Original Article

Development and validation of a prognostic model for death 30 days after adult emergency laparotomy

N. Eugene,[1] **K. Angela,**[2] **P. Martin,**[3] **C. M. Oliver,**[4] **M. Berry,**[5] **I. K. Moppett,**[6] **C. Johnston,**[7] **S. Hare,**[8] **S. Lockwood,**[9] **D. Murray,**[10] **K. Walker**[11] and **D. A. Cromwell,**[12] **on behalf of the NELA Project Team**

1 Research Fellow, 2 Quantitative Analyst, Clinical Effectiveness Unit, Royal College of Surgeons of England, London, UK
3 Associate Professor, Department of Applied Health Research, University College London, London, UK
4 Associate Professor, UCL Division of Surgery and Interventional Science, University College London Hospitals NHS Foundation Trust, London, UK
5 Consultant, Critical Care, King's College Hospital NHS Foundation Trust, London, UK
6 Professor, Anaesthesia and Critical Care Section, Academic Unit of Injury, Inflammation and Repair, University of Nottingham, Nottingham, UK
7 Consultant, Department of Anaesthesia, St George's Hospital, London, UK
8 Consultant, Department of Anaesthesia, Medway Maritime Hospital, Gillingham, Kent, UK
9 Consultant, Colorectal Surgery Department, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK
10 Consultant, Department of Anaesthesia, James Cook University Hospital, Middlesbrough, UK
11 Associate Professor, 12 Professor, Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

Summary

The probability of death after emergency laparotomy varies greatly between patients. Accurate pre-operative risk prediction is fundamental to planning care and improving outcomes. We aimed to develop a model limited to a few pre-operative factors that performed well irrespective of surgical indication: obstruction; sepsis; ischaemia; bleeding; and other. We derived a model with data from the National Emergency Laparotomy Audit for patients who had emergency laparotomy between December 2016 and November 2018. We tested the model on patients who underwent emergency laparotomy between December 2018 and November 2019. There were 4077/40,816 (10%) deaths 30 days after surgery in the derivation cohort. The final model had 13 pre-operative variables: surgical indication; age; blood pressure; heart rate; respiratory history; urgency; biochemical markers; anticipated malignancy; anticipated peritoneal soiling; and ASA physical status. The predicted mortality probability deciles ranged from 0.1% to 47%. There were 1888/11,187 deaths in the test cohort. The scaled Brier score, integrated calibration index and concordance for the model were 20%, 0.006 and 0.86, respectively. Model metrics were similar for the five surgical indications. In conclusion, we think that this prognostic model is suitable to support decision-making before emergency laparotomy as well as for risk adjustment for comparing organisations.

Correspondence to: D. A. Cromwell
Email: david.cromwell@lshtm.ac.uk
Accepted: 25 June 2023
Keywords: emergency laparotomy; postoperative mortality; prediction; prognosis
Twitter: @pmstatistics; @CMOliver_; @IainMoppett; @DrCJohn; @sarah_hare; @SoniaLockwood79; @DaveMurrayAnaes

Introduction

Five years ago, we developed and published a model to produce adjusted rates of 30-day mortality after emergency laparotomy for individual hospitals [1]. The model supported the National Emergency Laparotomy Audit (NELA) to provide hospitals with comparative benchmarking. That model (the NELA casemix model) was not designed primarily to predict postoperative mortality for individual patients, and it had a number of weaknesses as a prognostic model. Some of the variables in the model were only available during or after surgery, which limited its use before surgery (to use it in this way requires the peri-operative data to be estimated). It also contained 21 variables, which placed a burden on staff in terms of data collection.

An accurate prognostic model of postoperative mortality could help clinicians discuss treatment options with patients and plan peri-operative care to limit mortality and morbidity [2]. Various models are used by clinicians to predict mortality after emergency laparotomy [3–7]. Some models have the same prognostic limitations as the NELA casemix model, and include many variables, which may mean not all of the required information is available at the time of decision-making [8]. Ideally, a prognostic model should produce accurate, clinically-relevant predictions while keeping the number of variables to a minimum [9, 10]. Casemix models perform well in the overall cohort but not necessarily well for subgroups, for instance for different indications for emergency laparotomy, such as viscus perforation, ischaemia, sepsis, bleeding or obstruction [11].

We aimed to develop a model that uses a limited number of pre-operative variables to predict the probability of dying within 30 days of emergency laparotomy. Our objective was for the model to perform well in the overall cohort and for different indications for emergency laparotomy.

Methods

We analysed data submitted to NELA by NHS hospitals in England and Health Boards in Wales for patients undergoing emergency laparotomy between 1 December 2016 and 30 November 2019. The study focused on procedures after an emergency admission and excluded unplanned surgery after an elective procedure (approximately 6% of the emergency laparotomies during the 3 years). Data collection and linkage were approved by the Confidentiality Advisory Group under section 251 of the NHS Act 2006. We identified postoperative deaths by linking NELA records to the Office for National Statistics death register. We identified 6% of deaths from unlinked hospital records.

We generated the model with emergency laparotomies performed from December 2016 to November 2018. We tested the model on emergency laparotomies performed from December 2018 to November 2019. The prognostic model was developed from the 21 variables contained in the NELA risk adjustment model, with the addition of albumin [1]. The candidate variables included patient characteristics, pre-operative laboratory tests and other clinical measurements (Table 1). We summarised the numerous surgical indications in the NELA dataset with five categories: obstruction; sepsis; ischaemia; bleeding; and other (online Supporting Information Table S1).

We recategorised respiratory history, ASA physical status and Glasgow coma scale (GCS) score because of small numbers (see Table 1 for groups). We log-transformed values for urea, white blood cell count and creatinine as their distributions were highly skewed. We winsorised the 1st and 99th percentiles from physiological variables (detailed in online Supporting Information Appendix S1 and Table S2). We used pre-operative anticipated values of four intra-operative variables: peritoneal soiling; blood loss; extent of malignancy; and operative severity.

We developed the model in two steps. First, we generated 1000 random bootstrap samples for each of four indications (`other´ had too few patients). We used backward variable selection to identify a subset from 22 potential risk factors, forcing the inclusion of patient age because of its known relationship with postoperative mortality. We included a variable in the `basic´ logistic regression model if it was selected in at least 2800/4000 (70%) bootstrap samples; a `bootstrap inclusion fraction´ (BIF) $\geq$ 70% [12]. The backward elimination p value thresholds chosen to exclude variables from each sample were 0.001 for the obstruction, sepsis and ischaemia subgroups and 0.05 for the bleeding subgroup, due to its smaller sample size.

The ordered secondary additions to the basic model were: the surgical indication variable; pre-defined interaction terms; and risk factors with BIF 50–70%. We changed the model if these additions improved the integrated calibration index for the whole cohort and for the four surgical indications.

We measured the model's accuracy with a scaled Brier score and decision curve analysis [13, 14]. We used Harrell's concordance to measure discrimination and the integrated calibration index to measure calibration [15]. We used 200 bootstrap samples (with replacement) to derive optimism-adjusted performance measures for the final model [16]. We calculated the predicted mortality probability for patients who had emergency laparotomy between December 2018

Table 1 Twenty-two pre-operative variables assessed as components of a model to predict death 30 days after emergency laparotomy. Continuous variables were analysed as such: ordered categories are for illustration and comparison [7, table 1]. Values are number (proportion).

Variable	Total n = 40,816	Deaths	Variable	Total n = 40,816	Deaths
Age; y			Sex		
18–39	4584 (11%)	92 (2%)	Female	21,279 (52%)	2112 (10%)
40–49	3807 (9%)	141 (4%)	Male	19,537 (48%)	1965 (10%)
50–59	6023 (15%)	333 (6%)	Albumin; $g.l^{-1}$		
60–69	8050 (20%)	716 (9%)	<35	19,083 (47%)	2916 (15%)
70–79	10,297 (25%)	1369 (13%)	35–50	21,227 (52%)	1135 (5%)
80–89	7107 (17%)	1243 (17%)	>50	506 (1%)	26 (5%)
90+	948 (2%)	183 (19%)	Creatinine; $\mu mol.l^{-1}$		
ECG			Men < 59; women < 45	3700 (9%)	350 (9%)
No abnormalities	33,176 (81%)	2498 (8%)	Men 59–104; women 45–84	25,589 (63%)	1539 (6%)
Atrial fibrillation 60–90.min^{-1}	1832 (5%)	351 (19%)	Men > 104; women > 84	11,527 (28%)	2188 (19%)
Other arrhythmia	5808 (14%)	1228 (21%)	Urea; $mmol.l^{-1}$		
Cardiac signs			<2.5	1779 (4%)	66 (4%)
No failure	30,059 (74%)	2137 (7%)	2.5–7.0	22,105 (54%)	1147 (5%)
Diuretic, digoxin, antihypertensive drug	8387 (21%)	1310 (16%)	>7.0	16,932 (42%)	2864 (17%)
Peripheral oedema, warfarin or CXR: borderline cardiomegaly	1995 (5%)	514 (26%)	White blood cells; $10^9.l^{-1}$		
Raised jugular venous pressure or CXR: cardiomegaly	375 (1%)	116 (31%)	<4.0	1519 (4%)	300 (20%)
Systolic blood pressure; mmHg			4.0–11.0	18,892 (46%)	1641 (9%)
<90	1585 (4%)	550 (35%)	>11.0	20,405 (50%)	2136 (10%)
90–120	15,903 (39%)	1860 (12%)	Sodium; $mmol.l^{-1}$		
>120	23,328 (57%)	1667 (7%)	<136	14,117 (35%)	1774 (13%)
Heart rate; min^{-1}			136–142	23,926 (59%)	1873 (8%)
<60	1055 (3%)	55 (5%)	>142	2773 (7%)	430 (16%)
60–100	29,669 (73%)	2221 (7%)	Potassium; $mmol.l^{-1}$		
>100	10,092 (25%)	1801 (18%)	<3.5	4352 (11%)	541 (12%)
Dyspnoea			3.5–5.0	34,106 (84%)	3017 (9%)
None	29,672 (73%)	1865 (6%)	>5.0	2358 (6%)	519 (22%)
On exertion or CXR: mild COAD	6787 (17%)	1081 (16%)	Haemoglobin; $g.l^{-1}$		
Limits exertion or at rest or CXR: moderate COAD/fibrosis/ consolidation	4357 (11%)	1131 (26%)	Men < 130; women < 115	14,895 (37%)	1993 (13%)
Glasgow coma scale			Men 130–180; women 115–165	24,889 (61%)	1923 (8%)
<14	1035 (3%)	465 (45%)	Men > 180; women > 165	1032 (3%)	161 (16%)
14	1638 (4%)	495 (30%)			
15	38,143 (94%)	3117 (8%)	Surgical severity		
Surgery during this admission			Major	25,945 (64%)	2079 (8%)
1	38,193 (94%)	3743 (10%)	Major+	14,871 (36%)	1998 (13%)
2	2409 (6%)	292 (12%)	Urgency of surgery		
>2	214 (1%)	42 (20%)	Expedited; >18 h	7100 (17%)	406 (6%)

(continued)

Table 1 (*continued*)

Variable	Total n = 40,816	Deaths	Variable	Total n = 40,816	Deaths
ASA physical status			Urgent; 6–18 h	13,903 (34%)	885 (6%)
1 or 2	18,739 (46%)	360 (2%)	Urgent; 2–6 h	15,374 (38%)	1673 (11%)
3	14,706 (36%)	1389 (9%)	Immediate; <2 h	4439 (11%)	1113 (25%)
4	6698 (16%)	1928 (29%)			
5	673 (2%)	400 (59%)			
Diagnosis indicating surgery			Anticipated blood loss; ml		
Obstruction	20,338 (50%)	1349 (7%)	<100	17,896 (44%)	1327 (7%)
Sepsis	14,047 (34%)	1524 (11%)	101–500	21,004 (52%)	2350 (11%)
Ischaemia	5245 (13%)	1024 (20%)	501–999	1505 (4%)	301 (20%)
Bleeding	981 (2%)	150 (15%)	>999	411 (1%)	99 (24%)
Other	205 (1%)	30 (15%)			
Anticipated peritoneal soiling			Anticipated malignancy		
None	16,152 (40%)	1143 (7%)	None	31,964 (78%)	3081 (10%)
Serous fluid	11,454 (28%)	1142 (10%)	Solitary	4732 (12%)	406 (9%)
Localised pus	4298 (11%)	273 (6%)	Nodal metastases	1678 (4%)	171 (10%)
Free bowel content, pus or blood	8912 (22%)	1519 (17%)	Distant metastases	2442 (6%)	419 (17%)

CXR, chest x-ray; COAD, chronic obstructive airways disease.

Table 2 Number of times that the predictor variables were included in prognostic models developed using bootstrap sampling for each patient subgroup. Each model was fitted in 1000 samples. The top 10 variables were selected in at least 2800 (70%) of the models developed for all four subgroups, which was the threshold set for inclusion in the initial prognostic model.

Pre-operative variable	Obstruction	Sepsis	Ischaemia	Bleeding	Proportion of samples variable selected
ASA physical status	1000	1000	1000	998	100%
Albumin	1000	1000	1000	971	99%
Heart rate	1000	1000	932	749	92%
Glasgow coma scale	849	1000	996	801	91%
White blood cell count	935	909	988	589	86%
Systolic blood pressure	994	1000	891	502	85%
Malignancy	1000	1000	546	767	83%
Dyspnoea	1000	1000	772	367	79%
[Urea]	999	1000	969	127	78%
Surgical urgency	632	1000	1000	208	71%
[Sodium]	1000	906	207	411	63%
ECG	719	652	941	199	63%
Peritoneal soiling	645	994	585	220	61%
[Potassium]	592	566	957	329	61%
Creatinine	982	520	268	612	60%
Blood loss	605	615	345	296	47%
Surgical severity	738	100	799	70	43%
Cardiac history	119	948	151	276	39%
Number of operations	99	801	58	218	30%
[Haemoglobin]	131	633	291	55	28%
Sex	214	113	286	199	20%

and November 2019 using the model coefficients estimated from the development dataset. The performance of the model was again summarised using the scaled Brier score, integrated calibration index and the C-statistic. We analysed data with Stata® version 17 (StataCorp LP, College Station, TX, USA).

Table 3 Thirteen pre-operative variables and 17 terms selected for the logistic regression model to predict survival 30 days after emergency laparotomy (online Supporting Information Appendix S1 provides precise coefficients).

Model term	Coefficient	Odds ratio (95%CI)	p value
Constant	-3.047		
Age; y	0.067	1.079 (1.060–1.078)	<0.001
ASA (Ref 1 or 2)			
3	1.130	3.096 (2.650–3.616)	<0.001
4	1.763	5.830 (4.959–6.853)	
5	2.553	12.85 (10.13–16.31)	
Age*ASA interaction; y			
3	-0.030	0.970 (0.961–0.980)	<0.001
4	-0.034	0.967 (0.958–0.977)	
5	-0.047	0.954 (0.941–0.968)	
[Albumin]; $g.l^{-1}$	-0.043	0.958 (0.953–0.962)	<0.001
Heart rate; min^{-1}			
Linear term	0.013	1.013 (1.010–1.015)	<0.001
Squared term	-0.0001	0.9999 (0.9998–1.0000)	0.002
Glasgow coma scale (15 ref)			
14	0.416	1.515 (1.332–1.724)	<0.001
<14	0.645	1.906 (1.622–2.239)	
White blood cells; $ln(10^9.l^{-1})$			
Linear term	0.020	1.021 (0.954–1.092)	0.551
Squared term	0.242	1.273 (1.178–1.376)	<0.001
Systolic blood pressure; mmHg			
Linear term	-0.007	0.993 (0.992–0.995)	<0.001
Squared term	0.0001	1.0001 (1.0001–1.0002)	<0.001
Anticipated malignancy (Ref none)			
Solitary	0.192	1.212 (1.071–1.371)	<0.001
Nodal metastases	0.506	1.659 (1.381–1.993)	
Distant metastases	0.943	2.568 (2.250–2.931)	
Dyspnoea (Ref none)			
On exertion or CXR: mild COAD	0.354	1.424 (1.300–1.561)	<0.001
Limits exertion or at rest or CXR: moderate COAD/fibrosis/consolidation	0.607	1.835 (1.665–2.022)	
[Urea]; $ln(mmol.l^{-1})$	0.380	1.462 (1.370–1.560)	<0.001
Urgency (Ref expedited)			
6–18 h	0.038	1.039 (0.910–1.185)	<0.001
2–6 h	0.148	1.159 (1.016–1.323)	
<2 h	0.573	1.774 (1.518–2.072)	
Indication (Ref obstruction)			
Sepsis	0.028	1.029 (0.922–1.147)	<0.001
Ischaemia	0.569	1.767 (1.575–1.983)	
Bleeding	-0.406	0.666 (0.530–0.837)	
Anticipated soiling	0.295	1.343 (1.215–1.483)	<0.001

ASA, ASA physical status; Ref, reference; ln, natural logarithm; CXR, chest X-ray; COAD, chronic obstructive airways disease.

Results

Values for all pre-operative variables and survival at 30 postoperative days were available for 62,394/69,370 (90%) emergency laparotomies performed between December 2016 and November 2019: 4695 (7%) albumin values were missing; 1084 (2%) creatinine values were missing; and 691 (1%) urea values were missing. We developed the prognostic model with the 40,816 laparotomies performed from December 2016 to November 2018 (Table 1).

After the 40,816 laparotomies, 4077 (10%) patients died within 30 days (Table 1). Mortality was associated with increasing age, lower systolic blood pressure, higher ASA physical status and lower GCS score. Physiological factors typically exhibited a U-shaped relationship with postoperative mortality, with albumin and urea being the principal exceptions.

Ten of the 21 pre-operative variables fulfilled our criterion for inclusion in the prognostic model (Table 2) and, with age, these formed the basic model. Four factors frequently associated with postoperative mortality for all indications for surgery: ASA physical status; albumin; heart rate; and GCS. Malignancy, respiratory history and surgical urgency met the 70% BIF threshold for two of four indications (see online Supporting Information Figure S1 for further details on how models containing different selections of risk factors compare to a model containing all 22 risk factors).

In the overall cohort, the scaled Brier score, integrated calibration index and concordance for the model with 11 risk factors were similar to the values for a model with age and all 21 factors: 21% vs. 21%; 0.008 vs 0.005; and 0.866 vs. 0.863, respectively (online Supporting Information Table S3). However, the basic model was only moderately calibrated for the ischaemia and bleeding subgroups (see online Supporting Information Figure S2 for the calibration plots).

Model calibration was increased by including the indication for surgery, age-ASA physical status interaction and anticipated peritoneal soiling. We characterised anticipated peritoneal soiling as `free bowel content, pus or blood´ or not, as agreement of anticipated with observed soiling was 88% for the dichotomy but 70% for the original four categories. Online Supporting Information Table S3 summarises performance metrics for each iteration of the prognostic model and online Supporting Information Figure S3 shows the results of a decision curve analysis for the final model and NELA risk adjustment models in the whole cohort.

Table 3 lists the coefficients and odds ratios (95%CI) of the final prognostic model. The non-linear relationships and interaction effect are shown in Figure 1. The distribution of

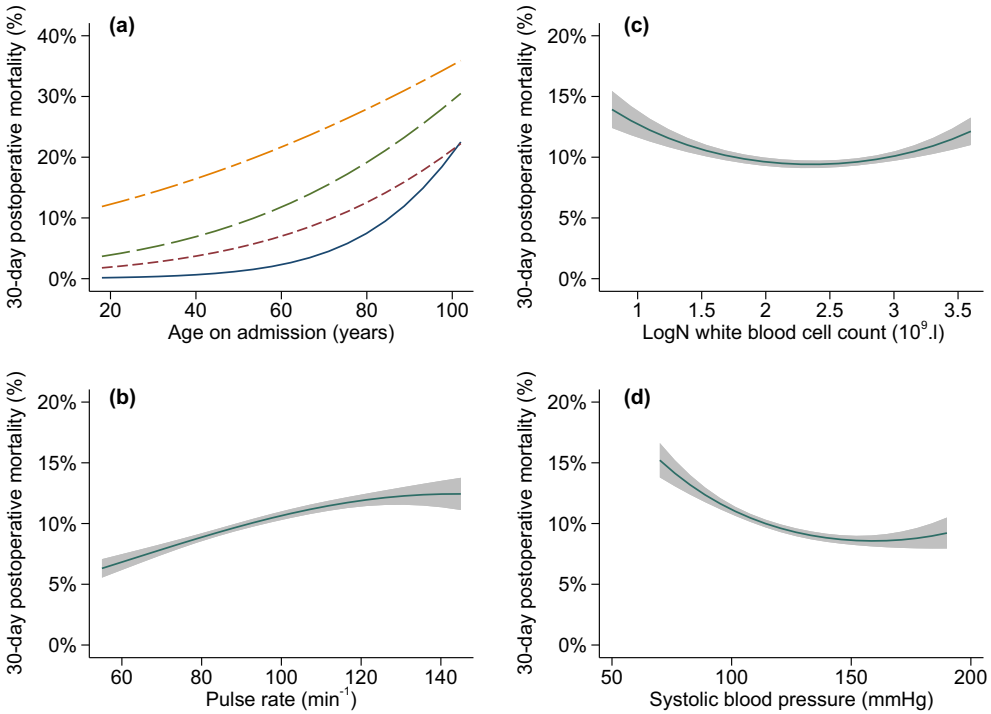

Figure 1 The non-linear association of four pre-operative continuous variables with 30-day postoperative mortality, estimated using margins command with other variables given their mean value: (a) age with ASA physical status 1 or 2 (solid black line), 3 (dashed dark red line), 4 (long dashed green line) or 5 (variable length dashed orange line); (b) heart rate; (c) log white blood cell count $10^9.l^{-1}$; (d) systolic blood pressure. Shaded areas are 95%CI.

predicted mortality probability deciles ranged from 0.1% to 47%. The differences between the predictions from the prognostic model and the NELA risk adjustment model were generally small, with a median difference (IQR) of -0.03% (-1.04–0.72%). The median difference was most pronounced for the ischaemia (1.4%) and bleeding (-1.0%) indication groups (see online Supporting Information Figure S4).

The scaled Brier score, integrated calibration index and concordance for this final model were 22%, 0.006 and 0.87, respectively. Figure 2 is the calibration plots for the final model for the overall cohort and the clinical indication subgroups. Internal validation with bootstrap resampling generated similar adjusted metrics: scaled Brier 22%; integrated calibration index 0.006; concordance 0.87.

We tested the model on 21,578 emergency laparotomies performed from December 2018 to November 2019: 11,187 (52%) due to obstruction; 7096 (33%) due to sepsis; 2751 (13%) due to ischaemia; 463 (2%) due to bleeding; and 81 other indications. There were 1888 (9%) deaths in 30 postoperative days. The scaled Brier score, integrated calibration index and concordance for the model were 20%, 0.006 and 0.86, respectively (online Supporting Information Figure S5). Figure 3 illustrates the variation in predicted mortality when stratified by age and ASA physical status for four surgical indications.

Discussion

We developed a pre-operative model to predict individual probabilities of mortality 30 days after emergency laparotomy that contained 13 risk factors. We tested the model for five surgical indications. Our model will support critical care triage and multidisciplinary discussions with patients as we think performance metrics were satisfactory at 5% and 25% hospital mortality thresholds, respectively [2].

Various risk models are available to estimate early mortality in patients after emergency laparotomy [11]. Most models used many variables; for instance, the National Surgical Quality Improvement Program uses at least 39 variables. Some variables in these models were context-sensitive, for instance

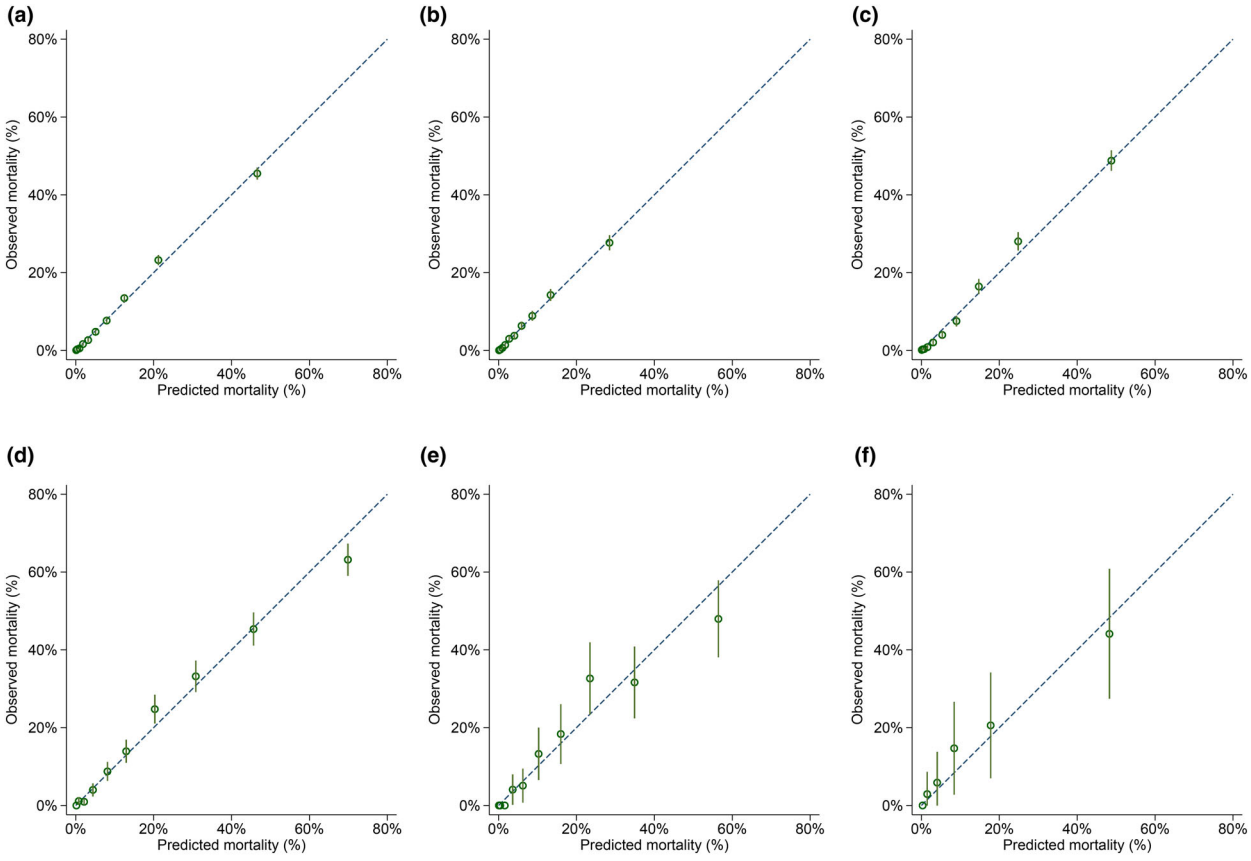

Figure 2 Calibration plots for the final NELA prognostic model, categorised by surgical indication: (a) whole cohort; (b) obstruction; (c) sepsis; (d) ischaemia; (e) bleeding; (f) other. The circles and vertical lines are the point estimates (95%CI). The dashed line is the reference line of agreement.

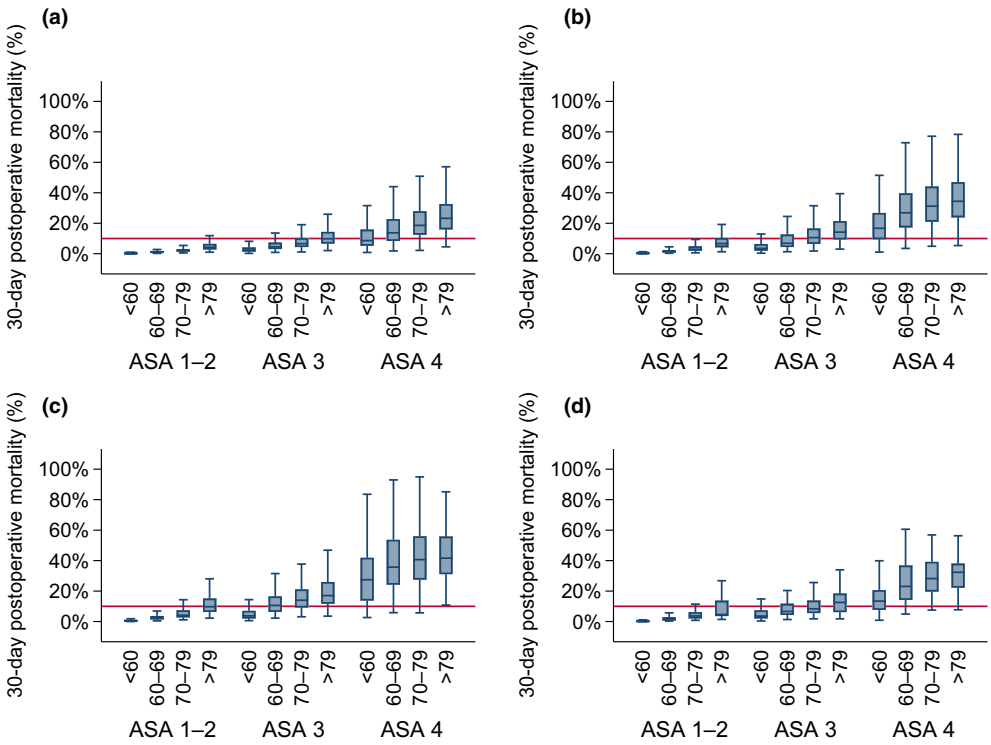

Figure 3 Distribution of predicted individual mortality probability categorised by age and ASA physical status (ASA 5 not shown as it was infrequent) and plotted by surgical indication: (a) obstruction; (b) sepsis; (c) ischaemia; (d) bleeding. The upper limit shows the maximum or the Tukey outlier limit (outliers are not shown). The red horizontal line is a mortality probability of 10%.

specific to a particular country. While the cohort of patients in the UK is likely to be similar to emergency laparotomy cohorts in other countries, we recommend recalibration of the model for other populations. The NELA team will assess the model's performance periodically to monitor whether recalibration for UK patients is required.

The discrimination of some models, for instance P-POSSUM and CR-POSSUM, can be worse for some groups – such as patients with colorectal cancer – than other groups [11]. The interaction of model performance with surgical indication is important for patients having emergency laparotomy [17]. We adopted a pragmatic definition of surgical indication as there is no agreed set of indications for which a model should be evaluated. We did not identify significant interactions between surgical indication and the other factors in our model, although it remains unclear whether model components should vary with indication.

We included pre-operative estimates of some intra-operative variables, for instance peritoneal soiling or estimated blood loss. There was only moderate agreement between anticipated vs. observed intra-operative events which might explain why estimated peritoneal soiling did not qualify for inclusion in the `basic´ model. This lack of agreement was reduced when we contracted the four categories of soiling to two. We think that predictions by this pre-operative model should be modified to accommodate new information, for instance generated during and after surgery.

We think we have generated an accurate model because we studied a large representative cohort of patients, we used Office for National Statistics registered deaths, we limited over-fitting and we tested the model on a separate cohort. We used various methods to limit data error. We excluded patients with missing data for one or more risk factors; most noticeably we excluded 7% of patients because of missing albumin values. We think it unlikely that these patients were systematically different to patients represented by the model because the distribution of other risk factors were similar. Postoperative mortality was similar in patients with missing data compared with included patients: 522/5177 (10%) vs. 4087/40,816 (10%) in the derivation cohort, $p = 0.83$; and 163/1737 (9%) vs. 1888/21,578 (9%) in the test cohort, $p = 0.37$. Some factors in our model relied on clinical judgement, such as indication for surgery and ASA physical status. Their categorisation might vary systematically between hospitals. There is no single method for measuring comorbidities such as cardiac or respiratory conditions, and the definition of surgical

urgency is UK-specific. Nonetheless, most variables in the model are commonly collected and have agreed definitions, and so the model should work similarly in other countries. We did not include subjective variables for frailty [18]. Age was an important predictor of mortality and at least some of its predictive power is likely to stem from the increasing prevalence of frailty among older patients. It is unclear the degree to which including frailty might affect the model's performance. We have started to collect pre-operative frailty measures. Finally, we did not adjust the model for seasonal variation, the effects of which vary substantially with latitude, longitude and geography.

In conclusion, we developed a prognostic model for patients undergoing emergency laparotomy with only 13 pre-operative factors. It discriminates well between low- and high-risk patients across a range of clinical indications for surgery. As such, it should better aid clinicians in determining appropriate peri-operative management and patient communication, compared with existing prognostic models. It can also be used for risk adjustment of 30-day postoperative mortality rates for comparing organisational outcomes.

Acknowledgements

We are grateful to the NHS hospital staff across England and Wales who submitted data to the Audit since 2013. Membership of the NELA Project Team during the period of data collection can be found here: https://www.nela.org.uk/downloads/List%20of%20NELA%20Project%20Team%20Members.pdf. DM's Trust is reimbursed for the time committed to NELA and he is reimbursed for NELA roles. NELA is commissioned by the Healthcare Quality Improvement Partnership as part of the National Clinical Audit Programme. The analysis and interpretation of data and the writing and submission of this manuscript were independent of the funder. The funder has no responsibility or liability for the accuracy, currency or reliability of this study. The study is exempt from UK National Research Ethics Committee approval as it involved data collected for the purposes of clinical audit. Office for National Statistics date of death were made available by NHS Digital (Copyright© 2022, reused with the permission of NHS Digital. All rights reserved). No other conflicting interests.

References

1. 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. *British Journal of Anaesthesia* 2018; **121**: 739–48.
2. Royal College of Surgeons of England. The High-Risk General Surgical Patient: Raising the Standard. 2018. https://www.rcseng.ac.uk/-/media/files/rcs/news-and-events/media-centre/2018-press-releases-documents/rcs-report-the-highrisk-general-surgical-patient--raising-the-standard--december-2018.pdf (accessed 26/06/2023).
3. Walker K, Finan PJ, van der Meulen JH. Model for risk adjustment of postoperative mortality in patients with colorectal cancer. *British Journal of Surgery* 2015; **102**: 269–80.
4. Tekkis PP, Prytherch DR, Kocher HM, Senapati A, Poloniecki JD, Stamatakis JD, Windsor ACJ. Development of a dedicated risk-adjustment scoring system for colorectal surgery (colorectal POSSUM). *British Journal of Surgery* 2004; **91**: 1174–82.
5. Sreeharsha H, Sp R, Sreekar H, Reddy R. Efficacy of POSSUM score in predicting the outcome in patients undergoing emergency laparotomy. *Polish Journal of Surgery* 2014; **86**: 159–65.
6. Prytherch DR, Sirl JS, Weaver PC, Schmidt P, Higgins B, Sutton GL. Towards a national clinical minimum data set for general surgery. *British Journal of Surgery* 2003; **90**: 1300–5.
7. Protopapa KL, Simpson JC, Smith NC, Moonesinghe SR. Development and validation of the surgical outcome risk tool (SORT). *British Journal of Surgery* 2014; **101**: 1774–83.
8. Wood AM, Royston P, White IR. The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. *Biometrical Journal* 2015; **57**: 614–3.
9. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. *Applied Statistics* 1999; **48**: 313–29.
10. Ambler G, Brady AR, Royston P. Simplifying a prognostic model: a simulation study based on clinical data. *Statistics in Medicine* 2002; **21**: 3803–22.
11. Oliver CM, Walker E, Giannaris S, Grocott MP, Moonesinghe SR. Risk assessment tools validated for patients undergoing emergency laparotomy: a systematic review. *British Journal of Anaesthesia* 2015; **115**: 849–60.
12. Royston P, Sauerbrei W. *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Continuous Variables*. Chichester: Wiley, 2008.
13. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21**: 128–38.
14. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making* 2006; **26**: 565–74.
15. Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Statistics in Medicine* 2019; **38**: 4051–65.
16. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology* 2001; **54**: 774–81.
17. Mathiszig-Lee JF, Catling FJR, Moonesinghe SR, Brett SJ. Highlighting uncertainty in clinical risk prediction using a model of emergency laparotomy mortality risk. *Nature Portfolio Journal Digital Medicine* 2022; **5**: 70.
18. Parmar KL, Law J, Carter B, et al. Frailty in older patients undergoing emergency laparotomy: results from the UK observational emergency laparotomy and frailty (ELF) study. *Annals of Surgery* 2021; **273**: 709–18.

Supporting Information

Additional supporting information may be found online via the journal website.

Appendix S1. Full details of model development and performance assessment.

Figure S1. R-square values of the simplified models when compared to the prognostic index calculated from the full model with all 22 variables.

Figure S2. Calibration plots for the basic model for the whole cohort and indication group.

Figure S3. Decision analysis curves for the NELA risk adjustment model and the final risk model.

Figure S4. Distribution of predicted risk of 30-day mortality using the NELA risk adjustment model and the final risk model stratified by the four categories of clinical indication.

Figure S5. Temporal validation of final model.

Table S1. Definition of clinical indication categories.

Table S2. Preparation of continuous physiological variables.

Table S3. Performance statistics of the full model, and the iterations of the pre-operative NELA prognostic model on the development dataset.