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 internal validation of a pre-operative machine-learning model for prolonged postoperative stay in older adults undergoing colorectal cancer surgery

Bartlomiej Roj
Southend University Hospital
Cerys Prosser
cerys.prosser@nhs.net

Southmead Hospital
David Abe
Southend University Hospital
Eleanor Stock
Southend University Hospital
Myuran Kaneshamoorthy
Southend University Hospital
James Jegard
Southend University Hospital

Research Article

Keywords:

Posted Date: August 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10645896/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

Purpose

Older adults undergoing colorectal cancer surgery frequently have frailty, multimorbidity and social-discharge complexity, and prolonged admission exposes them to hospital-associated disability, deconditioning and delirium. We developed and internally validated a pre-operative machine-learning model to identify those at risk of prolonged postoperative stay.

Methods

Retrospective single-centre cohort of 197 consecutive patients aged ≥ 65 years assessed in a combined frailty–anaesthesia clinic; 149 were operated with recorded length of stay (LOS). The primary outcome was prolonged stay (LOS > 10 days). Pre-operative predictors included age, Clinical Frailty Scale (CFS), comorbidity, polypharmacy, peri-operative risk scores, and social and operative factors. A gradient-boosted (XGBoost) classifier was evaluated by repeated stratified five-fold cross-validation; exact-LOS regression and illustrative cost scenarios were secondary.

Results

In cross-validation the classifier achieved an AUROC of 0.75 (95% CI 0.69–0.79); at a screening threshold, sensitivity was 0.66, specificity 0.75 and positive predictive value 0.47 (Brier score 0.16). A single 80/20 split gave a more optimistic AUROC of 0.86 with 100% sensitivity but only 10 test events. The strongest predictors were CFS and NSQIP risk, followed by age and comorbidity; day-level LOS prediction was poor (test $R^2 \leq 0.06$). Of 149 operated patients, 38 (25.5%) had prolonged stay; nationally, a 1-day reduction in 10% of the 20,977 annual resections would save ~ 2,098 bed-days (£1.89–£2.51 million).

Conclusion

A model using routine pre-operative data identified older colorectal cancer patients at risk of prolonged stay with moderate discrimination; binary classification was more robust than exact-LOS prediction. External validation and prospective evaluation are required before clinical use.

Introduction

Colorectal cancer is largely a disease of older age. In SEER data the median age at diagnosis is 66 years and incidence peaks between 65 and 74 years [1], so as the population ages the number of older patients coming to colorectal cancer surgery is set to rise. Peri- and postoperative care has improved, but older patients are a heterogeneous group with wide variation in physiological reserve. Many have

features that complicate recovery, including frailty, multimorbidity, poor mobility, polypharmacy and dependence on social support, whereas others remain fit and functionally independent.

These characteristics affect both physiological recovery and the practical business of getting home once a patient is medically fit for discharge. Prolonged admission is often framed as a pathway or service problem, but it is also a clinically important outcome in its own right, because it prolongs exposure to the hospital environment in a vulnerable group. Frailty, a state of reduced physiological reserve and heightened vulnerability to stressors, is now recognised as a major determinant of outcome after colorectal surgery, and current guidance treats it as more informative than chronological age [2,3].

Guidance therefore recommends looking beyond age. Comprehensive Geriatric Assessment (CGA) is widely regarded as the reference standard for finding modifiable medical, functional, cognitive, nutritional and social problems, which are then used to build an individualised plan. For older adults living with frailty (usually a Clinical Frailty Scale ≥ 5) undergoing major surgery, CGA is recommended within structured optimisation pathways, and it has been linked to fewer complications, less delirium and shorter stay, although the size of the benefit varies between surgical populations [3–5].

Prolonged hospitalisation matters because admission itself is not benign. Hospital-associated deconditioning, the decline in physical or cognitive function attributable to being in hospital, is particularly relevant to older people and those living with frailty [6]. A meta-analysis put the pooled prevalence of hospital-associated disability at 30% in hospitalised older adults [7], and more recent work describes functional decline in 13.8% and delirium in 8.8% [8]. Delirium both signals and drives deconditioning by limiting mobility and rehabilitation, and because it lengthens admission it feeds back into the immobility, sleep disruption and disorientation that provoke it. Frailty adds vulnerability to other hospital-acquired harms such as infection, falls and pressure injury [6,8].

In colorectal surgery, frailty tracks with worse postoperative outcomes, longer stay and more readmissions and medical complications [9–11]. This reflects reduced physiological reserve and a blunted response to surgical stress, including diminished inflammatory and metabolic resilience. Frailty also tends to accompany low baseline function and sarcopenia, which slow mobilisation and rehabilitation and lengthen recovery. One prospective colorectal cancer study found longer admissions and higher readmission rates in frail than non-frail patients [11], and reviews of older colorectal populations describe the same pattern of longer stay, more complications and worse outcomes [2,3,9].

Prolonged stay also carries a cost. The British Geriatrics Society cites £901 per non-elective bed day, and a recent parliamentary answer put the elective inpatient bed-day cost at £1,198 in 2024/25 [12,13]. The National Bowel Cancer Audit recorded 20,977 major bowel cancer resections in England and Wales in 2023, with a median stay of 6 days, 30-day unplanned readmission of 11.9% and 90-day mortality of 2.5% [14]. At that scale, even small reductions in admission duration translate into large bed-day and cost savings.

Length of stay in older adults is therefore not simply a function of surgical recovery. It reflects physiological reserve, frailty, postoperative complications, cognitive vulnerability and the social factors that govern discharge readiness [4], and it mixes biological recovery with the broader functional decline and hospital-associated disability to which older inpatients are exposed [5]. That complexity makes high-risk patients hard to spot with conventional perioperative tools, which tend to assess a single domain or focus on dichotomous outcomes such as mortality and major complications [6]. A more integrated approach is needed to stratify risk in this group and to support timely planning and resource allocation.

Conventional perioperative risk tools mainly estimate mortality or major complications and are poorly suited to predicting who will need a prolonged admission. Day-level length of stay is itself shaped by many interacting factors, among them operative complexity, physiological reserve, frailty, social-discharge issues and tumour burden. Machine-learning methods are attractive here because they can accommodate nonlinear effects and interactions across these domains [15–18].

We therefore set out to develop and internally validate a pre-operative machine-learning model to identify older adults at risk of prolonged postoperative stay, defined as more than 10 days after colorectal cancer surgery. Secondary aims were to compare binary prolonged-stay prediction with exact length-of-stay regression, to identify the most influential predictors, and to estimate the possible clinical and economic implications of prolonged admission.

Methods

Study design and setting

This was a retrospective observational cohort study using routinely collected data from a consultant-led combined frailty–anaesthesia clinic for older adults being considered for colorectal cancer surgery. The study cohort comprised 197 consecutive patients aged 65 years or older assessed between 23 September 2021 and 28 December 2023. The current modelling extract contained 149 operated patients with recorded postoperative LOS.

Outcomes

The primary outcome was prolonged postoperative stay, defined a priori as LOS > 10 days. This threshold was chosen to distinguish clearly extended admissions from the more typical postoperative pathway and to provide a clinically meaningful binary target for pre-operative risk stratification.

Secondary outcomes were:

- exact postoperative LOS in days
- exploratory multiclass LOS:
- short stay (≤ 3 days)

- moderate stay (4–10 days)
- prolonged stay (> 10 days)

- Candidate predictors

Only variables available before surgery were used.

Numeric predictors

- Age
- Clinical Frailty Scale (CFS)
- comorbidity burden
- a unified peri-operative risk variable
- parsed P-POSSUM, NSQIP, and SORT scores
- medication count
- binary indicators for frequently occurring PMH-derived conditions, including hypertension, type 2 diabetes, osteoarthritis, atrial fibrillation, chronic obstructive pulmonary disease, chronic kidney disease, and ischaemic heart disease

Categorical predictors

- proposed operation type
- operative approach
- mobility
- location of living
- package of care
- grouped procedure category
- tumour stage (T, N, M)
- Comorbidity burden was defined as the number of conditions recorded in the past medical history. Medication count was defined as the number of recorded pre-clinic medications.

Data preprocessing

Categorical variables were one-hot encoded. Missing numeric values were imputed with medians and missing categorical values with the most frequent category. Risk percentage fields were converted to numeric form. PMH-derived features were generated from common tokens within the free-text history.

Machine-learning model development

Primary binary classifier: XGBoost

The primary model was a calibrated XGBoost classifier. Gradient-boosted trees suit structured clinical data: they handle nonlinear effects and interactions between predictors without pre-specification and perform well on mixtures of clinical, functional and social variables [15]. The model builds decision trees in sequence, each correcting the errors of the last.

All pre-operative variables were entered together: age, frailty, peri-operative risk, operative category, social-discharge factors and the PMH-derived features. Because prolonged stay was the minority class, class weights were adjusted to reduce bias toward the majority. Predicted probabilities were calibrated, and an operating threshold was chosen to favour sensitivity, as suits a screening tool.

The classifier outputs the predicted probability of prolonged stay, which becomes a binary label once the threshold is applied. The main metric was the AUROC, which captures how well the model separates prolonged from non-prolonged stays across thresholds; we also report recall, precision, F1 and accuracy.

Secondary regression models for exact LOS

Elastic Net

Elastic Net is a regularised linear regression [17], included as an interpretable benchmark. It models length of stay as a linear combination of predictors and shrinks unstable coefficients through combined L1 and L2 penalties, which helps when predictors are correlated. It returns a predicted length of stay in days, assessed with R^2, MAE and RMSE.

Random Forest

Random Forest is a tree-based ensemble [16]. It fits many trees to bootstrapped samples and averages their predictions, with each tree seeing a random subset of predictors; this dampens the instability of a single tree while still capturing nonlinear effects. It also returns a predicted length of stay in days.

XGBoost regressor

The XGBoost regressor is the continuous-outcome counterpart of the classifier [15]. It fits trees in sequence to reduce residual error and is among the stronger methods for tabular prediction, at some cost to interpretability.

Support Vector Regression

Support Vector Regression uses a kernel to model nonlinear relationships while limiting sensitivity to outliers [17]. Instead of fitting every point, it seeks a smooth function that keeps most observations within a tolerance band, and returns a predicted length of stay in days.

Exploratory multiclass LOS model

An exploratory multiclass analysis classified LOS into:

- short stay (≤ 3 days)
- moderate stay (4–10 days)
- prolonged stay (> 10 days)

We also tried a three-class version (short, moderate, prolonged), since categorical groupings can be clinically intuitive. Both a direct multinomial XGBoost and a one-vs-rest calibrated approach were explored, producing class labels or probabilities.

Model evaluation

Internal validation used repeated stratified five-fold cross-validation (50 repeats), preserving outcome prevalence in each fold; all preprocessing, class-imbalance weighting and model fitting were performed within the training folds only to avoid information leakage. The primary discrimination metric was the AUROC, summarised as the mean and 2.5th–97.5th percentile (95% CI) across repeats. At a screening-oriented probability threshold we report sensitivity, specificity, positive and negative predictive value, accuracy and F1. Calibration was assessed on pooled out-of-fold predictions using the Brier score and the calibration slope and intercept. For comparison, performance from a single stratified 80/20 train–test split is also reported, with a 95% CI for its AUROC derived by the method of Hanley and McNeil. Predictor contribution was assessed using permutation importance, which measures how much model performance declines when a predictor is randomly shuffled.

For the regression models, evaluation metrics were:

- train and test R2
- mean absolute error (MAE)
- root mean squared error (RMSE)

MAE is the average absolute error in days and is the most intuitive for planning; RMSE penalises large misses and so reflects long-stay outliers; R^2 is the share of length-of-stay variance explained on unseen data.

Cohort-based clinical impact analysis

The operated cohort was divided into:

- non-prolonged stay (≤ 10 days)
- prolonged stay (> 10 days)

The uploaded extract contained mortality fields but did not contain a sufficiently robust dedicated readmission variable stratified by prolonged-stay status. Accordingly, mortality was described cautiously and a definitive readmission analysis could not be performed from the available file.

Health-burden analysis

To illustrate the possible patient-level implications of prolonged admission, an exploratory burden analysis was undertaken using:

- the observed prolonged-stay rate in the cohort (38/149 = 25.5%)
- the national annual major resection volume (20,977)
- published older-inpatient rates for hospital-associated disability and hospital-associated complications [7,8]

Economic scenario analysis

Published cost estimates were used:

- £901 per non-elective bed day
- £1,198 per elective inpatient bed day [12,13]

The following scenarios were examined:

- 1-day reduction in stay in 10% of national cases
- 1-day reduction in stay in 20% of national cases
- 2-day reduction in stay in 10% of national cases
- 2-day reduction in stay in 20% of national cases

These estimates were intended as illustrative and do not represent a formal cost-effectiveness analysis.

Results

Cohort and prolonged-stay groups

Among the 149 operated patients with recorded LOS, 38 patients (25.5%) had prolonged stay > 10 days and 111 (74.5%) did not. Median LOS was 18 days in the prolonged group and 5 days in the non-prolonged group. Baseline characteristics by group are shown in Table 1; the prolonged-stay group was older (median 82 vs 80 years, $p = 0.03$), frailer (median Clinical Frailty Scale 4 vs 3, $p < 0.001$), had higher NSQIP risk (6% vs 2%, $p = 0.002$) and underwent more open or converted surgery ($p = 0.004$).

Table 1
Baseline characteristics of the operated cohort by prolonged-stay status. Continuous variables are median (IQR); categorical variables are n (%). p from Mann–Whitney U or Fisher/χ^2 test.

Variable	All (n = 149)	Non-prolonged (n = 111)	Prolonged (n = 38)	p
Age, years	80 (76–84)	80 (76–83)	82 (78–85)	0.030
Clinical Frailty Scale	3 (3–4)	3 (3–4)	4 (3–5)	0.000
Comorbidity burden (n conditions)	5 (3–6)	4 (3–6)	6 (4–6)	0.064
Medication count	5 (3–7)	5 (3–7)	5 (3–7)	0.880
P-POSSUM, %	2 (2–4)	2 (1–4)	3 (2–5)	0.045
NSQIP, %	3 (1–6)	2 (1–4)	6 (2–10)	0.002
Length of stay, days	6 (5–11)	5 (4–7)	18 (13–21)	0.000
Operative approach				0.004
Converted	16 (11%)	10 (9%)	6 (16%)	
Laparoscopic	122 (82%)	97 (87%)	25 (66%)	
Open	8 (5%)	2 (2%)	6 (16%)	
Other	3 (2%)	2 (2%)	1 (3%)	
Procedure category				0.024
Colectomy	11 (7%)	6 (5%)	5 (13%)	
Hemicolectomy	82 (55%)	64 (58%)	18 (47%)	
Hernia	2 (1%)	0 (0%)	2 (5%)	
Other	19 (13%)	12 (11%)	7 (18%)	
Rectal	35 (23%)	29 (26%)	6 (16%)	

Binary prediction of prolonged stay

On repeated stratified five-fold cross-validation (50 repeats), the prolonged-stay classifier achieved a mean AUROC of 0.75 (95% CI 0.69–0.79) (Fig. 1). At a screening-oriented probability threshold, cross-validated performance was: sensitivity 0.66, specificity 0.75, positive predictive value 0.47, negative predictive value 0.86, accuracy 0.72 and F1 0.55 (Table 2). Calibration was modest, with a Brier score of 0.16 and a calibration slope of 0.49, indicating that predicted probabilities were somewhat over-extreme (Fig. 2). For comparison, a single stratified 80/20 train–test split produced a higher AUROC of 0.86 (95% CI 0.71–1.00) with a recall of 1.00, but that test set contained only 10 prolonged-stay events; the gap

between the two estimates indicates that the single-split result was optimistic. The cross-validated estimates should therefore be regarded as the primary measure of internal performance, and all estimates require external validation.

Overall, the model discriminated moderately well between patients with and without prolonged stay and is best positioned as a risk-stratification aid rather than a definitive rule-in or rule-out test; at the operating threshold, roughly half of the patients flagged as high risk did not go on to have a prolonged stay (positive predictive value 0.47).

Table 2
Performance of the prolonged-stay classifier. Cross-validated estimates (repeated stratified 5-fold, 50 repeats) are primary; the single 80/20 split is shown for comparison. Operating threshold 0.25; Hanley–McNeil 95% CI for the single-split AUROC.

Metric	Cross-validation (primary)	Single 80/20 split
AUROC (95% CI)	0.75 (0.69–0.79)	0.86 (0.71–1.00)
Sensitivity	0.66	1.00
Specificity	0.75	0.68
Positive predictive value	0.47	0.53
Negative predictive value	0.86	1.00
Accuracy	0.72	0.76
F1 score	0.55	0.69
Brier score	0.16	—
Calibration slope	0.49	—
Test-set prolonged events	—	10 of 38

Permutation importance identified two comparably important leading predictors (Fig. 3):

- Clinical Frailty Scale — 0.043
- NSQIP risk score — 0.040

Other relevant contributors included:

- comorbidity burden
- operative category
- operative approach
- mobility
- living circumstances

- package of care
- medication count

Prolonged stay in this group therefore reflects both physiological and operative risk and frailty- or discharge-related complexity.

Continuous LOS prediction

Exact day-level length-of-stay prediction was less reliable than binary prolonged-stay classification (Table 3).

Table 3
Regression models for exact postoperative length of stay (test-set performance). High training but near-zero test R^2 indicates over-fitting and limited generalisation.

Model	Train R^2	Test R^2	Test MAE (days)	Test RMSE (days)
Elastic Net	0.40	0.06	6.50	13.31
Random Forest	0.86	-0.08	6.40	14.28
XGBoost	1.00	0.01	5.70	13.66
Support Vector Regression	0.47	0.03	5.37	13.56

All four models return a predicted length of stay in days. SVR had the lowest average error, missing by about 5.4 days, but every model explained little of the test-set variance, so exact length of stay could not be estimated reliably from pre-operative data alone.

Exploratory multiclass LOS modelling

The three-class model was less consistent than the binary classifier, with an accuracy of about 0.63 and better identification of moderate and prolonged stays than the under-represented short-stay group. Binary prolonged-stay prediction was the more stable target in this cohort.

Recorded postoperative outcomes in the uploaded cohort

Using the mortality fields available in the modelling extract:

- in the non-prolonged group (n = 111), recorded 30-day mortality was 1/111 (0.9%)
- in the prolonged group (n = 38), recorded 30-day mortality was 0/38

The same pattern was observed in the available 90-day mortality field. These numbers were very small and are reported descriptively only. A robust stratified readmission analysis could not be performed from the current extract because a validated readmission field by LOS category was not available.

Illustrative health-burden analysis

If the observed cohort prolonged-stay rate (25.5%) is applied to the 20,977 major bowel cancer resections performed nationally each year, it corresponds to roughly 5,349 prolonged-stay cases annually. Set against published complication rates in older inpatients, this maps to the burden shown in Table 4.

Table 4
Illustrative health burden: published complication rates in older inpatients applied to the estimated ~ 5,349 prolonged-stay cases arising from 20,977 national resections per year. Illustrative only; not causal estimates.

Hospital-associated complication	Published rate	Estimated cases per year
Hospital-associated disability	30%	1,605
Functional decline	13.8%	738
Delirium	8.8%	471
Incontinence	5.6%	300
Pressure injury	0.7%	37

These estimates are illustrative and do not imply that prolonged stay itself caused these outcomes; rather, they quantify the possible health burden associated with extended hospital exposure in an older high-risk population [6,7,8].

Economic scenario analysis

Applied to 20,977 national resections each year [14], the four illustrative scenarios give the bed-day and cost savings shown in Table 5.

Table 5
Economic scenarios: bed-days and direct bed-day cost saved under illustrative reductions in length of stay, applied to 20,977 national resections. Illustrative only; not a formal cost-effectiveness analysis.

Scenario	Bed-days saved	Saving at £901/day	Saving at £1,198/day
1-day reduction in 10% of cases	2,098	£1.89 million	£2.51 million
1-day reduction in 20% of cases	4,195	£3.78 million	£5.03 million
2-day reduction in 10% of cases	4,195	£3.78 million	£5.03 million
2-day reduction in 20% of cases	8,391	£7.56 million	£10.05 million

Discussion

This was a predictive modelling study, but its interest goes beyond raw performance. The main finding was that binary prediction of prolonged stay held up better than exact day-level prediction in older colorectal cancer patients. That difference matters clinically, because the decision-relevant question is

usually whether a patient faces a prolonged, higher-risk admission that consumes bed-days, not the precise number of days they will stay.

The binary classifier showed moderate discrimination. On repeated cross-validation the AUROC was 0.75 (95% CI 0.69–0.79), so routinely collected pre-operative data carried a useful but incomplete signal for identifying patients likely to stay beyond 10 days; the apparently excellent single-split result (AUROC 0.86, recall 1.00) did not survive cross-validation and should be read as optimistic given only 10 events in that test set. The strongest signals were frailty (Clinical Frailty Scale) and peri-operative physiological risk (NSQIP), consistent with prolonged stay arising from the interplay of physiological risk, frailty, multimorbidity and discharge complexity rather than age or operative category alone [1,2,6–8,20].

We compared several methods to see whether the modelling assumptions changed predictive value. Tree-based models, XGBoost in particular, were chosen for their handling of nonlinear interactions in structured data [15]; Elastic Net gave an interpretable linear benchmark [17]; Random Forest a robust ensemble comparator [16]; and SVR was included because kernel regression can do well in moderate-sized datasets with nonlinear structure [17]. The binary problem proved more tractable than exact regression: the regression models achieved reasonable average day-level error but explained little variance, which makes them unsuitable for patient-level forecasting.

The clinical value of this work rests on what a prolonged admission does to an older patient. Hospital-associated disability affects around 30% of older inpatients, with functional decline and delirium quantified at 13.8% and 8.8% among hospital-associated complications [7,8], and frailty compounds these risks [3,5]. A prolonged stay is not a neutral service metric; it is extended exposure to immobility, cognitive stress, infection and loss of independence. A tool that flags the patients likely to stay longest may be useful even if its only current role is pre-operative risk stratification.

The health-burden figures are illustrative rather than causal. Applying the cohort prolonged-stay rate to national resection volume gives roughly 5,349 prolonged-stay cases a year; set against published complication rates in older inpatients, that maps to hundreds of possible cases of functional decline and delirium linked to longer admission. These are not direct causal estimates, but they give a sense of why prolonged stay matters in this population.

The economic figures point the same way. On conservative assumptions, cutting stay by one day in just 10% of major bowel cancer cases nationally would save roughly £1.9–£2.5 million a year in bed-days, rising to about £3.8–£5.0 million at 20% of cases. Because they count only direct bed-day cost, these numbers probably understate the real saving, which would also take in effects on complications, rehabilitation, community support and delayed discharge [9–11].

The study has clear limitations. It was retrospective, single-centre and only internally validated, and the held-out test set was small. Some predictors were engineered pragmatically from free text, which may limit portability. The extract did not support a robust readmission analysis stratified by length of stay, so we could not quantify the link between prolonged stay and readmission, which should be a priority for

future work. Calibration was modest (slope 0.49), and discrimination depended heavily on the validation scheme, with a single split markedly overstating performance relative to cross-validation. Larger, ideally multicentre and prospective, datasets will be needed to show whether performance holds and stays adequately calibrated across different populations.

The study also has strengths. It drew on consecutive patients from a real-world frailty–anaesthesia service, used only routinely available pre-operative variables, targeted a clinically meaningful outcome, and compared several modelling approaches before settling on the final classifier.

Priorities for further work are external validation of the classifier, prospective testing of whether acting on the risk score reduces prolonged stay, and linkage to richer outcomes such as readmission, discharge destination, functional decline and patient-reported recovery.

Conclusion

Pre-operative machine-learning risk stratification shows promise as an adjunct to perioperative planning for older adults undergoing colorectal cancer surgery, but it needs multicentre external validation and prospective evaluation. Frailty and peri-operative risk were the strongest signals, with further contributions from comorbidity and operative, functional and social-discharge variables. Binary prediction of prolonged stay was more robust than exact length-of-stay prediction. The work is clinically relevant because prolonged admission exposes older patients to hospital-associated disability, functional decline and delirium, and economically relevant because even modest reductions in stay could yield meaningful national savings. The next step is to test prospectively whether acting on this risk information actually reduces prolonged stay and related harm.

Declarations

Ethics approval.

This study was registered with, and approved by, the local clinical governance board at Southend University Hospital. It was conducted in accordance with the Declaration of Helsinki. Because the study used retrospectively collected, de-identified clinical data, individual patient consent was waived.

Consent to participate.

Individual consent to participate was waived owing to the retrospective use of anonymised, routinely collected clinical data.

Consent for publication.

Not applicable; the study contains no identifiable individual patient information.

Competing interests.

The authors declare no competing interests.

Funding.

No specific funding was received for this work.

Author Contribution

Authors' contributions. BR and CP conceived and designed the study. BR curated the data and performed the analysis. BR and CP drafted the manuscript. DA, HG, ES, MK and JJ contributed to data acquisition and interpretation and critically revised the manuscript for important intellectual content. All authors approved the final version and agree to be accountable for all aspects of the work.

Acknowledgements.

The authors thank the clinical teams involved in the care of the patients included in this study and the departmental staff who supported case identification and record retrieval.

Data Availability

he datasets analysed during the current study are not publicly available because they contain patient-level clinical data subject to institutional data-governance restrictions; de-identified data may be available from the corresponding author on reasonable request, subject to institutional approval and applicable data-sharing agreements.

References

1. National Cancer Institute (2026) SEER Cancer Stat Facts: colorectal cancer. Bethesda (MD): National Cancer Institute. https://seer.cancer.gov/statfacts/html/colorect.html (accessed 5 August
2. Rostoft S, Hamaker ME (2020) Basic geriatric principles for colorectal surgeons: how to optimize assessment and care of older patients in the perioperative period. Eur J Surg Oncol 46:310–315
3. Saur NM, Davis BR, Montroni I, Shahrokni A, Rostoft S, Russell MM et al (2022) The American Society of Colon and Rectal Surgeons clinical practice guidelines for the perioperative evaluation and management of frailty among older adults undergoing colorectal surgery. Dis Colon Rectum 65:473–488
4. National Institute for Health and Care Excellence (2016) Multimorbidity: clinical assessment and management (NG56). NICE, London

5. Centre for Perioperative Care (2021) British Geriatrics Society. Guideline for the perioperative care of people living with frailty undergoing elective and emergency surgery. CPOC and BGS, London
6. Welch C, Chen Y, Hartley P, Naughton C, Martinez-Velilla N, Stein D et al (2024) New horizons in hospital-associated deconditioning: a global condition of body and mind. Age Ageing 53:afae241
7. Loyd C, Markland AD, Zhang Y, Fowler M, Harper S, Wright NC et al (2020) Prevalence of hospital-associated disability in older adults: a meta-analysis. J Am Med Dir Assoc 21:455–61e5
8. Mitsutake S et al (2025) Predictive validity of hospital-associated complications of older people identified using Diagnosis Procedure Combination data from an acute care hospital in Japan: observational study. JMIR Aging 8:e68267
9. Soler-González G, Sastre-Valera J, Viana-Alonso A, Aparicio-Urtasun J, García-Escobar I, Gómez-España MA et al (2024) Update on the management of elderly patients with colorectal cancer. Clin Transl Oncol 26:69–84
10. Xu JY, Madden HE, Martínez-Camblor P, Deiner SG (2025) Frailty as an independent risk factor for prolonged postoperative length of stay: a retrospective analysis of 2015–2019 ACS NSQIP data. J Clin Anesth 101:111730
11. Ommundsen N et al (2018) Post-discharge complications in frail older patients after surgery for colorectal cancer. Eur J Surg Oncol 44:1542–1547
12. British Geriatrics Society. Key messages: older people's healthcare. London: British Geriatrics Society (2025) https://www.bgs.org.uk/keymsgs (accessed 5 August 2026)
13. UK Parliament (2025) Hospital beds: costs. Written question UIN 100046, tabled 15 December 2025. UK Parliament, London. https://questions-statements.parliament.uk/written-questions/detail/2025-12-15/100046
14. National Bowel Cancer Audit (2025) State of the nation report 2025: an audit of care received by people diagnosed with or undergoing surgery for bowel cancer in England and Wales. National Cancer Audit Collaborating Centre, Royal College of Surgeons of England, London
15. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, ; 785–94
16. Breiman L (2001) Random forests. Mach Learn 45:5–32
17. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, New York
18. Dreiseitl S, Ohno-Machado L (2002) Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inf 35:352–359
19. Seow-En I et al (2021) Results of a colorectal enhanced recovery after surgery (ERAS) programme and a qualitative analysis of healthcare workers' perspectives. Asian J Surg 44:307–312
20. Li RD, Joung RH, Brajcich BC, Schlick CJR, Yang AD, McGee MF et al (2022) Comprehensive evaluation of the trends in length of stay and post-discharge complications after colon surgery in

the USA. J Gastrointest Surg 26:2184–2192

21. Cathomas M et al (2024) Adherence to enhanced recovery after surgery (ERAS) in older adults following colorectal resection. J Geriatr Oncol 15:102062
22. Chan KS, Lee DJK, Kok YD, Chen WW, Lim XC, Tan L et al (2024) A 10-year review on older patients undergoing colorectal cancer surgery: surgeon-led geriatric service and good quality surgery are drivers of good outcomes. J Gastrointest Surg 28:40–46
23. Robinson TN, Wu DS, Sauaia A et al (2011) Frailty predicts increased hospital and six-month healthcare cost following colorectal surgery in older adults. Am J Surg 202:511–514

Figures

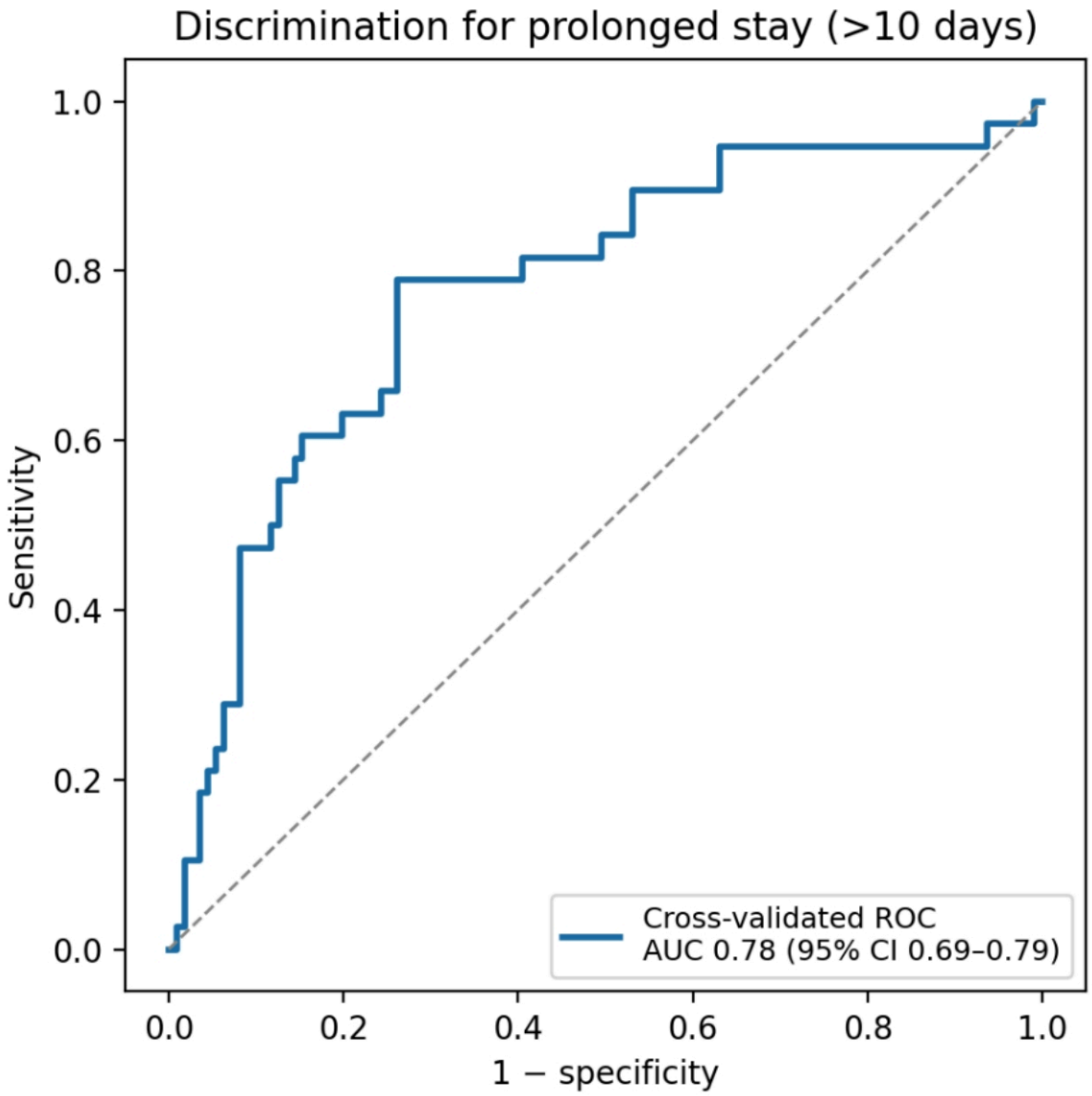


Figure 1

Cross-validated receiver operating characteristic curve for prediction of prolonged stay (>10 days).

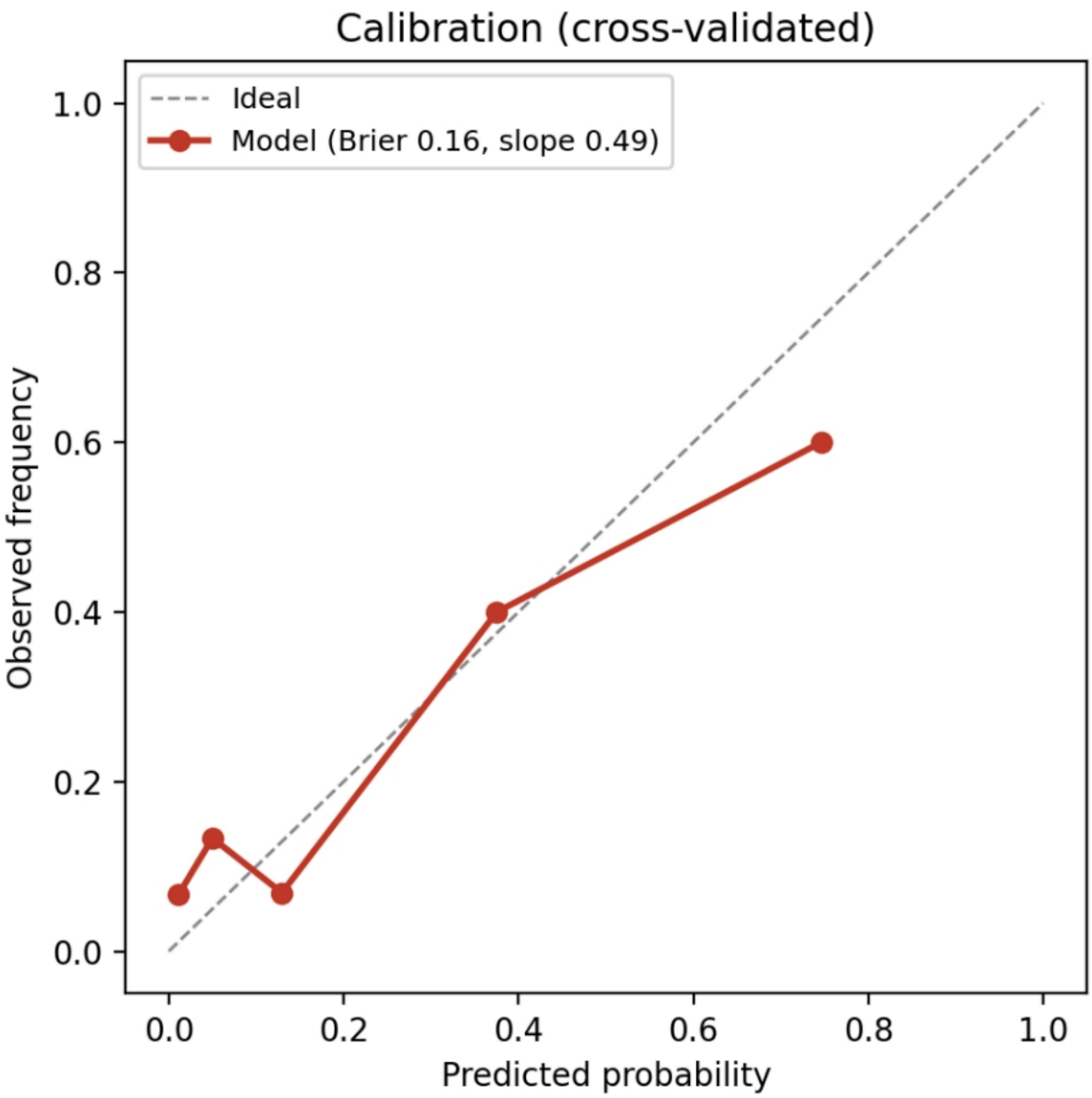


Figure 2

Calibration of predicted probabilities against observed frequencies (cross-validated).

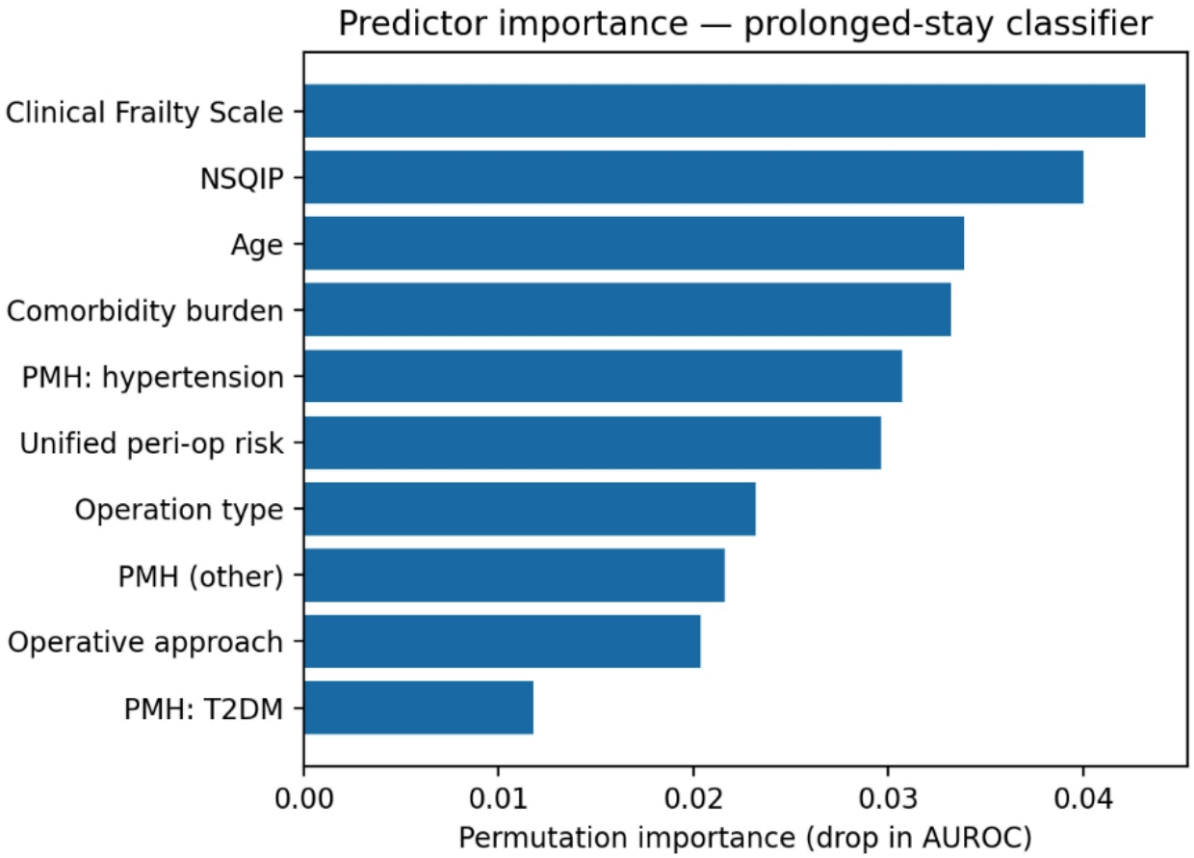


Figure 3

Grouped permutation importance of the leading predictors (drop in AUROC when each predictor is shuffled).