Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning Models for Predicting Bacteraemia in Paediatric Febrile Neutropenia: Development and External Multicentre Validation

Srujan Mallagundla

srujanm682@gmail.com

Ballari Medical College and Research Centre, Karnataka, India https://orcid.org/0009-0006-0774-7088

Research Article

Keywords: febrile neutropenia, bacteraemia, machine learning, paediatric oncology, prediction model, external validation, SHAP, decision curve analysis

Posted Date: June 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9893508/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background.

Bacteraemia complicates approximately one in five febrile neutropenia (FN) episodes in children receiving chemotherapy, and existing clinical decision rules perform inconsistently outside their derivation settings. This study developed and externally validated two complementary machine learning models predicting bacteraemia at two clinically distinct time points: before and at FN onset.

Methods.

Two cohorts from the open-access Bern FN Core dataset were used for derivation: Project 17B (observation-period level, 712 FN episodes; pre-episode Model A) and Project 17C (episode level, 703 FN episodes in 291 patients; at-onset Model B), both 1993–2012. The prospective SPOG 2015 FN study (360 episodes, six Swiss centres, 2016–2018) served as external validation. Four algorithms were tested: Random Forest, XGBoost, LightGBM, and Logistic Regression. Performance was assessed by AUC with 95% bootstrap confidence intervals, Brier score, and decision curve analysis.

Results.

Bacteraemia occurred in 148 of 703 (21.1%) Bern 17C episodes and 56 of 360 (15.6%) SPOG episodes. Model A achieved internal AUC 0.840–0.877; chemotherapy course duration was the dominant predictor (SHAP 0.861). On external validation using portable features, AUC fell to 0.566–0.664. Model B achieved internal AUC 0.636–0.722 and external AUC 0.607–0.671. Platelet count (SHAP 0.50) and haemoglobin (SHAP 0.45) were the dominant at-onset predictors. Logistic Regression showed near-zero generalisation loss (internal 0.636; external 0.635).

Conclusions.

Platelet count and haemoglobin at FN diagnosis are the strongest at-onset bacteraemia predictors. Chemotherapy course duration is a novel pre-episode predictor warranting prospective collection. Logistic Regression maintains consistent performance across institutions and supports low-risk discharge decision support.

Introduction

Febrile neutropenia (FN) is among the most frequent and serious complications of cytotoxic chemotherapy in children with cancer. The seminal work of Bodey et al. in 1966 established that infection risk rises sharply as the neutrophil count falls below 500 per cubic millimetre and that prolonged neutropenia compounds this risk in a quantitative manner (1). Bacteraemia, confirmed by a

positive blood culture within seven days of FN onset, complicates 15–25% of paediatric FN episodes and remains a leading contributor to treatment-related mortality (2). Current international guidelines recommend empiric broad-spectrum intravenous antibiotics and inpatient admission for all children with FN, regardless of individual clinical profile (3). While this conservative approach is safe, it generates a large proportion of low-risk admissions that expose patients to catheter-related complications and disrupt chemotherapy scheduling. A systematic review of outpatient and oral antibiotic management by Manji et al. found that carefully selected low-risk children can be managed safely outside hospital, yet clinical tools to identify such patients reliably are lacking (4).

Efforts to stratify paediatric FN patients into risk groups have produced several clinical decision rules over the past three decades. The Klaassen index, derived from a Canadian single-centre cohort, uses cancer type, ANC, and clinical appearance to assign low-risk status (5). The Hakim model, derived at St Jude Children's Research Hospital, incorporates temperature and monocyte count alongside cancer diagnosis (6). The SPOG 2003 FN score, derived from a prospective multicentre Swiss cohort, identified bacteraemia-associated predictors including temperature, CRP, and bone marrow involvement (7). The PICNICC collaboration assembled individual participant data from thirteen countries to derive a pooled prediction rule using temperature, ANC, and cancer type (8). Despite their variety, these tools share a fundamental limitation: a systematic review by Phillips et al. found that discriminatory performance falls consistently when these rules are applied outside their derivation settings, with most achieving AUC values below 0.70 on independent validation (9). An individual participant data meta-analysis of the PICNICC model specifically reported a pooled external AUC of only 0.59 (10). Work by Haeusler et al. demonstrated that even after statistical recalibration to a new population, performance remained modest and required repeated updating to retain clinical utility (11).

Part of the reason for this persistent failure is that virtually all published decision rules were built using episode-level data collected at the time of FN presentation, restricting predictors to variables observable only once fever has occurred. This leaves untested a potentially informative class of pre-episode treatment characteristics, including the planned length of a chemotherapy course. Longer courses produce more sustained neutropenia and a wider window of bacteraemia risk, yet chemotherapy course duration has not been evaluated in any published paediatric FN model. Machine learning algorithms, which can identify non-linear interactions across many candidate variables, offer a systematic framework for evaluating both episode-level and pre-episode predictors within the same study (12).

The Bern FN Core provides a unique resource for this purpose. Project 17B provides observation-period-level data including planned treatment characteristics at the start of each course (13), and Project 17C provides episode-level clinical and laboratory data at FN diagnosis (14). Together they support development of two conceptually distinct prediction models. The prospective SPOG 2015 FN definition study (15), conducted across six Swiss centres between 2016 and 2018, provides a temporally and geographically independent cohort for external validation.

In this study, we derived and externally validated two complementary machine learning models for bacteraemia prediction in paediatric FN. Model A uses observation-period-level characteristics to estimate risk before FN occurs, supporting risk stratification at the start of a chemotherapy course. Model B uses acute clinical and laboratory variables at FN onset to support triage at presentation. Both models were evaluated using SHAP-based interpretability analysis and decision curve analysis, using only openly available data that make the analysis fully reproducible.

Methods

Study Design and Reporting

This prediction model development and prospective external validation study was reported following the TRIPOD statement (16). All datasets used are openly available; no ethics approval was required.

Data Sources

Bern FN Core Project 17B (doi:10.6084/m9.figshare.4765216). An observation-period-level dataset from the University Children's Hospital Bern covering 583 patients and 2,113 chemotherapy observation periods from 1993 to 2012. Variables describe patient demographics, cancer type, treatment intensity, CVAD presence, and prior FN and bacteraemia history. The primary endpoint was the occurrence of at least one bacteraemia episode per observation period, encoded as a binary variable (13).

Bern FN Core Project 17C (doi:10.6084/m9.figshare.5769732). An episode-level dataset from the same institution covering 703 FN episodes in 291 patients from 1993 to 2012. Variables include maximum temperature, ANC, CRP, monocyte count, platelet count, haemoglobin, and clinical examination findings at FN diagnosis. The bacteraemia outcome is the binary variable *cultpos* (14).

SPOG 2015 FN Definition Study (doi:10.6084/m9.figshare.22337248). A prospective multicentre cohort of 360 FN episodes in 158 patients across six Swiss paediatric oncology centres, 2016–2018. The bacteraemia outcome is the binary variable *d_sre_bac*. This cohort was used exclusively for external validation and was not involved in any training step (15).

Feature Engineering

Model A features (n = 17) included sex, age group, cancer type (one-hot encoded), relapsed status, chemotherapy intensity (ordinal 1–4), bone marrow involvement, CVAD presence, prior FN history, prior bacteraemia history, chemotherapy course length in days (CHEMO.LEN.D), and calendar year period. Model B features (n = 22) additionally included temperature, ANC, monocyte count, platelet count, haemoglobin, CRP, chills, clinically detectable infection focus, general condition at presentation, and episode number per patient as a proxy for cumulative FN history. Missing values were imputed with

training-set medians. For SPOG external validation, features unavailable in SPOG (CRP, chills, infection focus) were excluded, leaving a portable feature set of 19 variables.

Model Development

Four algorithms were evaluated: Random Forest, XGBoost, LightGBM, and Logistic Regression. Class imbalance was addressed using balanced class weights (Random Forest, LightGBM, Logistic Regression) or the scale_pos_weight parameter (XGBoost). A temporal split was applied: episodes from 1993–2008 formed the derivation set and 2009–2012 the internal validation set. This design simulates prospective deployment and avoids the optimism associated with random cross-validation when multiple FN episodes per patient are present (17). For external validation, models were retrained on the complete derivation cohort using portable features and applied to SPOG without modification. The imputer was fit on training data only.

Performance Evaluation

The primary metric was AUC with 95% bootstrap confidence intervals (1,000 resamples). Secondary metrics included sensitivity, specificity, PPV, NPV, and positive and negative likelihood ratios at the Youden Index threshold. The Brier score was used as the primary calibration statistic (18). Decision curve analysis quantified net benefit across threshold probabilities (19). Variable importance was quantified using SHAP TreeExplainer (20). Analyses were performed in Python 3.12 using scikit-learn 1.3, XGBoost 1.7, LightGBM 4.0, and SHAP 0.42.

Model Framework

Results

Cohort Characteristics

The Bern 17C derivation cohort comprised 703 FN episodes in 291 patients (median 2 episodes per patient, range 1–9). ALL or lymphoma was the most frequent diagnosis (n = 385, 54.8%), followed by solid tumours (n = 165, 23.5%) and AML (n = 92, 13.1%). Bacteraemia was confirmed in 148 episodes (21.1%). The SPOG external validation cohort comprised 360 episodes in 158 patients across six centres, with bacteraemia in 56 episodes (15.6%). The two cohorts differed substantially: CVAD prevalence was 66.3% in Bern 17C versus 99.4% in SPOG; bone marrow involvement was 16.6% versus 6.7%; median platelet count was 25 versus 47 $x10^9$/L; and median temperature at FN onset was 39.1 degrees C versus 38.9 degrees C. Full characteristics are presented in Table 1.

Table 1

Cohort characteristics. Values are n (%) unless stated. CVAD = central venous access device; ANC = absolute neutrophil count; CRP = C-reactive protein; IQR = interquartile range. p-values: chi-squared for proportions; Mann-Whitney U for continuous variables.

Characteristic	Bern 17C (n = 703)	SPOG 2015 FN (n = 360)	p
Unique patients	291	158	
Episodes per patient, median (range)	2 (1–9)	2 (1–7)	
Male sex, n (%)	371 (52.8)	187 (51.9)	0.79
Age group at diagnosis, n (%)			
0-3.99 years	228 (32.4)	73 (20.3)	
4-7.99 years	171 (24.3)	98 (27.2)	
8-11.99 years	132 (18.8)	90 (25.0)	
12–17 years	172 (24.5)	99 (27.5)	< 0.001
Cancer type, n (%)			
ALL / lymphoma	385 (54.8)	227 (63.1)	
AML / mixed leukaemia	92 (13.1)	16 (4.4)	
CNS tumour	59 (8.4)	32 (8.9)	
Solid tumour	165 (23.5)	85 (23.6)	< 0.001
Relapsed malignancy, n (%)	126 (17.9)	22 (6.1)	< 0.001
Bone marrow involvement, n (%)	117 (16.6)	24 (6.7)	< 0.001
CVAD present, n (%)	466 (66.3)	358 (99.4)	< 0.001
Prior FN episode, n (%)	412 (58.6)	202 (56.1)	0.43
Intensive chemotherapy (> 10 d neutropenia or myeloablative), n (%)	138 (19.6)	119 (33.1)	< 0.001
Temperature at FN onset, median (IQR), °C	39.1 (39.0-39.4)	38.9 (38.5–39.2)	< 0.001
ANC at FN onset, median (IQR), ×10⁹/L	0.04 (0.00-0.19)	0.10 (0.03–0.31)	< 0.001
Platelet count, median (IQR), ×10⁹/L	25 (8–79)	47 (22–100)	< 0.001

Characteristic	Bern 17C (n = 703)	SPOG 2015 FN (n = 360)	p
CRP at FN onset, median (IQR), mg/L	46 (18–90)	Not available	-
Bacteraemia, n (%)	148 (21.1)	56 (15.6)	0.03

Model A: Pre-Episode Prediction (Bern 17B)

Internal temporal validation on the 2009–2012 Bern 17B test set showed strong discrimination (Table 2). XGBoost achieved the highest AUC of 0.877, followed by LightGBM (0.856), Random Forest (0.840), and Logistic Regression (0.805). LightGBM offered the most balanced profile: sensitivity 87%, specificity 81%, and NPV 0.97. XGBoost and Random Forest each achieved 100% sensitivity at the Youden threshold, with specificity of 69% and 65% respectively. Feature ablation confirmed that removing CHEMO.LEN.D alone reduced internal AUC by 0.197–0.321, accounting for most of the performance gap seen on external validation. SHAP analysis identified CHEMO.LEN.D as the dominant predictor (mean |SHAP| 0.861), followed by prior FN history (0.268), CNS tumour diagnosis (0.230), and bone marrow involvement (0.215).

Table 2
Model A internal temporal validation (Bern 17B, test period 2009–2012). Sens = sensitivity; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value; LR + and LR- = positive and negative likelihood ratios at Youden threshold; Brier = Brier score.

Model	AUC	Sens	Spec	PPV	NPV	LR+	LR−	Brier
XGBoost	**0.877**	100%	69%	0.39	1.00	3.2	0.00	0.153
LightGBM	**0.856**	87%	81%	0.48	0.97	4.6	0.16	0.159
Random Forest	**0.840**	100%	65%	0.37	1.00	2.8	0.00	0.149
Logistic Regression	**0.805**	87%	70%	0.37	0.96	2.9	0.19	0.208

On external validation against SPOG using the portable feature set (CHEMO.LEN.D excluded), AUC fell across all algorithms. Logistic Regression showed the smallest generalisation loss, from internal AUC 0.805 to external 0.664 (Table 3).

Table 3
Model A external validation (SPOG 2015 FN, n = 360). Portable feature set used; CHEMO.LEN.D excluded. Confidence intervals by 1,000-sample bootstrap.

Model	AUC (95% CI)	Sens	Spec	PPV	NPV	LR+	LR−	Brier
Logistic Regression	**0.664 [0.583–0.742]**	79%	47%	0.22	0.92	1.5	0.45	0.237
Random Forest	**0.633 [0.536–0.719]**	71%	52%	0.22	0.91	1.5	0.55	0.201
XGBoost	**0.610 [0.524–0.692]**	62%	57%	0.21	0.89	1.5	0.66	0.246
LightGBM	**0.566 [0.474–0.652]**	23%	94%	0.41	0.87	3.7	0.82	0.254

Model B: At-FN-Onset Prediction (Bern 17C)

Internal temporal validation on the 2009–2012 Bern 17C test set yielded AUC values of 0.720 for Random Forest (95% CI 0.562–0.869), 0.722 for XGBoost (0.548–0.866), 0.711 for LightGBM (0.563–0.856), and 0.636 for Logistic Regression (0.471–0.788). NPV ranged from 0.93 to 0.96 (Table 4).

Table 4
Model B internal temporal validation (Bern 17C, test period 2009–2012). Confidence intervals by 1,000-sample bootstrap.

Model	AUC (95% CI)	Sens	Spec	PPV	NPV	LR+	LR−	Brier
Random Forest	**0.720 [0.562–0.869]**	71%	73%	-	0.95	-	-	0.137
XGBoost	**0.722 [0.548–0.866]**	57%	86%	-	0.93	-	-	0.155
LightGBM	**0.711 [0.563–0.856]**	86%	51%	-	0.96	-	-	0.163
Logistic Regression	**0.636 [0.471–0.788]**	71%	60%	-	0.93	-	-	0.227

On external validation against SPOG, Random Forest achieved AUC 0.671 (95% CI 0.581–0.752), XGBoost 0.641 (0.561–0.715), Logistic Regression 0.635 (0.538–0.720), and LightGBM 0.607 (0.528–0.689). Logistic Regression showed near-zero generalisation loss: internal AUC 0.636 to external 0.635. All models except LightGBM had confidence intervals entirely above 0.50 (Table 5).

Table 5
Model B external validation (SPOG 2015 FN, n = 360; 56 bacteraemia events). Portable feature set applied; CRP, chills, and infection focus excluded. Confidence intervals by 1,000-sample bootstrap.

Model	AUC (95% CI)	Sens	Spec	PPV	NPV	LR+	LR−	Brier
Random Forest	**0.671 [0.581–0.752]**	50%	79%	0.22	0.90	2.4	0.63	0.134
XGBoost	**0.641 [0.561–0.715]**	75%	52%	0.21	0.92	1.6	0.48	0.158
Logistic Regression	**0.635 [0.538–0.720]**	61%	66%	0.22	0.90	1.8	0.59	0.198
LightGBM	**0.607 [0.528–0.689]**	57%	63%	0.21	0.89	1.5	0.68	0.166

Table 6

AUC summary across all four analyses. Model A = pre-episode (17B derivation); Model B = at-onset (17C derivation). Internal = Bern 2009–2012 test set; External = SPOG 2015 FN validation cohort.

Model	Model A Internal	Model A External	Model B Internal	Model B External
XGBoost	**0.877**	0.610	**0.722**	0.641
LightGBM	**0.856**	0.566	**0.711**	0.607
Random Forest	**0.840**	0.633	**0.720**	0.671
Logistic Regression	**0.805**	0.664	**0.636**	0.635

SHAP Variable Importance

XGBoost was selected for SHAP analysis given its highest Model B internal AUC. Platelet count at FN onset was the dominant predictor (mean |SHAP| 0.50), followed by haemoglobin (0.45), age group (0.35), maximum temperature (0.33), and CRP (0.28). ANC ranked only 11th (0.15), and AML diagnosis ranked near the bottom at 0.04, indicating that its contribution to bacteraemia risk is captured by the associated laboratory abnormalities (Fig. 2).

Logistic Regression coefficients identified bone marrow involvement as the strongest categorical predictor (OR 1.88), followed by chemotherapy intensity (OR 1.66 per step) and temperature at FN onset (OR 1.52 per degree C). Monocyte count was the strongest protective predictor (OR 0.63 per 10^9/L), followed by ANC (OR 0.72). The discrepancy between SHAP ranking (platelet count first) and the per-unit LR odds ratio (0.997 per unit) reflects scale: platelet count ranges from approximately 1 to 500 + in this cohort, so the per-unit coefficient produces a large total contribution across the physiological range (Table 7).

Table 7
Logistic Regression coefficients and odds ratios, Model B portable feature set (trained on complete Bern 17C derivation cohort). Top nine predictors ranked by absolute coefficient.

Variable	Coefficient	Odds Ratio	Interpretation
Bone marrow involvement	+ 0.632	**1.88**	Strongest single categorical predictor
Chemotherapy intensity (per step)	+ 0.509	**1.66**	Higher intensity increases risk
Temperature at FN onset (per °C)	+ 0.416	**1.52**	Each degree carries incremental risk
ALL diagnosis	+ 0.207	**1.23**	Mild increase versus other cancers
CVAD present	+ 0.180	**1.20**	Central line as bacteraemia portal
Reduced general condition	+ 0.127	**1.13**	Severity marker at presentation
Relapsed malignancy	+ 0.105	**1.11**	Proxy for disease intensity
ANC at FN onset (per $\times 10^9$/L)	-0.336	**0.71**	Higher ANC is protective
Monocyte count (per $\times 10^9$/L)	-0.469	**0.63**	Strongest protective predictor

ROC Curves and Calibration

ROC curves and the four-analysis AUC comparison are shown in Fig. 3. Random Forest showed the best absolute calibration, with near-identical Brier scores internally (0.137) and externally (0.134), indicating stable probability estimation. Logistic Regression showed moderate internal calibration (Brier 0.227) but improved on external validation (Brier 0.198). Calibration plots (Fig. 4) showed irregular patterns at extreme predicted probabilities, attributable to sparse events per bin given the limited bacteraemia count in both the internal test set and the SPOG validation cohort; Brier scores are therefore the primary calibration statistic.

Discussion

This study developed and externally validated two machine learning models for bacteraemia prediction in paediatric FN using open-access datasets spanning three decades and two distinct patient cohorts. Four principal findings emerge: chemotherapy course duration is the dominant pre-episode bacteraemia predictor and has not been reported in any prior model; platelet count and haemoglobin at FN onset are the strongest at-onset predictors, outranking temperature and ANC; Logistic Regression achieves near-zero generalisation loss across institutions; and clinical at-onset features generalise substantially better than pre-episode treatment characteristics.

The dominance of chemotherapy course duration in Model A has a coherent mechanistic basis. Longer courses accumulate greater cumulative bone marrow toxicity, extending both the depth and duration of

neutropenia and widening the window of bacteraemia risk, as first quantified by Bodey et al. (1). Prior work from the prospective SPOG 2003 FN study showed that chemotherapy intensity, measured categorically as expected neutropenia duration, predicted bacteraemia risk (7), but the actual planned course length in days has not been tested as a continuous predictor in any published model. Our Model A addresses a different clinical question from most published work: which patients should receive enhanced monitoring at the start of a chemotherapy course? A bacteraemia prediction model at that decision point, identifying patients prospectively at the time of treatment planning rather than reactively at FN presentation, has practical value for prophylaxis decisions and infection surveillance protocols. The SPOG 2003 bacteraemia study by Agyeman et al. similarly demonstrated that pre-episode characteristics were informative predictors (21), a finding our Model A extends in a systematic machine learning framework.

The finding that platelet count at FN onset outranks ANC as the dominant at-onset predictor in Model B represents a departure from established thinking. The ANC-centred framework has been foundational since Bodey et al. (1), and it underpins the Klaassen index (5), the Hakim model (6), the SPOG 2003 score (7), and the PICNICC model (8). Platelet count, by contrast, integrates cumulative cytotoxic exposure, haematopoietic reserve, and underlying disease burden in ways that a single ANC measurement does not. Alali et al. identified thrombocytopenia below 50 $x10^9$/L as a significant bacteraemia predictor (OR 2.1) in their US single-centre cohort (22). Haemoglobin was associated with adverse FN outcomes in a retrospective analysis of the Bern cohort by Stergiotis et al. (23). The present study extends both findings by showing that platelet count and haemoglobin together provide the strongest continuous predictive signal in a multivariate SHAP framework, independent of ANC.

The external performance of Model B compares reasonably with published benchmarks. The PICNICC individual participant data meta-analysis reported a pooled external AUC of 0.59 across multiple validation cohorts (10). Hakim et al. reported a model AUC of 0.64–0.66 in their derivation sample (6). Alali et al. achieved an internal AUC of 0.79 using Random Forest, without external validation (22). Our Logistic Regression AUC of 0.635 externally, derived from data collected 20–30 years before the validation cohort in a different institutional context, represents stable and genuinely transportable performance. The NPV of 0.90–0.96 across models supports use in low-risk discharge protocols, where a post-test bacteraemia probability below 10% has been proposed as a threshold for safe outpatient management by Ammann et al. (24). The near-identical generalisation for Logistic Regression confirms that simpler models transport more reliably across settings than complex tree-based ensembles, consistent with the external validation framework described by Moons et al. (17).

Several limitations apply. Both derivation cohorts come from a single Swiss institution; local pathogen distributions, antimicrobial prophylaxis practices, and resistance profiles have changed substantially since the collection period of 1993–2012. CRP, which ranked fifth in the SHAP analysis, was absent from the SPOG dataset and had to be excluded from the portable feature set; its inclusion would likely close part of the internal-to-external AUC gap. Acute vital signs such as hypotension and heart rate, significant predictors in Alali et al. (22), were not available in Bern 17C. Confidence intervals for the internal Model B

validation are wide, reflecting the small 2009–2012 test set. The SPOG cohort differs from Bern 17C in CVAD prevalence (99.4% vs 66.3%) and bacteraemia rate (15.6% vs 21.1%), which may affect calibration in ways not fully captured by the Brier score.

In clinical terms, the two models address distinct decision points. Model A, requiring only planned treatment data before fever occurs, can guide monitoring intensity at the start of a chemotherapy course. Model B, requiring platelet count, haemoglobin, temperature, and bone marrow involvement at FN onset, supports real-time triage. The portable Logistic Regression score needs no specialist software and is compatible with standard hospital laboratory workflows. Prospective validation in non-European and resource-limited settings, where unnecessary hospitalisation is most costly, is the most important next step for clinical implementation.

References

1. Bodey GP, Buckley M, Sathe YS, Freireich EJ (1966) Quantitative relationships between circulating leukocytes and infection in patients with acute leukemia. Ann Intern Med 64(2):328–340. 10.7326/0003-4819-64-2-328
2. Freifeld AG, Bow EJ, Sepkowitz KA et al (2011) Clinical practice guideline for the use of antimicrobial agents in neutropenic patients with cancer: 2010 update. Clin Infect Dis 52(4):e56–e93. 10.1093/cid/cir073
3. Lehrnbecher T, Robinson P, Fisher B et al (2017) Guideline for the management of fever and neutropenia in children with cancer and hematopoietic stem-cell transplantation recipients: 2017 update. J Clin Oncol 35(18):2082–2094. 10.1200/JCO.2016.71.7017
4. Manji A, Beyene J, Dupuis LL, Phillips R, Lehrnbecher T, Sung L (2012) Outpatient and oral antibiotic management of low-risk febrile neutropenia are effective in children: a systematic review of prospective trials. Support Care Cancer 20(6):1135–1145. 10.1007/s00520-012-1425-8
5. Klaassen RJ, Goodman TR, Pham B, Doyle JJ (2000) Low-risk prediction rule for pediatric oncology patients presenting with fever and neutropenia. J Clin Oncol 18(5):1012–1019. 10.1200/JCO.2000.18.5.1012
6. Hakim H, Flynn PM, Srivastava DK, Knapp KM, Li C, Okuma J, Gaur AH (2010) Risk prediction in pediatric cancer patients with fever and neutropenia. Pediatr Infect Dis J 29(1):53–59. 10.1097/INF.0b013e3181c3f6f0
7. Ammann RA, Bodmer N, Hirt A et al (2010) Predicting adverse events in children with fever and chemotherapy-induced neutropenia: the prospective multicenter SPOG 2003 FN study. J Clin Oncol 28(12):2008–2014. 10.1200/JCO.2009.23.5499
8. Phillips RS, Sung L, Ammann RA et al (2016) Predicting microbiologically defined infection in febrile neutropenic episodes in children: global individual participant data multivariable meta-analysis. Br J Cancer 114(6):623–630. 10.1038/bjc.2016.28

9. Phillips B, Wade R, Stewart LA, Sutton AJ (2010) Systematic review and meta-analysis of the discriminatory performance of risk prediction rules in febrile neutropaenic episodes in children and young people. Eur J Cancer 46(16):2950–2964. 10.1016/j.ejca.2010.05.026
10. Phillips B, Morgan JE, Haeusler GM, Riley RD, PICNICC Collaborative (2020) Individual participant data validation of the PICNICC prediction model for febrile neutropenia. Arch Dis Child 105(7):664–669. 10.1136/archdischild-2019-317308
11. Haeusler GM, Phillips R, Slavin MA et al (2020) Re-evaluating and recalibrating predictors of bacterial infection in children with cancer and febrile neutropenia. EClinicalMedicine 23:100394. 10.1016/j.eclinm.2020.100394
12. Obermeyer Z, Emanuel EJ (2016) Predicting the future: big data, machine learning, and clinical medicine. N Engl J Med 375(13):1216–1219. 10.1056/NEJMp1606181
13. Ammann RA, Zermatten MG, Koenig C, von Allmen A (2017) Project 17B: Bern FN Core: Observation periods in pediatric patients in Bern, Switzerland, 1993–2012 [dataset]. 10.6084/m9.figshare.4765216. figshare
14. Ammann RA, Zermatten MG, Koenig C, von Allmen A (2019) Project 17C: Bern FN Core: Episodes of fever in chemotherapy-induced neutropenia in pediatric patients in Bern, Switzerland, 1993–2012 [dataset]. 10.6084/m9.figshare.5769732. figshare
15. Santschi M, Ammann RA, Agyeman PKA et al (2023) Outcome prediction in pediatric fever in neutropenia: development of clinical decision rules and external validation based on data from the SPOG 2015 FN definition study. PLoS One. ;18(8):e0287233. 10.1371/journal.pone.0287233. Dataset: doi:10.6084/m9.figshare.22337248
16. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). Ann Intern Med 162(1):55–63. 10.7326/M14-0697
17. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M (2012) Risk prediction models: II. External validation, model updating, and impact assessment. Heart 98(9):691–698. 10.1136/heartjnl-2011-301247
18. Van Calster B, McLernon DJ, van Smeden M et al (2019) Calibration: the Achilles heel of predictive analytics. BMC Med 17(1):230. 10.1186/s12916-019-1466-7
19. Vickers AJ, van Calster B, Steyerberg EW (2016) Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ 352:i6. 10.1136/bmj.i6
20. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 30:4765–4774
21. Agyeman P, Aebi C, Hirt A et al (2011) Predicting bacteremia in children with cancer and fever in chemotherapy-induced neutropenia: results of the prospective multicenter SPOG 2003 FN study. Pediatr Infect Dis J 30(7):e114–e119. 10.1097/INF.0b013e318215a290
22. Alali M, Mayampurath A, Dai Y, Bartlett AH (2022) A prediction model for bacteremia and transfer to intensive care in pediatric and adolescent cancer patients with febrile neutropenia. Sci Rep

12(1):7429. 10.1038/s41598-022-11576-z

23. Stergiotis M, Ammann RA, Droz S, Koenig C, Agyeman PKA (2021) Pediatric fever in neutropenia with bacteremia: pathogen distribution and antibiotic susceptibility patterns over time. PLoS ONE 16(2):e0246654. 10.1371/journal.pone.0246654
24. Ammann RA, Hirt A, Luthy AR, Aebi C (2003) Identification of children presenting with fever in chemotherapy-induced neutropenia at low risk for severe bacterial infection. Med Pediatr Oncol 41(5):436–443. 10.1002/mpo.10360

Figures

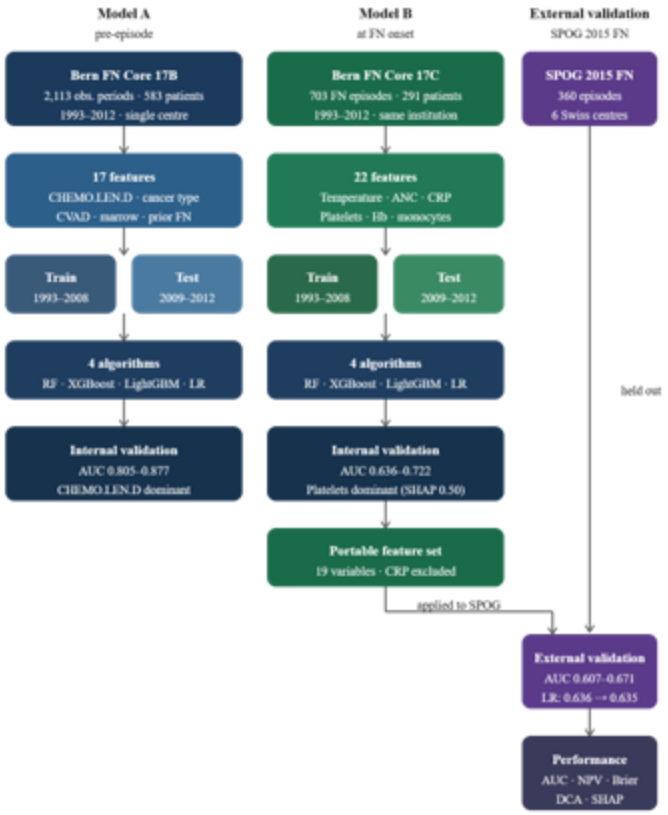


Figure 1

Study design and model development framework. Model A (left column) was derived from observation-period-level data (Project 17B) for pre-episode risk stratification. Model B (centre column) was derived from FN-episode-level data (Project 17C) for at-onset triage. The SPOG 2015 FN study (right column) served as the prospective external validation cohort for both models.

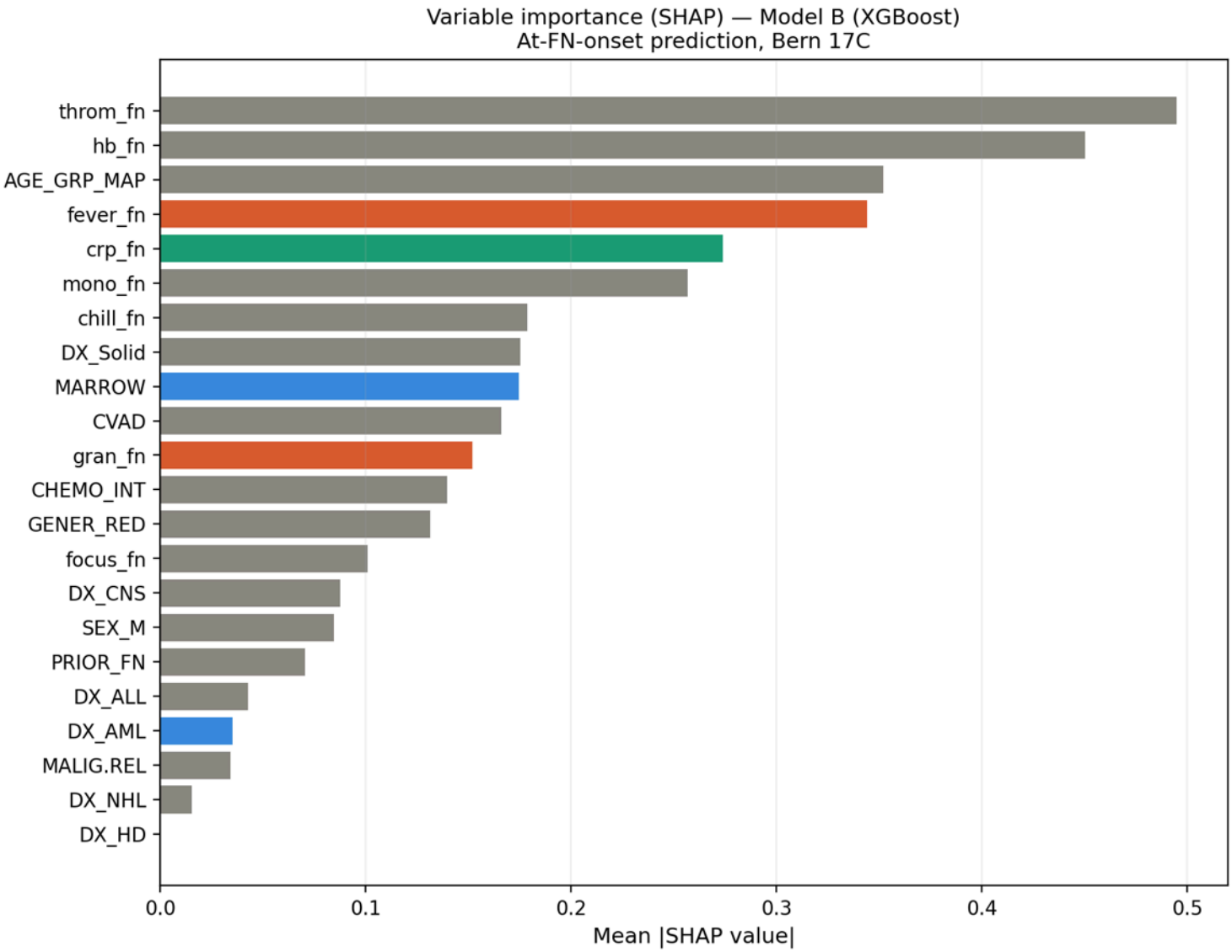


Figure 2

SHAP variable importance for Model B (XGBoost, at-FN-onset prediction, evaluated on the internal 2009-2012 test set). Bars represent mean absolute SHAP values. Orange = key vital signs; green = CRP; blue = structural cancer factors.

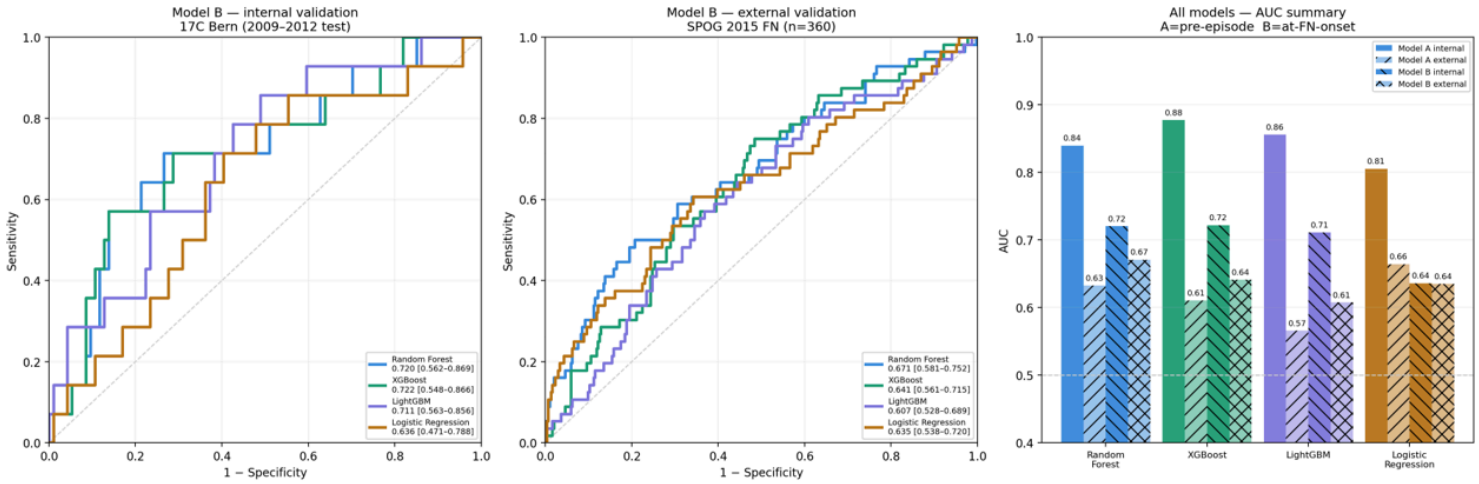


Figure 3

ROC curves and AUC comparison across all four analyses. Left: Model B external validation (SPOG 2015 FN, n=360). Centre: Model B internal validation (Bern 17C, 2009-2012). Right: AUC summary (A-Int = Model A internal; A-Ext = Model A external; B-Int = Model B internal; B-Ext = Model B external). Hatched bars indicate evaluation on the portable feature set.

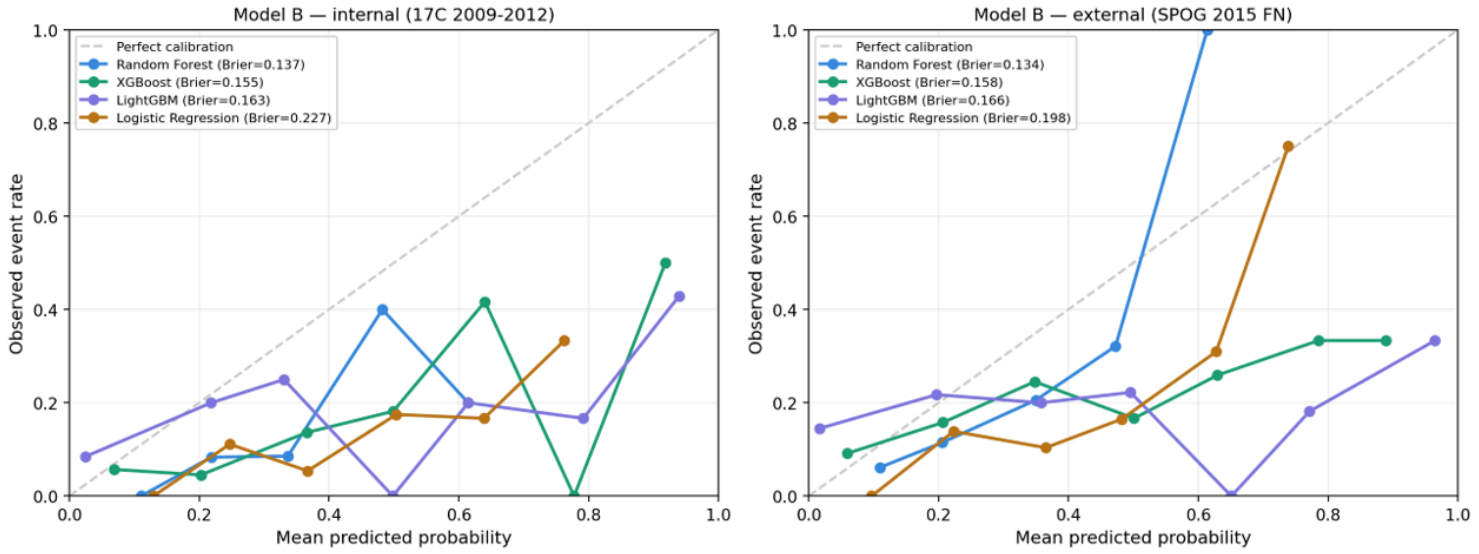


Figure 4

Calibration plots for Model B. Left: internal validation. Right: external SPOG validation. Brier scores in parentheses. Dashed diagonal = perfect calibration.