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.

Development and temporal external validation of a prediction model for unplanned readmission in hospitalized patients with cancer and coronary heart disease

Nuoyi Wu
Army Medical University
Xinlian Zhang
Army Medical University
Mei Yang
Army Medical University
Wenyan Yang
Army Medical University
Wenli Yu
Army Medical University
Yuyi Yang
Army Medical University
Juan Liu
Army Medical University
Qian Wen
wenqian0312@tmmu.edu.cn
Army Medical University

Research Article

Keywords: cancer, coronary heart disease, unplanned readmission, prediction model, nomogram

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11046775/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

Unplanned readmission is a substantial burden in patients with cancer, particularly when coronary heart disease (CHD) coexists, yet no risk tool has been developed for this population. We developed and temporally validated a prediction model for unplanned readmission based on routine data available at admission.

Methods

We conducted a single-center retrospective cohort study of adults with cancer and CHD discharged between January 2022 and June 2024 (438 admissions; 407 unique patients). The outcome was unplanned readmission (33 events, 7.5%). Thirtyfour candidate predictors with < 5% missing data were screened by univariable analysis and least absolute shrinkage and selection operator (LASSO) regression with 10-fold × 5-repeat cross-validation (1-SE rule) and refitted in an unpenalized multivariable logistic model. Discrimination was assessed with the c-index and 1000-bootstrap optimism correction; calibration with the Hosmer-Lemeshow test, calibration slope, and intercept; and clinical utility with decision curve analysis. The model was applied without refitting to a temporally non-overlapping cohort discharged from October 2023 to June 2024 (n = 118; 6 events).

Results

The final model comprised serum sodium, diastolic blood pressure, hemoglobin, and digestive cancer (overall likelihood-ratio test $\chi 2 = 16.55$, $P = 0.002$). The apparent and optimism-corrected c-indices were 0.709 and 0.682 (95% confidence interval [CI] 0.599–0.766), and the temporal validation c-index was 0.841 (bootstrap 95% CI 0.643–0.985). Calibration was adequate in both cohorts (Hosmer-Lemeshow $P = 0.932$ and 0.606). At the Youden-optimal threshold (predicted probability 0.089), sensitivity and specificity were 0.61 and 0.75. The model outperformed the Padua score (c-index 0.538) and NRS2002 (0.502).

Conclusion

A four-variable model using routine admission data provides moderate discrimination with adequate calibration for unplanned readmission in patients with cancer and CHD and remained

1. Introduction

Unplanned readmission after hospital discharge is a leading indicator of care quality and a major driver of health care costs. In the US Medicare population, approximately one in five hospitalizations is

followed by a readmission within 30 days, at an annual cost of billions of dollars [1]. Readmissions reflect not only progression of the underlying disease but also failures in discharge planning, medication management, and social support; Krumholz has conceptualized this vulnerability as the "post-hospital syndrome," an acquired, transient state of generalized risk [2]. Patients with cancer are particularly vulnerable: treatmentrelated myelosuppression, malnutrition, and comorbidity burden markedly increase the risk of hospitalization and readmission compared with general medical populations [3]. Existing readmission tools, most notably the LACE index, were developed in general internal medicine populations, and their performance in oncology patients is limited [4].

Patients with coexisting cancer and coronary heart disease (CHD) represent a distinctive subgroup at the intersection of cardiology and oncology. The 2022 European Society of Cardiology (ESC) guidelines on cardio-oncology explicitly note that comorbidity management in these patients is poorly supported by evidence and recommend multidimensional risk assessment-including nutrition, thrombosis, and cardiac status-during hospitalization [5]. However, no prediction tool has been developed or validated specifically for unplanned readmission in this population. The scores in routine clinical use were designed for different outcomes: the Padua prediction score targets venous thromboembolism [6] and the Nutritional Risk Screening 2002 (NRS2002) targets nutritional risk [7]. Whether these instruments carry any predictive information for readmission in cancer patients with CHD is untested, and their role in readmission risk stratification is unclear.

Against this background, we aimed to (i) develop a prediction model for unplanned readmission using only data available at admission in a cohort of hospitalized patients with cancer and CHD, and (ii) evaluate its performance in a temporally non-overlapping cohort from the same institution, following the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) reporting standard [8, 9]. We also benchmarked the model against the Padua score and NRS2002.

2. Methods

2.1 Study design and population

We conducted a single-center retrospective cohort study at the Department of Special Medical Service, Army Medical Center of PLA, Chongqing, China. Consecutive adults (≥ 18 years) with (i) a pathologically or radiologically confirmed malignant tumor and (ii) established CHD-defined as prior myocardial infarction, coronary revascularization, or coronary stenosis ≥ 50% on angiography-discharged between January 2022 and June 2024 were eligible. Patients who died in hospital, discharged against medical advice, or lacked outcome or discharge date data were excluded. The cohort comprised 438 admissions from 407 unique patients (26 patients admitted ≥ 2 times during the study window). The study was approved by the Ethics Committee of the Army Medical University (Medical Research Ethics Review No. 271, 2023); informed consent was waived owing to the retrospective design. The study protocol was registered at the Chinese Clinical Trial Registry (ChiCTR2400088966).

2.2 Outcome

The outcome was unplanned readmission to the same institution after discharge, recorded as a

binary indicator in the medical records. Planned admissions (e.g., scheduled chemotherapy cycles) were not counted as events.

2.3 Predictors and data sources

Data were extracted from the electronic medical record. The candidate set comprised 34 variables available within 24 h of admission with < 5% missing data: demographics and lifestyle (age, sex, body mass index, smoking, alcohol use), comorbidities (hypertension, diabetes, prior surgery, prior antitumor therapy), vital signs (systolic and diastolic blood pressure, heart rate), commonly used scores (Padua score, NRS2002), blood counts (white blood cells, hemoglobin, platelets, lymphocytes), biochemistry (total protein, albumin, globulin, total bilirubin, creatinine, uric acid, urea, potassium, sodium, chloride), coagulation (fibrinogen, D-dimer, prothrombin time, activated partial thromboplastin time, thrombin time), and cancer site grouped by organ system (lung, digestive tract, or other). Laboratory tests with > 20% missing data (e.g., tumor markers, cardiac biomarkers, left ventricular ejection fraction) were excluded a priori. Systematic data-quality checks identified 22 erroneous or implausible values across 10 variables: 15 decimal-point or unit errors were corrected on the basis of internal consistency (e.g., sodium 1374.4 mmol/L to 137.44 mmol/L; cross-checks between cell counts and derived indices); 5 values were unrecoverable and set to missing; and 2 D-dimer values reported as "<21" were assigned the lower detection limit. The complete log is provided as **Supplementary Table S1.**

2.4 Statistical analysis

Continuous variables are summarized as medians (interquartile range [IQR]) and compared with the Mann-Whitney *U* test; categorical variables are summarized as frequencies (percentages) and compared with the χ^2 or Fisher exact test. Proportions are reported with exact (Clopper-Pearson) 95% confidence intervals.

2.4.1 Variable selection

All 34 candidates were first examined by univariable logistic regression. All candidates (standardized) then entered a least absolute shrinkage and selection operator (LASSO) regression [10, 11]; the penalty was chosen by 10-fold × 5-repeat stratified cross-validation minimizing binomial deviance, applying the 1-SE rule (the largest penalty whose deviance is within one standard error of the minimum). A pre-specified rule stipulated that if LASSO retained more than six variables (exceeding the events-per-variable constraint) or none, the candidate set would revert to variables with univariable $P < 0.05$. The selected variables were refitted in an unpenalized multivariable logistic model on the complete data (n = 435) to avoid shrinkage-induced bias in the coefficients. Events per variable (EPV) was 33/4 = 8.25 and is reported in accordance with current sample-size recommendations [12]. Model fit was assessed with likelihood-ratio tests for the full model and for the joint contribution of predictor subsets.

2.4.2 Model performance and validation

Discrimination was quantified with the c-index (equivalent to the area under the receiver operating characteristic [ROC] curve [13]); overall accuracy with the Brier score. Internal validation used 1000 bootstrap resamples with optimism correction [14]: in each resample the model was refit and the difference between apparent and bootstrap-predicted discrimination on the original data was averaged. Calibration was evaluated with calibration curves (deciles internally; quartiles with Wilson 95% CIs in the validation cohort), the calibration slope, the calibration-in-the-large intercept, and the Hosmer-Lemeshow test [15]. Clinical utility was examined with decision curve analysis (DCA) [16], and the Youden index defined the optimal probability threshold [13].

2.4.3 Temporal validation

Discharges from October 2023 to June 2024 (n = 118; 6 events; 117 complete cases) constituted a temporally non-overlapping validation cohort. Model coefficients were applied without modification; the c-index (bootstrap percentile 95% CI), calibration slope and intercept, Brier score, and DCA were computed in this cohort. Baseline characteristics were compared between the two periods to characterize temporal drift.

2.4.4 Sensitivity analyses

(i) Restriction to first admissions per patient (n = 407; 27 events) to address correlated repeated admissions; (ii) forced adjustment for age and sex; (iii) tests of linearity of continuous predictors via quadratic Wald tests. Analyses used Python 3.10 (pandas, statsmodels, scikit-learn, SciPy); random seeds were fixed. Two-sided $P < 0.05$ was considered significant. Reporting follows the TRIPOD statement [8, 9].

3. Results

3.1 Cohort and baseline characteristics

Table 1
Baseline characteristics of the development period and the temporal validation cohort.

Variable	Development period (Jan 2022-Sep 2023; n = 320)	Temporal validation (Oct 2023-Jun 2024; n = 118)	*P*
Age (years)	70.0 (62.5, 77.0)	70.0 (61.0, 77.0)	0.987
Male sex	201 (62.8%)	73 (61.9%)	0.944
BMI (kg/m2)	23.3 (21.1, 25.2)	23.7 (21.5, 26.1)	0.278
Smoking (former/current)	143 (44.7%)	49 (41.5%)	0.629
Hypertension	198 (61.9%)	70 (59.3%)	0.707
Diabetes mellitus	84 (26.2%)	36 (30.5%)	0.444
Prior surgery	247 (77.2%)	93 (78.8%)	0.816
Prior antitumor therapy	94 (29.4%)	48 (40.7%)	0.033
Systolic BP (mmHg)	128.0 (115.0, 142.0)	127.5 (114.5, 140.0)	0.567
Diastolic BP (mmHg)	75.0 (65.0, 84.0)	78.0 (69.0, 83.0)	0.393
Heart rate (bpm)	80.0 (72.0, 90.0)	79.0 (71.0, 88.0)	0.432
Padua score	3.0 (1.0, 4.0)	3.0 (2.0, 4.0)	0.002
NRS2002	1.0 (1.0, 2.0)	1.0 (1.0, 2.0)	0.505
White blood cells (x10^9/L)	6.3 (5.0, 7.9)	6.1 (4.9, 7.5)	0.372
Hemoglobin (g/L)	128.0 (115.0, 140.0)	128.5 (114.2, 141.0)	0.617
Platelets (x10^9/L)	204.0 (154.2, 254.0)	200.0 (164.0, 247.0)	0.980
Lymphocytes (x10^9/L)	1.4 (1.0, 1.7)	1.4 (1.0, 1.7)	0.902
Total protein (g/L)	67.5 (62.0, 72.7)	66.3 (61.8, 71.5)	0.264
Albumin (g/L)	38.5 (35.1, 41.9)	38.3 (35.8, 41.6)	0.941
Total bilirubin (umol/L)	11.2 (9.0, 14.9)	12.2 (9.3, 15.7)	0.115
Creatinine (umol/L)	71.0 (57.4, 87.1)	74.0 (60.1, 92.9)	0.365
Urea (mmol/L)	5.7 (4.4, 7.2)	6.1 (4.6, 7.4)	0.518
Potassium (mmol/L)	4.1 (3.8, 4.3)	4.2 (3.9, 4.4)	0.119
Serum sodium (mmol/L)	139.9 (137.9, 141.6)	140.0 (138.1, 142.7)	0.155
Chloride (mmol/L)	104.3 (101.7, 106.8)	103.6 (101.7, 106.1)	0.068

Variable	Development period (Jan 2022-Sep 2023; n = 320)	Temporal validation (Oct 2023-Jun 2024; n = 118)	*P*
D-dimer (ng/mL)	187.4 (99.0, 354.2)	196.5 (96.0, 352.8)	0.941
Fibrinogen (g/L)	3.3 (2.9, 3.8)	3.4 (2.9, 4.1)	0.070
Cancer type			0.017
lung	134 (42%)	40 (34%)	
digestive	81 (25%)	33 (28%)	
other	46 (14%)	8 (7%)	
reproductive	24 (8%)	19 (16%)	
urinary	19 (6%)	11 (9%)	
hematologic	16 (5%)	7 (6%)	

Continuous variables are medians (interquartile range); categorical variables are n (%). *P* values: Mann–Whitney U test for continuous, χ^2 test for categorical variables.

Of 438 admissions, 33 (7.5%, 95% CI 5.2–10.4) ended in unplanned readmission: 27 of 320 (8.4%, 95% CI 5.6–12.0) in the development period and 6 of 118 (5.1%, 95% CI 1.9–10.7) in the temporal validation cohort (Fig. 1). The median age was 70 years (IQR 62–77) and 274 patients (62.6%) were men. Lung cancer (174, 39.7%) and digestive cancers (114, 26.0%) were the most common tumor types. Baseline characteristics of the two periods are shown in Table 1. Apart from prior antitumor therapy (29.4% vs. 40.7%, *P* = 0.033), Padua score (median 3 vs. 3, *P* = 0.002), and cancer-type composition (*P* = 0.017), the cohorts were similar; this temporal drift is precisely what temporal validation must contend with.

3.2 Univariable analysis

Four variables reached P < 0.05 in univariable logistic regression (Table 2): serum sodium (odds ratio [OR] 0.888 per mmol/L, 95% CI 0.809–0.975), diastolic blood pressure (OR 0.965 per mmHg, 95% CI 0.937–0.993), hemoglobin (OR 0.983 per g/L, 95% CI 0.969–0.998), and digestive cancer (OR 2.247, 95% CI 1.087–4.646). Hypertension (OR 0.501, *P* = 0.058), creatinine (*P* = 0.082), and heart rate (*P* = 0.093) showed borderline associations. The Padua score (*P* = 0.410), NRS2002 (*P* = 0.505), BMI (*P* = 0.261), systolic blood pressure (*P* = 0.380), and white blood cell count (*P* = 0.589) showed no association with the outcome.

Table 2
Univariable logistic regression for unplanned readmission among all candidates

Variable	OR	95% CI	*P*
Serum sodium (per mmol/L)	0.888	0.809–0.975	0.013
Diastolic BP (per mmHg)	0.965	0.937–0.993	0.015
Hemoglobin (per g/L)	0.983	0.969–0.998	0.022
Digestive cancer	2.247	1.087–4.646	0.029
Hypertension	0.501	0.245–1.023	0.058
Creatinine (per umol/L)	1.005	0.999–1.010	0.082
Heart rate (per bpm)	0.978	0.954–1.004	0.093
Fibrinogen (per g/L)	1.309	0.936–1.832	0.116
Prior antitumor therapy	0.538	0.228–1.272	0.158
Albumin (per g/L)	0.952	0.884–1.025	0.189
Potassium (per mmol/L)	0.602	0.262–1.385	0.233
Chloride (per mmol/L)	0.945	0.861–1.038	0.239
BMI (per kg/m2)	0.943	0.852–1.044	0.261
D-dimer (per ng/mL)	1.000	1.000–1.000	0.299
Male sex	1.432	0.701–2.925	0.325
Smoking (former/current)	0.714	0.342–1.491	0.370
Systolic BP (per mmHg)	0.991	0.973–1.011	0.380
Padua score (per point)	1.093	0.885–1.349	0.410
APTT (per s)	1.008	0.988–1.028	0.455
Urea (per mmol/L)	1.036	0.941–1.141	0.472
Total protein (per g/L)	0.982	0.935–1.032	0.479
Prothrombin time (per s)	0.928	0.747–1.153	0.498
Alcohol use (former/current)	0.776	0.351–1.715	0.530
White blood cells (per 10^9/L)	1.026	0.935–1.127	0.589
Lung cancer	0.857	0.410–1.790	0.682
Total bilirubin (per umol/L)	1.005	0.976–1.035	0.725

Variable	OR	95% CI	*P*
Age (per year)	1.005	0.971–1.040	0.789
Thrombin time (per s)	1.002	0.990–1.013	0.792
Globulin (per g/L)	0.996	0.956–1.037	0.839
Prior surgery	1.076	0.453–2.561	0.868
Lymphocytes (per 10^9/L)	0.974	0.707–1.343	0.874
NRS2002 (per point)	1.015	0.669–1.541	0.943
Platelets (per 10^9/L)	1.000	0.996–1.004	0.945
Uric acid (per umol/L)	1.000	0.996–1.003	0.969
Diabetes mellitus	0.993	0.448–2.203	0.987

ORs are per unit increment (continuous) or exposure vs.no exposure (categorical).

3.3 Model development

Under the 1-SE rule, LASSO retained 25 of 34 candidates-too many for the available events and indicative of unstable regularization paths in a small sample **(Supplementary Fig. S1)**. Per the pre-specified rule, the model reverted to the univariable P < 0.05 set, yielding four predictors: **serum sodium, diastolic blood pressure, hemoglobin, and digestive cancer (**Table 3; Fig. 2**).** The model equation is:

logit(P) = 13.142 − 0.0913 × Na − 0.0271 × DBP − 0.0102 × HB + 0.819 × digestive cancer,

where Na is serum sodium (mmol/L), DBP diastolic blood pressure (mmHg), and HB hemoglobin (g/L). Only digestive cancer reached conventional significance in the refitted model (OR 2.27, 95% CI 1.08–4.78, *P* = 0.031); serum sodium (*P* = 0.064) and diastolic blood pressure (*P* = 0.089) were marginal and hemoglobin (*P* = 0.200) was not. There was no evidence of collinearity (variance inflation factors 1.00-1.09) and linear functional forms were supported for all continuous predictors (quadratic Wald tests *P* = 0.142–0.423). Two points warrant emphasis. First, predictor selection in this study was based on cross-validated predictive performance and a pre-specified selection rule, not on hypothesis testing; with 33 events, individual Wald statistics are estimated imprecisely and are not a valid criterion for judging a predictor's contribution [12, 17]. Second, the model as a whole is strongly significant (likelihood-ratio χ^2 = 16.55, df = 4, *P* = 0.002), and the three continuous predictors jointly add significant information beyond digestive cancer (χ^2 = 11.98, df = 3, *P* = 0.008). Collapsing the model to conventionally significant variables would leave only digestive cancer, whose univariable discrimination (c-index 0.589) is far below that of the full model (0.709). A nomogram based on the exact model coefficients maps total scores of 0-300 to predicted risks of 0.01–0.70 and supports bedside risk estimation **(**Fig. 3**).**

Table 3
Final multivariable logistic regression model for unplanned readmission

Variable	beta	SE	OR	95% CI	*P*	Scaled OR (95% CI)
Intercept	13.142	6.700	-	-	0.050	-
Serum sodium (per 5 mmol/L)	-0.0913	0.0492	0.913	0.829–1.005	0.064	0.63 (0.39–1.03)
Diastolic BP (per 10 mmHg)	-0.0271	0.0159	0.973	0.943–1.004	0.089	0.76 (0.56–1.04)
Hemoglobin (per 10 g/L)	-0.0102	0.0080	0.990	0.975–1.005	0.200	0.90 (0.77–1.06)
Digestive cancer (vs others)	0.8187	0.3799	2.268	1.077–4.775	0.031	2.27 (1.08–4.78)

(n = 435,33 events; EPV = 8.25). ORs per unit and per clinically meaningful increment are shown; 95%CIs in parentheses.

3.4 Internal performance and calibration

The apparent c-index was 0.709; after 1000-bootstrap optimism correction it was 0.682 (95% CI 0.599–0.766), with an optimism of 0.027 and an internal Brier score of 0.067. The optimism-corrected calibration slope was 0.885, indicating mild overfitting. Calibration was good: decile observations tracked the ideal line closely **(**Fig. 4A**)**, the Hosmer-Lemeshow test was not significant (χ^2 = 3.03, *P* = 0.932), and the calibration-in-the-large intercept was − 0.000 (*P* = 1.000). At the Youden-optimal threshold (predicted probability 0.089), sensitivity and specificity were 0.61 and 0.75. As reference instruments, the Padua score (c-index 0.538) and NRS2002 (0.502) performed near chance.

3.5 Temporal validation

Applied without refitting to the temporally non-overlapping cohort (117 complete cases, 6 events), the model yielded a c-index of 0.841 (bootstrap 95% CI 0.643–0.985) and a Brier score of 0.042. Calibration slope (1.918) and intercept (+ 1.483; *P* = 0.271) suggested that the model underestimated risk in the higher-risk range of the validation cohort, although estimates derived from six events are inherently imprecise; the Hosmer-Lemeshow test did not indicate miscalibration (χ^2 = 6.37, *P* = 0.606; Fig. 4B). ROC curves for both cohorts are shown in Fig. 6.

3.6 Clinical utility

Decision curve analysis (Fig. 5) showed that the model yielded higher net benefit than both "intervene in all" and "intervene in none" across threshold probabilities of approximately 3–20%, with maximum net benefit around 5–10%, in both the development and validation cohorts.

3.7 Sensitivity analyses

All sensitivity analyses supported the primary results **(Supplementary Table S2)**: restriction to first admissions (407 patients; 27 events) gave a c-index of 0.710 with serum sodium (OR 0.900, *P* = 0.047) and digestive cancer (OR 2.335, *P* = 0.041) remaining significant; forced adjustment for age and sex left the estimates essentially unchanged (sodium OR 0.904, *P* = 0.045; digestive cancer OR 2.335, *P* = 0.027), with neither age (*P* = 0.333) nor sex (*P* = 0.411) significant.

4. Discussion

In 438 admissions of patients with cancer and CHD, we developed a four-variable model-serum sodium, diastolic blood pressure, hemoglobin, and digestive cancer-for unplanned readmission using only data available at admission. The model showed moderate discrimination (optimism-corrected c-index 0.682) with adequate calibration, was stable when applied to a temporally non-overlapping cohort of 118 patients (c-index 0.841), and clearly outperformed the Padua score and NRS2002, which performed at chance level in this setting.

The comparison with existing instruments is the most direct clinical message. Padua and NRS2002 were designed for venous thromboembolism and nutritional risk, respectively [6, 7]; our results confirm that scores targeting other outcomes cannot be assumed to carry predictive information for readmission, and that their use in readmission risk stratification is not supported. Against readmission models from general medical populations (typically c-index 0.60–0.70 [18, 4, 19]), our model performs comparably internally and is, to our knowledge, the first reported specifically for cancer patients with CHD. The high temporalvalidation c-index must be interpreted cautiously: with six events, the 95% CI is extremely wide (0.643–0.985), and the point estimate may partly reflect chance. The calibration slope above 1 and positive intercept in the validation cohort, though not statistically significant, suggest that the model may underestimate risk in later patients and should be re-examined in larger samples [20].

All four predictors have plausible pathophysiological links to readmission in this population.

Hyponatremia is among the most common electrolyte disturbances in cancer-arising from syndrome of inappropriate antidiuretic hormone secretion, chemotherapy-related renal injury, and volume management-and is a recognized marker of frailty and adverse outcomes [21]. Low diastolic blood pressure may reflect impaired vascular reserve from cumulative cardiotoxic therapy or reduced cardiac reserve, consistent with the circulatory vulnerability described in the post-hospital syndrome [2]. Anemia, a common complication of cancer and its treatment, is associated with fatigue, functional limitation, and rehospitalization [3]. Digestive cancers carried the highest risk (OR 2.27), plausibly through malabsorption, treatment-related toxicity (nausea, diarrhea, infection), and postoperative complications. Notably, none of the three continuous variables reached conventional significance in the multivariable model; the value of the model lies in their joint information rather than in any single marker, which is precisely the rationale for a prediction model rather than a single biomarker.

On the question of statistical significance, a methodological clarification is warranted. In prediction modeling, variables are selected on predictive performance under cross-validation, not on hypothesis

tests; with an EPV of 8.25, individual Wald statistics have limited power and are unstable, and using P values as an inclusion criterion systematically discards predictors that are genuinely informative but imprecisely estimated [12, 17]. Here, the model as a whole is strongly significant ($P = 0.002$), the continuous predictors contribute significant joint information ($P = 0.008$), and the full model's discrimination clearly exceeds that of any conventionally significant subset (0.709 vs. 0.589–0.669). We therefore report the complete four-variable model with all P values transparently displayed.

The model is intended to support-not replace-clinical judgment in discharge planning and follow-up intensity. Its practical advantages are that all variables derive from routine admission workup, with no additional cost, and that the nomogram enables risk estimation within 24 h of admission. Decision curve analysis suggests net benefit when patients with predicted probabilities of roughly 5–10% are offered intensified post-discharge support (telephone follow-up, nutritional intervention, shortened review intervals). However, moderate discrimination means the model is only an aid to stratification, and its generalizability to other centers-where treatment protocols, discharge processes, and readmission policies differ-remains to be established.

Several limitations deserve acknowledgment. First, this is a single-center study with 33 events; the EPV of 8.25 falls below the commonly recommended threshold of 10, variable selection was constrained, and potentially informative predictors (tumor stage, treatment phase, discharge medications) were unavailable [12, 17]. Second, the temporal validation cohort contained only six events; the wide confidence interval and unstable calibration estimates mean the validation should be regarded as preliminary support rather than confirmation [20]. Third, the outcome excludes death as a competing event and does not capture out-of-hospital death or admissions to other institutions; socioeconomic status, medication adherence, and distance to the hospital were not available. Fourth, 26 patients contributed repeated admissions; a first-admission-only analysis yielded concordant results, suggesting this did not materially affect the findings. Fifth, we used complete-case analysis for the primary model; missingness in the final model variables was minimal (< 5%).

5. Conclusion

A logistic regression model based on four routine admission variables-serum sodium, diastolic blood pressure, hemoglobin, and digestive cancer-has moderate discrimination and adequate calibration for unplanned readmission in hospitalized patients with cancer and CHD, and remained stable under temporal validation. Independent validation in larger, multi-center cohorts, ideally with competing-risk accounting, is the necessary next step before clinical deployment.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Army Medical University (Medical Research Ethics Review No. 271, 2023). Informed consent was waived because of the retrospective design. The study protocol was registered at the Chinese Clinical Trial Registry (registration number: ChiCTR2400088966, registration date:August 29, 2024).

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Nursing Innovation Fund Project of the Army Medical Centre (Grant Number: 2023HLCXZ02) .

Authors' contributions

NW, JL, and QW participated in the study design. WY assessed the patients; WYu, XZ, and YY collected and organized the patient data. NW and MY performed the data analysis. NW drafted the initial manuscript, and JL and QW reviewed it. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Acknowledgements

None.

References

1. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medi- care fee-for-service program[J]. N Engl J Med. 2009;360(14):1418–28.

2. Krumholz HM. Post-hospital syndrome-an acquired, transient condition of generalized risk. N Engl J Med. 2013;368(2):100–2.
3. Manzano JG, Luo R, Elting LS, George M, Suarez-Almazor ME. Patterns and predic- tors of unplanned hospitalization in a population-based cohort of elderly patients with gastrointestinal cancer. J Clin Oncol. 2014;32(32):3527–33.
4. van Walraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ. 2010;182(6):551–7.
5. Lyon AR, López-Fernández T, Couch LS, Asteggiano R, Aznar MC, Bergler-Klein J, et al. 2022 ESC Guidelines on cardio-oncology developed in collaboration with the Euro- pean Hematology Association (EHA), the European Society for Therapeutic Radiology and Oncology (ESTRO) and the International Cardio-Oncology Society (IC-OS). Eur Heart J. 2022;43(41):4229–361.
6. Barbar S, Noventa F, Rossetto V, Ferrari A, Brandolin B, Perlati M, et al. A risk assessment model for the identification of hospitalized medical patients at risk for venous thromboembolism: the Padua Prediction Score. J Thromb Haemost. 2010;8(11):2450–7.
7. Kondrup J, Rasmussen HH, Hamberg O, Stanga Z, Ad Hoc ESPEN, Working Group. Nutritional risk screening (NRS 2002): a new method based on an analysis of controlled clinical trials. Clin Nutr. 2003;22(3):321–36.
8. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55–63.
9. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1–73.
10. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol. 1996;58(1):267–88.
11. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22.
12. Riley RD, Ensor J, Snell KI, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441.
13. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36.
14. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361–87.
15. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230.

16. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74.
17. van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, et al. Analytic approaches to clinical prediction models in the presence of a small number of events: a simulation study. J Clin Epidemiol. 2019;117:64–76.
18. Donzé J, Aujesky D, Williams D, Schnipper JL. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. JAMA Intern Med. 2013;173(8):632–8.
19. van Walraven C, Wong J, Forster AJ. LACE+ index: extension of a validated index to include more comorbidity and functional information. J Clin Epidemiol. 2012;65(4):436–46.
20. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353:i3140.
21. Wald R, Jaber BL, Price LL, Upadhyay A, Madias NE. Impact of hospital-associated hyponatremia on selected outcomes. Arch Intern Med. 2010;170(3):294–302.

Figures

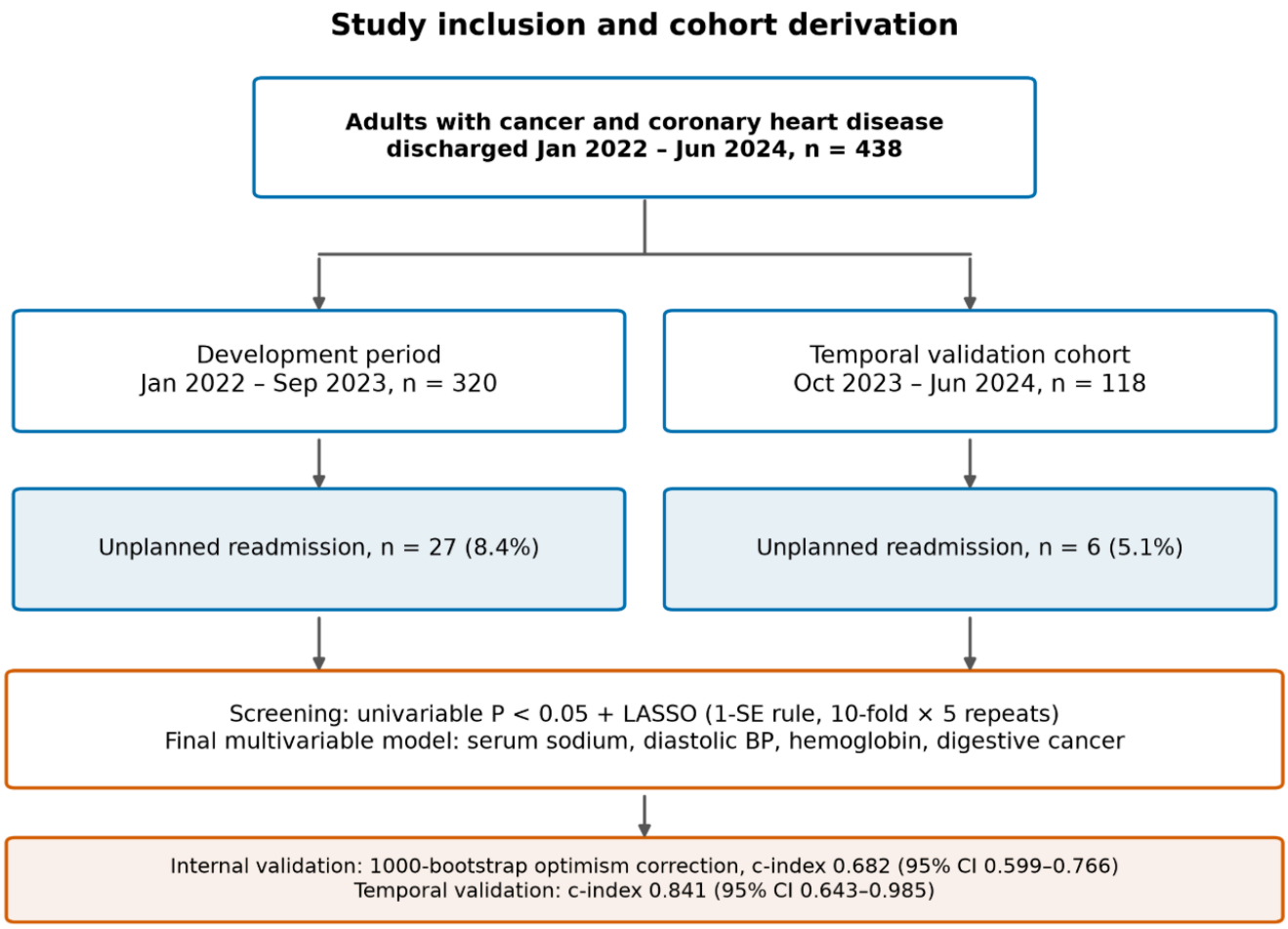


Figure 1

Study inclusion and cohort derivation. The cohort comprises 438 admissions (407 unique patients); discharges before October 2023 constitute the development period and later discharges the temporally non-overlapping validation cohort.

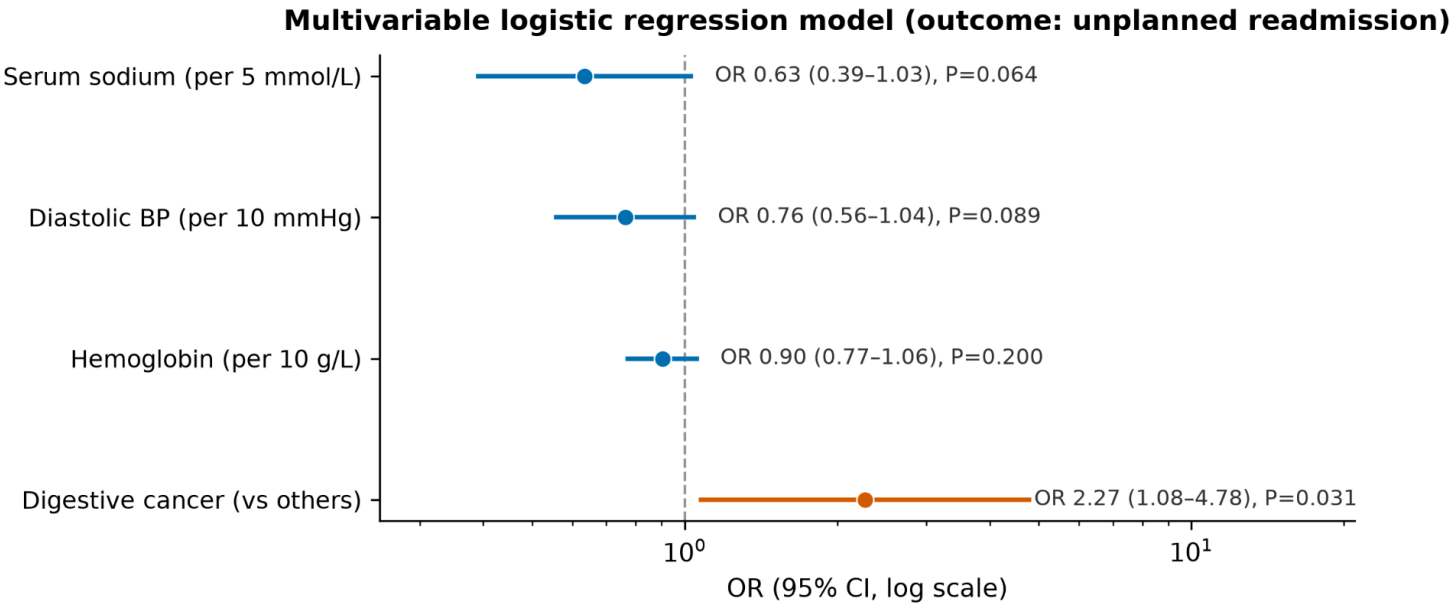


Figure 2

Forest plot of the multivariable model. ORs are scaled to clinically meaningful increments (serum sodium per 5 mmol/L, diastolic blood pressure per 10 mmHg, hemoglobin per 10 g/L); exact P values are shown.

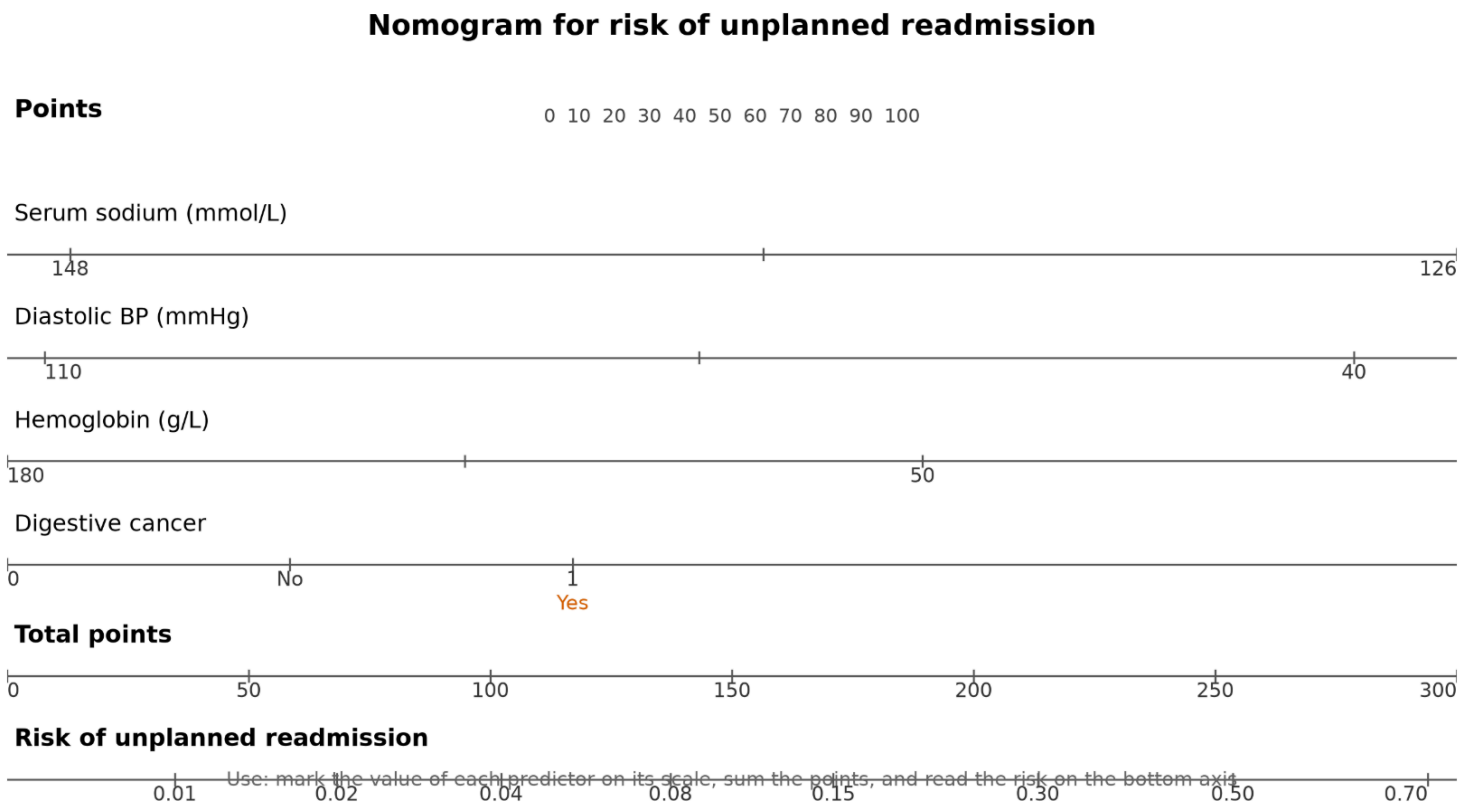


Figure 3

Nomogram for the risk of unplanned readmission. Total scores of 0–300 map to predicted risks of 0 .01

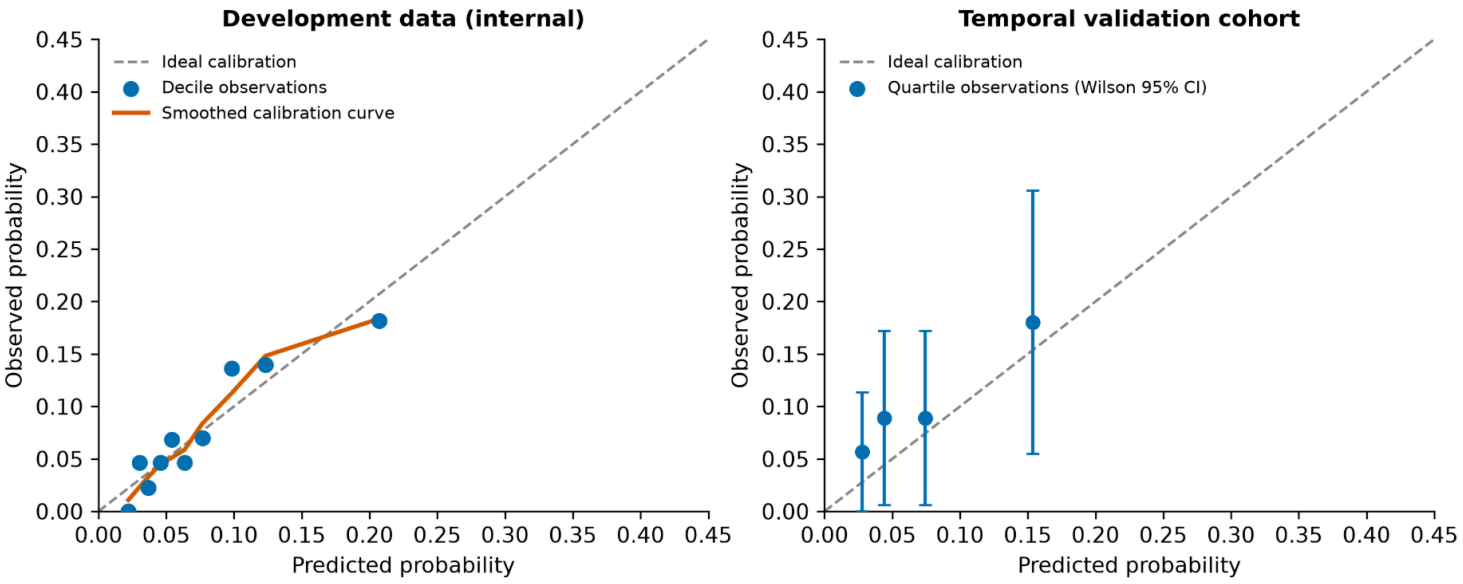


Figure 4

Calibration curves. (A) Development data: decile observations with a smoothed calibration curve. (B) Temporal validation cohort: quartile observations with Wilson 95%CIs. The dashed line is the ideal

calibration.

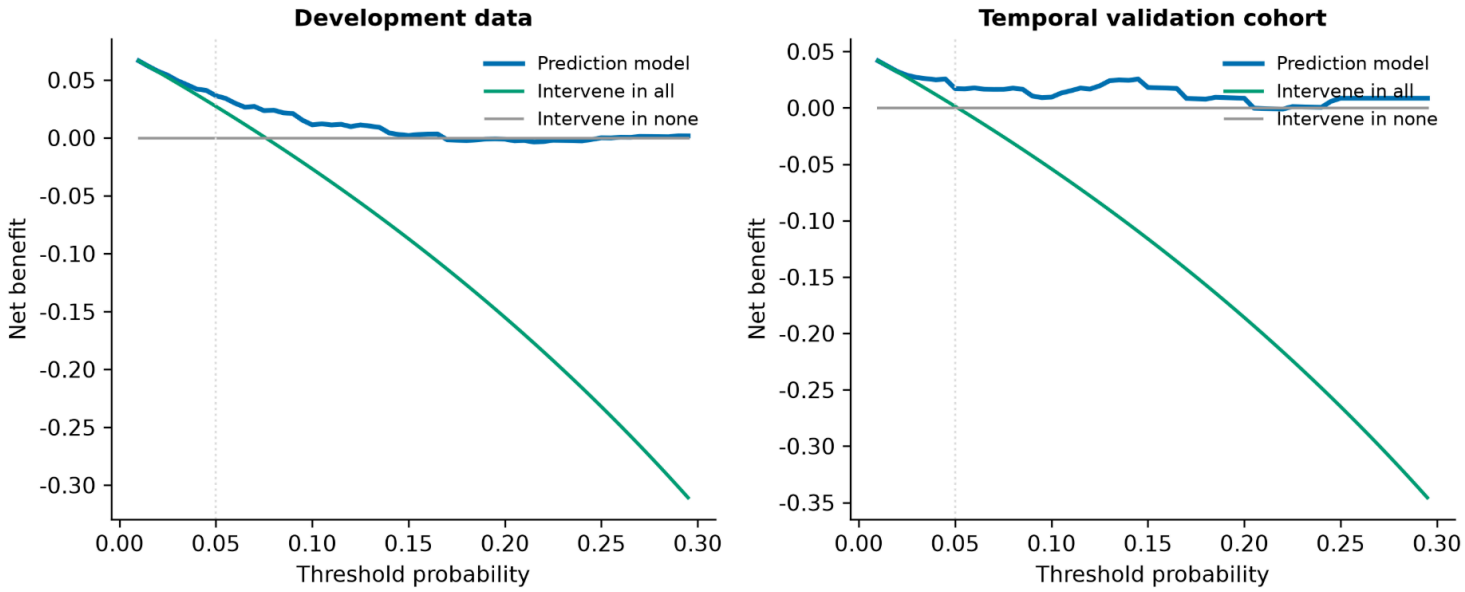


Figure 5

Decision curve analysis. Net benefit of the prediction model versus intervening in all or none, across threshold probabilities of 1–30%, in the development data (left)and the temporal validation cohort (right). The dotted line marks a threshold of 5%.

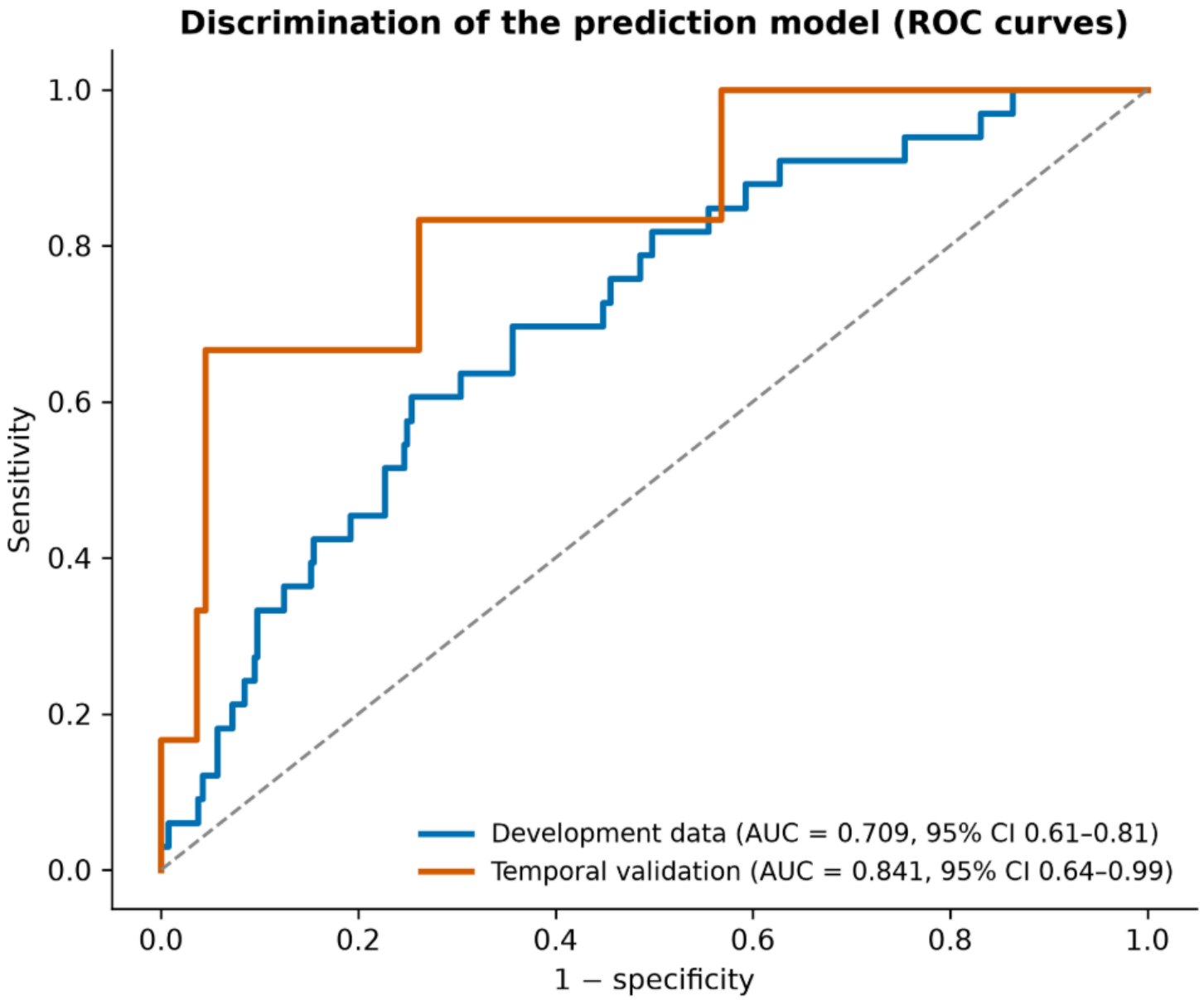


Figure 6

ROC curves of the model in the development data (AUC =0.709, Hanley– McNeil 95% CI 0.61–0.81) and the temporal validation cohort(AUC = 0.841, bootstrap percentile 95%CI 0.64–0.99).

Supplementary Files

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

- supplementmaterial.docx