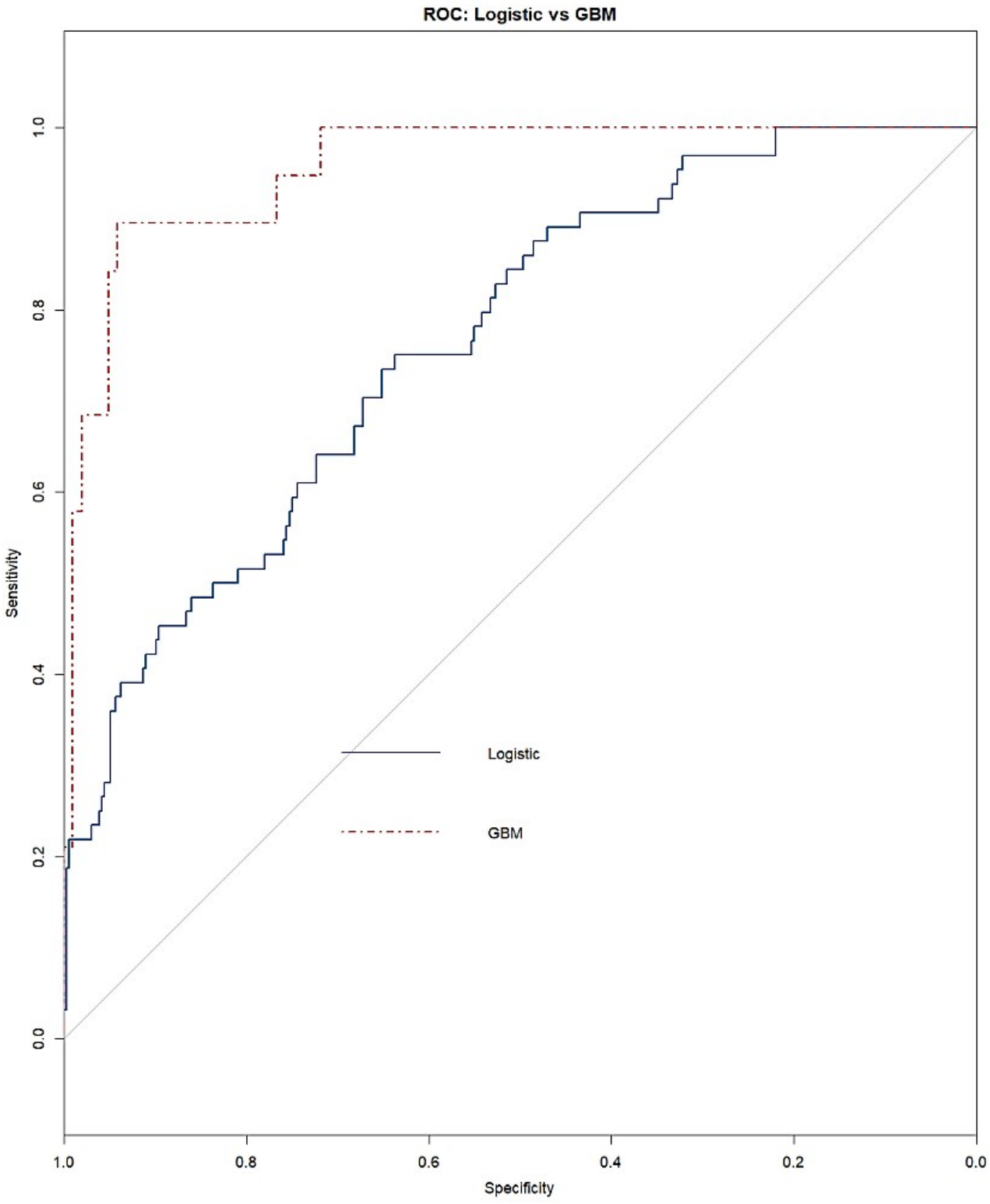

Figure 2: ROC Plot of Logistic and GBM Models in Predicting GTN

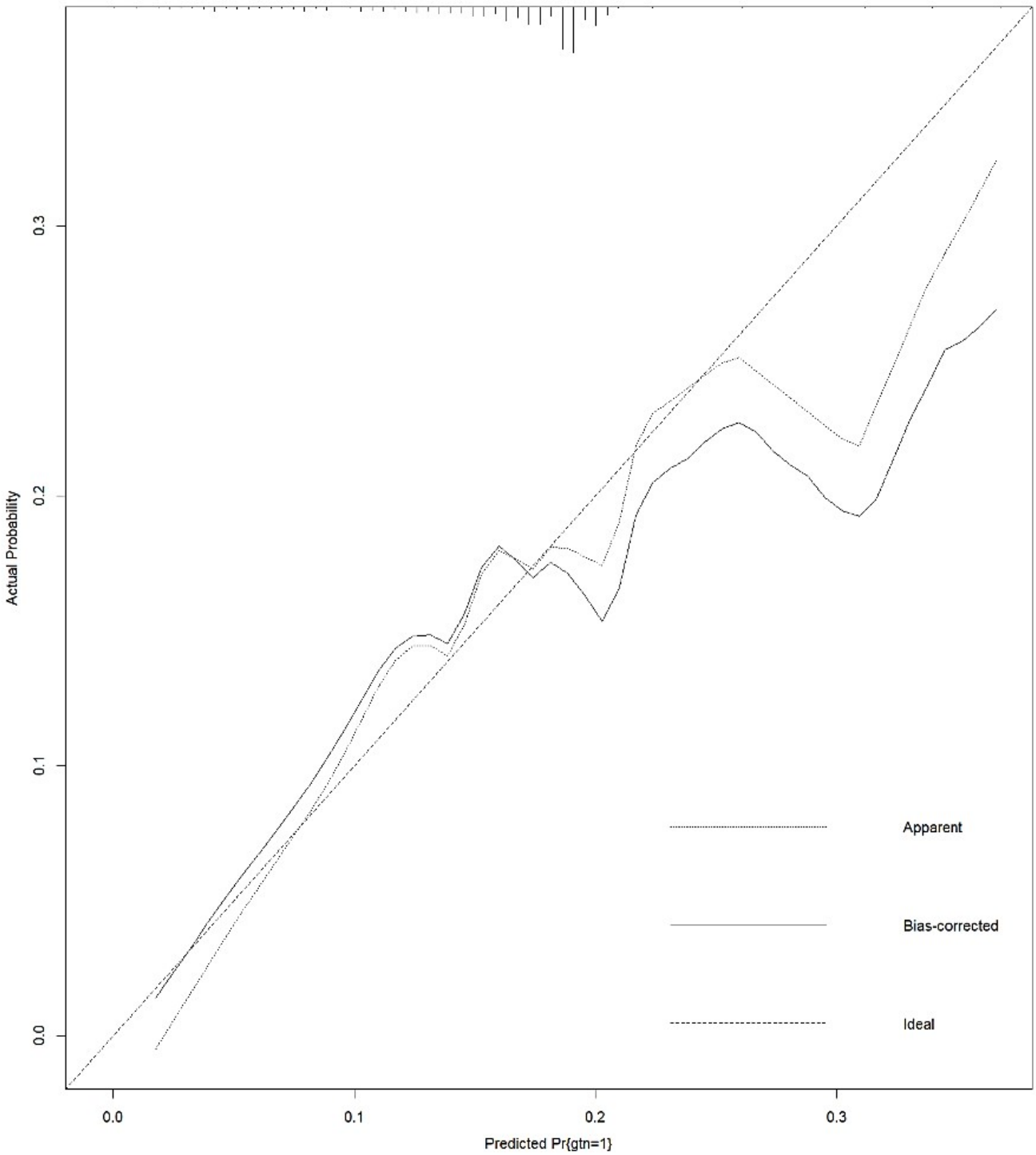

Figure 3: Calibration Plot of the Adjusted Logistic Model