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.

Risk-based preoperative red-cell ordering in abdominal and pelvic cancer surgery: a retrospective model development and temporal evaluation study

Jingjing Zheng
Ningbo Medical Center Lihuili Hospital
Ying Wang
Ningbo Medical Center Lihuili Hospital
Qingwei zhou
673847620@qq.com

Ningbo Medical Center Lihuili Hospital

Research Article

Keywords: Red-cell crossmatching, Perioperative transfusion, Preoperative risk prediction, Temporal evaluation, Patient blood management, Cancer surgery

Posted Date: August 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10401779/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Procedure-based blood-ordering schedules may not capture patient-level variation in perioperative transfusion risk. We aimed to develop and temporally evaluate a preoperative red-cell transfusion model and quantify the consequences of development-derived crossmatching thresholds.

Methods

This single-centre retrospective study included adults undergoing elective abdominal or pelvic cancer surgery in 2024–2025. Patients treated in 2024 formed the development cohort and those treated in 2025 formed the temporal evaluation cohort. The outcome was any red-cell transfusion from the intraoperative period through 72 h after surgery. Ridge-penalised logistic regression used routinely available preoperative patient and surgical variables. Performance was assessed using discrimination, Brier score, and calibration. Sensitivity-based thresholds were derived only in the development cohort and applied unchanged to the temporal cohort.

Results

Among 1,068 patients, 541 were included in the development cohort and 527 in the temporal cohort; 115 (21.3%) and 86 (16.3%), respectively, received red-cell transfusion within 72 h. The primary continuous-haemoglobin model had an optimism-corrected area under the curve of 0.812 in development and a temporal area under the curve of 0.790 (95% confidence interval 0.738–0.844), compared with 0.713 for the binary-anaemia model and 0.622 for the procedure-only model. The temporal Brier score was 0.114, with a calibration intercept of − 0.480 and slope of 0.781, indicating overall risk overestimation and overly dispersed predictions. The spline model had a temporal area under the curve of 0.808. At the primary 95% sensitivity target, the locked 3.83% threshold recommended crossmatching for 421 patients (79.9%), with sensitivity 97.7% and negative predictive value 98.1%. In simulation, this represented 95 fewer recommended crossmatches than observed practice, while two of 86 transfusion events occurred below the threshold.

Conclusions

A parsimonious model based on routinely available preoperative variables retained discrimination in a subsequent-year cohort, but temporal miscalibration limits direct implementation. Risk-based crossmatching warrants further evaluation as a decision-support strategy, with local recalibration, external validation, and prospective impact assessment required before clinical adoption.

Introduction

Preoperative red-cell ordering is intended to ensure timely access to compatible blood for patients at risk of perioperative transfusion. However, routine crossmatching based on broad surgical categories may exceed subsequent transfusion requirements and increase pretransfusion testing and blood-bank workload.[1–3]

Institution-specific and data-driven maximum surgical blood-ordering schedules have standardised practice and reduced unnecessary blood orders and crossmatching, demonstrating the value of aligning blood preparation with observed surgical utilisation.[1–3] Nevertheless, these schedules remain primarily anchored to average procedure-level demand.[1, 4]

Transfusion risk may vary substantially among patients undergoing the same category of surgery because of differences in preoperative haemoglobin, clinical status, and the planned operative approach.[4–7] Recent patient-specific blood-ordering approaches suggested that incorporating individual clinical characteristics could improve prediction beyond procedure-based schedules.[4–6] Less well characterised are the temporal performance and calibration of such approaches, and the clinical consequences of applying development-derived decision thresholds to subsequent patients.[4–6, 8] A clinically useful strategy therefore needs not only to estimate transfusion risk, but also to translate that estimate into an explicit crossmatching decision while preserving a high level of event capture.[8]

We therefore developed and temporally evaluated a parsimonious model for red-cell transfusion within 72 h among adults undergoing elective abdominal or pelvic cancer surgery. We compared procedure-based and patient-level specifications, including binary, continuous linear, and restricted cubic spline representations of preoperative haemoglobin. We further derived sensitivity-based crossmatching thresholds exclusively in the development cohort and applied them unchanged to a subsequent-year cohort to quantify the trade-off between crossmatching requirements and transfusion events occurring below the threshold.

Methods

Study design and setting

This single-centre retrospective model development and temporal evaluation study was conducted at Ningbo Medical Center Lihuili Hospital. Consecutive adults undergoing elective surgery for an abdominal or pelvic malignant tumour between 1 January 2024 and 31 December 2025 were screened. Patients undergoing emergency surgery, those with benign tumours, and those with incomplete data for any prespecified predictor or the primary outcome were excluded. When a patient underwent more than one eligible operation during the study period, only the first operation was included.

Patients undergoing surgery in 2024 formed the development cohort, whereas those undergoing surgery in 2025 formed the temporal evaluation cohort. The temporal cohort was not used for predictor specification, model fitting, penalty selection, model selection, or risk-threshold determination. No formal sample-size calculation was performed because all eligible patients within the predefined study period were included.

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Ningbo Medical Center Lihuili Hospital (Approval No. KY2025SL438–01). The requirement for informed consent was waived because of the retrospective nature of the study.

Data sources and data quality

Data were obtained from the institutional electronic medical record, anaesthesia information system, laboratory information system, and transfusion-service records. Extracted variables included demographic characteristics, preoperative clinical status, planned and final surgical approach, laboratory measurements, preoperative red-cell crossmatching, and perioperative transfusion. Data extraction and coding were checked by two investigators against the source records, and discrepancies were resolved through joint review.

For haemoglobin, platelet count, and international normalised ratio, the most recent value recorded before surgery was used. Missingness was defined as an absent or null source field. Patients with incomplete data for any prespecified predictor or the outcome were excluded; the resulting analytical dataset was complete, and no statistical imputation was performed.

Outcome

The primary outcome was any red-cell transfusion greater than 0 units during the intraoperative period or within 72 h after surgery. Intraoperative and postoperative red-cell units were combined to define the binary outcome of any versus no transfusion. Preoperative red-cell crossmatching in observed practice was identified from transfusion-service records.

Candidate predictors and model specifications

Predictors were selected on the basis of availability before the intended crossmatching decision and clinical relevance to perioperative transfusion risk. The primary predictor set comprised age, sex, planned surgical approach, previous abdominal or pelvic surgery, American Society of Anesthesiologists physical status, surgical category, and preoperative haemoglobin.

Age and haemoglobin were modelled as continuous variables and entered per 10-year and 10-g/L increments, respectively. Sex was coded as male versus female. Previous abdominal or pelvic surgery was coded as yes versus no, and ASA physical status was coded as III versus I–II. Surgical categories were colorectal, gastric, liver, biliary, pancreatic, and gynaecological surgery, with colorectal surgery as the reference category. Planned surgical approach was coded as open versus minimally invasive according to the intended preoperative operative plan. Patients converted from a minimally invasive to an open approach remained classified according to the planned approach in the primary analysis.

Five candidate models were considered. The procedure-only model included surgical category and planned approach. The binary-anaemia model replaced continuous haemoglobin with anaemia defined as haemoglobin < 130 g/L in men or < 120 g/L in women.[9] The primary model included haemoglobin as a continuous linear predictor. The restricted cubic spline model used the same predictors but represented haemoglobin using a three-knot restricted cubic spline, with knots at 87, 124, and 148 g/L, corresponding to the 10th, 50th, and 90th percentiles in the development cohort.[10] The expanded laboratory model added platelet count, INR, and any preoperative anticoagulant or antiplatelet medication use to the primary model.

Irregular-antibody screening was considered a transfusion-safety and implementation variable and was not included as a predictor of transfusion occurrence.

Model development and selection

All model development and selection procedures were conducted exclusively in the 2024 development cohort. Ridge-penalised logistic regression was used for all candidate models. Continuous variables and indicator variables were standardised using means and standard deviations estimated from the development cohort. The same transformations were subsequently applied unchanged to the temporal cohort. The model intercept was not penalised.

For each candidate model, the ridge penalty parameter, expressed as the inverse regularisation strength C, was selected using fivefold stratified cross-validation in the development cohort by minimising the mean cross-validated log loss over a logarithmically spaced grid. The selected value of C was then held fixed.

Development performance was estimated using stratified tenfold out-of-fold predictions generated with the selected penalty parameter. Candidate models were compared using the mean cross-validated log loss. The one-standard-error eligibility limit was defined as the lowest mean log loss plus one standard error of the model achieving the lowest value. Among models within this limit, the simpler specification was retained. On this basis, the continuous linear-haemoglobin model was selected as the primary model, while the three-knot haemoglobin spline model was retained as a sensitivity analysis. The temporal cohort was not examined during model or threshold selection.

Internal validation and temporal evaluation

The final primary and spline models were fitted using the entire development cohort. Apparent performance was assessed using the area under the receiver operating characteristic curve, Brier score, calibration intercept, and calibration slope. Calibration intercept and slope were estimated by regressing the observed outcome on the logit of the predicted probability.

Internal validation used 500 non-parametric bootstrap resamples.[11] In each resample, preprocessing and model fitting were repeated, while the development-selected penalty parameter was held fixed. Optimism was calculated as the difference between performance in the bootstrap sample and performance of the bootstrap-fitted model in the original development cohort. Mean optimism was subtracted from apparent performance to obtain optimism-corrected estimates.

The fitted model coefficients, predictor transformations, and penalty parameter were then locked and applied unchanged to the 2025 temporal evaluation cohort. Temporal performance was assessed using AUC, Brier score, calibration intercept, calibration slope, and graphical calibration. Ninety-five per cent confidence intervals for temporal AUC, Brier score, calibration intercept, and calibration slope were obtained using 1,000 bootstrap resamples. Differences in AUC and Brier score between the linear and spline models were assessed using paired bootstrap resampling.

Clinical utility was examined using decision curve analysis by comparing the net benefit of each candidate model with crossmatch-all and crossmatch-none strategies across the evaluated range of risk thresholds.

Selection and evaluation of crossmatching thresholds

Development-cohort risk thresholds were derived from the tenfold out-of-fold predictions of the primary model. Target sensitivities of 90%, 95%, 97.5%, and 99% were specified, with 95% designated as the primary operating target. For each target, the largest predicted-risk threshold achieving at least the specified sensitivity in the development cohort was selected.

All thresholds were locked before temporal evaluation and applied unchanged to the 2025 cohort. Preoperative crossmatching was considered model-recommended when the predicted probability was greater than or equal to the corresponding threshold. For each operating point, we calculated the number and proportion of patients recommended for crossmatching, sensitivity, negative predictive value, and the number of transfusion events below the threshold.

The difference in the number of crossmatches was calculated relative to observed preoperative crossmatching practice in the temporal cohort and was also expressed per 100 temporal-cohort patients. Wilson 95% confidence intervals were calculated for sensitivity and negative predictive value.

Sensitivity analyses

Four sensitivity analyses were conducted. First, surgical approach was recoded according to the actual or final approach, with conversions to open surgery classified as open; this was considered exploratory because final approach was not available at the intended prediction time. Second, preoperative haemoglobin was modelled using the three-knot restricted cubic spline. Third, platelet count, INR, and preoperative anticoagulant or antiplatelet medication use were added to the primary predictor set. Each alternative model used its own development-selected ridge penalty and development-locked threshold corresponding to the same target-sensitivity procedure.

Finally, the temporal evaluation of the primary model was repeated after excluding one transfusion event attributed to clearly unanticipated postoperative haemorrhage. For this analysis, the coefficients and primary 3.83% risk threshold remained unchanged.

Statistical analysis

Continuous variables were summarised as mean (standard deviation) or median (interquartile range), and categorical variables as number (percentage). Differences between the development and temporal cohorts were described using absolute standardised mean differences, with values ≥ 0.10 considered potentially meaningful; no significance testing was used to assess cohort comparability.

Ridge-penalised logistic regression models were developed in the 2024 cohort, with penalty parameters selected by cross-validation and internal performance assessed using bootstrap optimism correction. The final model and development-derived thresholds were applied unchanged to the 2025 temporal cohort. Performance was evaluated using the area under the receiver operating characteristic curve, Brier score, calibration intercept and slope, and decision-curve analysis. Threshold performance was summarised using recommended crossmatching, sensitivity, negative predictive value, and transfusion events below the threshold. Ninety-five per cent confidence intervals were reported where applicable. Analyses were performed

using Python 3.14 and relevant scientific-computing packages; software versions and the random seed will be specified after verification of the final analysis code.

Results

Study population and temporal differences

Of 1,351 screened patients, 1,068 were included in the final analysis: 541 in the 2024 development cohort and 527 in the 2025 temporal evaluation cohort (Fig. 1). Red-cell transfusion within 72 h occurred in 115 patients (21.3%) in the development cohort and 86 patients (16.3%) in the temporal evaluation cohort. All prespecified predictors and the primary outcome were complete, and no imputation was required (Supplementary Table S1).

Compared with the development cohort, the temporal evaluation cohort had higher preoperative haemoglobin, lower anaemia prevalence and platelet count, greater preoperative anticoagulant or antiplatelet medication use, and a lower 72-h red-cell transfusion rate. Differences were also observed in the distribution of ASA physical status. Absolute SMDs were < 0.10 for age, sex, previous abdominal or pelvic surgery, surgical category, planned surgical approach, INR, and irregular-antibody screening.

Among 1,351 consecutive adults undergoing abdominal or pelvic tumour surgery between January 2024 and December 2025, 214 with benign tumours and 69 with incomplete predefined variables were excluded. The final analytical cohort included 1,068 patients, comprising 541 in the 2024 development cohort and 527 in the 2025 temporal evaluation cohort.

Table 1
Characteristics of the development and temporal evaluation cohorts

Characteristic	Development cohort (n = 541)	Temporal evaluation cohort (n = 527)	Absolute SMD
Patient characteristics			
Age, years, mean (SD)	67.6 (11.6)	67.3 (11.3)	0.028
Male sex, n (%)	316 (58.4%)	302 (57.3%)	0.022
Previous abdominal or pelvic surgery, n (%)	155 (28.7%)	159 (30.2%)	0.033
Surgical characteristics			
Surgical category, n (%)			
Colorectal	125 (23.1%)	124 (23.5%)	0.010
Gastric	125 (23.1%)	125 (23.7%)	0.014
Liver	109 (20.1%)	113 (21.4%)	0.032
Biliary	28 (5.2%)	21 (4.0%)	0.057
Pancreatic	88 (16.3%)	85 (16.1%)	0.004
Gynaecological	66 (12.2%)	59 (11.2%)	0.031
Planned open surgical approach, n (%)	134 (24.8%)	135 (25.6%)	0.020
ASA physical status, n (%)			
I	25 (4.6%)	9 (1.7%)	0.167
II	460 (85.0%)	475 (90.1%)	0.155
III	56 (10.4%)	43 (8.2%)	0.076
Preoperative laboratory variables			
Preoperative haemoglobin, g/L, mean (SD)	120.6 (22.8)	123.9 (22.4)	0.145
Preoperative anaemia, n (%)	278 (51.4%)	242 (45.9%)	0.110
Platelet count, $\times 10^9$/L, median (IQR)	227.0 (171.0–282.0)	211.0 (165.5–265.0)	0.175
INR, median (IQR)	1.00 (0.95–1.05)	0.99 (0.94–1.05)	0.008
Anticoagulant or antiplatelet medication, n (%)	56 (10.4%)	76 (14.4%)	0.124
Positive irregular-antibody screen, n (%)	2 (0.4%)	3 (0.6%)	0.029

Characteristic	Development cohort (n = 541)	Temporal evaluation cohort (n = 527)	Absolute SMD
Red-cell transfusion within 72 h, n (%)	115 (21.3%)	86 (16.3%)	0.127

Data are presented as mean (SD), median (IQR), or n (%). Standardized mean differences are reported as absolute values. The SMD for continuous variables was calculated using the difference in means divided by the pooled standard deviation, including variables presented as median (IQR). An absolute SMD ≥ 0.10 was considered to indicate a potentially meaningful temporal difference. Preoperative anaemia was defined as haemoglobin < 130 g/L in men and < 120 g/L in women. Planned surgical approach was based on the preoperative operative plan.

Model development and internal validation

The primary continuous linear-haemoglobin Ridge model had an apparent AUC of 0.829 and Brier score of 0.113 in the development cohort. After bootstrap optimism correction, the AUC was 0.812 and the Brier score was 0.119, with a calibration intercept of − 0.024 and slope of 0.977. The restricted cubic spline haemoglobin model had an apparent AUC of 0.850 and an optimism-corrected AUC of 0.832; the corresponding Brier score, calibration intercept, and slope were 0.116, − 0.019, and 0.978, respectively (Table 2). The full coefficient set and clinical-scale implementation formula are provided in Supplementary Table S2. Candidate-model comparisons are presented in Supplementary Table S3; the expanded laboratory model had a temporal AUC of 0.788 and a Brier score of 0.114.

Table 2
Performance of the transfusion-risk models

Model	Evaluation	AUC	Brier score	Calibration intercept	Calibration slope
Continuous linear Hb Ridge	Development cohort, apparent	0.829	0.113	0.067	1.066
	Bootstrap optimism-corrected	0.812	0.119	−0.024	0.977
	2025 temporal evaluation	0.790 (0.738–0.844)	0.114 (0.094–0.131)	−0.480 (− 0.806 to − 0.154)	0.781 (0.596–0.966)
Hb RCS Ridge	Development cohort, apparent	0.850	0.109	0.080	1.076
	Bootstrap optimism-corrected	0.832	0.116	−0.019	0.978
	2025 temporal evaluation	0.808 (0.758–0.858)	0.107 (0.088–0.125)	−0.482 (− 0.810 to − 0.153)	0.808 (0.624–0.991)

Both models included age, sex, planned surgical approach, previous abdominal or pelvic surgery, ASA physical status ≥ III, surgical category, and preoperative haemoglobin. The primary model treated haemoglobin as a

continuous linear predictor. The sensitivity model used a restricted cubic spline with knots at 87, 124, and 148 g/L. Ridge penalties were selected by fivefold cross-validation in the development cohort. Ideal values for the calibration intercept and slope are 0 and 1, respectively.

Temporal model performance and clinical utility

In the 2025 temporal evaluation cohort, the primary linear-haemoglobin Ridge model had an AUC of 0.790, compared with 0.713 for the binary-anaemia model and 0.622 for the procedure-only model. The restricted cubic spline haemoglobin model had an AUC of 0.808. For the primary model, the calibration intercept was −0.480 and the calibration slope was 0.781 (Table 2; Fig. 2A and B). Decision curve analysis compared the candidate models with the crossmatch-all and crossmatch-none strategies (Fig. 2C). Predicted-risk distributions according to observed transfusion status are shown in Supplementary Figure S1, and marginal predicted risks across preoperative haemoglobin values are shown in Supplementary Figure S2.

A, Receiver operating characteristic curves in the 2025 temporal evaluation cohort. B, Calibration of the primary continuous linear-haemoglobin Ridge model; the dashed line indicates ideal calibration. C, Decision curve analysis comparing the candidate models with crossmatch-all and crossmatch-none strategies. AUC, area under the receiver operating characteristic curve.

Clinical consequences of risk-based thresholds

Across the prespecified targets, recommended crossmatching ranged from 294 patients (55.8%) at the 90% target to 501 (95.1%) at the 99% target, while missed transfusion events decreased from 10 to 1. At the prespecified 95% target, the locked 3.83% threshold recommended crossmatching for 421 patients (79.9%), with a sensitivity of 97.7%, an NPV of 98.1%, and two missed events. This corresponded to 95 fewer crossmatches than observed practice, or 18.0 fewer per 100 patients (Table 3). The corresponding trade-off between crossmatching requirements and missed events across sensitivity targets is shown in Fig. 3.

Table 3
Clinical consequences of development-locked, risk-based crossmatching thresholds in the 2025 temporal evaluation cohort

Target sensitivity	Risk threshold	Crossmatched, n (%)	Sensitivity, % (95% CI)	NPV, % (95% CI)	Missed events, n	Crossmatches avoided, n(%)	Avoided per 100 patients
90%	8.21%	294 (55.8%)	88.4 (79.9–93.6)	95.7 (92.3–97.7)	10	222 (43.0%)	42.1
95%	3.83%	421 (79.9%)	97.7 (91.9–99.4)	98.1 (93.4–99.5)	2	95 (18.4%)	18
97.5%	3.05%	453 (86.0%)	97.7 (91.9–99.4)	97.3 (90.7–99.3)	2	63 (12.2%)	12
99%	1.69%	501 (95.1%)	98.8 (93.7–99.8)	96.2 (81.1–99.3)	1	15 (2.9%)	2.8

Risk thresholds were selected exclusively in the 2024 development cohort using tenfold out-of-fold predictions. For each prespecified target, the largest threshold achieving at least the target sensitivity was locked and applied unchanged to the 2025 temporal evaluation cohort. Crossmatching was recommended when predicted risk was greater than or equal to the locked threshold. Reductions were calculated relative to the 516 patients who underwent preoperative red-cell crossmatching in observed 2025 practice. The reduction per 100 patients used all 527 temporal-cohort patients as the denominator. Confidence intervals for sensitivity and negative predictive value were calculated using the Wilson method.

A, Proportion of patients in the 2025 temporal evaluation cohort recommended for crossmatching at each development-locked sensitivity target. B, Proportion of observed transfusion events occurring below the corresponding risk threshold. The shaded band indicates the prespecified primary target of 95%. Labels show percentages and corresponding counts.

Sensitivity analyses

Exploratory coding based on the actual or final surgical approach yielded a temporal AUC of 0.796 and a Brier score of 0.111, with 418 patients recommended for crossmatching and two transfusion events below the model-specific threshold. The expanded laboratory model yielded an AUC of 0.788 and a Brier score of 0.114, with 432 patients recommended for crossmatching and two events below the threshold. At its model-specific development-locked threshold, the restricted cubic spline model recommended crossmatching for 343 patients (65.1%), with six transfusion events below the threshold. After exclusion of one clearly unanticipated postoperative haemorrhage, the primary model had an AUC of 0.799 and a Brier score of 0.112, with one event below the unchanged primary threshold (Supplementary Table S4).

Discussion

In this retrospective model development and temporal evaluation study, a parsimonious model based on routinely available preoperative variables retained discrimination for red-cell transfusion within 72 h in the subsequent-year cohort. The primary model outperformed procedure-only and binary-anaemia approaches, whereas the more flexible spline specification provided limited incremental gain. Development-locked thresholds identified distinct trade-offs between crossmatching requirements and event capture. These findings support further evaluation of risk-based crossmatching as a perioperative decision-support strategy.

Procedure-based maximum surgical blood-ordering schedules have standardised preoperative preparation and reduced unnecessary crossmatching, but remain anchored to average procedure-level blood utilisation.[12, 13] Earlier patient-specific approaches suggested that incorporating clinical characteristics could improve prediction beyond procedure alone.[14–16] Building on this work, we combined surgical category and planned approach with routinely available preoperative variables, and evaluated temporal performance and the consequences of development-locked thresholds. The model is intended to support decisions about whether red cells should be crossmatched before surgery, rather than to determine transfusion indications or dose, and should therefore complement existing institutional blood-ordering pathways.

Treating haemoglobin as a continuous predictor preserved information that was lost through binary anaemia classification. Although the spline specification showed modestly better overall performance, the linear model

remained eligible under the development-defined one-standard-error rule and was retained as the simpler specification. The small temporal difference in AUC did not provide a compelling basis for accepting the additional spline complexity. Adding platelet count, INR, and preoperative antithrombotic medication did not improve temporal performance, further supporting a parsimonious predictor set. In the subsequent-year cohort, discrimination was retained, but the negative calibration intercept indicated overall risk overestimation and the slope below unity indicated overly dispersed predictions. Because crossmatching decisions depend on absolute probabilities rather than ranking alone, local recalibration and threshold reassessment may be required before use in later periods or other institutions.[17, 19]

The threshold analysis is best interpreted as a policy trade-off rather than as identification of a universally applicable cut-off. At the prespecified 95% operating point, the model would have recommended 95 fewer crossmatches than observed practice, equivalent to 18.0 fewer per 100 patients, while two of 86 transfusion events occurred below the threshold. Lower sensitivity targets produced larger reductions in crossmatching but more events below the threshold, whereas the 99% target approached near-universal crossmatching. The preferred operating point will therefore depend on local blood-bank capacity, turnaround times, and the acceptable balance between resource use and event capture.[18] Accordingly, the model should support rather than replace clinical judgement, predefined override criteria, and rapid access to red cells when requirements change unexpectedly.

This study has several strengths, including the use of predictors available at the intended preoperative decision point, coding based on the planned rather than final surgical approach, temporal evaluation in a complete subsequent-year cohort, and operating thresholds locked in the development data. However, this was a single-centre retrospective study, and both the observed transfusion outcome and the simulated consequences of risk-based crossmatching reflect local clinical practice and blood-bank logistics. The heterogeneous surgical cohort broadened the intended clinical scope but did not permit precise assessment of performance within individual procedure categories. The estimated reduction in crossmatching was simulated and does not demonstrate actual changes in workload, cost, blood availability, or treatment delays. Temporal miscalibration further limits direct transportability of the model. External validation, local recalibration, and prospective impact evaluation are therefore required to determine whether this approach can improve crossmatching efficiency while maintaining timely access to red cells.[17, 19, 20]

Conclusions

A parsimonious model based on routinely available preoperative variables retained discrimination for red-cell transfusion in a subsequent-year cohort. Development-derived thresholds demonstrated explicit trade-offs between crossmatching requirements and event capture, but temporal miscalibration limits direct implementation. External evaluation, local recalibration, and prospective impact assessment are required before clinical adoption.

Declarations

Human Ethics and Consent to Participate declarations

This retrospective cohort study was conducted in compliance with the principles of the Declaration of Helsinki. The study protocol was approved by the Ethics Committee of Ningbo Medical Center Lihuili Hospital (approval no. No. KY2025SL438–01). Given the retrospective nature of the study, the requirement for informed consent was waived by the ethics committee.

The name of ethics committe: Ningbo Medical Center Lihuili Hospital Ethics Committee.

The institute: Ningbo Medical Center Lihuili Hospital.

Consent for Publication

Not Applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Competing Interests

We have reviewed the journal's policy on competing interests and hereby confirm the following. All authors of this manuscript declare that they have no financial or personal relationships with other people or organizations that could inappropriately influence (bias) their work. This includes, but is not limited to, employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and grants or other funding relevant to the subject of this study.

Clinical trial number: not applicable.

Funding

None.

Author Contributions

Conception and design: Q W.

Administrative support: J Z.

Provision of study materials or patients: X Y, Q W.

Collection and assembly of data: J Z, Y W.

Data analysis and interpretation: J Z, Y W.

Manuscript writing: All authors.

Final approval of manuscript: All authors.

References

1. Benzil DL, Auron M, Otrock ZK, Lallo D, Flowers N, Cummings K 3, et al. A maximum surgical blood ordering schedule: does it add value? Vox Sang. 2025;120:411–8.
2. Guduri PR, Shastry S, Raturi M, Shenoy A. Surgical blood ordering schedule for better inventory management: an experience from a tertiary care transfusion center. Med J Armed Forces India. 2022;78:283–90.
3. Tan PP, Abdul Rahman J, Mat Noh S, Mohd Yasin I, Mohd Noor S. Implementation of maximum surgical blood ordering schedule in a tertiary hospital in Malaysia during COVID-19 pandemic. Transfus Apher Sci. 2021;60:103280.
4. Lou SS, Liu H, Lu C, Wildes TS, Hall BL, Kannampallil T. Personalized surgical transfusion risk prediction using machine learning to guide preoperative type and screen orders. Anesthesiology. 2022;137:55–66.
5. Zapf MAC, Fabbri DV, Andrews J, Li G, Freundlich RE, Al-Droubi S, et al. Development of a machine learning model to predict intraoperative transfusion and guide type and screen ordering. J Clin Anesth. 2023;91:111272.
6. Hur S, Yoo J, Min JY, Jeon YJ, Cho JH, Seo JY, et al. Development, validation, and usability evaluation of machine learning algorithms for predicting personalized red blood cell demand among thoracic surgery patients. Int J Med Inf. 2024;191:105543.
7. Shander A, Corwin HL, Meier J, Auerbach M, Bisbe E, Blitz J et al. Recommendations from the International Consensus Conference on Anemia Management in Surgical Patients (ICCAMS). Ann Surg. 2023;277:581–90.
8. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17:230.
9. World Health Organization. Guideline on haemoglobin cutoffs to define anaemia in individuals and populations. Geneva: World Health Organization; 2024.
10. Ma J, Dhiman P, Qi C, Bullock G, van Smeden M, Riley RD, et al. Poor handling of continuous predictors in clinical prediction models using logistic regression: a systematic review. J Clin Epidemiol. 2023;161:140–51. 10.1016/j.jclinepi.2023.07.017.
11. Iba K, Shinozaki T, Maruo K, Noma H. Re-evaluation of the comparative effectiveness of bootstrap-based optimism correction methods in the development of multivariable clinical prediction models. BMC Med Res Methodol. 2021;21:9. 10.1186/s12874-020-01201-w.
12. Shaikh OH, Bhattarai S, Shankar VG, Basavarajegowda A. Blood ordering and utilization in patients undergoing elective general surgery procedures in a tertiary care hospital: a prospective audit. Natl Med J India. 2022;35(2):68–73. 10.25259/NMJI_543_19.
13. Zhao Y, Li X, Wang Y, Chen Y, Li D, Jiang Q, et al. Maximum surgical blood order schedule for flap reconstruction in oral and maxillofacial cancer patients. BMC Oral Health. 2022;22:322. 10.1186/s12903-022-02357-1.
14. Huang H, Cao M. Development and validation of a nomogram to predict intraoperative blood transfusion for gastric cancer surgery. Transfus Med. 2021;31(4):250–61. 10.1111/tme.12777.

15. Zhang Y, Fu X, Xie X, Yan D, Wang Y, Huang W, et al. A novel model forecasting perioperative red blood cell transfusion. Sci Rep. 2022;12:16127. 10.1038/s41598-022-20543-7.
16. Zang H, Hu A, Xu X, Ren H, Xu L. Development of machine learning models to predict perioperative blood transfusion in hip surgery. BMC Med Inf Decis Mak. 2024;24:158. 10.1186/s12911-024-02555-7.
17. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ. 2024;384:e074819. 10.1136/bmj-2023-074819.
18. Sadatsafavi M, Adibi A, Puhan M, Gershon A, Aaron SD, Sin DD. Moving beyond AUC: decision curve analysis for quantifying net benefit of risk prediction models. Eur Respir J. 2021;58(5):2101186. 10.1183/13993003.01186-2021.
19. Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. BMC Med. 2023;21:70. 10.1186/s12916-023-02779-w.
20. Arshi B, Cowley LE, Rijnhart E, Reeve K, Smits LJ, Wynants L. External validation, impact assessment and clinical utilization of clinical prediction models: a prospective cohort study. J Clin Epidemiol. 2025;186:111902. 10.1016/j.jclinepi.2025.111902.

Figures

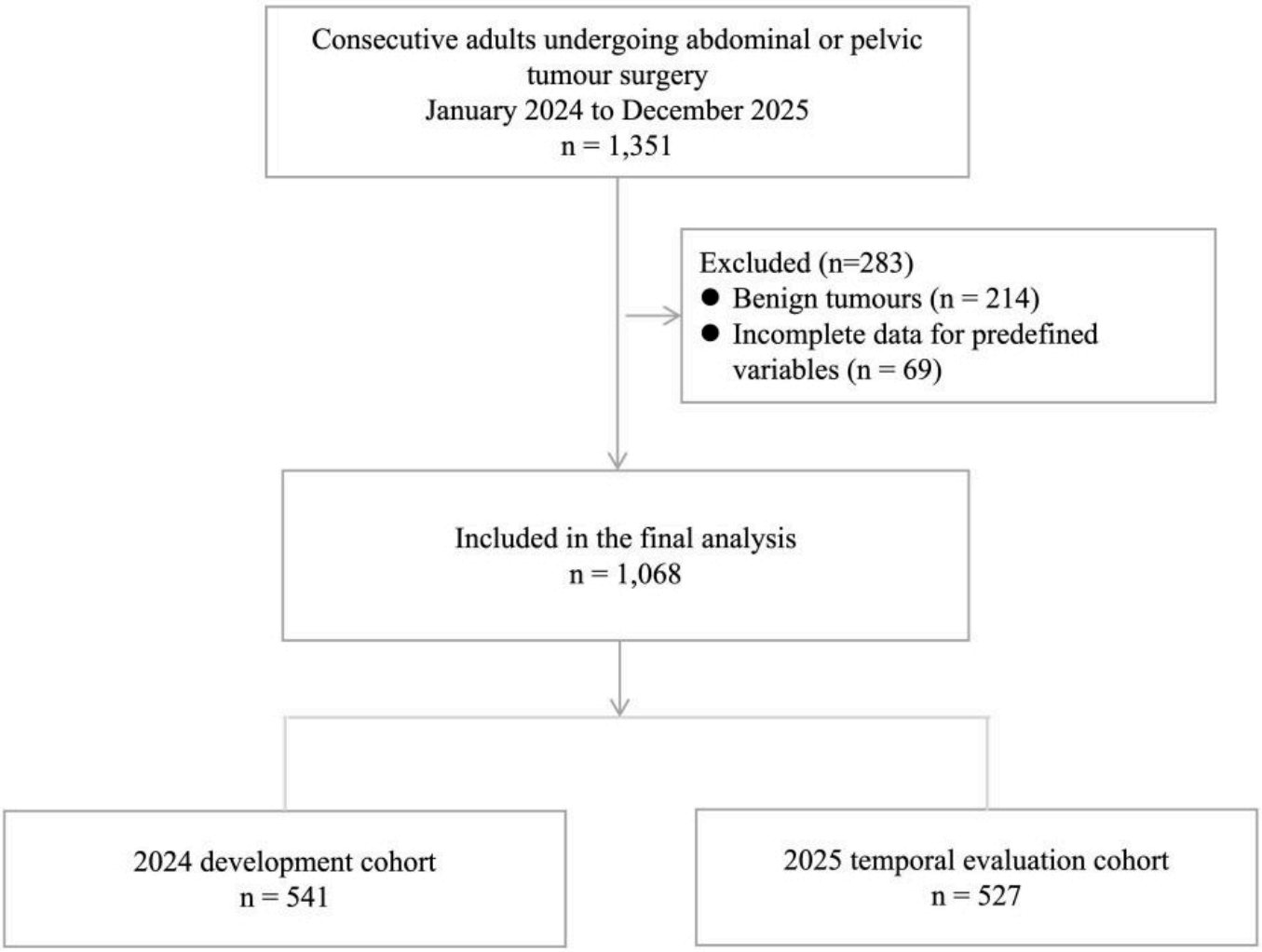


Figure 1

Flowchart of study population selection and cohort allocation.

Among 1,351 consecutive adults undergoing abdominal or pelvic tumour surgery between January 2024 and December 2025, 214 with benign tumours and 69 with incomplete predefined variables were excluded. The final analytical cohort included 1,068 patients, comprising 541 in the 2024 development cohort and 527 in the 2025 temporal evaluation cohort.

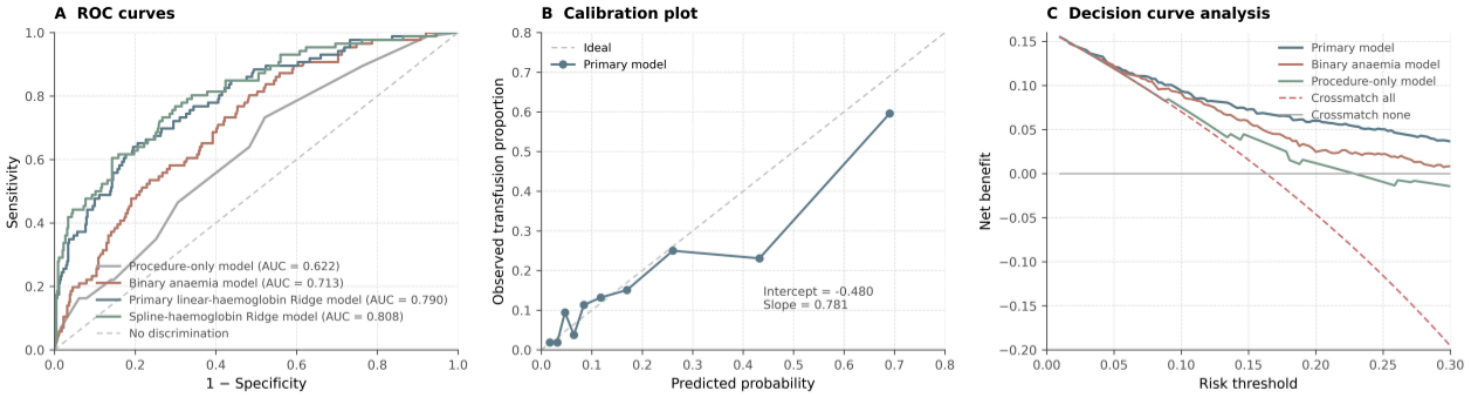


Figure 2

Temporal performance and clinical utility of the transfusion-risk models.
A, Receiver operating characteristic curves in the 2025 temporal evaluation cohort. B, Calibration of the primary continuous linear-haemoglobin Ridge model; the dashed line indicates ideal calibration. C, Decision curve analysis comparing the candidate models with crossmatch-all and crossmatch-none strategies. AUC, area under the receiver operating characteristic curve.

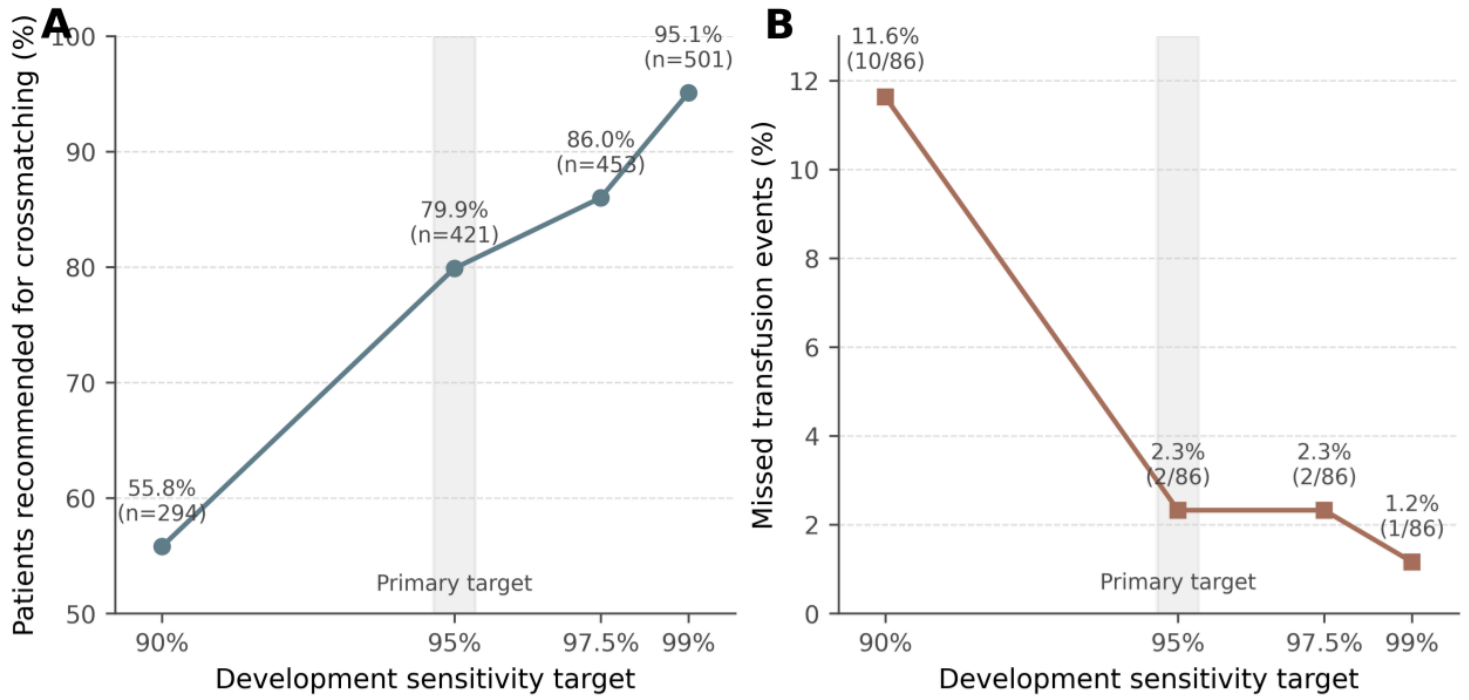


Figure 3

Crossmatching requirements and missed transfusion events across development sensitivity targets.
A, Proportion of patients in the 2025 temporal evaluation cohort recommended for crossmatching at each development-locked sensitivity target. B, Proportion of observed transfusion events occurring below the corresponding risk threshold. The shaded band indicates the prespecified primary target of 95%. Labels show percentages and corresponding counts.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable20260715.docx