INTERNATIONAL JOURNAL OF SURGERY

OPEN

Improving risk models for patients having emergency bowel cancer surgery using linked electronic health records: a national cohort study

Helen A. Blake, PhD[a,b,c,*], Linda D. Sharples, PhD[d], Jemma M. Boyle, PhD, MBChB[b], Angela Kuryba, MSc[b], Suneetha R. Moonesinghe, MD (Res)[e], Dave Murray, FRCA[f], James Hill, FRCS[g], Nicola S. Fearnhead, DM[h], Jan H. van der Meulen, PhD[a,b], Kate Walker, PhD[a,b]

Background: Life-saving emergency major resection of colorectal cancer (CRC) is a high-risk procedure. Accurate prediction of postoperative mortality for patients undergoing this procedure is essential for both healthcare performance monitoring and preoperative risk assessment. Risk-adjustment models for CRC patients often include patient and tumour characteristics, widely available in cancer registries and audits. The authors investigated to what extent inclusion of additional physiological and surgical measures, available through linkage or additional data collection, improves accuracy of risk models.

Methods: Linked, routinely-collected data on patients undergoing emergency CRC surgery in England between December 2016 and November 2019 were used to develop a risk model for 90-day mortality. Backwards selection identified a 'selected model' of physiological and surgical measures in addition to patient and tumour characteristics. Model performance was assessed compared to a 'basic model' including only patient and tumour characteristics. Missing data was multiply imputed.

Results: Eight hundred forty-six of 10 578 (8.0%) patients died within 90 days of surgery. The selected model included seven preoperative physiological and surgical measures (pulse rate, systolic blood pressure, breathlessness, sodium, urea, albumin, and predicted peritoneal soiling), in addition to the 10 patient and tumour characteristics in the basic model (calendar year of surgery, age, sex, ASA grade, TNM T stage, TNM N stage, TNM M stage, cancer site, number of comorbidities, and emergency admission). The selected model had considerably better discrimination compared to the basic model (C-statistic: 0.824 versus 0.783, respectively).

Conclusion: Linkage of disease-specific and treatment-specific datasets allowed the inclusion of physiological and surgical measures in a risk model alongside patient and tumour characteristics, which improves the accuracy of the prediction of the mortality risk for CRC patients having emergency surgery. This improvement will allow more accurate performance monitoring of healthcare providers and enhance clinical care planning.

Keywords colorectal cancer, emergency surgery, risk model, postoperative mortality, record linkage, electronic health records

Introduction

Major resection is a common treatment for patients diagnosed with colorectal cancer (CRC), and is associated with a high-risk of death when undertaken in the emergency setting[1]. Comparisons of postoperative mortality among hospitals, teams or surgeons, or over time, are important for quality assessment and quality improvement of CRC services and risk-adjustment is needed to ensure fair comparisons. Risk models are also important for preoperative risk assessment, which can aid clinical care planning and inform the counselling of patients for emergency CRC surgery. Clinical guidelines recommend that all high-risk surgical patients should receive certain standards of care, such as direct transfer to critical care and presence of a consultant surgeon and

anaesthetist in theatre[2]. Accurate risk prediction is therefore vital to ensure the highest risk patients receive appropriate care.

There are four major arguments to develop a new extended risk model for emergency CRC surgery. The first argument is that recent reviews have identified a substantial number of models for CRC surgery[3], and for emergency bowel surgery[4,5], but only a few models for the intersection between these two groups. Using either a CRC risk model or an emergency bowel surgery risk model may lead to inaccurate predictions of risk for patients undergoing emergency CRC surgery.

Second, the available models that focus on emergency CRC surgery were specifically developed using data from a single provider[6,7], or from a limited geographical area[8]. A risk prediction model developed in a national population of patients undergoing emergency CRC surgery will have increased precision in the model estimates and will be more widely applicable.

Third, risk models for CRC surgery tend to include patient and tumour characteristics, which are widely available in cancer registries and audits of care for patients with CRC. For this new risk model, we also considered the inclusion of physiological measures (e.g. measurements of organ function and overall health, such as serum creatinine level, breathlessness history, etc.) and surgical measures (e.g. peritoneal soiling, intraoperative blood loss, etc.), which are less readily available. Their inclusion may improve accuracy of risk prediction in the emergency setting.

Fourth, increasing the number of measures included in a risk model may also limit the model's utility in clinical practice given the potentially greater impact of missing data, mismeasurement, and misclassification. We therefore used an explicit model development approach to ensure that the improvement in prediction accuracy by including physiological and surgical measures is balanced against utility of the model.

The development of such an extended risk model for patients undergoing major emergency CRC resection is possible due to availability of detailed patient, tumour, surgical, and physiological information from a large cohort of patients recorded in linked electronic health databases[9]. In the study described in this paper, we investigated whether accuracy of a basic risk model including only patient and tumour characteristics was improved by inclusion of physiological and surgical measures. We aimed to develop and validate such an extended risk model that can be used for both *risk-adjustment* for performance monitoring of healthcare providers, and *risk prediction* for clinical care planning.

Methods

Datasets and linkage

Data from disease, treatment, and administrative hospital databases for patients in England were linked in a national cohort study. For information on patient characteristics, hospital admissions, and outcomes, we used an administrative hospital database, Hospital Episode Statistics Admitted Patient Care (HES), which contains information on all hospital admissions in the English National Health Service (NHS)[10,11]. For information on patient and tumour characteristics, we used a disease-specific dataset collected by the National Bowel Cancer Audit (NBOCA), containing data on patients diagnosed with CRC[1]. For physiological and surgical measures, we used a treatment-specific dataset collected by the National Emergency Laparotomy

HIGHLIGHTS

- We developed and validated a risk model for emergency colorectal cancer surgery.
- Our risk model can be used for risk-adjustment or clinical care planning.
- Colorectal cancer surgery risk models often use patient and tumour characteristics from routine data.
- Physiological and surgical measures improved accuracy of mortality risk predictions.

Audit (NELA), containing data on patients having emergency bowel surgery[12]. Each of these datasets contains information on mortality provided via national UK mortality statistics[13].

The three datasets were linked using a spine linkage approach, where the administrative hospital database was designated the 'spine dataset' and the other two datasets were linked to it[9]. Linkage used deterministic rules regarding agreement on patient identifiers (NHS number, sex, date of birth, and residential post code)[9,14].

Patients were eligible for analysis if they were in the spine dataset, were recorded as undergoing an emergency major resection for CRC between 1 December 2016 and 30 November 2019 in at least one dataset (Appendix: Table A.1, Figure A.1, Supplemental Digital Content 1, http://links.lww.com/JS9/B778), and had complete data for mortality.

The eligible cohort was split into development and validation datasets based on date of surgery. The development cohort included patients having surgery between 1 December 2016 and 30 November 2018. The validation cohort included patients having surgery between 1 December 2018 and 30 November 2019.

Identification and definition of variables

The risk model for 90-day mortality was developed to be used for both risk-adjustment and for clinical care planning. For both risk-adjustment and clinical care planning, variables need to capture a patient's risk immediately before surgery. For risk-adjustment, variables included need to be factors outside of the control of the provider, so that due merit is given to providers with high-quality care. For clinical care planning, information that is reflected by the included variables needs to be available before surgery. Variables to be considered for inclusion were identified from existing risk models for CRC surgery or emergency bowel surgery[3–5,15–21].

A 'basic model' included all patient and tumour characteristics included in these models that met the inclusion criteria and were routinely available in electronic health records (i.e. without any variable selection). Physiological and surgical measures included in models identified in the literature were candidate variables for selection if they met the inclusion criteria, and were available through linkage to external databases. Candidate surgical measures were preoperative estimates of operative severity (i.e. type of procedure to be undertaken), expected peritoneal soiling, and expected intraoperative blood loss, as these variables reflect clinicians' assessments of patients' health just before surgery[16].

For continuous variables, we used fractional polynomials with functional forms established in an existing model developed and validated in 38 830 patients (e.g. age was modelled as linear plus quadratic)[16]. To reduce the influence of outliers, continuous variables were Winsorised, by setting observed values beyond the

Blake et al. International Journal of Surgery (2024)

International Journal of Surgery

1st and/or the 99th percentile as the value of that percentile (Appendix Table A.2, Supplemental Digital Content 1, http://links.lww.com/JS9/B778). Levels of categorical variables were combined if any categories had <10 patients with complete data in either the development dataset or validation dataset. We considered all interactions that had been identified by existing models: an interaction between age and metastases, an interaction between age and ASA grade, and an interaction between ASA grade and respiratory history[15,16].

Handling missing data

Multiple imputation with chained equations was used to handle missing data, under the assumption of missing-at-random conditional on mortality and other observed information[22]. The imputation procedure was undertaken separately in the development and validation datasets, and for the full dataset as a whole. The number of imputations was set at 20. Year of procedure, sex, and procedure type were complete for all patients. Eight patients with missing mortality were excluded from analysis (Appendix Figure A.1, Supplemental Digital Content 1, http://links.lww.com/JS9/B778). Binary variables were imputed using logistic regression, categorical variables were imputed using multinomial logistic regression, continuous variables were imputed using predictive mean matching with values drawn at random from a pool of k = 3 observations with similar predicted values[23,24].

Parameters and performance measures (on the appropriate scale) were pooled over imputed datasets using Rubin's rules[25]. As the χ^2 statistic from Hosmer–Lemeshow (H–L) tests cannot be pooled using Rubin's rules, we calculated a F-statistic to account for between-imputation variation[26].

Potential bias from nonlinkage, the main source of missing data, was assessed by comparing patient characteristics (recorded in the administrative spine dataset) and 90-day mortality between linked and unlinked patients[27].

Model development

Using the development dataset, a 'full model' was fitted including all patient and tumour characteristics in the basic model and all candidate physiological and surgical measures. To assess whether a simpler model could achieve similarly high prediction performance, backwards selection was used with criteria defined by model R^2 (measure of variation in outcomes explained). At each step, R^2 values were pooled over the imputed datasets using Rubin's rules[25,28]. Candidate physiological and surgical variables were excluded in turn, with those resulting in the smallest reduction in model R^2 excluded first[29], until the basic model was reached.

The R^2 value from the full model was used to define a threshold for choosing a model containing only the most important additional physiological and surgical variables. The 'selected model' was the simplest model with an R^2 value greater than 95% relative to the full model R^2 value.

Once the selected model was finalised in the development dataset, this model was refitted using the full dataset (i.e. both developmental and validation datasets). Estimated model coefficients were reported in the form of an equation and the corresponding odds ratios (ORs) were also reported.

Model performance and validation

Other performance measures calculated were the C-statistic, and the scaled Brier score (SBS). The C-statistic quantifies the discrimination of the model, ranging from 0.5 (noninformative) to 1 (perfect discrimination)[30–33]. The Brier score quantifies the average prediction error (accuracy) of the model predictions, compared to a naïve noninformative model. We calculated SBS ranging from 0 (noninformative) to 100% (perfect predictions)[30,34,35]. Verburg et al.[36] gave the following rule-of-thumb for interpreting a SBS for binary outcomes: <0.04 representing very weak predictions, 0.04–0.15 weak, 0.16–0.35 moderate, 0.36–0.62 strong, and >0.63 very strong.

Using the development dataset, we calculated the R^2 value, C-statistic, and SBS for the full model, selected model, basic model, and all models in between.

Using the validation dataset, we refitted the full model, selected model, and basic model and assessed calibration of the selected model compared to the full model and basic model by plotting observed versus predicted mortality by deciles of risk. We also calculated the R^2 value, C-statistic, and SBS for the three models.

Table 1

Data sources for the patient and tumour characteristics included in all models, and the physiological and surgical measures considered for variable selection.[a]

Source of data items	NBOCA	NELA	HES
Outcome:			
90-day mortality[b]	✓[3]	✓[2]	✓[1]
Patient and tumour characteristics:			
Calendar period	✓[2]	✓[1]	✓[3]
Sex	✓[2]	✓[1]	✓[3]
Age	✓[2]	✓[1]	✓[3]
ASA grade	✓[2]	✓[1]	
Pretreatment TNM staging	✓		
Cancer site	✓		
Comorbidities			✓
Emergency admission	✓[3]	✓[2]	✓[1]
Physiological and surgical measures:			
ECG		✓	
Cardiac signs		✓	
Systolic BP		✓	
Pulse		✓	
Breathlessness history		✓	
Glasgow coma score		✓	
Urea		✓	
White blood cell count		✓	
Creatinine		✓	
Sodium		✓	
Potassium		✓	
Albumin		✓	
Predicted peritoneal soiling		✓	
Predicted intraoperative blood loss		✓	
Number of operations in admission		✓	
Surgical urgency		✓	
Preoperative severity		✓	

[a]Where data was available from more than one source, order of preference for reconciliation rules are given using numbers in superscript.
[b]For each dataset, mortality information was obtained from the Office of National Statistics.
NBOCA.
HES, Hospital Episode Statistics.
NBOCA, National Bowel Cancer Audit.
NELA, National Emergency Laparotomy Audit.

Table 2

Distribution of patient and tumour characteristics, and 90-day mortality percentages. Using full dataset (December 2016 to November 2019).[a]

	n (%)	Mortality		n (%)	Mortality
Total	10 578		Died within 90 days of surgery	846 (8.0)	
Patient and tumour characteristics			Calendar year		
Age (years)			2016	302 (2.9)	7.9%
< 50	971 (9.2)	2.9%	2017	3612 (34.1)	8.4%
50–59	1494 (14.1)	4.1%	2018	3480 (32.9)	7.9%
60–74	4082 (38.6)	6.9%	2019	3184 (30.1)	7.6%
75–84	3008 (28.4)	10.9%	ASA grade		
≥ 85	1008 (9.5)	14.4%	1	1009 (9.5)	1.7%
Missing	15 (0.1)	< 0.1%	2	4466 (42.2)	3.1%
Sex			3	3721 (35.2)	9.8%
Female	5082 (48.0)	7.7%	4 or 5	1042 (9.9)	27.8%
Male	5496 (52.0)	8.3%	Missing	340 (3.2)	10.6%
TNM T stage			Cancer site		
T1 or T2 (inner layer of bowel or into muscle layer)	548 (5.2)	3.3%	Appendix/caecum/ascending colon	3083 (29.1)	6.8%
T3 (beyond muscle layer)	3140 (29.7)	5.3%	Hepatic flexure	505 (4.8)	7.3%
T4 (breached outer lining or invaded adjacent organs)	4375 (41.4)	7.4%	Transverse colon	916 (8.7)	7.1%
			Splenic flexure/descending colon	1047 (9.9)	6.3%
Missing	2515 (23.8)	13.4%	Sigmoid colon	2214 (20.9)	6.3%
TNM N stage			Rectosigmoid/Rectal	912 (8.6)	5.4%
N0 (0 nodes involved)	3314 (31.3)	5.9%	Missing	1901 (18.0)	14.8%
N1 (1–3 nodes involved)	2626 (24.8)	5.6%	Comorbidities		
N2 (4 or more nodes involved)	2107 (19.9)	7.8%	0	5122 (48.4)	4.4%
Missing	2531 (23.9)	13.4%	1	3171 (30.0)	8.4%
TNM M stage			2 +	1836 (17.4)	15.5%
M0 (no metastases)	6519 (61.6)	5.7%	Missing	449 (4.2)	15.8%
M1 (metastases)	1293 (12.2)	8.4%	Emergency admission		
Missing	2766 (26.1)	13.2%	No	2843 (26.9)	4.3%
			Yes	7728 (73.1)	9.4%
			Missing	7 (0.1)	14.3%

[a]Missing categories for variables include patients who had missing values in all linked data sources as well as patients who did not link to a data source containing that information.

Using the full dataset, we used funnel plots to visually explore hospital trust-level variation in 90-day mortality for the basic and selected models to determine whether variation between hospital trusts was greater than expected by chance alone[37]. A hospital trust is an organisational unit in the English NHS that can include more than one hospital.

For sensitivity analysis, we produced estimates of adjusted ORs for the selected model using complete case analysis instead of using multiple imputation. A further sensitivity analysis explored using postoperative measures of actual operative severity, peritoneal soiling, and intraoperative blood loss instead of their preoperative estimates. We used Stata 17.0 for all linkage, imputation, and analysis of data[38] and reviewed the TRIPOD guidelines for reporting on multivariable prediction models (Appendix B, Supplemental Digital Content 1, http://links.lww.com/JS9/B778)[39]. This work has also been reported in line with the strengthening the reporting of cohort, cross-sectional, and case–control studies in surgery (STROCSS) criteria[40] (Supplemental Digital Content 2, http://links.lww.com/JS9/B779).

Results

Overall, 846 of 10 578 (8.0%) patients died within 90 days of surgery, with 90-day mortality in the development and validation datasets of 8.2 and 7.6%, respectively, (Appendix Table A.3, Supplemental Digital Content 1, http://links.lww.com/JS9/B778). 10 441 (98.7%) of the analysis cohort linked to the disease-specific dataset (NBOCA), the treatment-specific dataset (NELA) or both (Appendix Figure A.1, Supplemental Digital Content 1, http://links.lww.com/JS9/B778). 5,803 (54.9%) patients linked to both the disease-specific and treatment-specific datasets, providing information on patient and tumour characteristics as well as physiological and surgical measures.

Table 1 shows the patient and tumour characteristics in the basic model and the candidate physiological and surgical measures, and the source of information for each variable. Where information on a variable was available from multiple datasets, reconciliation of conflicting information was undertaken (Table 1)[9].

Patient and tumour characteristics, physiological and surgical measures, and 90-day mortality are summarised in Tables 2 and 3 for the whole analysis cohort. The missing categories in the tables include patients with missing data due to nonlinkage and, to a lesser extent, patients with missing values in the datasets. 55.5% of the analysis cohort had at least one missing variable. Distributions of patient characteristics were similar for the development dataset and the validation dataset (Appendix Table A.3, Supplemental Digital Content 1, http://links.lww.com/JS9/B778).

The selected model included seven physiological and surgical measures (pulse rate, systolic blood pressure, breathlessness,

Blake et al. International Journal of Surgery (2024)

International Journal of Surgery

Table 3

Distribution of physiological and surgical measures, and 90-day mortality percentages. Using full dataset (December 2016 to November 2019).[a]

	n (%)	Mortality		n (%)	Mortality
Total	10 578		Died within 90 days of surgery	846 (8.0)	
Physiological and surgical measures			Sodium (mmol l^-1)		
Electrocardiogram			Low (< 133)	1198 (11.3)	14.6%
No abnormalities	6269 (59.3)	7.2%	Normal (133–146)	6317 (59.7)	7.8%
AF rate 60–90	343 (3.2)	15.5%	High (> 146)	44 (0.4)	22.7%
AF rate > 90 or other abnormal rhythm	934 (8.8)	17.7%	Missing	3019 (28.5)	5.7%
Missing	3032 (28.7)	5.7%	Potassium (mmol l^-1)		
Cardiac signs			Low (< 3.5)	769 (7.3)	12.4%
No failure	5719 (54.1)	7.3%	Normal (3.5–5.3)	6620 (62.6)	8.2%
Diuretic, digoxin, antianginal Or antihypertensive therapy	1499 (14.2)	12.1%	High (> 5.3)	3189 (30.1)	6.5%
Borderline cardiomegaly Or cardiomegaly	329 (3.1)	21.6%	Albumin (g/l)		
			Low (< 35)	3136 (29.6)	13.5%
Missing	3031 (28.7)	5.8%	Normal (35–50)	3926 (37.1)	5.3%
Systolic BP (mmHg)			High (> 50)	66 (0.6)	6.1%
Low (< 90)	159 (1.5)	29.6%	Missing	3450 (32.6)	6.1%
Normal (90–120)	2618 (24.7)	10.4%	Haemoglobin (g/l)		
High (> 120)	4761 (45.0)	7.5%	Low (male <130/female <115)	3417 (32.3)	10.0%
			Normal (male 130–180/female 115–165)	4061 (38.4)	8.0%
Missing	3040 (28.7)	5.6%	High (male > 180/female > 165)	3100 (29.3)	5.7%
Pulse rate (beats per min)			Surgical urgency		
Low (< 60)	206 (1.9)	5.8%	Expedited (> 18 h)	1890 (17.9)	6.9%
Normal (60–100)	6002 (56.7)	7.6%	Urgent (6–18 h)	3187 (30.1)	7.2%
High (> 100)	1332 (12.6)	15.5%	Urgent (2–6 h)	2127 (20.1)	11.7%
Missing	3038 (28.7)	5.6%	Immediate (< 2 h) or emergency (resus of > 2 h possible)	353 (3.3)	19.5%
Breathlessness history			Missing	3021 (28.6)	5.7%
No breathlessness	5760 (54.5)	6.5%	Number of operations within		
Breathlessness on exertion or CXR	1228 (11.6)	13.4%	Admission 1	7304 (69.0)	8.8%
Breathlessness limiting exertion and at rest	565 (5.3)	24.1%	2 +	257 (2.4)	13.6%
Missing	3025 (28.6)	5.7%	Missing	3017 (28.5)	5.7%
Glasgow coma score			Preoperative severity		
15	7344 (69.4)	8.3%	Major	2732 (25.8)	7.8%
14 or less	205 (1.9)	31.2%	Major +	4829 (45.7)	9.6%
Missing	3029 (28.6)	5.6%	Missing	3017 (28.5)	5.7%
Urea (mmol L^-1)			Predicted peritoneal soiling		
Low (< 2.5)	299 (2.8)	5.0%	None	3492 (33.0)	6.7%
Normal (2.5–7.8)	4992 (47.2)	6.1%	Serous fluid	2161 (20.4)	8.6%
High (> 7.8)	5287 (50.0)	9.9%	Localised pus	781 (7.4)	7.0%
White blood cell count (× 10^9 l^-1)			Free pus, blood, or bowel contents	1123 (10.6)	17.9%
Low (< 3.6)	162 (1.5)	17.3%	Missing	3021 (28.6)	5.7%
Normal (3.6–11.0)	4224 (39.9)	8.0%	Predicted intraoperative blood loss		
High (> 11.0)	6192 (58.5)	7.7%	< 100	2363 (22.3)	6.9%
Serum creatinine (mu mol l^-1)			100–500	4831 (45.7)	9.6%
Low (male <59/female <45)	665 (6.3)	11.0%	> 500	366 (3.5)	12.8%
Normal (male 59–104/female 45–84)	5198 (49.1)	6.5%	Missing	3018 (28.5)	5.7%
High (male > 104/female > 84)	4715 (44.6)	9.2%			

[a]Missing categories for variables include patients who had missing values in all linked data sources as well as patients who did not link to a data source containing that information.

sodium, urea, albumin, and prediction of peritoneal soiling) in addition to the ten patient and tumour characteristics in the basic model (calendar year of surgery, age, sex, ASA grade, TNM T stage, TNM N stage, TNM M stage, cancer site, number of comorbidities, emergency admission).

Figure 1 shows model performance measured at each stage of the variable selection process, using the development cohort only. More details can be found in Appendix Table A.4 (Supplemental Digital Content 1, http://links.lww.com/JS9/B778). The full model (all patient, tumour, surgical, and physiological variables) had an R^2 of 0.235 (blue markers). The percentage decrease in R^2 is shown for the remaining model after each variable is removed until we are left with the basic model (patient and tumour characteristics only).

Of all physiological and surgical measures, serum albumin was the most predictive, contributing most to the R^2, therefore being the last variable removed in the variable selection process. The selected model had an R^2 value of 0.225, 95.7% relative to the full model, whilst the basic model had an R^2 value of 0.163, 69.2% relative to the full model (Fig. 1). The selected model had a
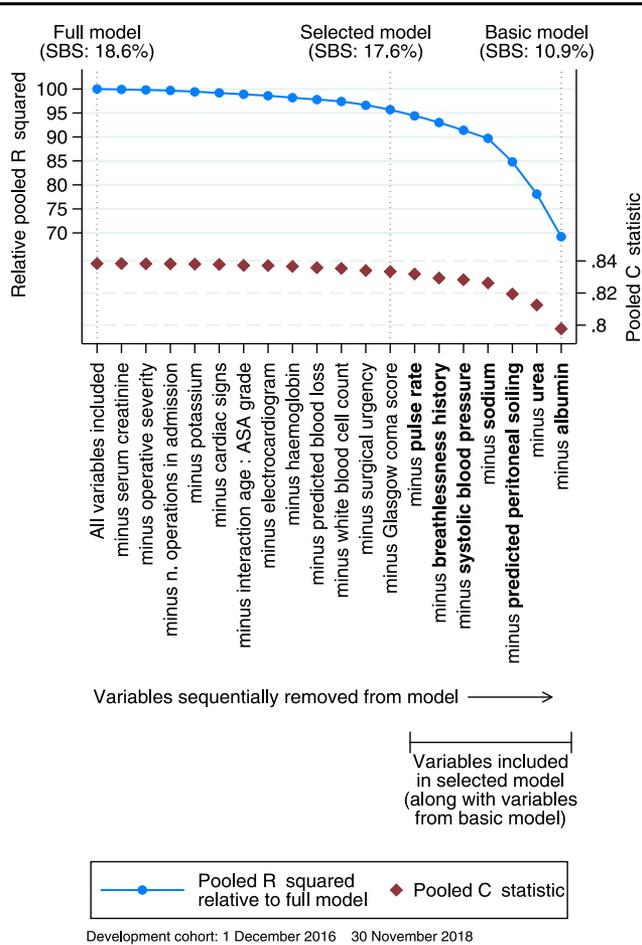
Figure 1. Pooled C-statistic and relative pooled R^2 for the full model, basic model, and all intermediate models.* Scaled pooled Brier score (SBS) also given for the full model, selected model, and basic model. Using development dataset (December 2016 to November 2018).*The full model includes all patient and tumour characteristics, and physiological and surgical measures. The basic model includes all patient and tumour characteristics. Each intermediate model is defined by removing one of the physiological and surgical measures from the previous model, according to which would result in the smallest difference in pooled R^2. For example, model 2 includes all variables included in the full model except operative severity. The selected model includes all patient and tumour characteristics, and the physiological and surgical measures highlighted in bold.

moderate SBS of 17.6%[36], slightly lower than the full model SBS (18.6%), whereas the basic model had a substantially weaker SBS (10.9%) (Fig. 1). All fitted models had good discrimination in the development cohort, with C-statistics, 0.84 for the full model, 0.83 for the selected model, and 0.80 for the basic model (Fig. 1).

The measures of model performance (R^2, C-statistic, HL P-value, SBS) for the full model, the selected model, and the basic model in the validation cohort (Table 4) were slightly lower than those in the development cohort. Discrimination was lower for the basic model in the validation cohort (0.78) whilst the full model and selected model had similar discrimination as in the development cohort (0.84 and 0.83, respectively).

Figure 2 displays calibration of the full model, selected model, and basic model using data from the validation cohort, by plotting observed mortality versus the model predicted risk by deciles of predicted risk. All three models demonstrate good calibration.

There was no evidence of poor fit as HL test P-values were large for all three models (Table 4).

Figure 3 explores whether the basic model lacks calibration because it does not include the seven physiological and surgical measures included in the selected model. It displays observed and predicted mortality, using the basic model and the validation cohort, by categories of the physiological and surgical measures that were included in the selected model. The basic model had poorer calibration for patients with abnormal levels of serum albumin, urea, systolic blood pressure, or history of breathlessness at rest. On average across imputations, 8.3% of patients in the validation cohort had abnormal levels in at least one of these.

Table 5 gives adjusted ORs and 95% CIs calculated using the selected model, for all variables included in the selected model. Appendix Table A.5 (Supplemental Digital Content 1, http://links.lww.com/JS9/B778) gives the equation for the selected model using the whole analysis cohort. As the effect of a continuous risk factor on mortality is not easily expressed when the relationship is non-linear, Table 5 presents ORs for selected values of the continuous factors. The adjusted odds of death within 90 days of major resection increased with worse breathlessness history, lower systolic blood pressure, higher pulse rate, higher urea levels, and lower albumin. The adjusted odds of death with serum albumin levels of 25 g/l was 1.66 (CI: 1.48–1.86) times the adjusted odds at serum albumin 35 g/L (reference value). The adjusted odds of death with urea levels of 20 mmol L-1 or 30 mmol L-1 were 1.57 (CI: 1.32–1.88) and 2.18 (CI: 1.54–3.10) times the adjusted odds at urea 10 mmol L-1 (reference value), respectively.

Funnel plots of adjusted 90-day mortality by hospital trust, risk-adjusted using (i) the basic model and (ii) the selected model, are shown in Figure 4. Using the selected model, five hospital trusts were above the inner limit (i.e. identified as having outlying performance). If instead the basic model was used for risk-adjustment, one of these hospital trusts would not have been identified as having outlying performance (a 'false negative') and one further hospital trust would been incorrectly identified as having outlying performance (a 'false positive').

Fewer patients would be classified as high-risk and therefore receive care recommended for high-risk patients using the selected model compared to the basic model (22.9% of patients had predicted risk > 0.1 and 18.5% had predicted risk between 0.05 to 0.1 using the selected model, versus 25.4 and 21.1%, respectively, for the basic model). The selected model classified 15.2% of patients into a lower risk category than the basic model, and 7.8% into a higher risk category.

In sensitivity analysis, a similar pattern of associations between risk factors and mortality was seen in the complete case analysis, only with greater uncertainty (Appendix Table A.6, Supplemental Digital Content 1, http://links.lww.com/JS9/B778). Furthermore, results were not sensitive to using postoperative surgical measures in variable selection instead of preoperative estimates of operative severity, peritoneal soiling, and intraoperative blood loss, and prediction did not improve substantially (results not shown).

Appendix Table A.7 (Supplemental Digital Content 1, http://links.lww.com/JS9/B778) compares characteristics of patients according to which datasets they were linked between. Patient characteristics were broadly similar, comparing those linked from the administrative dataset (HES) to only the disease-specific dataset (NELA), only the treatment-specific dataset (NBOCA), those linked to both, and those linked to neither. However, the

Blake et al. International Journal of Surgery (2024)

International Journal of Surgery

Table 4

Model performance comparing the full model, to the selected model, and to the basic model. Using validation dataset (December 2018 to November 2019).

$N = 3467$	Pooled R^2 (as % of full model)	Pooled C-statistic (95% CI)	HL test P-value	Scaled Brier score
Full model	0.230 (100.0)	0.843 (0.816–0.869)	0.801	16.6%
Selected model	0.200 (87.1)	0.827 (0.800–0.853)	0.758	13.7%
Basic model	0.149 (65.0)	0.784 (0.755–0.813)	0.657	10.2%

HL, Hosmer–Lemeshow.

patients that only linked between the spine (HES) dataset and the treatment-specific (NELA) dataset had much higher mortality compared to other groups.

Discussion

We have developed and validated an accurate and relatively simple risk model for patients with CRC undergoing emergency surgery, that can be used for risk-adjustment or clinical care planning. We identified seven important physiological and surgical measures which we recommend should be included along with patient and tumour characteristics to accurately predict postoperative mortality. These measures are either easy to capture (pulse rate, blood pressure, and breathlessness history) or estimate (peritoneal soiling), or are completely objective (urea, albumin, and sodium). This extended risk model, including these additional physiological and surgical measures, allows more accurate prediction than available basic risk models, particularly for patients with abnormal values of these additional measures.

Comparison to other risk models

All seven of the physiological and surgical measures included in our selected model have been found to be important predictors in other models, including the surgical risk score P-POSSUM[41,42]. The discrimination of our selected model compares very favourably to existing risk models (which have C-statistics ranging from 0.732 to 0.861[7,8,15–21] whilst including a relatively small set of risk factors that are all routinely available in patients undergoing emergency CRC surgery.

Of the seven measures we selected, four (serum albumin, urea, sodium, and breathlessness history) are composite markers of pre-existing chronic disease, and impact of current acute status. For example, breathlessness could be attributable to either cardiac or respiratory disease and cardiorespiratory insufficiency (poor functional capacity) is a known major risk factor for poor surgical outcome[43]. Breathlessness may also be a symptom of metabolic acidosis in an unwell patient with sepsis. Urea is a biomarker for kidney function, but is also affected by diet, dehydration, and proximal colonic bleeding[44]. Serum albumin falls in the acute phase response of sepsis but may also be a marker of chronic malnutrition.

Strengths, limitations, and opportunities for further work

The model was developed in a large representative national study using linked electronic health records, and candidate variables were drawn from the literature. The variables identified by the
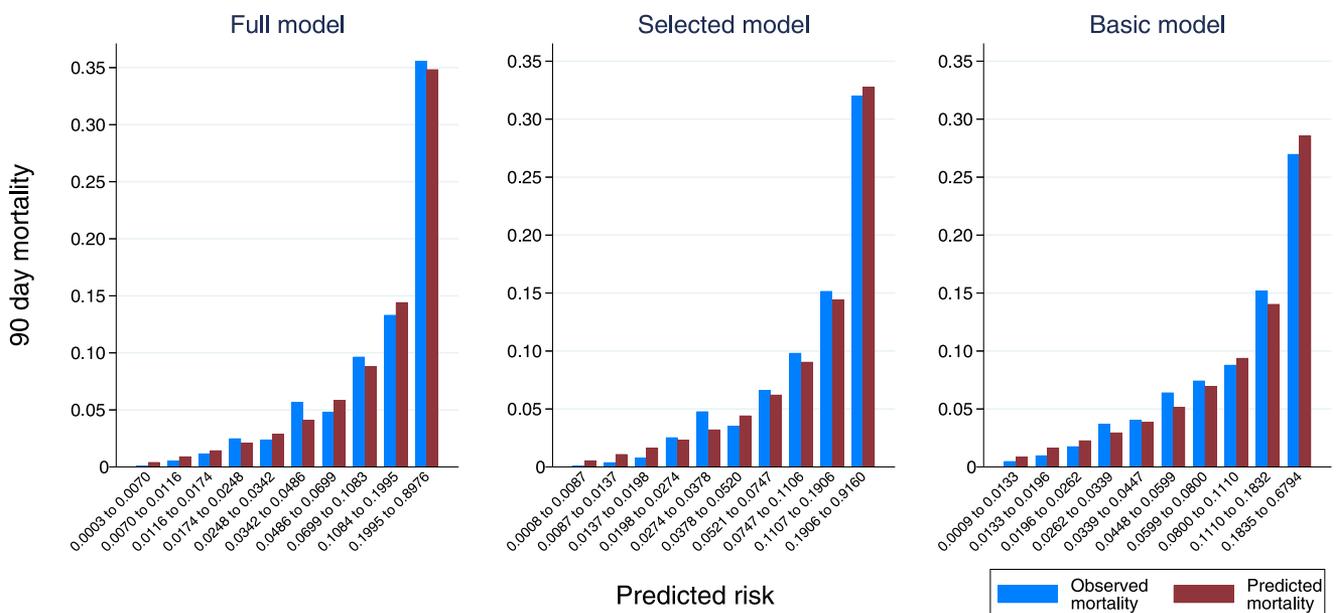
Figure 2. Calibration plot showing proportions of patients that died within 90 days (observed mortality) versus mortality predicted using the full model, selected model, and basic model, by deciles of predicted risk. Using validation dataset (December 2018 to November 2019).
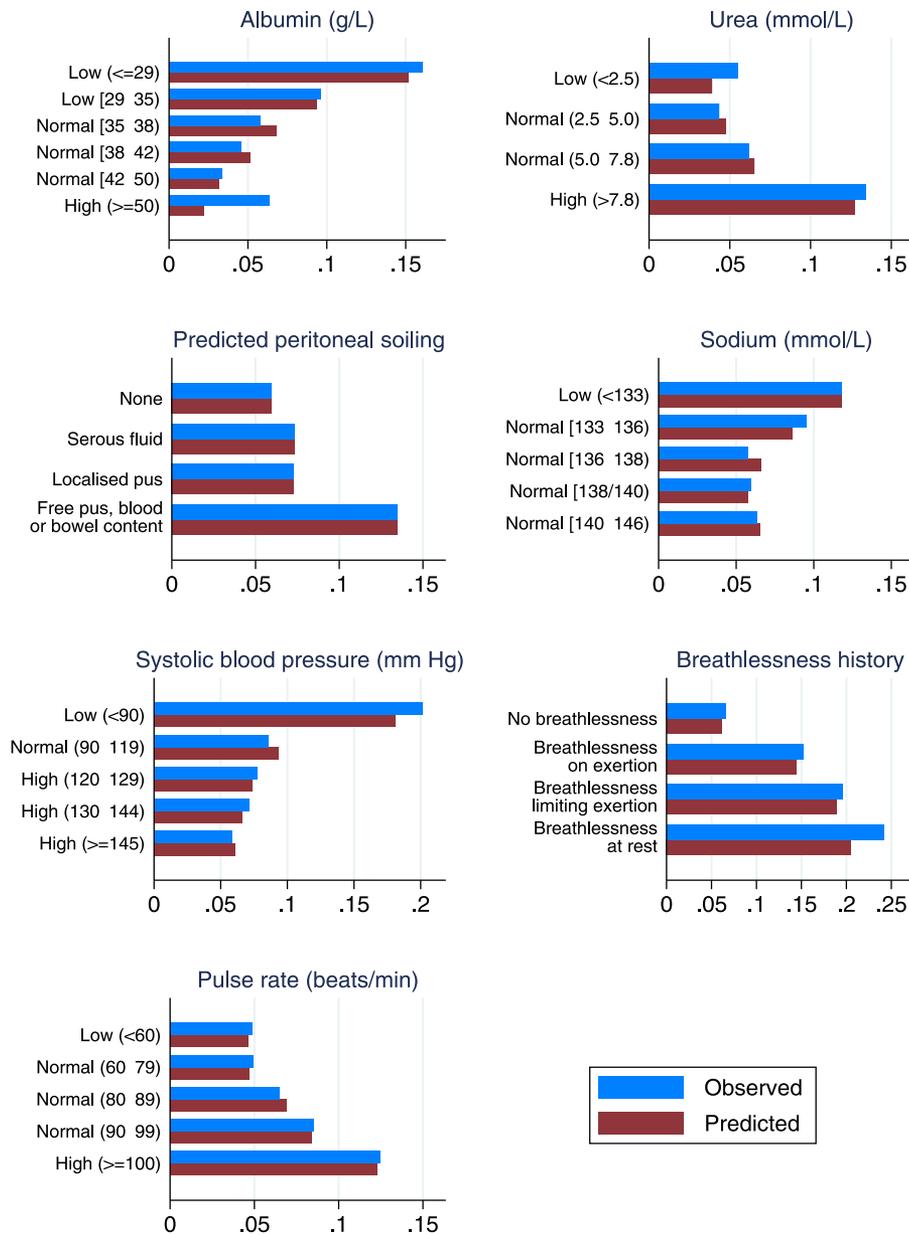
Figure 3. Calibration plot showing proportions of patients that died within 90 days (observed mortality) versus mortality predicted using the basic model, by categories of variables that are included in the selected model but not the basic model. Using validation dataset (December 2018 to November 2019).

selected model are all recorded as close as possible to the time that a patient was booked for theatre. During the development of our selected model, we included the physiological measures as continuous variables. Other risk models have categorised these variables, which leads to loss of information that could mask or exaggerate relationships[45].

We split the data into development and validation datasets by date of surgery (temporal validation), since this provides a stronger test of the validity of the predictions than splitting the data at random[46]. External validation of the selected risk model

using data from other emergency CRC surgery populations would further increase confidence in predictive ability[47].

Our study period ends in November 2019, meaning that all patients in the cohort had their emergency CRC surgery before the start of the COVID-19 pandemic. During the early pandemic period, there were rapid changes in national guidelines for cancer services, and it has been shown that postoperative mortality increased for patients having emergency CRC surgery[48]. For these reasons, we decided against using data from the early pandemic period for model development and validation. The selected

Blake et al. International Journal of Surgery (2024)

International Journal of Surgery

Table 5

Adjusted odds ratios (ORs), standard errors (SEs), and 95% CIs for the variables in the selected model (equation given in Appendix Table A.5), using multiple imputation to handle missing data. Using full dataset (December 2016 to November 2019).

	OR	SE	95% CI		OR	SE	95% CI
Calendar year				Comorbidities			
2016	1.00		(Reference)	0	1.00		(Reference)
2017	1.23	0.30	(0.76– 2.00)	1	1.20	0.13	(0.98–1.49)
2018	1.17	0.29	(0.72–1.90)	2 +	1.81	0.20	(1.45–2.26)
2019	1.17	0.29	(0.72–1.92)	Breathlessness history			
Age (years) × no metastases				No breathlessness	1.00		(Reference)
50	0.61	0.09	(0.46–0.81)	On exertion	1.18	0.14	(0.94–1.48)
60	0.77	0.04	(0.69–0.86)	Limiting exertion and at rest	1.52	0.20	(1.17–1.96)
70	1.00		(Reference)	Systolic blood pressure			
80	1.32	0.07	(1.19–1.47)	80 mmHg	1.60	0.27	(1.15–2.23)
90	1.79	0.25	(1.36–2.36)	100 mmHg	1.23	0.08	(1.08–1.39)
Age (years) × metastases				120 mmHg	1.00		(Reference)
50	1.06	0.21	(0.72–1.58)	150 mmHg	0.82	0.06	(0.72–0.95)
60	1.18	0.21	(0.83–1.68)	180 mmHg	0.78	0.16	(0.52–1.17)
70	1.40	0.23	(1.01–1.94)	Pulse			
80	1.76	0.30	(1.27–2.45)	60 beats per min	0.62	0.09	(0.46–0.84)
90	2.37	0.61	(1.43–3.92)	70 beats per min	0.75	0.06	(0.63–0.88)
Sex				90 beats per min	1.00		(Reference)
Female	1.00		(Reference)	120 beats per min	1.28	0.12	(1.06–1.55)
Male	1.10	0.09	(0.94–1.29)	140 beats per min	1.33	0.30	(0.86–2.08)
ASA grade				Urea			
1	1.00		(Reference)	2 mmol l^{-1}	0.58	0.12	(0.39–0.86)
2	1.31	0.35	(0.77–2.20)	5 mmol l^{-1}	0.72	0.04	(0.65–0.81)
3	2.52	0.67	(1.49–4.25)	10 mmol l^{-1}	1.00		(Reference)
4 or 5	4.67	1.31	(2.70–8.08)	20 mmol l^{-1}	1.57	0.14	(1.32–1.88)
Tumour stage				30 mmol l^{-1}	2.18	0.39	(1.54–3.10)
T1 or T2	1.00		(Reference)	Sodium			
T3	1.14	0.32	(0.66–1.99)	125 mmol l^{-1}	1.95	0.33	(1.40–2.73)
T4	1.25	0.36	(0.71–2.20)	130 mmol l^{-1}	1.42	0.14	(1.17–1.71)
Node stage				140 mmol l^{-1}	1.00		(Reference)
N0	1.00		(Reference)	150 mmol l^{-1}	1.73	0.82	(0.69–4.36)
N1	1.02	0.13	(0.79–1.32)	Albumin			
N2	1.51	0.22	(1.13–2.02)	25 g/l	1.66	0.10	(1.48–1.86)
Cancer site				30 g/l	1.29	0.04	(1.21–1.36)
Appendix/caecum/ascending	1.00		(Reference)	35 g/l	1.00		(Reference)
Hepatic flexure	1.28	0.26	(0.86–1.90)	40 g/l	0.78	0.02	(0.73–0.82)
Transverse colon	1.01	0.16	(0.74–1.38)	50 g/l	0.47	0.04	(0.39–0.56)
Splenic flexure/descending	1.08	0.17	(0.80–1.48)	Predicted peritoneal soiling			
Sigmoid colon	0.98	0.12	(0.77–1.25)	None	1.00		(Reference)
Rectosigmoid/Rectal	0.97	0.19	(0.66–1.42)	Serous fluid	1.08	0.12	(0.87–1.34)
Emergency admission				Localised pus	0.74	0.13	(0.52–1.03)
No	1.00		(Reference)	Free pus, blood or, bowel contents	1.54	0.20	(1.20–1.98)
Yes	1.80	0.21	(1.43–2.28)				

risk model should be recalibrated using data from after the recovery of cancer services in England, once it becomes available.

We used logistic regression to model risk of mortality. Machine learning is an alternative option, but a comparison of methods is not the focus of the current paper. Rigorous comparisons of machine learning and logistic regression have found similar performance in patients with CRC and in other patient groups. For example, a study comparing logistic regression and machine learning models to predict mortality after hospital admission using the same large national datasets demonstrated that their performance was comparable[49]. This is fully in line with a recent systematic review that found no evidence of superior performance of machine learning over logistic regression, when both were applied according to recommended analysis strategies[50]. Therefore, it is very unlikely that a comparison between statistical and machine learning methods here would provide additional clinical insight into the risks for CRC patients undergoing emergency surgery.

We have not accounted for clustering by hospital trust in our model. In order to develop a risk model that accounts for clustering, we would need to implement this into multiple imputation, which adds considerably to the complexity of the procedure[51]. Furthermore, since the within-hospital trust intraclass correlation coefficient was estimated as 3.8% (95% CI: 1.2–10.8) for the selected model using complete case analysis, accounting for
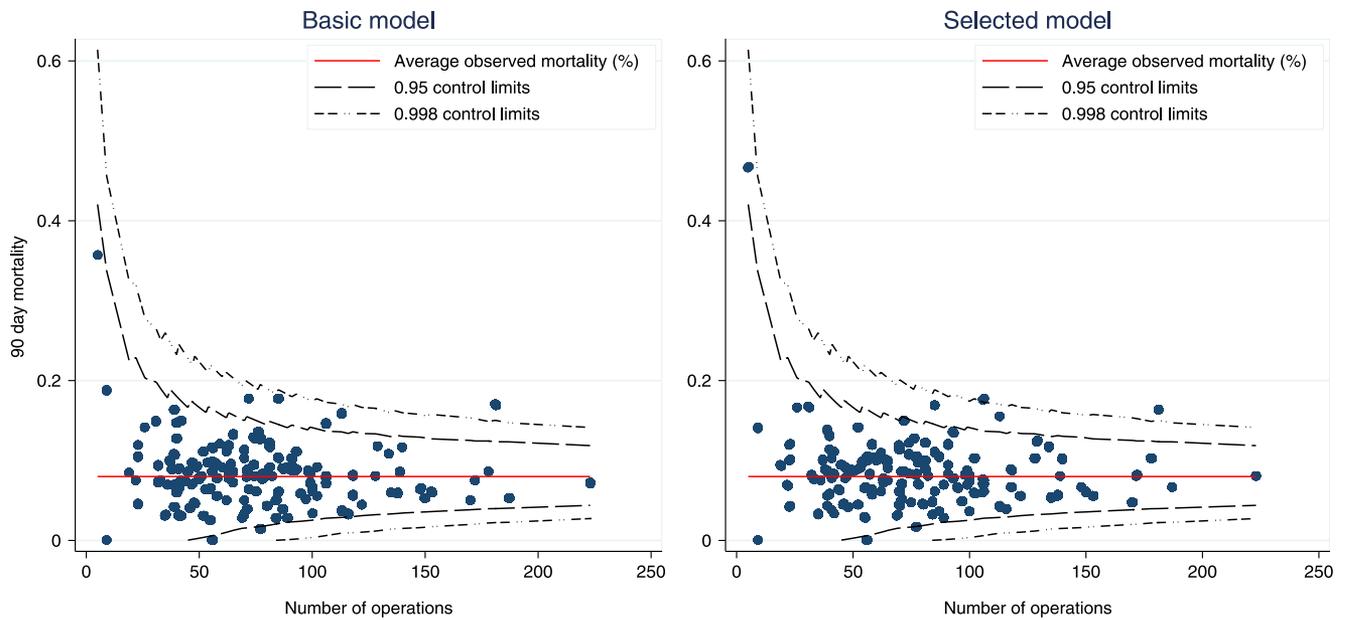
Figure 4. Funnel plots of adjusted 90-day mortality risk-adjusted using the basic model and the selected model, by hospital trust. Using full dataset (December 2016 to November 2019).

clustering would have little impact on the model intercept or the weights attached to risk measures.

We used spine linkage to construct the analysis cohort used here. In a recent paper, comparing spine linkage with using all pairwise linkages, we have shown that this spine approach is appropriate in this setting[9].

In this study, a substantial proportion of patients had missing data, the majority of which was due to nonlinkage. We used the same multiple imputation procedure for all missing data. Although we did not treat missingness in datasets differently to missingness due to nonlinkage, we did compare characteristics of patients by linkage group to assess potential for bias due to nonlinkage. Patients linked only between the administrative spine dataset and the treatment-specific dataset had higher mortality. Multiple imputation methods result in unbiased estimates of model weights, provided all important variables associated with missing values are included in the imputation models[22]. By also including mortality in these imputation models, we allowed for differences in missing measurements between survivors and nonsurvivors, which is much more plausible than assuming that missing measurements are similar in survivors and nonsurvivors. It is also important to note that a sensitivity analysis, excluding patients with missing values regardless of the reason for missingness, produced a similar pattern of results, which provides further evidence for the validity of our approach.

Implications

This new accurate and relatively simple risk prediction model for patients undergoing emergency CRC surgery is recommended for risk-adjustment and to aid clinical care planning within the context of shared decision-making with patients and their families. The model is ready to be translated into a risk calculator that can be used to predict preoperative risk for an individual patient. Since HES, the administrative dataset used as the spine dataset to define our patient cohort, contains information on hospital admissions for the whole of England[11], there was no selection of patients other than the eligibility criteria described in the methods section. This should mean that our risk model is transportable to other healthcare settings with similar patients and populations. The model should be recalibrated periodically because overall mortality is likely to change over time. Recalibration may also be necessary before being applied in healthcare settings that differ substantially from England. For settings with very different background mortality, the weights for different factors should be reliable but the intercept may need to be adjusted to ensure good overall calibration.

Inclusion of the seven physiological and surgical measures identified in our study in the risk prediction model improves the accuracy of prediction. This will, for example, help to ensure that the patients at the highest risk can be identified so that they can benefit from appropriate interventions such as the presence of a consultant surgeon and anaesthetist in theatre and direct transfer to critical care in the postoperative phase.

Whereas patient and tumour characteristics are routinely available in national clinical datasets, this is not usually the case for physiological and surgical measures. Thus, for accurate risk-adjustment of CRC patients undergoing emergency surgery, it is important either to link to existing data, capture existing hospital data into databases, or to collect the key physiological and surgical measures identified in our model.

We recommend using multiple imputation to handle missing data when using the risk model for case-mix adjustment for a cohort of patients. When using the risk model for preoperative care planning for an individual patient, it may be appropriate to assume the highest risk values of variables that are unknown.

Although the purpose of this exercise was to develop a model for risk-adjustment or clinical care planning, the findings that there are some potentially modifiable risk factors for survival may

Blake et al. International Journal of Surgery (2024)

International Journal of Surgery

have implications for future research. However, we have only established associations and predictive ability. Further work is needed to investigate whether modifying these risk factors before surgery is feasible in the emergency setting and would improve patient outcomes.

We demonstrated an approach for developing a risk model for patients undergoing emergency resection for CRC for which there is an intersection of two sources of national clinical electronic health records, and where data linkage can be used to obtain all of the required risk factors. We included approaches which others can follow for variable selection, dealing with missing data, and assessing bias from missing data due to nonlinkage. In our study, there was information available on the patient group of interest from a disease-based dataset and a treatment-based dataset. Similar situations in which both types of datasets may be available are, for example, solid organ transplantation or cardiac surgery[52,53].

Conclusion

We showed that including seven physiological and surgical measures which are easy to measure or predict from preoperative imaging, in addition to patient and tumour characteristics, improve the performance of models predicting risk for patients undergoing emergency CRC surgery. Inclusion of these additional measures, in our study available through linkage of electronic health records, will lead to more accurate performance monitoring of hospitals providing CRC surgery and enhance clinical care planning for individual patients.

Ethical approval

As the National Bowel Cancer Audit involves analysis of data for service evaluation, it is exempt from UK National Research Ethics Committee approval. Section 251 approval was obtained from the Ethics and Confidentiality Committee for the collection of personal health data without the consent of patients. The study was performed in accordance with the Declaration of Helsinki.

Consent

As the National Bowel Cancer Audit involves analysis of data for service evaluation, it is exempt from UK National Research Ethics Committee approval. Section 251 approval was obtained from the Ethics and Confidentiality Committee for the collection of personal health data without the consent of patients. The study was performed in accordance with the Declaration of Helsinki.

Sources of funding

Author contribution

H.A.B.: data curation, methodology, data analysis and interpretation, writing – original draft, review and editing; L.D.S.: funding acquisition, methodology, writing – original draft, review and editing; J.M.B.: data interpretation, writing – review and editing; A.K.: data interpretation, writing – review and editing; S.R.M.: data interpretation, writing – review and editing; D.M.: data interpretation, writing – review and editing; J.H.: data interpretation, writing – review and editing; N.S.F.: data interpretation, writing – review and editing; J.v.d.M.: conceptualisation, methodology, data interpretation, funding acquisition, writing – original draft, review and editing; K.W.: conceptualisation, methodology, data interpretation, funding acquisition, writing – original draft, review and editing.

Conflicts of interest disclosure

The authors declare no conflicts of interest.

Research registration unique identifying number (UIN)

This research was registered at ResearchRegistry.com under research registration number: 4616.

Guarantor

Helen A. Blake and Kate Walker.

Data availability statement

The data used in this study are available from NHS Digital and Public Health England's Office for Data Release but restrictions apply to the availability of these data, which were used under licence for the current study, and so are not publicly available. We do not have permission to share the patient-level records used in our analysis.

Provenance and peer review

Paper was not invited or commissioned by IJS.

Acknowledgements

individual projects, other devolved administrations and crown dependencies: hqip.org.uk/national-programmes. This work uses data provided by patients and collected by the NHS as part of their care and support. As the National Bowel Cancer Audit involves analysis of data for service evaluation, it is exempt from UK National Research Ethics Committee approval. Section 251 approval was obtained from the Ethics and Confidentiality Committee for the collection of personal health data without the consent of patients. The study was performed in accordance with the Declaration of Helsinki. The data used in this study are available from NHS Digital and Public Health England's Office for Data Release but restrictions apply to the availability of these data, which were used under licence for the current study, and so are not publicly available. The authors do not have permission to share the patient-level records used in our analysis. https://www.researchregistry.com/browse-the-registry#home/registrationdetails/5c335cb81c8513085846d2e9/.

References

[1] National Bowel Cancer Audit. Annual Report 2021. Accessed 4 May 2022. www.nboca.org.uk/reports/annual-report-2021/

[2] Royal College of Surgeons of England. The Higher Risk General Surgical Patient: Raising the Standard; 2018. Accessed 14 September 2022. https://www.rcseng.ac.uk/news-and-events/media-centre/press-releases/high-risk-general-surgical/

[3] Mahar AL, Compton C, Halabi S, et al. Personalizing prognosis in colorectal cancer: a systematic review of the quality and nature of clinical prognostic tools for survival outcomes. J Surg Oncol 2017;116:969–82.

[4] Stonelake S, Thomson P, Suggett N. Identification of the high risk emergency surgical patient: which risk prediction model should be used? Ann Med Surg 2015;4:240–7.

[5] Oliver CM, Walker E, Giannaris S, et al. Risk assessment tools validated for patients undergoing emergency laparotomy: a systematic review. Br J Anaesth 2015;115:849–60.

[6] Skala K, Gervaz P, Buchs N, et al. Risk factors for mortality-morbidity after emergency-urgent colorectal surgery. Int J Colorectal Dis 2009;24:311–6.

[7] Farooq N, Patterson AJ, Walsh SR, et al. Predicting outcome following colorectal cancer surgery using a colorectal biochemical and haematological outcome model (Colorectal BHOM). Colorectal Dis 2011;13:1237–41.

[8] Tekkis PP, Prytherch DR, Kocher HM, et al. Development of a dedicated risk-adjustment scoring system for colorectal surgery (colorectal POSSUM). Br J Surg 2004;91:1174–82.

[9] Blake HA, Sharples LD, Harron K, et al. Linkage of multiple electronic health record datasets using a 'spine linkage' approach compared with all 'pairwise linkages'. Int J Epidemiol 2022;52:214–26.

[10] NHS Digital. Hospital Episode Statistics (HES). Accessed 25 May 2020. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics

[11] Herbert A, Wijlaars L, Zylbersztejn A, et al. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). [Journal Article] Int J Epidemiol 2017;46:1093–3i.

[12] National Emergency Laparotomy Audit. The Seventh Patient Report of the National Emergency Laparotomy Audit. Accessed 21 December 2021. https://www.nela.org.uk/reports

[13] Office for National Statistics. Deaths registered in England and Wales. Accessed 1 December 2020. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/ datasets/deathsregisteredinenglandandwalesseriesdrreferencetables

[14] Paixão, Harron ES, Andrade K, et al. Evaluation of record linkage of two large administrative databases in a middle income country: stillbirths and notifications of dengue during pregnancy in Brazil. BMC Med Inform Decis Mak 2017;17:108.

[15] Walker K, Finan PJ, van der Meulen JH. Model for risk adjustment of postoperative mortality in patients with colorectal cancer. Br J Surg 2015;102:269–80.

[16] Eugene N, Oliver CM, Bassett MG, et al. Development and internal validation of a novel risk adjustment model for adult patients undergoing emergency laparotomy surgery: the National Emergency Laparotomy Audit risk model. Br J Anaesth 2018;121:739–48.

[17] Heriot AG, Tekkis PP, Smith JJ, et al. Prediction of postoperative mortality in elderly patients with colorectal cancer. Dis Colon Rectum 2006;49:816–24.

[18] Baré M, Mora L, Torà N, et al. CCR-CARESS score for predicting operative mortality in patients with colorectal cancer. Br J Surg 2018;105:1853–61.

[19] García-Torrecillas JM, Olvera-Porcel MC, Ferrer-Márquez M, et al. Predictive model of the risk of in-hospital mortality in colorectal cancer surgery, based on the minimum basic data set. Int J Environ Res Public Health 2020;17:4216.

[20] Degett TH, Christensen J, Dalton SO, et al. Prediction of the postoperative 90-day mortality after acute colorectal cancer surgery: development and temporal validation of the ACORCA model. Int J Colorectal Dis 2021;36:1873–83.

[21] Zhu J, Hao J, Ma Q, et al. A novel prognostic model and practical nomogram for predicting the outcomes of colorectal cancer: based on tumor biomarkers and log odds of positive lymph node scheme. Front Oncol 2021;11:661040.

[22] White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30:377–99.

[23] Schenker N, Taylor JMG. Partially parametric techniques for multiple imputation. Comput Stat Data Anal 1996;22:425–46.

[24] Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Med Res Methodol 2014;14:75.

[25] Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol 2009;9:57.

[26] Li KH, Meng XL, Raghunathan TE, et al. Significance levels from repeated p values with multiply-imputed data. Statistica Sinica 1991;1:65–92.

[27] Harron KL, Doidge JC, Knight HE, et al. A guide to evaluating linkage quality for the analysis of linked data. Int J Epidemiol 2017;46:1699–710.

[28] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology (Cambridge, Mass) 2010;21:128.

[29] Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? Stat Med 2008;27:3227–46.

[30] Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updatingSpringer; 2019.

[31] Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007;115:928–35.

[32] Pencina MJ, D'Agostino RB Sr. Evaluating discrimination of risk prediction models: the C statistic. JAMA 2015;314:1063–4.

[33] Caetano SJ, Sonpavde G, Pond GR. C-statistic: a brief explanation of its construction, interpretation and limitations. Eur J Cancer 2018;90:130–2.

[34] Spiegelhalter D. The art of statistics: Learning from data. Penguin, UK. 2019.

[35] Li P, Taylor JMG, Spratt DE, et al. Evaluation of predictive model performance of an existing model in the presence of missing data. Stat Med 2021;40:3477–98.

[36] Verburg IWM, Holman R, Peek N, et al. Guidelines on constructing funnel plots for quality indicators: a case study on mortality in intensive care unit patients. Stat Methods Med Res 2017;27:3350–66.

[37] Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med 2005;24:1185–202.

[38] StataCorp Stata Statistical Software: Release 17 2021College Station, TX: StataCorp LLC.

[39] Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–73.

[40] Mathew G, Agha. for the STROCSS Group R. STROCSS 2021: strengthening the reporting of cohort, cross-sectional and case-control studies in surgery. Int J Surg 2021;96:106165.

[41] Mohil RS, Bhatnagar D, Bahadur L, et al. POSSUM and P-POSSUM for risk-adjusted audit of patients undergoing emergency laparotomy. Br J Surg 2004;91:500–3.

[42] Brooks MJ, Sutton R, Sarin S. Comparison of Surgical Risk Score, POSSUM and p-POSSUM in higher-risk surgical patients. Br J Surg 2005;92:1288–92.

Blake et al. International Journal of Surgery (2024)

International Journal of Surgery

[43] Moran J, Wilson F, Guinan E, *et al*. Role of cardiopulmonary exercise testing as a risk-assessment method in patients undergoing intra-abdominal surgery: a systematic review. Br J Anaesth 2016;116:177–91.

[44] Higgins C. Urea and the clinical value of measuring blood urea concentration; 2016. Accessed 21 July 2022. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics

[45] Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006;25:127–41.

[46] Ramspek CL, Jager KJ, Dekker FW, *et al*. External validation of prognostic models: what,Â why, how, when and where? Clin Kidney J 2020;14:49–58.

[47] Steyerberg EW, Harrell FE Jr, Borsboom GJ, *et al*. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54:774–81.

[48] National Bowel Cancer Audit. Annual Report 2022. Accessed 24 January 2023. www.nboca.org.uk/reports/annual-report-2022/

[49] Cowling TE, Cromwell DA, Bellot A, *et al*. Logistic regression and machine learning predicted patient mortality from large sets of diagnosis codes comparably. J Clin Epidemiol 2021;133:43–52.

[50] Christodoulou E, Ma J, Collins GS, *et al*. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12–22.

[51] Resche-Rigon M, White IR. Multiple imputation by chained equations for systematically and sporadically missing multilevel data. Stat Methods Med Res 2018;27:1634–49.

[52] Ivanics T, Wallace D, Abreu P, *et al*. Survival after liver transplantation: an international comparison between the United States and the United Kingdom in the years 2008–2016. Transplantation 2022;106:1390–400.

[53] Ohri SK, Benedetto U, Luthra S, *et al*. Coronary artery bypass surgery in the UK, trends in activity and outcomes from a 15-year complete national series. Eur J Cardiothorac Surg 2022;61:449–56.