Edited by Smith

Alexander R Darbyshire

Department of General Surgery, Portsmouth Hospitals University NHS Trust.

Southwick Hill Road, Portsmouth, PO6 3LY

alexander.darbyshire@nhs.net.

Mobile: 07980322141

Original article

Novel predictors of mortality in emergency bowel surgery: a single centre cohort study

A.R. Darbyshire[1], I. Kostakis[2], P. Meredith[3], S.K.C. Toh[4], D. Prytherch[5] and J. Briggs[6].

1 Registrar, 4 Consultant and Professor, Department of General Surgery, 2 Research Data Manager, Research Department, Portsmouth Hospitals University NHS Trust, Portsmouth, PO6 3LY. 2 Senior Research Analyst, 5 Professor of Health Informatics and 6 Professor of Informatics, Centre for Healthcare Modelling and Informatics, University of Portsmouth, Portsmouth, PO1 3HE.

Running title: Novel Predictors of Mortality in Emergency Laparotomy

Key words: emergency, laparotomy, laparoscopy, frailty, early warning score.

Twitter handle: @AlexDarbyshire2.

Summary

Pre-operative risk stratification is a key part of the care pathway for emergency bowel surgery, as it facilitates the identification of high-risk patients. Several novel risk scores have recently been published that are designed to identify patients who are frail or significantly unwell. They can also be calculated pre-operatively from routinely collected clinical data. This study aimed to investigate the ability of these scores to predict 30-day mortality after emergency bowel surgery.

A single centre cohort study was performed using our local data from the National Emergency Laparotomy Audit database (01/12/2013 to 31/01/2020). Further data were then extracted from electronic hospital records (n=1,508). The National Early Warning Score, Laboratory Decision Tree Early Warning Score and Hospital Frailty Risk Score were then calculated. The most abnormal National or Laboratory Decision Tree Early Warning Score in the 24 or 72 hours prior to surgery were used in analysis.

Individual scores were reasonable predictors of mortality (c-statistic 0.699-0.740) but all poorly calibrated. A National Early Warning Score ≥4 were associated with a high overall mortality rate (>10%). A logistic regression model was developed using age, National Early Warning Score, Laboratory Decision Tree Early Warning Score and Hospital Frailty Risk Score as predictor variables, and its performance compared to other established risk models. The model demonstrated good discrimination and calibration (c-statistic 0.827) but was marginally outperformed by the National Emergency Laparotomy Audit score (c-statistic 0.861). All other models compared performed less well (c-statistics 0.734-0.808).

Pre-operative patient vital signs, blood tests and markers of frailty can be used to accurately predict the risk of 30-day mortality after emergency bowel surgery.

Accepted: xx December 2022

Correspondence to: A. Darbyshire

Email: alexander.darbyshire@nhs.net.

Introduction

Emergency bowel surgery continues to be a common procedure performed by general surgeons, with a high overall mortality rate [1]. In England and Wales, outcomes are monitored by the National Emergency Laparotomy Audit (NELA) against recommended standards of care [2]. At the heart of this has been the identification and prioritisation of high-risk patients. Patients are typically risk-stratified by calculating their predicted mortality for surgery with a risk prediction model [2]. Several studies have demonstrated the effectiveness of these care processes, and there has been a modest but sustained reduction in mortality (11.8% to 9.3%) [2–5]. It is evident that risk stratification remains a key part of the care pathway for emergency bowel surgery.

Separate to this, several risk scores have been developed to predict a range of adverse clinical outcomes. Vital signs are regularly measured on hospital inpatients and have been successfully incorporated into early warning scores that identify patients who are unwell with a high probability of deterioration and death [6]. The National Early Warning Score (NEWS) is a one of the most accurate scores available and has been validated in both acute medical and surgical populations [7,8]. While a variety of vital signs have been included in the risk models validated in emergency laparotomy, none have contained the full seven variables used by NEWS [9].

Routine blood tests have been built into a score similar to NEWS, called the Laboratory Decision Tree Early Warning Score (LDTEWS), using decision tree modelling. Although LDTEWS has not been adopted into clinical practice, it is equally good at detecting the same outcomes [10]. These scores have also been combined into the composite LDTEWS-NEWS Risk Score, to further improve its accuracy [11]. Patients requiring emergency bowel surgery are usually unwell, and it seems intuitive that the severity of their illness and operative risk might be reflected in their blood tests too. It is therefore reasonable to suggest that NEWS and LDTEWS could be accurate predictors of 30-day mortality after emergency bowel surgery, meriting further investigation.

Patient frailty has been shown to be associated with increased mortality after emergency bowel surgery in several observational studies [12,13]. Frailty describes a reduced physiological reserve and vulnerability to illness because of age-related accumulated deficits across multiple organ systems [14]. The Hospital Frailty Risk Score (HFRS) is designed to provide an objective measure of frailty in emergency hospital admissions in older people [15]. It utilises specific ICD-10 diagnosis codes (International Statistical Classification of Diseases and Related Health Problems, 10th revision) as surrogate markers of frailty. High scores are associated with increased odds of 30-day mortality and prolonged hospital stay;

and HFRS also correlates well with two established clinical frailty scores [15]. We felt that the Hospital Frailty Risk Score also warranted further investigation as a possible predictor of mortality after emergency bowel surgery. We therefore aimed to evaluate the predictive ability of the above scores using our local emergency laparotomy patient characteristics and outcomes data.

Methods

Health Research Authority (HRA) approval was obtained for this study from an NHS Research Ethics Committee and the HRA Confidentiality Advisory Group.

This single centre retrospective cohort study used the local National Emergency Laparotomy Database (NELA) data from Portsmouth Hospitals University NHS Trust from 01/12/2013 to 31/01/2020 as the study population (n=1,508). NELA includes adults undergoing most types of emergency bowel surgery [2]. Patient identifiers were used to extract the relevant demographic and admission data, vital signs, laboratory test results, operating theatre and intensive care data from electronic hospital records for each patient episode.

The primary outcome modelled was 30-day mortality after emergency bowel surgery.

NEWS, LDTEWS and LDTEWS-NEWS risk scores were calculated as previously described (see online Supporting Information Appendix 1) [8,10,11]. The most abnormal (i.e., highest) NEWS and LDTEWS score in the respective 24- and 72-hour periods prior to surgery were selected for the final analysis. The HFRS was calculated using ICD-10 diagnosis codes for each patient episode and applying the appropriate weightings to each unique code (see online Supporting Information Appendix 2) [15]. APACHE II (Acute Physiology And Chronic Health Evaluation) predicted mortality was extracted from the intensive care dataset for those patients who were admitted to ICU (n=725). APACHE II is a risk model developed for patients on admission to ICU, which has been widely validated in emergency laparotomy with reasonable performance [9,16]. We were not able to calculate it for the remaining patients because arterial pO_2 and pH were not available. Mortality predictions for P-POSSUM (Portsmouth-Physiological and Operative Score for the enUmeration of Mortality and morbidity score) and the NELA score were extracted from the NELA dataset. P-POSSUM was available for the total cohort, and NELA score from 2016 onwards (n=649). The SORT (Surgical Outcomes Research Tool) predicted mortality was calculated using the algorithm from the first version of the model[17]. The recently updated model has an additional variable for a clinician's subjective assessment of risk, which we do not have data to calculate it [18].

The sample size was calculated with the "pmsampsize" package in R, using by *Riley et al's* methodology [19]. With the disease prevalence of 0.1, a minimum R^2 of 0.2 and a margin of error of 0.05, a minimum sample size of 538 participants would be required for a model using a maximum of 15 predictor variables.

The amount and distribution of missing data was examined and is presented in online Supporting Information Appendices 3 and 4. The pattern of 'missingness' visually appeared to be random over time and for a limited number of variables or patient episodes. Sensitivity analysis revealed subtle differences between episodes with complete/incomplete data but was not felt to introduce significant selection bias.

The entire available data was used to externally validate NEWS (n=1 342), LDTEWS (n=1,385), LDTEWS-NEWS risk score (n=1,252) and HFRS (n=1 505) as predictors of 30-day mortality. In a second step, the complete available data (n=1,237) was randomly split into training and test datasets for PRE-OP model development and internal validation. An 80/20 split was performed and stratified by 30-day mortality.

In exploratory data analysis, the relationships between each available vital sign, blood test, risk score, age and 30-day mortality were visualised by plotting mortality against each variable with a smoothed-regression line to display any trends (data not shown). Patient age, NEWS, LDTEWS, LDTEWS-NEWS and HFRS risk score displayed clear linear associations with mortality. The individual vital signs and blood tests all displayed varying degrees of non-linearity/U-shaped relationships. Based on these observations, we opted to further investigate only performance of aggregate risk scores.

Score and model performance was assessed in terms of discrimination and calibration. Discrimination demonstrates how likely the model is to correctly classify whether a patient is dead or alive at the defined endpoint. It is assessed by plotting receiver operating characteristic (ROC) curves and calculating the c-statistic (equivalent to the area under the curve). A c-statistic ≥0.7 is considered adequate, ≥0.8 to have clinical utility and ≥0.9 excellent; with 0.5 no better at predicting the outcome than random chance [20]. Calibration is assessed by evenly dividing the data into deciles of predicted risk, then plotting the observed vs predicted mortality rate for each decile. A line of best fit is plotted through the data points to help visualise how well the model's predictions match observed outcomes as risk increases [20,21]. The relationship between each of the scores was also plotted against mortality rate, to visualise the thresholds of risk for each possible score.

Logistic regression models were developed on the training dataset using patient age, NEWS, LDTEWS, LDTEWS-NEWS and HFRS risk scores as predictor variables. The performance of models using either NEWS and LDTEWS as separate variables, or the composite LDTEWS-NEWS risk score, was compared using 10-fold cross validation. The final model's performance was internally validated on the test data set, and compared to the performance of SORT, P-POSSUM, the NELA score and APACHE II on the entire data available for each of these models.

Data analysis was performed in R Studio: R Foundation for Statistical Computing 2020 (Vienna, Austria).

Results

Overall, 199 of the 1,362 patients died within 30 days of emergency bowel surgery (13.2%). A summary of demographic, operative data and outcomes for survivors and non-survivors is provided in Table 1, and the most frequently performed procedures in Table 2.

All the four scores demonstrated fair to adequate discrimination for 30-day mortality (Fig. 1). NEWS, LDTEWS and LDTEWS-NEWS risk score did not perform as well as predictors of post-operative mortality as they did for 24-h mortality in the original published studies, the respective c-statistics (with 95%CI when available) being 0.708 (0.663-0.753) vs. 0.894 (0.887–0.902) for NEWS; 0.727 (0.688-0.765) vs. 0.801 (0.776–0.826) for LDTEWS; 0.740 (0.695-0.784) vs. 0.916 (0.911–0.921) for the LDTEWS-NEWS risk score [8,10,11]. HFRS performed slightly better as a predictor of post-operative mortality than for 30-day mortality in emergency medical admissions, with a c-statistic (95% CI) of 0.699 (0.660-0.739) vs. 0.600 [15].

Calibration plots demonstrated that all scores were to some degree mis-calibrated (Fig. 2). The HFRS calibration line most closely followed the line of best fit, though did underpredict mortality in lower risk deciles. This was followed by NEWS and LDTEWS-NEWS risk score, but they increasingly overpredicted mortality in all but the low-risk deciles. LDTEWS score had the worst calibration, consistently overpredicting mortality across all risk deciles.

Figure 3 shows the distribution of each NEWS value as a proportion (blue bars), together with the mortality rate for each score (black points) and a line of best fit to display the trend. Beyond a threshold NEWS of 4, the mortality rate increased above the average for the dataset (13.2%). While there was some variability in mortality rate for the higher NEWS scores because of the low number of cases, the trend of increasing mortality is very clear. It suggests that patients requiring emergency bowel surgery with a NEWS of ≥4 could all be considered high-risk. Plots of LDTEWS and LDTEWS-NEWS risk score using the same methodology had similar findings (see online Supporting Information Appendix 5).

The prediction model developed using the individual NEWS and LDTEWS scores (which we termed PRE-OP) had slightly better performance than using the combined LDTEWS-NEWS risk score (c-statistic 0.802 vs 0.793). Thus, the individual scores were included in the final PRE-OP model. The model coefficients are displayed in online Supporting Information Appendix 6. PRE-OP demonstrated good discrimination in the test dataset, with a c-statistic (95%CI) of 0.827 (0.749-0.903), Fig 4. It was also well calibrated with the calibration plot closely overlying the line of best fit; though it did over- and under-predict mortality for the two highest-risk deciles (Fig. 5).

The NELA score had the best discrimination, the c-statistic (95%CI) being 0.861 (0.823-0.899), with SORT and P-POSSUM demonstrating similar levels of good discrimination at 0.808 (0.778-0.837) and 0.796 (0.767-0.826) respectively). APACHE II had the weakest discrimination of the established models, its c-statistic (95%CI) being 0.734 (0.684-0.784), Fig. 4). In terms of calibration, the NELA score performed well, though it did slightly underpredict mortality in the higher-risk deciles. SORT was not particularly well calibrated, with the lowest risk deciles concentrated below 5% predicted mortality, and then underpredicting mortality beyond this. P‐POSSUM was overall poorly calibrated, increasingly overpredicting mortality beyond 20%, but was accurate in the deciles below this threshold. APACHE II appeared to have overall good calibration but exhibited considerable variation at lower risk (Fig. 5).

Discussion

This study has been the first to investigate pre-operative NEWS, LDTEWS, LDTEWS-NEWS and HFRS risk scores as predictors of 30-day mortality after emergency bowel surgery. While originally developed to identify unwell patients at risk of clinical deterioration, or frailty on hospital admission, they are strongly associated with this outcome. The individual scores had acceptable discrimination but predicted outcomes were poorly calibrated. Combining NEWS, LDTEWS, HFRS and age into a new, simple logistic regression model (PRE-OP), has overcome this problem, and proved that together they are accurate predictors of 30-day mortality.

The importance of these findings is that vital signs and blood tests are monitored by clinicians as part of routine care, and patient frailty assessments are now widely used [2,22]. This study suggests that these variables, in addition to their current use, have a combined prognostic value in estimating post-operative mortality. Furthermore, these associations are independent to the underlying surgical pathology or the procedure being performed, something which typically underpins the degree of operative risk [23]. The fact that PRE-OP's performance is comparable to excellent models designed for emergency bowel surgery like the NELA score, or more general risk predictor SORT, also lends significant credibility to this study.

The NELA score was the best performing model, with the highest c-statistic (0.861) and well calibrated throughout most risk deciles. This is not surprising given that it was developed on a large modern dataset of patients undergoing emergency bowel surgery. It also utilises data on operative findings, urgency of surgery and ASA physical status grade, in addition to laboratory tests, some vital signs and markers of chronic diseases [23]. These diverse predictor variables represent many of the factors that influence operative risk, and partly explain its good performance in this and other studies [24,25]. The fact that PRE-OP can perform nearly as well as the NELA score without including any operative data (c-statistic 0.827), strongly suggests that NEWS, LDTEWS and HFRS are robust predictors of 30-day mortality.

PRE-OP outperformed SORT, but only by a small margin (c-statistic 0.808). However, SORT was not well calibrated and consistently underestimated risk beyond 5% predicted mortality. Recent external validation of SORT by Wong et al. in their study of the epidemiology of critical care provision after surgery (EPICCS), found it had excellent discrimination (c-statistic 0.90) and was well calibrated [18]. However, their study included a range of procedures from all surgical specialties, and low mortality rate (1.4%) [18]. It is possible that SORT underestimated risk in this study because of the higher outcome rate and being restricted to emergency gastrointestinal procedures.

P-POSSUM had a similar c-statistic to SORT (0.796) but over-predicted mortality; this was consistent with previous reports. This can be explained by the fact the model was developed around 20 years ago when surgical mortality was higher [23,26]. P-POSSUM had much better discrimination in the EPICCS study (c-statistic 0.89) but was poorly calibrated. Perhaps it shares a similar problem to SORT in that it is poorer at differentiating outcomes in the higher-risk emergency laparotomy population. APACHE II did not perform as well as in other studies, which is reasonable given that it was created in the 1980s to predict outcomes for a different population of patients on admission to ICU [16].

This study's findings are limited by their single centre nature, making them less generalisable to other hospitals. There will potentially be some differences in the characteristics of the study population that could result in PRE-OP being less accurate when externally validated in other hospitals/geographical areas. For example, the rates of attempted/successful laparoscopy are much higher in Portsmouth University Hospitals than the national average (55.9%/37.5% vs 20.2%/10.0%)[2]. There is now good evidence to show that performing emergency bowel surgery laparoscopically confers a reduction in mortality compared to an open approach [27]. This could therefore reduce PRE-OP accuracy in data on exclusively open surgery. However, a recently published analysis of national NELA data has found that P-POSSUM and the NELA score consistently over-predict mortality for laparoscopic surgery [26]. Thus, it may be of benefit that PRE-OP has been developed on a more even mix of operative approaches, allowing it to provide a more balanced, if slightly less accurate, prediction.

Any modelling study that does not validate the model on an external dataset is at risk of unrecognised 'over-fitting'. Over-fitting is when the model fits the training dataset so closely that it cannot be well applied to new data [20]. We limited this as much as possible by creating a simple logistic regression model, using clinically interpretable predictor variables, that has been trained and tested on separate, randomly-divided datasets. External validation on data from another centre is required before PRE-OP can be considered a reliable risk predictor.

This is a retrospective observational study and thus brings the potential for selection bias. However, the patient data were recorded prospectively into both the NELA audit and electronic hospital records, improving their reliability and reducing the chance of selection bias. We have also used a robust and transparent methodology. Unfortunately, we could only compare NELA score and APACHE II performance for approximately half the patients in the dataset. Nonetheless, we feel that the findings from available data are still valid and useful.

The strength of this exploratory study is the finding that routinely-recorded clinical data can accurately predict mortality after emergency bowel surgery. Clinicians are now widely familiar with early warning scores like NEWS, as they have been mandatory in the UK since 2007 [22]. The fact that a NEWS ≥4 is associated with a higher-than-average mortality rate alone is an important finding for clinicians. It may seem intuitive that patients who are unwell prior to surgery are at increased operative risk compared to those who are not. This raises the question of whether there are certain surgical conditions or patient groups for which high NEWS/LDTEWS scores confer particularly poor outcomes, or whether medical optimisation prior to surgery for patients with high scores leads to improved outcomes.

One could argue that we have simply dismantled and re-assembled some of the components of the NELA score or P-POSSUM. However, both models only include eight of the 14 physiological variables utilised in NEWS and LDTEWS. We therefore feel that PRE-OP captures significantly more data on a patient's physiological condition and justifies its value.

Frailty is now well recognised as being associated with increased risk of post-operative morbidity and mortality, independent of patient age [2,28]. This study builds on the findings of two previous studies [12,13] and is the first to validate the Hospital Frailty Risk Score in a surgical population. Another strength of PRE-OP's use of pre-operatively available variables is that it could be validated on patients with an intestinal emergency who do not undergo surgery. Research suggests this may be as high as one-third of admissions, and the lack of risk stratification tools for this population has been highlighted by the UK Royal College of Surgeons [29,30]. A risk model that can provide risk predictions for both emergency bowel surgery and non-operative management would be highly useful for clinicians in the decision making, and counselling of patients and their relatives.

We have shown that a patient's NEWS score, routine blood tests and frailty scores prior to surgery are associated with 30-day mortality and can be used in a simple regression model to provide accurate risk predictions. Future studies could use exclusively pre-operative variables to provide mortality predictions for patients who are candidates for emergency bowel surgery but who do not undergo an operation.

Acknowledgments

This study was funded by the Research Department of Portsmouth Hospitals University NHS Trust. The authors declare no conflicts of interest. The Research Ethics Committee reference for the study is 20/SC/0156 and the HRA Confidentiality Advisory Group reference is 20/CAG/0055.

References

1. Wohlgemut JM, Ramsay G, Jansen JO. The Changing Face of Emergency General Surgery: A 20-year Analysis of Secular Trends in Demographics, Diagnoses, Operations, and Outcomes. *Annals of Surgery* 2020; 271: 581–9.

2. The NELA Project Team. *Sixth Patient Report of the National Emergency Laparotomy Audit*. London: Royal College of Anaesthetists, 2020:1–51.

3. Aggarwal G, Peden CJ, Mohammed MA et al. Evaluation of the collaborative use of an evidence-based care bundle in emergency laparotomy. *JAMA Surgery* 2019; 154: e190145-10.

4. Peden CJ, Stephens T, Martin G et al. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery (EPOCH): a stepped-wedge cluster-randomised trial. *Lancet* 2019; 393: 2213–21.

5. Oliver CM, Bassett MG, Poulton TE et al. Organisational factors and mortality after an emergency laparotomy: multilevel analysis of 39 903 National Emergency Laparotomy Audit patients. *British Journal of Anaesthesia* 2018; 121: 1346–56.

6. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS--Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation* 2010; 81: 932–7.

7. Kovacs C, Jarvis SW, Prytherch DR et al. Comparison of the National Early Warning Score in non-elective medical and surgical patients. *British Journal of Surgery* 2016; 103: 1385–93.

8. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation* 2013; 84: 465–70.

9. Oliver CM, Walker E, Giannaris S, Grocott MPW, Moonesinghe SR. Risk assessment tools validated for patients undergoing emergency laparotomy: a systematic review. *British Journal of Anaesthesia* 2015; 115: 849–60.

10. Jarvis SW, Kovacs C, Badriyah T et al. Development and validation of a decision tree early warning score based on routine laboratory test results for the discrimination of hospital mortality in emergency medical admissions. *Resuscitation* 2013; 84: 1494–9.

11. Redfern OC, Pimentel MAF, Prytherch D et al. Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: Development and validation of a multivariable model. *Resuscitation* 2018; 133: 75–81.

12. 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* 2019; 273: 709–18.

13. Body S, Ligthart MAP, Rahman S et al. Sarcopenia and myosteatosis predict adverse outcomes after emergency laparotomy: a multi-centre observational cohort study. *Annals of Surgery* 2021; e - Publish Ahead of Print.

14. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet* 2013; 381: 752–62.

15. Gilbert T, Neuburger J, Kraindler J et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet* 2018; 391: 1775–82.

16. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Critical Care Medicine* 1985; 13: 818–29.

17. Protopapa KL, Simpson JC, Smith NCE, Moonesinghe SR. Development and validation of the Surgical Outcome Risk Tool (SORT). *British Journal of Surgery* 2014; 101: 1774–83.

18. Wong DJN, Harris S, Sahni A et al. Developing and validating subjective and objective risk-assessment measures for predicting mortality after major surgery: An international prospective cohort study. *PLOS Medicine* 2020; 17: e1003253-22.

19. Riley RD, Ensor J, Snell KIE et al. Calculating the sample size required for developing a clinical prediction model. *British Medical Journal* 2020; 368: m441-12.

20. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European Heart Journal* 2014; 35: 1925–31.

21. Calster BV, McLernon DJ, Smeden M van, Wynants L, Steyerberg EW, initiative TG 'Evaluating diagnostic tests and prediction models' of the S. Calibration: the Achilles heel of predictive analytics. *BMC Medicine* 2019; 17: 230–7.

22. National Institute for Health and Clinical Excellence. *Acutely Ill Patients in Hospital: Recognition of and Response to Acute Illness in Adults in Hospital*. London: National Institute for Health and Clinical Excellence, 2007.

23. 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.

24. Barazanchi A, Bhat S, Palmer-Neels K et al. Evaluating and improving current risk prediction tools in emergency laparotomy. *The journal of trauma and acute care surgery* 2020; 89: 382–7.

25. Boyd-Carson H, Shah A, Sugavanam A, Reid J, Stanworth SJ, Oliver CM. The association of pre-operative anaemia with morbidity and mortality after emergency laparotomy. *Anaesthesia* 2020; 75: 904–12.

26. Darbyshire AR, Kostakis I, Pucher PH, Prytherch D, Mercer SJ. P-POSSUM and the NELA Score Overpredict Mortality for Laparoscopic Emergency Bowel Surgery: An Analysis of the NELA Database. *World Journal of Surgery* 2021: 1–9.

27. Pucher PH, Mackenzie H, Tucker V, Mercer SJ. A national propensity score-matched analysis of emergency laparoscopic versus open abdominal surgery. *British Journal of Surgery* 2021: 1–7.

28. Carter B, Law J, Hewitt J et al. Association between preadmission frailty and care level at discharge in older adults undergoing emergency laparotomy. *British Journal of Surgery* 2020; 107: 218–26.

29. Royal College of Surgeons England. *The High-Risk General Surgical Patient: Raising the Standard.*.London, RCSE, 2018:1–64.

30. McIlveen EC, Wright E, Shaw M et al. A prospective cohort study characterising patients declined emergency laparotomy: survival in the 'NoLap' population. *Anaesthesia* 2020; 75: 54–62.

Figure legends

Figure 1 ROC curves for NEWS — LDTEWS — LDTEWS-NEWS — and HFRS — risk scores.

Figure 2 Calibration plots for NEWS — LDTEWS — LDTEWS-NEWS — and HFRS — risk scores.

Figure 3 Cumulative distribution NEWS scores as a proportion (blue bars), with the mortality rate for each score (black points) with line of best fit.

Figure 4 ROC curves for PRE-OP — SORT — NELA score — P-POSSUM — and APACHE II — risk models.

Figure 5 Calibration plots PRE-OP — SORT — NELA score — P-POSSUM — and APACHE II — risk models.

Supporting Information

Appendix 1 NEWS, LDTEWS and LDTEWS-NEWS Risk Score calculations.

Appendix 2 ICD 10 codes with scores used to calculate the HFRS scores.

Appendix 3 Missing data analysis.

Appendix 4 Map of missing data items.

Appendix 5 Mortality plots LDTEWS, LDTEWS-NEWS and HFRS risk scores.

Appendix 6 PRE-OP model coefficients.

Table 1 Summary of demographic, operative data and outcomes for survivors and non-survivors. Values are number (proportion) or median (IQR [range]).

	Survivor n=1,362	Non-survivor n=199
Age; y	66.0 (51.0-77.0 [18-100])	75.0 (67.0-82.5 [25-96])
Female	736 (54.0%)	101 (50.8%)
Admission type		
Elective	85 (6.2%)	16 (8.0%)
Emergency	1,277 (93.8%)	183 (92.0%)
Admission specialty		
General surgery	1,121 (82.3%)	141 (70.9%)
Other	241 (17.7%)	58 (29.1%)
Charlson Comorbidity Index	0.0 (0.0-8.0 [0.0-37.0])	11.0 (4.0-20.0 [0.0-37.0])
ASA physical status grade	2 (2-3 [1-5])	3 (3-4 [1-5])
Urgency of surgery		
<2 h	134 (9.8%)	51 (25.6%)
2-6 h	581 (42.7%)	75 (37.7%)
6-18 h	440 (32.3%)	53 (26.6%)
>18 h	207 (15.2%)	20 (10.1%)
Time to surgery, h	1.8 (0.8-4.0 [0.5-92.7])	1.7 (0.7-6.8 [0.1-53.2])
BUPA Operative Severity		
Major	1,098 (80.6%)	122 (61.3%)
Major Plus	264 (19.4%)	77 (38.7%)
Operative approach		
Open	548 (40.2%)	140 (70.4%)
Laparoscopic	559 (41.0%)	27 (13.6%)
Converted	197 (14.5%)	28 (14.1%)
Laparoscopically-assisted	58 (4.3%)	4 (2.0%)
Peritoneal soiling		
None	462 (33.9%)	52 (26.1%)
Serous fluid	415 (30.5%)	52 (26.1%)
Localised pus	98 (7.2%)	10 (5.0%)
Gross contamination	387 (28.4%)	85 (42.7%)
Malignancy		
None	1,128 (82.8%)	150 (75.4%)
Primary disease	122 (9.0%)	19 (9.5%)
Nodal metastases	34 (2.5%)	8 (4.0%)
Distant metastases	78 (5.7%)	22 (11.1%)
Return to theatre		
None	1,227 (90.1%)	143 (71.9%)
Planned	11 (0.8%)	2 (1.0%)

	Survivor n=1,362	Non-survivor n=199
Unplanned	124 (9.1%)	54 (27.1%)
ICU length of stay, days	3.0 (1.6-5.8 [0.0-89.0])	3.6 (1.5-8.6 [0.0-58.4])
Post-operative length of stay, days	10.6 (6.0-20.2 [1.5-166.5])	9.0 (2.4-22.8 [1.8-200.4])

ASA = American Association of Anaesthesiologists; BUPA = British United Provident Association; ICU = Intensive Care Unit.

Table 2 Most frequently performed procedures. Values are number (proportion).

Procedure	Survivor n=1,362	Non-survivor n=199
Adhesiolysis	299 (96.5%)	11 (3.5%)
Small bowel resection	217 (85.4%)	37 (14.6%)
Right colectomy	162 (85.7%)	27 (14.3%)
Hartmann's procedure	99 (78.0%)	28 (22.0%)
Stoma formation	95 (86.4%)	15 (13.6%)
Perforated peptic ulcer repair	68 (87.2%)	10 (12.8%)
Subtotal colectomy	52 (83.9%)	10 (16.1%)
Washout only	53 (94.6%)	3 (5.4%)
Drainage of abscess	47 (94.0%)	3 (6.0%)
Other	44 (88.0%)	6 (12.0%)
Left colectomy/anterior resection	34 (85.0%)	6 (15.0%)
Repair of intestinal perforation	29 (87.9%)	4 (12.1%)
Other colorectal resection	22 (78.6%)	6 (21.4%)
Gastric surgery	22 (88.0%)	3 (12.0%)

Figure 1

Click here to access/download;Figure;Figure 1. ROC Curves NEWS LDTEWS HFRS.jpeg

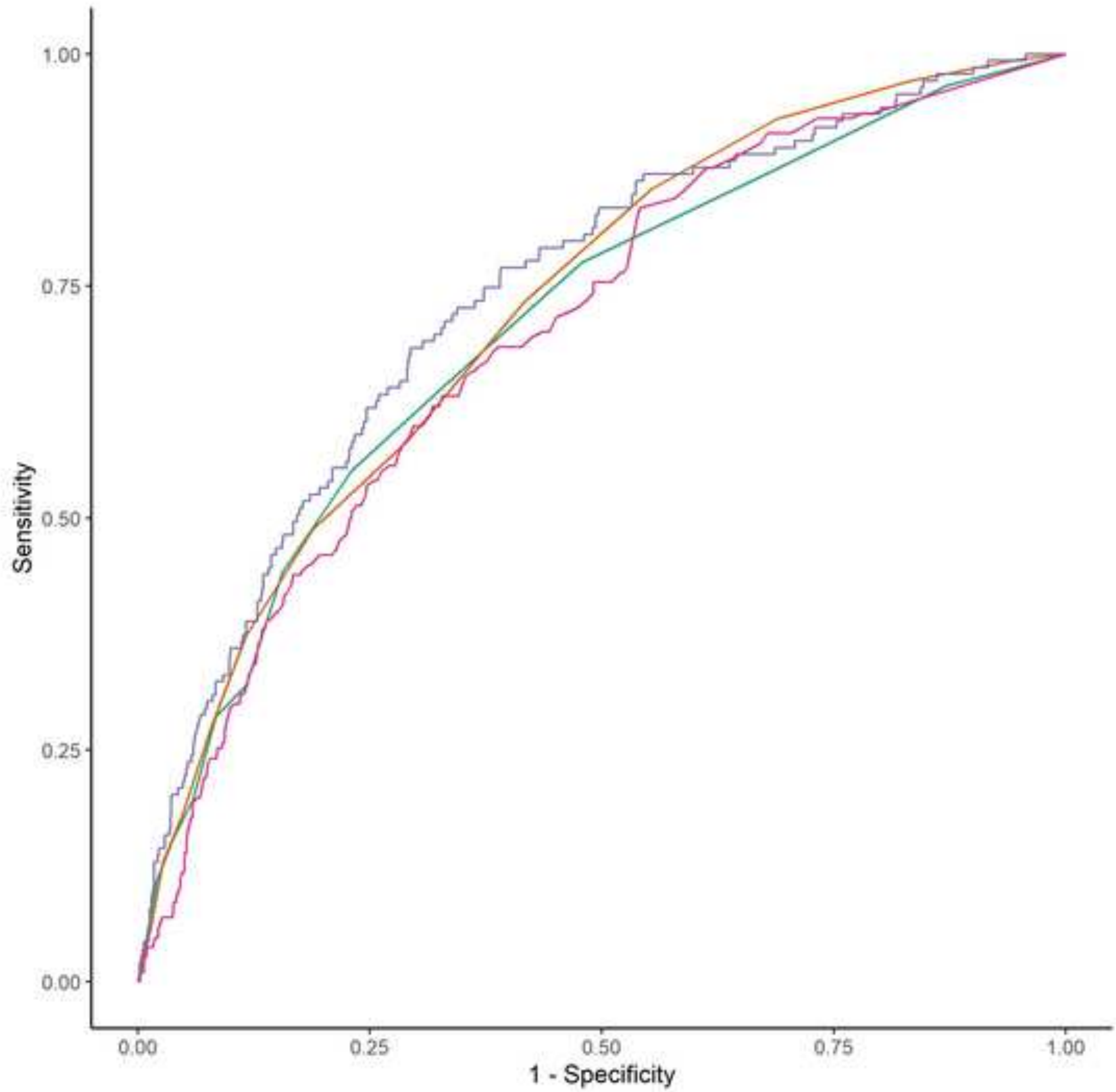


Figure 2

Click here to access/download;Figure;Figure 2. Calibration Plots NEWS LDTEWS HFRS.jpeg

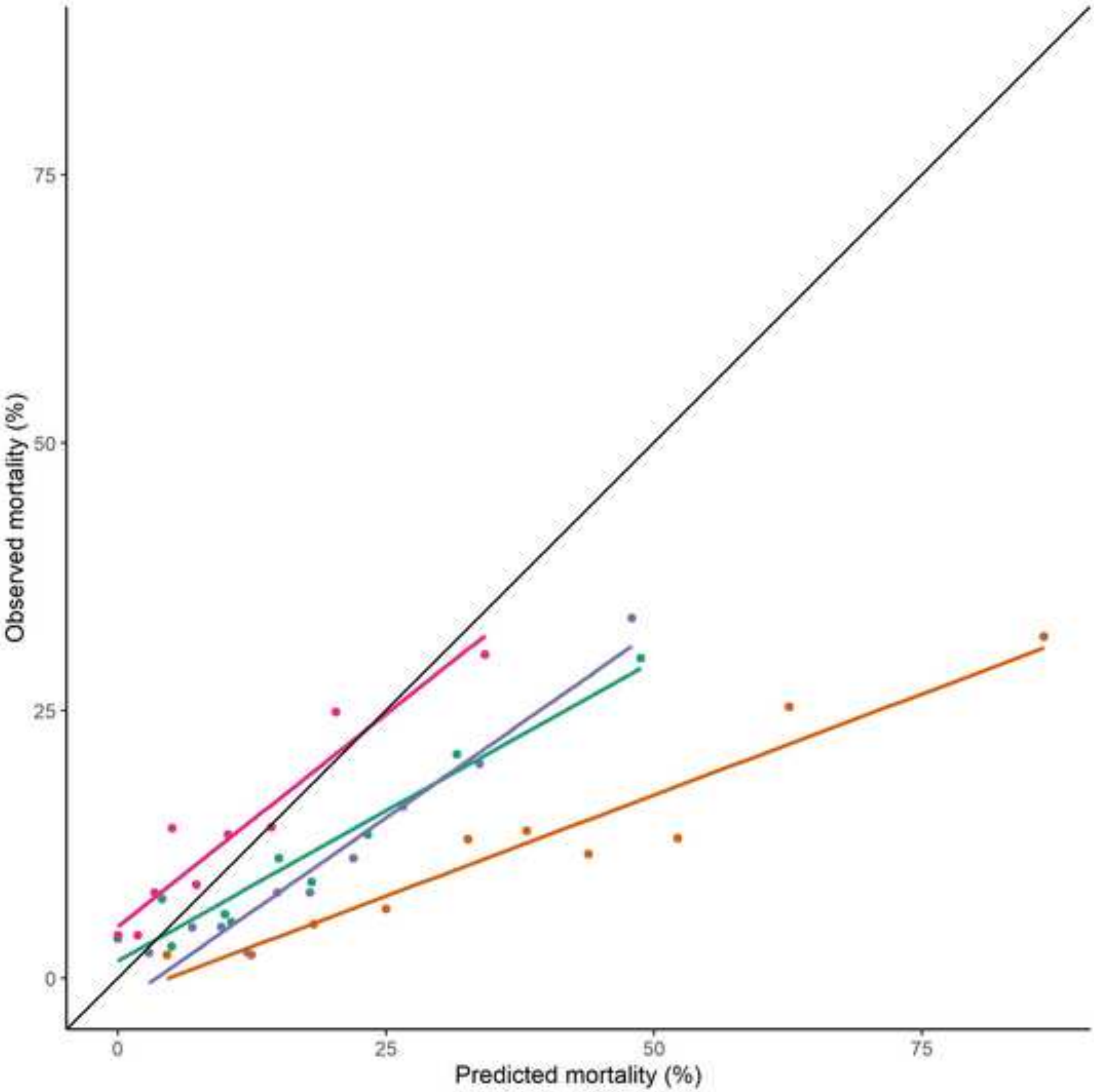
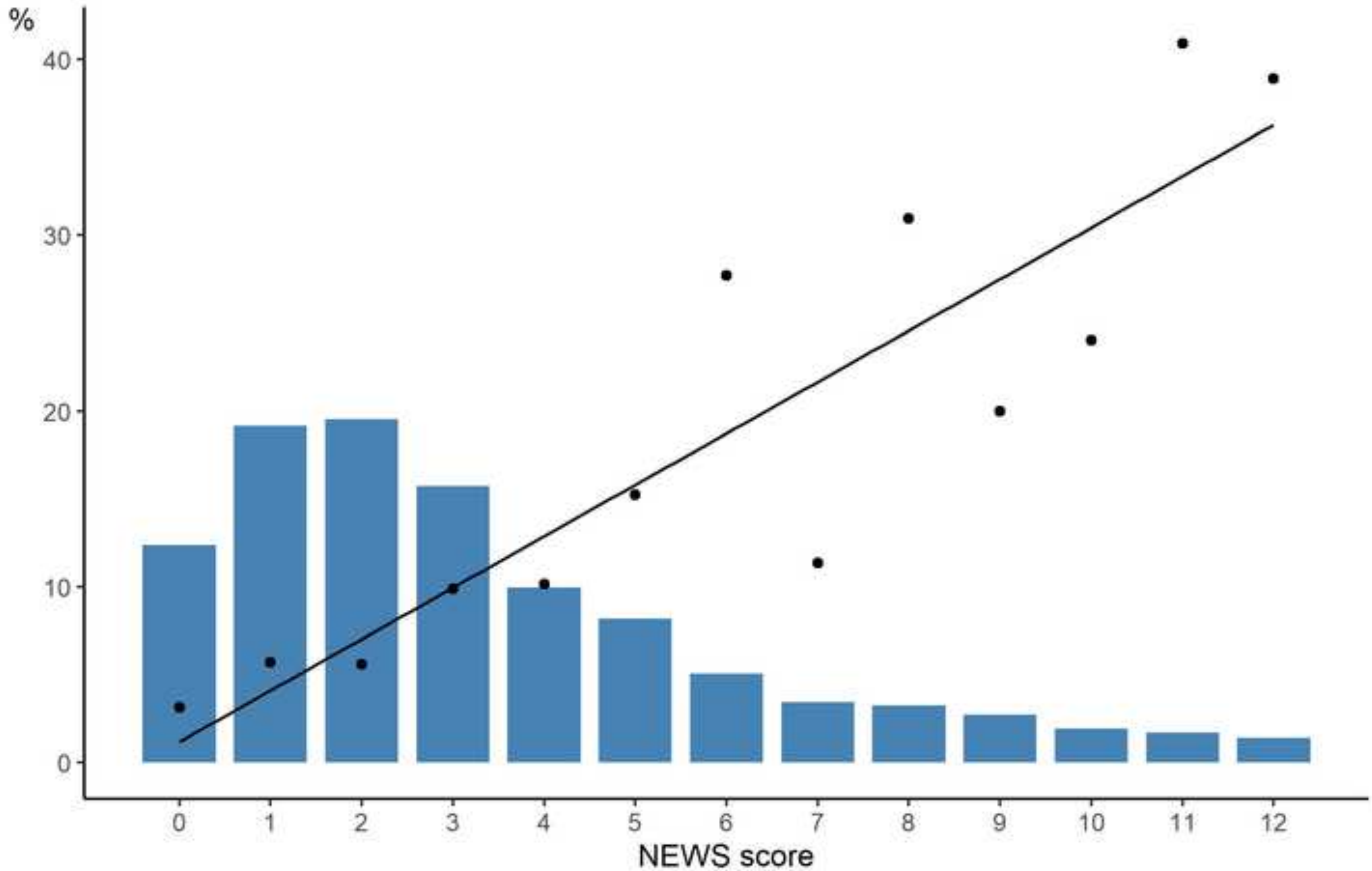


Figure 3

Click here to access/download;Figure;Figure 3. NEWS Mortality Plot.jpeg



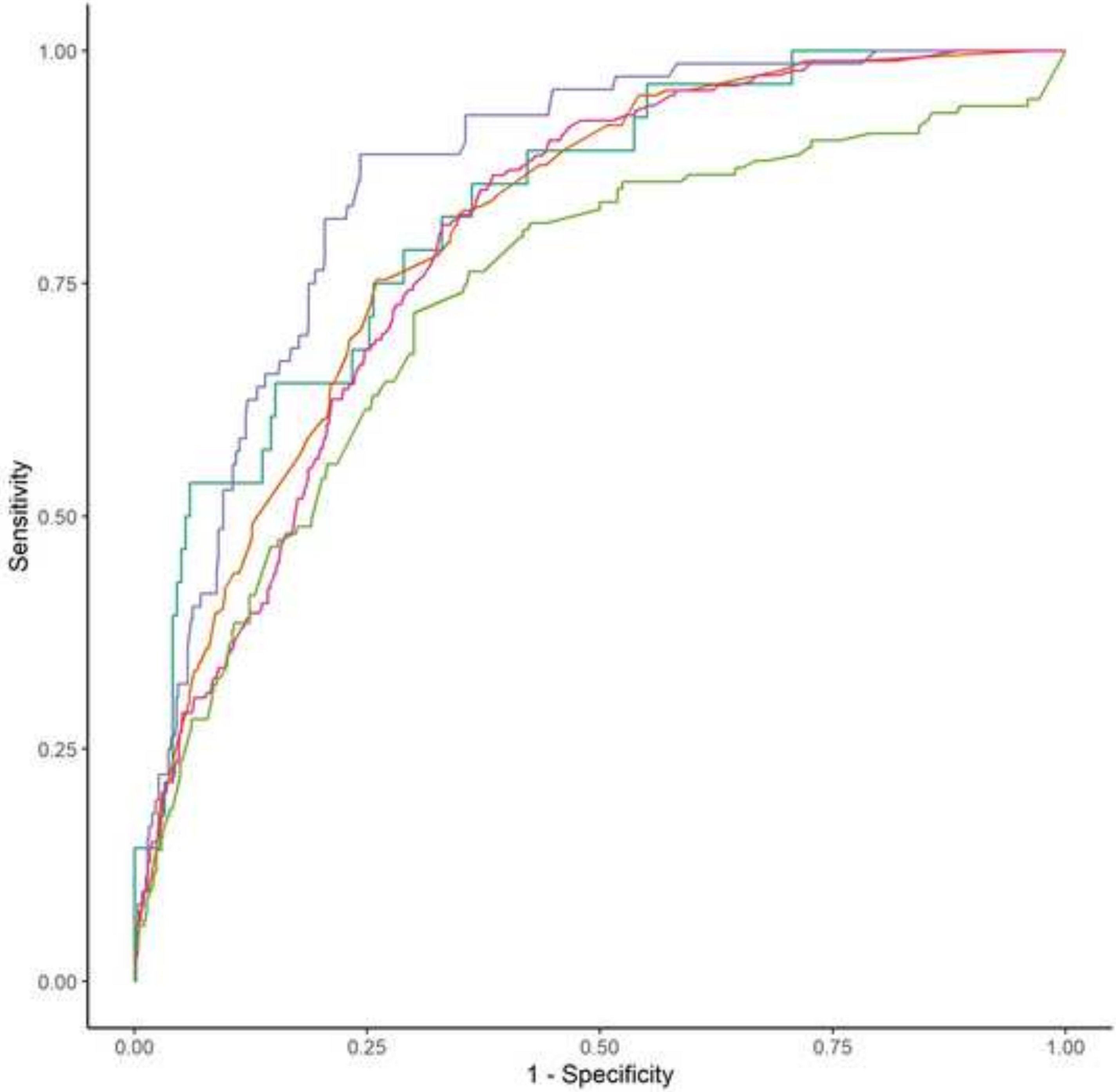
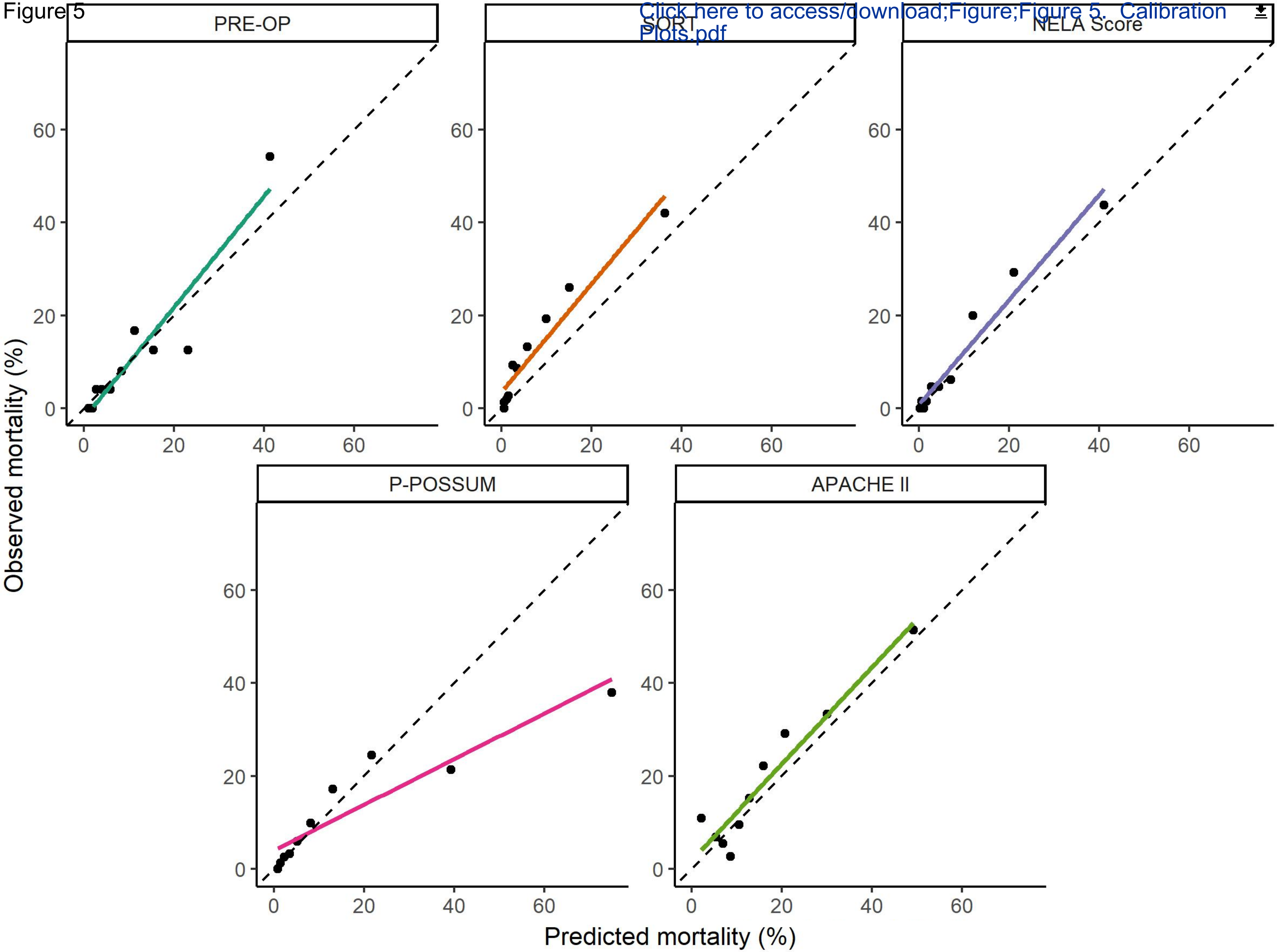

1.00
0.75
0.50
0.25
0.00
Sensitivity
0.00
0.25
0.50
0.75
1.00
1 - Specificity

Figure 5

Click here to access/download;Figure;Figure 5. Calibration Plots.pdf



Supplementary File

Click here to access/download;Supplementary File;Appendix 1.docx

Appendix 1 – Data Set Development and Calculation of Predictor Variables

Theatre Time Stamp

We identified the theatre episodes that matched each entry in the NELA dataset, to confirm the correct date/time into theatre. This time stamp is crucial to ensure only variables in the pre-operative period are used. Vital signs in the 24 hours and blood tests in the 72 hours prior to surgery were identified. These time periods were deemed clinically appropriate and reflective of a patient's acute physiology in the pre-operative period. Blood tests are taken less frequently than vital signs, so a longer time-period was allowed.

The National Early Warning Score - NEWS

	Score						
Vital Sign	**3**	**2**	**1**	**0**	**1**	**2**	**3**
Pulse (bpm)	≤40		41-50	51-90	91-110	111-130	≥131
Respiratory rate (bpm)	≤8		9-11	12-20		21-24	≥25
Temperature (˚C)	≤35		35.1-36.0	36.1-38.0	38.1-39	≥39.1	
Systolic BP (mmHg)	≤90	91-100	101-110	111-219			≥220
SaO_2 (%)	≤91	92-93	94-95	≥96			
Inspired O_2				Air		Any	
Conscious level (APVU scale)				Alert			Voice/Pain Unresponsive

The Laboratory Decision Tree Early Warning Score

Male	**2**	**1**	**0**	**1**	**2**	**3**
Hb	≤11.1	11.2-12.8	≥12.9			
WCC			≤9.3	9.4-16.6	16.7	
Na	≤132		133-140	≥141		
K		≤3.7	3.8-4.4	4.5-4.7	≥4.8	
Ur			≤9.4	9.5-13.7		13.8
Cre			≤114	115-179	180	
Alb	≤30	31-34	≥35			
Female	**2**	**1**	**0**	**1**	**2**	**3**
Hb		≤12	12.1-14.8	≥14.9		
WCC			≤12.6	12.7-14.8	≥14.9	
Na		≤134	135-140	≥141		
K		≤3.3	3.4-4.5	≥4.6		
Ur			≤8.4	8.5-13.8		≥13.9

Cre			≤91	92-157	≥158	
Alb	≤28	29-34	≥35			

The LDTEWS-NEWS Risk Score

The LDTEWS-NEWS risk score was calculated using the adapted methodology from its original paper[1]. To provide an overall score between 0 to 1, the NEWS and LDTEWS scores were divided by their maximum total score (20 and 15 respectively). Because blood tests are taken less frequently than vital signs, a weighting was created to progressively reduce the impact that bloods tests had in the algorithm, as the older they are the less relevant they are likely to be. This was calculated using the time elapsed from blood test results to the time the vital signs were taken, up to a 3-day period. The greater the time lapse the smaller the weighting LDTEWS is given. The weighting (ω), is calculated by:

$$\omega = 0.26(1 - \frac{TimeSinceLabs}{72}).$$

The score is calculated by:

$$LDTEWS\ NEWS\ Risk\ Index = \omega \times LDTEWS + (1 - \omega) \times NEWS$$

References

1. Redfern, O. C. *et al.* Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: Development and validation of a multivariable model. *Resuscitation* 133, 75–81 (2018).

Supplementary File Click here to access/download;Supplementary File;Appendix 2 -HFRS Scores.docx

DiagCode 3	Diagnosis_Name	HFRS
F00	Dementia in Alzheimer's disease	7.1
G81	Hemiplegia	4.4
G30	Alzheimer's disease	4
I69	Sequelae of cerebrovascular disease (secondary codes)	3.7
R29	Other symptoms and signs involving the nervous and musculoskeletal systems (R29∑6 Tendency to fall)	3.6
N39	Other disorders of urinary system (includes urinary tract infection and urinary incontinence)	3.2
F05	Delirium, not induced by alcohol and other psychoactive substances	3.2
W19	Unspecified fall	3.2
S00	Superficial injury of head	3.2
R31	Unspecified haematuria	3
B96	Other bacterial agents as the cause of diseases classified to other chapters (secondary code)	2.9
R41	Other symptoms and signs involving cognitive functions and awareness	2.7
R26	Abnormalities of gait and mobility	2.6
I67	Other cerebrovascular diseases	2.6
R56	Convulsions, not elsewhere classified	2.6
R40	Somnolence, stupor and coma	2.5
T83	Complications of genitourinary prosthetic devices, implants and grafts	2.4
S06	Intracranial injury	2.4
S42	Fracture of shoulder and upper arm	2.3
E87	Other disorders of fluid, electrolyte and acidbase balance	2.3
M25	Other joint disorders, not elsewhere classified	2.3
E86	Volume depletion	2.3
R54	Senility	2.2
Z50	Care involving use of rehabilitation procedures	2.1
F03	Unspecified dementia	2.1
W18	Other fall on same level	2.1
Z75	Problems related to medical facilities and other health care	2
F01	Vascular dementia	2
S80	Superficial injury of lower leg	2
L03	Cellulitis	2
H54	Blindness and low vision	1.9
E53	Deficiency of other B group vitamins	1.9
Z60	Problems related to social environment	1.8
G20	Parkinson's disease	1.8
R55	Syncope and collapse	1.8
S22	Fracture of rib(s), sternum and thoracic spine	1.8
K59	Other functional intestinal disorders	1.8

N17	Acute renal failure	1.8
L89	Decubitus ulcer	1.7
Z22	Carrier of infectious disease	1.7
B95	Streptococcus and staphylococcus as the cause of diseases classified to other chapters	1.7
L97	Ulcer of lower limb, not elsewhere classified	1.6
R44	Other symptoms and signs involving general sensations and perceptions	1.6
K26	Duodenal ulcer	1.6
I95	Hypotension	1.6
N19	Unspecified renal failure	1.6
A41	Other septicaemia	1.6
Z87	Personal history of other diseases and conditions	1.5
J96	Respiratory failure, not elsewhere classified	1.5
X59	Exposure to unspecified factor	1.5
M19	Other arthrosis	1.5
G40	Epilepsy	1.5
M81	Osteoporosis without pathological fracture	1.4
S72	Fracture of femur	1.4
S32	Fracture of lumbar spine and pelvis	1.4
E16	Other disorders of pancreatic internal secretion	1.4
R94	Abnormal results of function studies	1.4
N18	Chronic renal failure	1.4
R33	Retention of urine	1.3
R69	Unknown and unspecified causes of morbidity	1.3
N28	Other disorders of kidney and ureter, not elsewhere classified	1.3
R32	Unspecified urinary incontinence	1.2
G31	Other degenerative diseases of nervous system, not elsewhere classified	1.2
Y95	Nosocomial condition	1.2
S09	Other and unspecified injuries of head	1.2
R45	Symptoms and signs involving emotional state	1.2
G45	Transient cerebral ischaemic attacks and related syndromes	1.2
Z74	Problems related to care-provider dependency	1.1
M79	Other soft tissue disorders, not elsewhere classified	1.1
W06	Fall involving bed	1.1
S01	Open wound of head	1.1
A04	Other bacterial intestinal infections	1.1
A09	Diarrhoea and gastroenteritis of presumed infectious origin	1.1
J18	Pneumonia, organism unspecified	1.1
J69	Pneumonitis due to solids and liquids	1
R47	Speech disturbances, not elsewhere classified	1
E55	Vitamin D deficiency	1
Z93	Artificial opening status	1

R02	Gangrene, not elsewhere classified	1
R63	Symptoms and signs concerning food and fluid intake	0.9
H91	Other hearing loss	0.9
W10	Fall on and from stairs and steps	0.9
W01	Fall on same level from slipping, tripping and stumbling	0.9
E05	Thyrotoxicosis [hyperthyroidism]	0.9
M41	Scoliosis	0.9
R13	Dysphagia	0.8
Z99	Dependence on enabling machines and devices	0.8
U80	Agent resistant to penicillin and related antibiotics	0.8
M80	Osteoporosis with pathological fracture	0.8
K92	Other diseases of digestive system	0.8
I63	Cerebral Infarction	0.8
N20	Calculus of kidney and ureter	0.7
F10	Mental and behavioural disorders due to use of alcohol	0.7
Y84	Other medical procedures as the cause of abnormal reaction of the patient	0.7
R00	Abnormalities of heart beat	0.7
J22	Unspecified acute lower respiratory infection	0.7
Z73	Problems related to life-management difficulty	0.6
R79	Other abnormal findings of blood chemistry	0.6
Z91	Personal history of risk-factors, not elsewhere classified	0.5
S51	Open wound of forearm	0.5
F32	Depressive episode	0.5
M48	Spinal stenosis (secondary code only)	0.5
E83	Disorders of mineral metabolism	0.4
M15	Polyarthrosis	0.4
D64	Other anaemias	0.4
L08	Other local infections of skin and subcutaneous tissue	0.4
R11	Nausea and vomiting	0.3
K52	Other noninfective gastroenteritis and colitis	0.3
R50	Fever of unknown origin	0.1

Appendix 3 – Missing Data Analysis

While it is unusual for vital signs or blood tests not to be taken by clinical staff, they may not all be correctly recorded in electronic hospital records. The most frequently missing data items were vital signs (11.01%) and serum potassium (5.84%). The pattern of missingness visually appeared to be random over time and for a limited number of variables or patient episodes (see appendix 5 for missing data plot). We also compared patients with complete vital sign and biochemistry/haematology data to those with missing values. Patients with incomplete data had a significantly higher ASA grade, predicted and 30-day mortality, requiring more urgent or major surgery. This suggests that they may have been more unwell and needed more urgent surgery than their counterparts with complete data. It may be that in the rush to transfer the patient to theatre, some of the data items were therefore not taken or recorded. It is also probable that vital signs were missing for some patients because not all parts of PHU use the electronic vital signs system (VitalPAC^{TM}). The overall mortality rate for the cohort with complete data is comparable to those nationally (11.2%) even if the rate for the missing data cohort was higher (18.1%). Furthermore, no significant differences in age, Charlson Co-morbidity Index or HFRS score were observed. We therefore feel that the patients with complete data remain representative of the broad range of patients undergoing emergency bowel surgery.

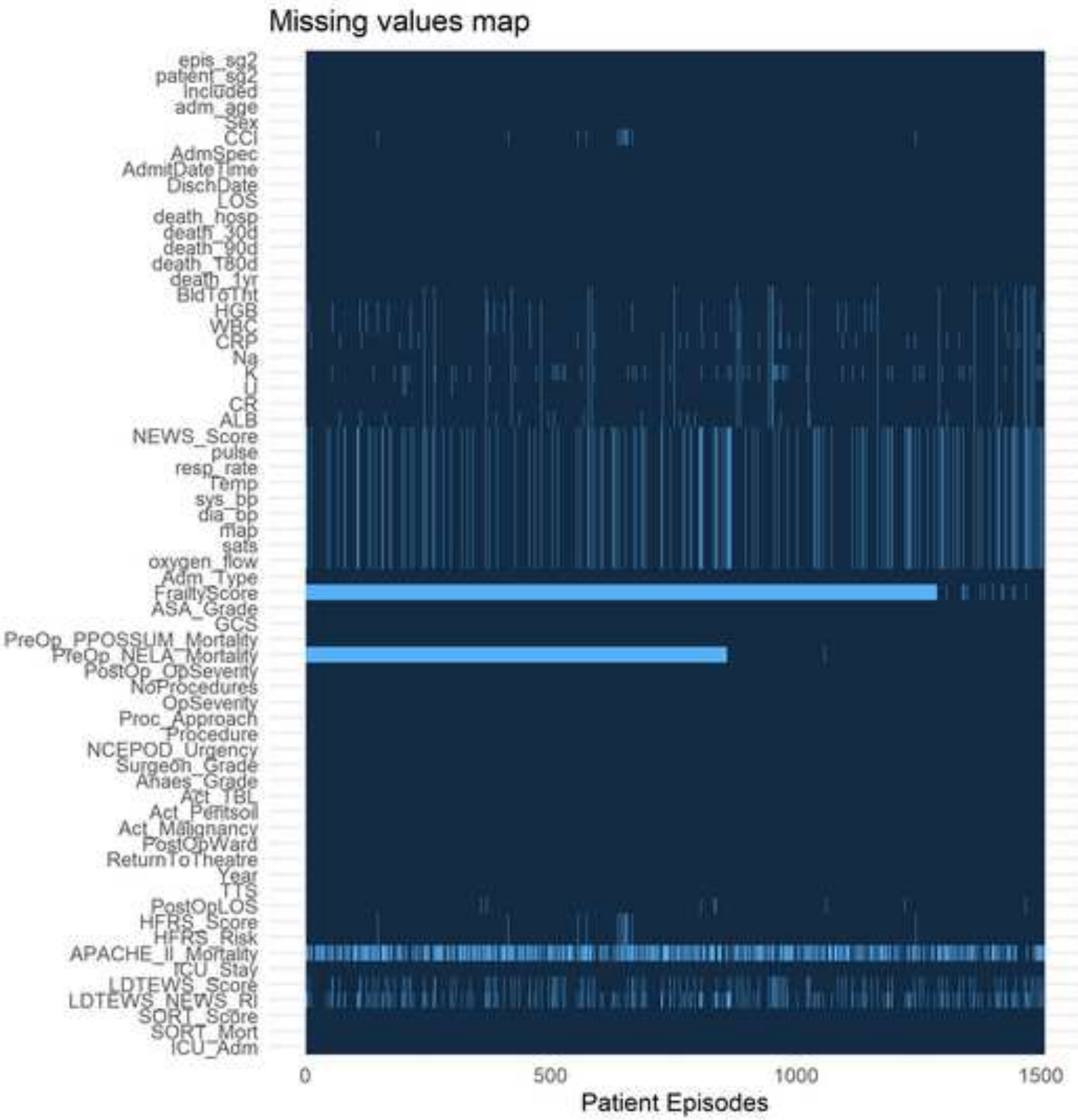
	Missing			Missing	
Variable	n	%	Variable	n	%
Clinical Frailty Scale	1302	86.34	LDTEWS Score	123	8.16
NELA Mortality	859	56.96	Potassium	88	5.84
APACHE II Mortality	783	51.92	CRP	58	3.85
ICU Stay	783	51.92	Haemoglobin	50	3.32
LDTEWS-NEWS Score	256	16.98	White Blood Cell Count	50	3.32
NEWS Score	166	11.01	Albumin	46	3.05
Pulse	166	11.01	Urea	32	2.12
Respiratory Rate	166	11.01	Sodium	28	1.86
Temperature	166	11.01	Creatinine	28	1.86
Systolic BP	166	11.01	Bloods to Theatre	24	1.59
Diasystolic BP	166	11.01	Post-op length of stay	11	0.73
Mean Arterial Pressure	166	11.01	CCI	13	0.86
Oxygen Saturations	166	11.01	HFRS Score	3	0.20
Oxygen Flow	166	11.01	HFRS Risk	3	0.20

Sensitivity analysis with demographic, operative and outcome data for patients with complete and missing vital sign, biochemistry and haematology data. Data presented as counts (%) or median (interquartile range). Comparative statistics we calculated using Chi Squared or Mann-Whitney U tests.

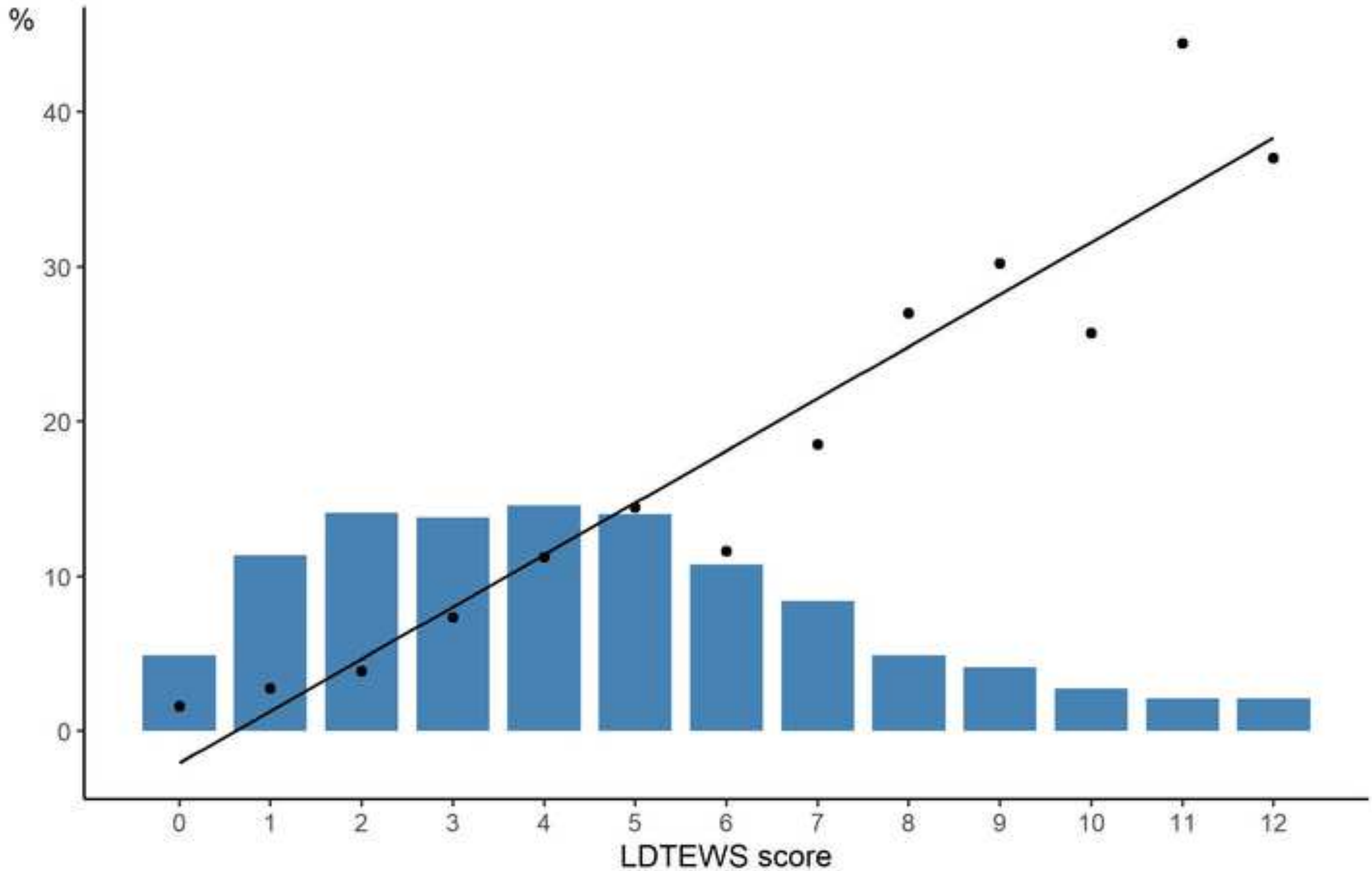
Complete Data	Yes n = 1,237	No n = 271	p
Patient Age	68.0 (53.0 to 78.0)	66.0 (51.0 to 77.0)	0.148
Female Sex	675 (54.6)	138 (50.9)	0.306
Admission Specialty			
General Surgery	995 (80.4)	225 (83.0)	0.370
Other	242 (19.6)	46 (17.0)	
Admission Type			
Emergency	1158 (93.6)	258 (95.2)	0.395
Elective	79 (6.4)	13 (4.8)	
Charslon Co-Morbidity Index	3.0 (0.0 to 9.0)	3.0 (0.0 to 9.0)	0.945
Clinical Frailty Score	0.0 (0.0 to 0.0)	0.0 (0.0 to 0.0)	0.075
HFRS Score	3.9 (1.1 to 9.3)	4.0 (1.4 to 10.7)	0.253
ASA Grade	2.0 (2.0 to 3.0)	3.0 (2.0 to 4.0)	**0.001**
APACHE II Mortality	10.7 (6.7 to 19.4)	12.7 (7.5 to 25.4)	**0.012**
P‐ POSSUM Mortality	5.7 (2.3 to 18.3)	13.1 (3.3 to 44.2)	**<0.001**
NELA Score Mortality	3.3 (0.9 to 10.2)	4.1 (1.2 to 19.7)	0.072
SORT Mortality	2.7 (1.3 to 8.1)	4.6 (1.3 to 14.0)	**<0.001**
Number of Procedures			
Primary	1076 (87.0)	232 (85.6)	**0.026**
Second	152 (12.3)	32 (11.8)	
Multiple	9 (0.7)	7 (2.6)	
BUPA Operative Severity			
Major	925 (74.8)	179 (66.1)	**0.004**
Major+	312 (25.2)	92 (33.9)	
Urgency of Surgery			
<2 hours	102 (8.2)	72 (26.6)	**<0.001**
2-6 hours	514 (41.6)	117 (43.2)	
6-18 hours	421 (34.0)	60 (22.1)	
>18 hours	200 (16.2)	22 (8.1)	
Blood Loss			
<100ml	769 (62.2)	153 (56.5)	0.141
101-500ml	389 (31.4)	93 (34.3)	
501-999ml	49 (4.0)	12 (4.4)	
>1000ml	29 (2.3)	13 (4.8)	
Degree of Peritoneal Soiling			
None	423 (34.2)	84 (31.0)	**<0.001**

Serous Fluid	393 (31.8)	70 (25.8)	
Localised Pus	91 (7.4)	9 (3.3)	
Gross Contamination	330 (26.7)	108 (39.9)	
Malignancy Present			
None	995 (80.4)	236 (87.1)	**0.005**
Primary Disease	115 (9.3)	22 (8.1)	
Nodal Metastases	33 (2.7)	8 (3.0)	
Distant Metastases	94 (7.6)	5 (1.8)	
Return To Theatre			
None	1116 (90.2)	232 (85.6)	**0.009**
Planned	7 (0.6)	6 (2.2)	
Unplanned	114 (9.2)	33 (12.2)	
Time To Theatre	2.0 (0.9 to 4.4)	0.5 (0.2 to 2.5)	**<0.001**
Length Of Stay	14.4 (8.6 to 27.5)	13.0 (6.3 to 27.1)	**0.009**
Post-Op Length of Stay	9.9 (5.4 to 18.6)	11.2 (5.2 to 22.2)	0.489
ICU Length of Stay	2.8 (1.4 to 5.4)	4.5 (1.8 to 8.1)	**<0.001**
30-day Mortality	138 (11.2)	49 (18.1)	**0.002**
1-year Mortality	249 (20.1)	67 (24.7)	0.109

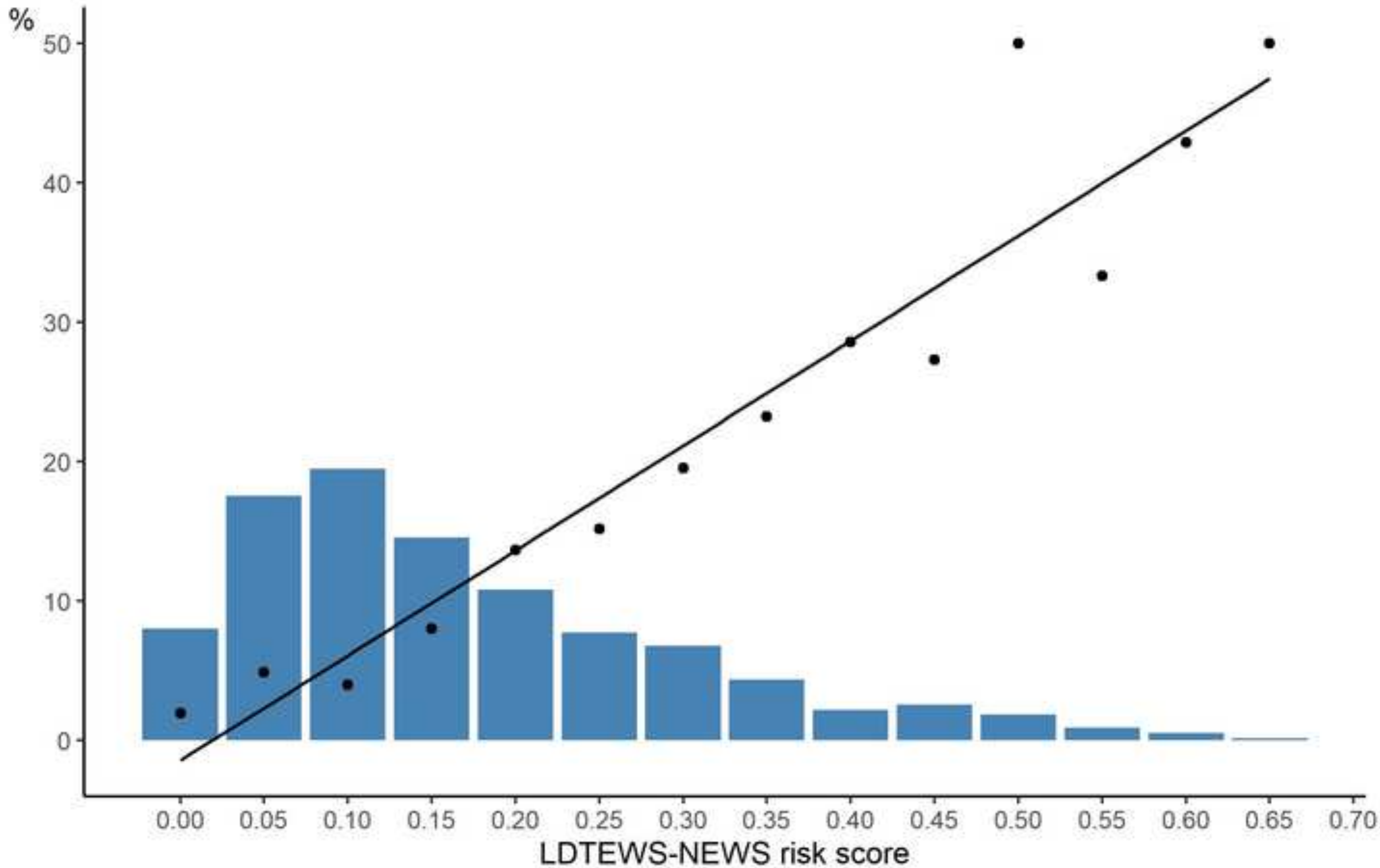
Click here to access/download;Supplementary File;Appendix 4.jpeg



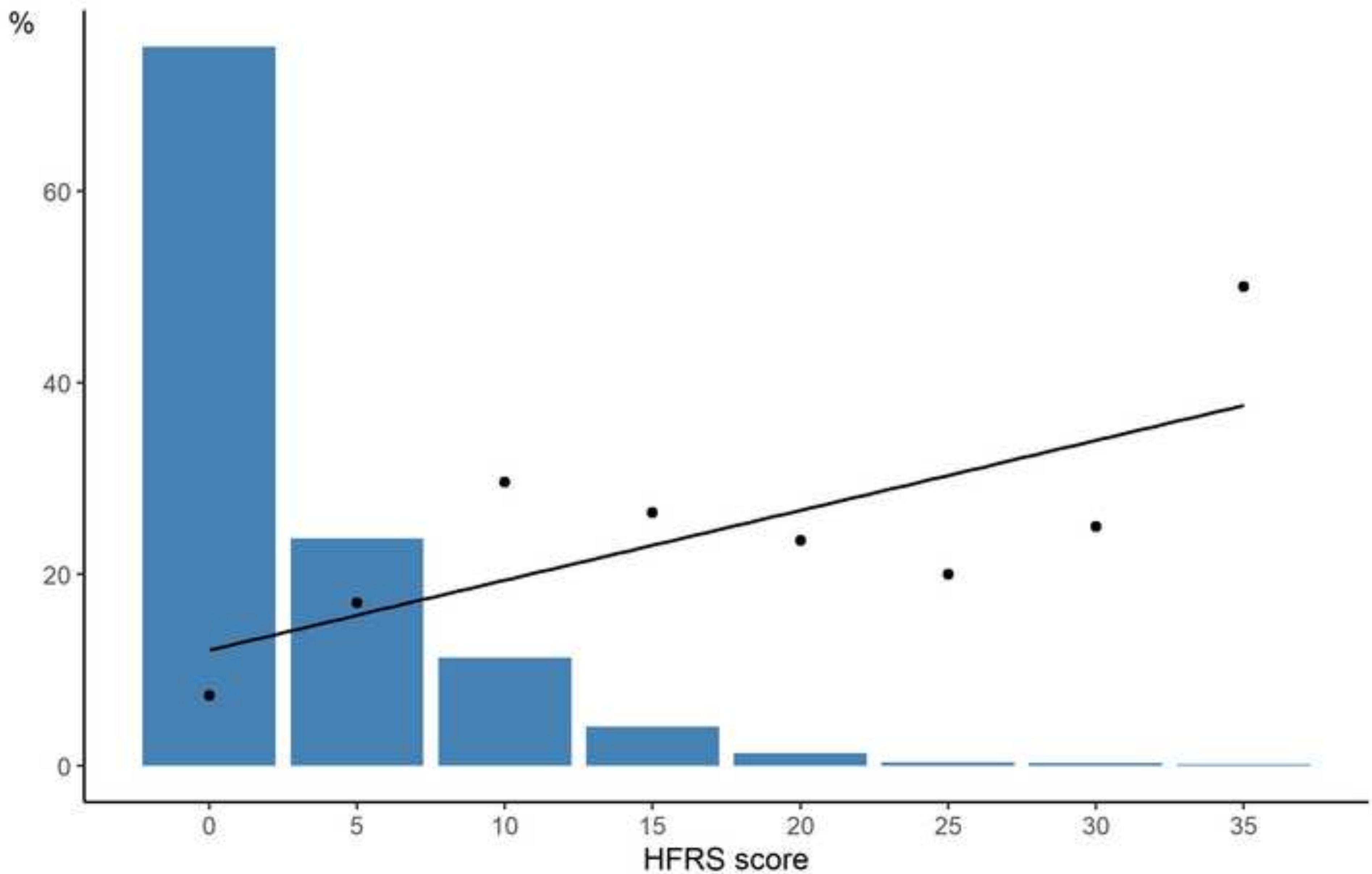
Click here to access/download;Supplementary File;Appendix 5.1 LDTEWS Mortality Plot.jpeg



Click here to access/download;Supplementary File;Appendix 5.2 LDTEWS-NEWS Mortality Plot.jpeg



Supplementary File
Click here to access/download;Supplementary File;Appendix 5.3 HFRS Mortality Plot.jpeg



Click here to access/download;Supplementary File;Appendix 6.docx

PRE-OP model coefficients

	β Coefficient	Standard Error	p value
Intercept	6.5253060	0.6793031	<0.001
Age	-0.0377797	0.0086576	<0.001
NEWS Score	-0.1460451	0.0351015	<0.001
LDTEWS Score	-0.1695041	0.0399985	<0.001
HFRS Score	-0.0541965	0.0183294	0.003

Supplementary File
Click here to access/download;Supplementary File;Appendix 7. Mortality over time.jpg