The Second Patient Report of the National Emergency Laparotomy Audit (NELA): Development of the Risk Adjustment Model

July 2016

This Report was produced by:

Ms Natalie Eugene, Statistician, Clinical Effectiveness Unit, Royal College of Surgeons of England
Dr Angela Kuryba, Statistician, Clinical Effectiveness Unit, Royal College of Surgeons of England
Dr Kate Walker, Statistician, Clinical Effectiveness Unit, Royal College of Surgeons of England

Development of the NELA risk adjustment model

Introduction

The National Emergency Laparotomy Audit (NELA) examines the delivery of emergency bowel surgery in hospitals within England and Wales. It is commissioned by the Healthcare Quality Improvement Partnership (HQIP) and forms part of the National Clinical Audit and Patient Outcomes Programme (NCAPOP).

This report describes the development of the risk model for the Second NELA Patient Report, to be published in 2016.

Patient Cohort

The development of the risk model was based on patients submitted to NELA who had their operation between 1st December 2013 and 30th November 2015. The records were linked to the Office for National Statistics (ONS) date of death based on their NHS number.

Records were included in the analysis if patients:

- Underwent surgery in an NHS hospital within England and Wales.
- Were aged between 18 and 105 years.
- Had a known mortality status at 30 days.

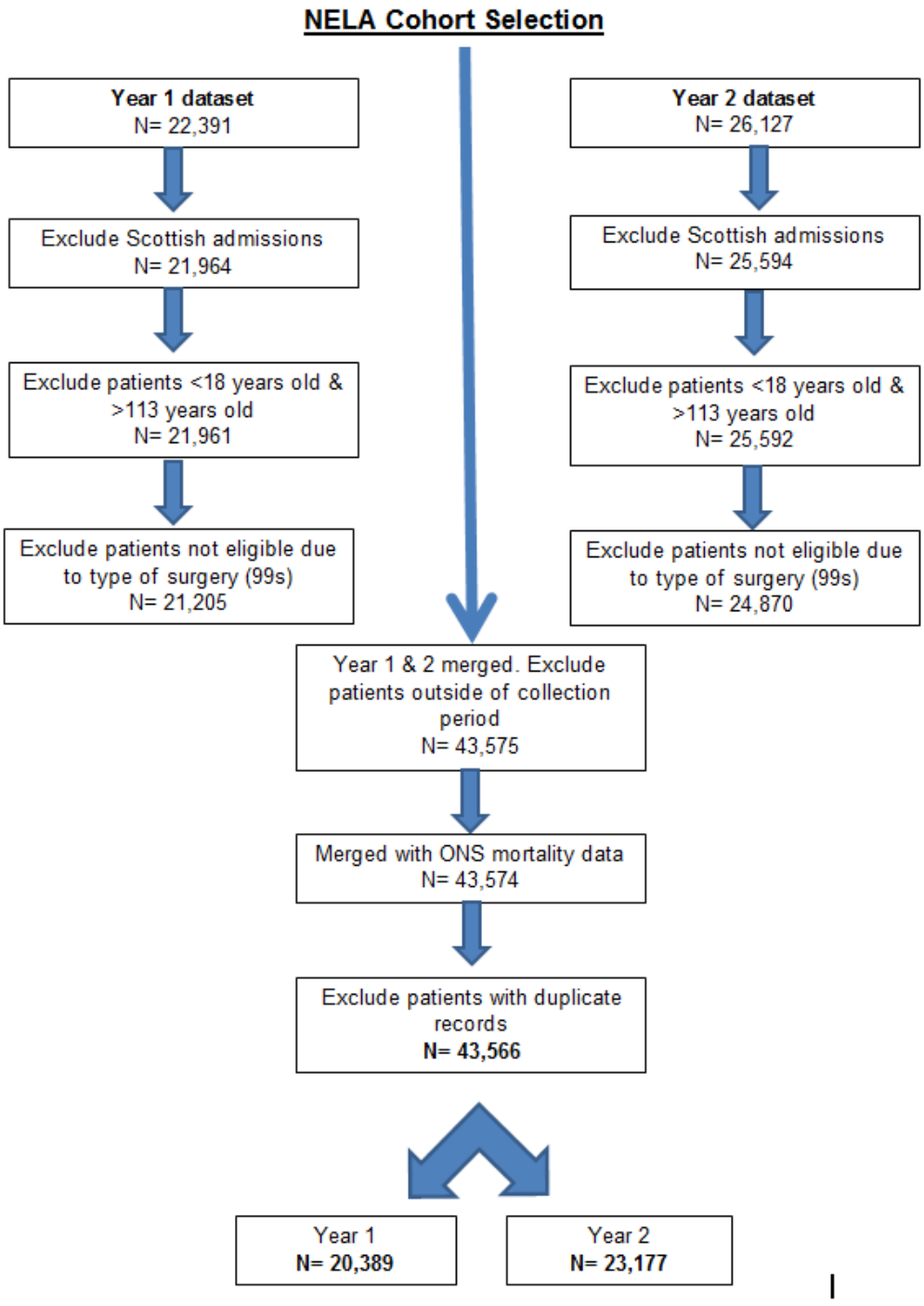
Exclusion criteria:

- Patients recorded as having a surgical procedure that was not eligible for inclusion in NELA.
- Potentially duplicate records.

879 patients were entered into the NELA dataset more than once (1,778 records). None of these multiple records had the same admission dates. However, records that had the same operation dates were assumed to be duplicate entries of the same emergency laparotomy and one of their entries was removed (10 patients). This left 1,768 records with multiple emergency laparotomies entered. Most of these were patients with two emergency laparotomies (854 patients, 1,708 records). There were also a rare few patients (60 records, 20 patients) that had 3 eligible procedures. As this process occurred after the initial cleaning of the dataset, some of the patients had already been dropped.

The cohort selection process is summarised in Figure 1 overleaf. It appears that a large number of observations have been dropped during the cleaning process, but this is not the case. Two separate extracts of patients were pulled from the online NELA resource. These two datasets included duplicate patients, as they had been pulled from a broad collection period. These duplicates were obviously dropped once the two datasets were merged.

Figure 1
Flowchart summarising the creation of the patient cohort



30-day mortality

Each patient's procedure was recorded as alive if they are alive at 30 days from that procedure and dead if there were not. The exceptions were patients with multiple procedures who died within 30 days of two procedures, in which case they were recorded as alive after their first procedure and dead after their second procedure.

There were 607 patients who were lost to follow-up or could not be traced by ONS. On review there were 328 patients in Year 1 distributed across 128 hospitals and 279 patients distributed across 123 hospitals in Year 2 (both years losing an average of two patients per hospital). This was not considered to be a systemic pattern of missingness. As the ONS mortality status was not recorded for these observations, it was necessary to use their inpatient mortality status from the NELA dataset instead.

Casemix variables

There were 22 variables describing different aspects of the patient's admission that were candidates for inclusion in the risk model. These included: patient demographic factors, preoperative factors including physiological measurements, and perioperative factors (see Table 1).

The distribution of values for a few variables posed a problem. For example, 92% of the values for the Glasgow Coma Score (GCS) had a score of 15, and consequently, the GCS variable included in the model was based on categorised scores: group 1 (3-8), group 2 (9-12) and group 3 (13-15). In addition, the variables for Urea and Creatinine were highly skewed and were therefore log-transformed in the model.

Some variables were categorical, and therefore restricted to limited number of values. There were fewer constraints placed on the values of the continuous variables, and the extremes of the distributions were unusual in some cases. All continuous physiological variables with extreme values at either or both ends of their distribution were winsorised at the 99th and/or 1st percentile (see Table 1).

It was deemed to be more clinically significant to use postoperative values of perioperative measures, such as intraoperative blood loss, peritoneal soiling, operative severity and malignancy, due to the accuracy of these measures. However it was necessary to use preoperative measures when these values were missing postoperatively.

The year of the patient's procedure (Year 1 or 2) was included as a variable in the risk adjustment to allow for organisational changes within hospitals. Although a hospital's caseload is likely to be relatively static from year to year, one hospital may undertake substantially more operations in Year 1 than in Year 2, and another vice versa. Including the year of treatment in the risk adjustment means that each patient's mortality is compared to the current background population mortality.

Table 1
Description of the candidate variable for the risk adjustment model

Variable	Type	Range of continuous (winsorised)	Transformations	Equation for continuous variables
Demographic				
Gender	Categorical			
Age	Continuous	18-105 years		Linear + Quadratic
Year of NELA audit	Ordinal			
Preoperative				
ASA score	Ordinal		ASA 1 and 2 combined	
Urgency of surgery	Ordinal			
ECG	Categorical			
Number of operations within admission	Ordinal			
Cardiac signs	Categorical			
Respiratory history	Categorical		limiting and at rest combined	
Physiological				
Creatinine	Continuous	0-1,200 umol/l (3.3-6.0)	log-transformed	Linear + Quadratic
Sodium	Continuous	100-180 mmol/l (124-148)		Fractional Poly (^3 + ^3*log)
Potassium	Continuous	1-10 mmol/l (2.8-5.9)		Linear + Quadratic
Urea	Continuous	1-273 mmol/l (0-3.7)	log-transformed	Linear + Quadratic
WBC	Continuous	0-200 $x10^9$/l (1-42.7)		Linear + Quadratic
Haemoglobin	Continuous	40-250 g/l (40-183)		N/A
Pulse	Continuous	10-200 bpm (55-145)		Linear + Quadratic
Systolic blood pressure	Continuous	10-240 mmHg (70-190)		Linear + Quadratic
Glasgow score	Continuous		Grouped (3-8/ 9-12/ 13-15)	
Perioperative				
Peritoneal soiling	Ordinal			
Intra-operative blood loss	Ordinal			
Malignancy	Ordinal			
Operative severity	Ordinal			

* ASA, American Society of Anaesthesiologists; ECG, Electrocardiogram; WBC, White Blood Count

The relationship between a particular continuous variable and 30-day mortality was not always linear. We therefore identified the best fitting fractional polynomials to select the appropriate shape for any non-linear relationships. For most variables, a linear plus quadratic term was sufficient to allow for any curvature in their relationship with mortality. The exception was sodium for which the best-fitting fractional polynomial, ($sodium^3$ + $sodium^3$ x log sodium) was a more clinically plausible fit. For other continuous variables, the best fitting fractional polynomials produced very similar relationships to the linear plus quadratic curves, but had features which did not seem to be clinically plausible and were therefore not selected. Haemoglobin was removed from the model as there was no statistical evidence of an association with mortality and the shape of the relationship was not clinically plausible. See Appendix 1 for the shape of the non-linear relationships.

A number of clinically plausible interactions between variables were examined. This process involved testing the strength of the interaction in 100 bootstrap samples. The criteria for the selection of interactions between variables was P<0.01 in at least 90% of bootstrap samples. This ensures that the interactions selected are the ones likely to be selected if another sample of data were used. This identified the following interactions, which were included in the model.

- ASA x respiratory signs
- ASA x age

Other interactions considered by not selected were:

- ASA x cardiac signs
- ASA x Glasgow coma scale
- ASA x presence of malignancy
- Systolic BP x age

Statistical analysis

Logistic regression was used to assess the relationship between 30-day mortality and the individual patient and treatment characteristics. All variables were included in the model initially, apart from interactions between variables. The best-fitting polynomial for each continuous variable was identified using maximum-likelihood techniques, whilst keeping all other variables in the model.

The performance of the model was assessed in terms of its calibration and discrimination. Calibration describes the level of agreement between the predicted and observed risks, comparing the predicted and observed mortality in deciles of predicted risk. If the Hosmer-Lemeshow test showed evidence of poor calibration *and* the size of the differences between observed and predicted risk were clinically meaningful was lack of calibration deemed to be an issue. Discrimination indicates the ability of a model to distinguish patients with a lower and higher risk of postoperative mortality. We evaluated this by using the C-statistic (equivalent to the area under the ROC curve (ROC AUC)).

Crude (unadjusted) and risk-adjusted rates of mortality within 30 days at each of the 193 hospitals, and nationally, were calculated. A number of variables contained missing data, but the records could not be dropped if an accurate risk-adjusted rate was to be derived. Consequently, we imputed values for the missing data with the Multiple Imputation using Chained Equations (MICE) technique. The imputation model included all risk factors, interactions and the indicator variable of 30-day mortality. The adjusted rate was derived with ten imputation sets.

The results are presented in a funnel plot. The control limits in the funnel plot were derived using binomial limits, and were defined to correspond to two and three standard deviations above and below the overall national average, respectively. The control limits can be described alternatively as 95% limits (two standard deviations) and 99.8% limits (three standard deviations). These limits indicate whether the difference between the mortality rate at a hospital and the national average is greater than would be expected from random fluctuations, and by how much.

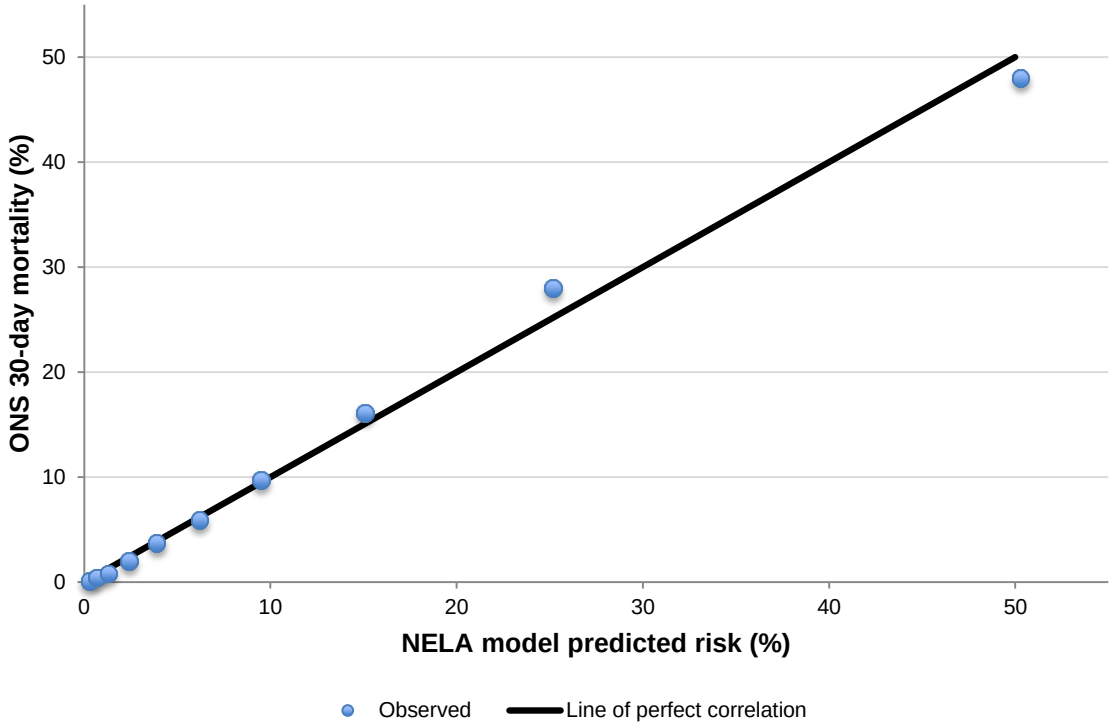
Results

The number of patients included in the analysis was 20,389 for the first year of the Audit, and 23,177 for the second. The overall mortality across the two years of the Audit were similar (11.7% in Year 1 v 11.0% in Year 2). There was also little difference in the risk profiles of the two years, with the adjusted mortality rates for Year 1 and 2 being 11.5% and 11.2% respectively, p-value = 0.2.

Performance of the risk adjustment model

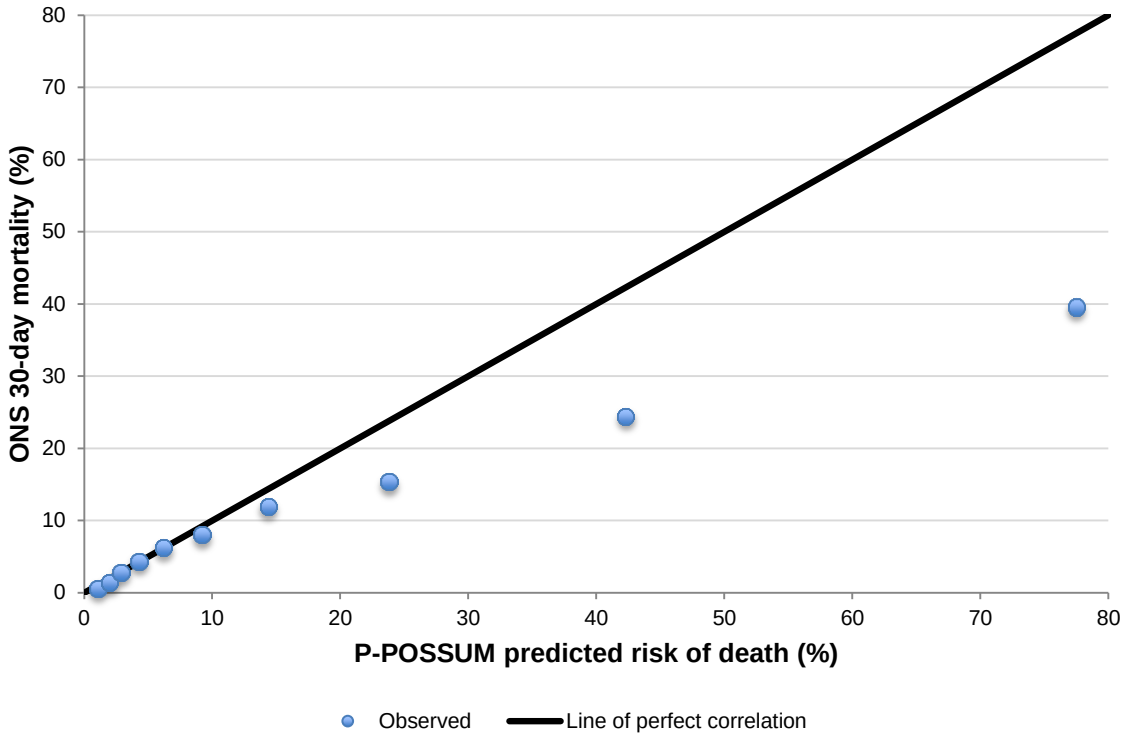
All variables contributed to the performance of the risk model, with the exception of haemoglobin which was dropped. The model proved to have excellent discrimination, with a C-index of 0.863 (95% CI: 0.858, 0.868) – the C-index ranges from 0.5 (no better than toss of coin) to 1 (perfect prediction). The model also demonstrated good calibration (See Figure 2). The calibration plot also highlights the considerable heterogeneity in risk faced by the patients undergoing emergency laparotomy. The average observed 30-day mortality in the two groups with the highest risk are around 28% and 48%.

Figure 2
Calibration plot comparing the observed 30-day mortality against that predicted by model in deciles of predicted risk. Points should lie on the line for perfect calibration



The discrimination of P-POSSUM is reasonably high, with a C-index of 0.803 (95% CI: 0.797, 0.809). However, the calibration is shown to be quite poor for patients with more than around 15% predicted risk (see Figure 3), at which point it overestimates their risk.

Figure 3
P-POSSUM calibration plot comparing the observed ONS 30-day mortality against that predicted by model in deciles of predicted risk



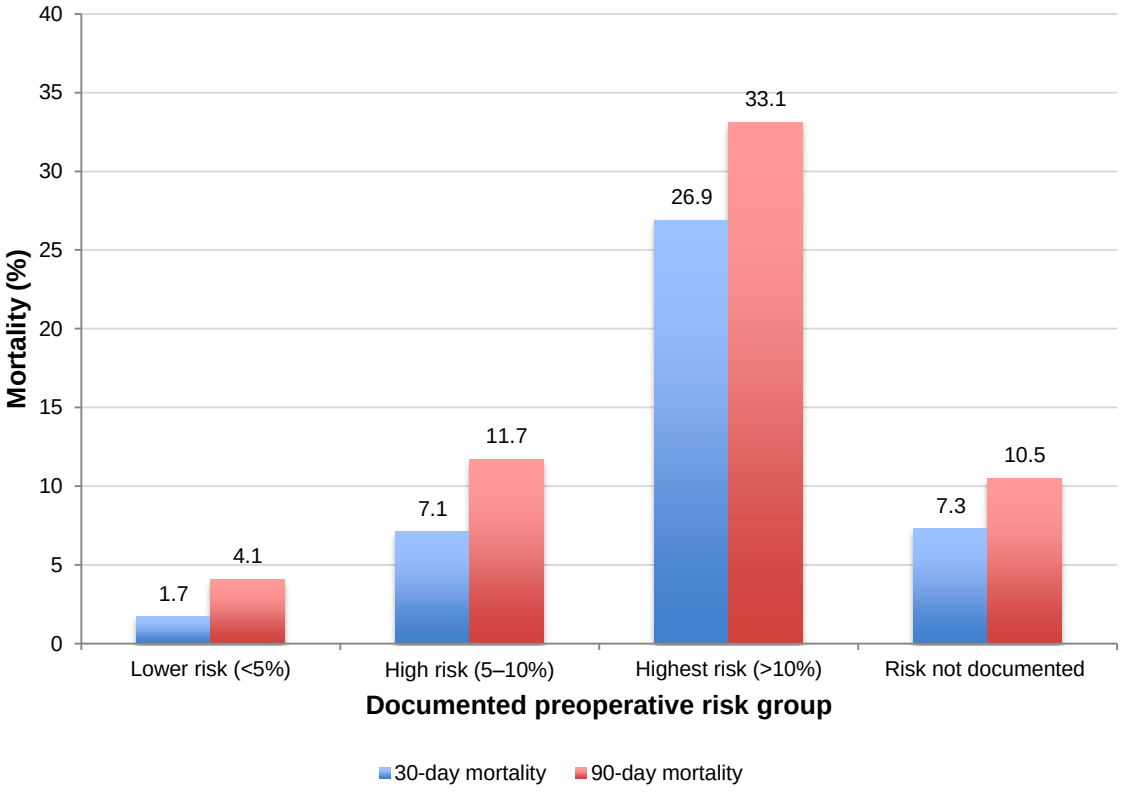
The unadjusted and risk adjusted mortality rates stratified by ASA and level of urgency are detailed in Table 2. The unadjusted rates show the influence of the other patient characteristics within the ASA and level of urgency variables. Not surprisingly, a high ASA grade or level or urgency are associated with values indicating worse health. The adjusted values remove the effect of these other factors to reveal the degree of severity that is not captured by the measured patient characteristics.

Table 2
Stratified unadjusted and adjusted 30-day mortality rates

	30-day unadjusted mortality rate- % (N) (Year 1)	30-day unadjusted mortality rate- % (N) (Year 2)	30-day adjusted mortality rate- % (N) (Year 1)	30-day adjusted mortality rate- % (N) (Year 2)
ASA Status				
None & mild systemic disease	2.6 (8,979)	2.4 (10,393)	6.5 (8,979)	6.1 (10,393)
Severe systemic disease	10.0 (7,184)	9.6 (8,170)	10.2 (7,184)	10.0 (8,170)
Severe, life-threatening	31.0 (3,785)	30.3 (4,149)	14.5 (3,785)	13.4 (4,149)
Moribund patient	56.2 (443)	60.4 (465)	15.8 (443)	15.0 (465)
Level of Urgency				
Immediate (<2hrs)	25.3 (3,047)	26.8 (2,958)	13.3 (3,047)	12.9 (2,958)
Urgent (2-6hrs)	11.9 (6,804)	11.4 (8,993)	11.3 (6,804)	10.5 (8,993)
Urgent (6-18hrs)	7.3 (5,090)	6.6 (7,281)	10.4 (5,090)	10.0 (7,281)
Expedited (>18hrs)	6.5 (2,827)	6.7 (3,904)	11.0 (2,827)	11.4 (3,904)
Missing	9.3 (2,622)	9.8 (41)	11.2 (2,622)	14.8 (41)

The model produces an objective assessment of 30-day mortality. The NELA dataset also includes a three-category rating by the surgeon. The consistency between the observed mortality and that predicted by the model is shown in Figure 2 and Table 3. Overall, the ratings allocated by surgeons are consistent with the actual outcomes. The mortality in those with no documented risk is relatively high.

Figure 4
Observed and expected 30-day mortality stratified by preoperative risk assessment



The use of linear and quadratic terms as well as fractional polynomials means that the effects of the individual risk factors are not easily interpretable from the standard statistical output. Consequently, we estimated the effect associated with specific values of the individual factors, and summarised this in Table 3. For the continuous variables, we selected a reference value for which the odds ratio was 1 (by definition). The odds ratio (OR) associated with the other values is derived compared to this reference value. For age and respiratory history (interaction terms) there are different ORs for each category of ASA (1/2, 3, 4, 5) because of the interaction in the model. The ORs for ASA category (the first row of Table 3) are for patients aged 70 (the baseline age) with no dyspnoea (the baseline category of respiratory history). See Appendix 2 for a graphical summary of the relationship between ASA category and 30-day mortality according to different ages and different categories of respiratory history.

Table 3
Model estimates

	ASA 1 or 2		ASA 3		ASA 4		ASA 5	
Variable	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
ASA (no resp history and age 70)	1		2.52	2.12 to 3.00	6.28	5.25 to 7.51	12.45	9.21 to 16.83
Age 50	0.48	.42 to .54	0.59	.52 to .67	0.70	.62 to .80	0.77	.68 to .88
Age 60	0.80	.78 to .82	0.86	.83 to .88	0.89	.87 to .92	0.91	.88 to .93
Age 70 (ref)	1		1		1		1	
Age 80	2.73	2.39 to 3.13	1.95	1.71 to 2.23	1.74	1.52 to 1.99	1.81	1.58 to 2.07
Age 90	5.59	4.23 to 7.38	3.08	2.33 to 4.07	2.68	2.03 to 3.54	3.09	2.34 to 4.08
No resp history (ref)	1		1		1		1	
Mild dyspnoea	1.97	1.53 to 2.53	1.37	1.2 to 1.56	1.22	1.06 to 1.39	1.03	.70 to 1.52
limiting & at rest	3.73	2.51 to 5.53	1.90	1.63 to 2.20	1.48	1.31 to 1.68	1.33	.95 to 1.86
Across all ASA categories								
NELA year1	1			Sodium 125mmol/l	1.53	1.37 to 1.71		
NELA year2	0.96	.90 to 1.03		Sodium 130 mmol/l	1.38	1.26 to 1.51		
Male	1			Sodium 140 mmol/l	1			
Female	1.04	.97 to 1.12		Sodium 150 mmol/l	2.99	2.20 to 4.07		
Blood loss <100ml	1			Systolic BP 80	1.75	1.57 to 1.94		
Blood loss (101-500ml)	1.02	.94 to 1.1		Systolic BP 100mmHg	1.26	1.22 to 1.32		
Blood loss (501-999ml)	1.04	.89 to 1.2		Systolic BP 120 mmHg	1			
Blood loss (>1,000ml)	0.85	.70 to 1.04		Systolic BP 150 mmHg	0.83	.79 to .87		
No cardiac failure	1			Systolic BP 180 mmHg	0.83	.72 to .96		
Antihypertensive therapy	1.07	.98 to 1.16		WBC 5x10^9/l	1.06	1.01 to 1.12		
Borderline cardiomegaly	1.33	1.17 to 1.51		WBC 10 x10^9/l	1			
Cardiomegaly	1.22	.99 to 1.52		WBC 20 x10^9/l	1.02	.97 to 1.08		
Glasgow score (13-15)	1			WBC 30 x10^9/l	1.26	1.16 to 1.38		
Glasgow score (9-12)	1.85	1.44 to 2.38		WBC 40 x10^9/l	1.89	1.59 to 2.24		
Glasgow score (3-8)	2.44	2.06 to 2.90		Urea 2 mmol/l	0.58	.47 to .70		

Malignancy (none)	1		Urea 5 mmol/l	0.80	.76 to .85
Malignancy (primary only)	1.10	.98 to 1.24	Urea 10 mmol/l	1	
Malignancy (nodal metastases)	1.54	1.30 to 1.82	Urea 20 mmol/l	1.21	1.13 to 1.29
Malignancy (distant metastases)	3.16	2.83 to 3.54	Urea 30 mmol/l	1.33	1.17 to 1.52
Number procedures (1)	1		Creatinine 40umol/l	1.16	1.03 to 1.31
Number procedures (2)	0.78	.70 to .87	Creatinine 70umol/l	1.02	.95 to 1.09
Number procedures (>2)	0.75	.56 to .99	Creatinine 100umol/l	1	
Operative severity (Major)	1		Creatinine 150umol/l	1.04	1.01 to 1.08
Operative (Major+)	1.17	1.09 to 1.26	Potassium 3mmol/l	1.36	1.23 to 1.51
ECG (no abnormalities)	1		Potassium 3.5 mmol/l	1.11	1.06 to 1.15
ECG (AF rate 60-90)	1.22	1.06 to 1.41	Potassium 4 mmol/l	1	
ECG (AF rate >90 or abnormal)	1.21	1.11 to 1.31	Potassium 4.5 mmol/l	1.01	.98 to 1.04
Peritoneal soiling (none)	1		Potassium 5 mmol/l	1.14	1.07 to 1.21
Peritoneal soiling (serous fluid)	1.20	1.09 to 1.31	Pulse 60bpm	0.64	.56 to .72
Peritoneal soiling (localised pus)	1.01	.87 to 1.16	Pulse 70bpm	0.76	.71 to .81
Peritoneal soiling (free bowel content)	1.41	1.28 to 1.55	Pulse 90bpm	1	
Surgical urgency (>18hrs)	1		Pulse 120bpm	1.30	1.23 to 1.38
Surgical urgency (6-18hrs)	0.91	.80 to 1.04	Pulse 140bpm	1.40	1.22 to 1.61
Surgical urgency (2-6hrs)	1.04	.91 to 1.18			
Surgical urgency (<2hrs)	1.58	1.37 to 1.82			

*WBC, white Blood Count; ASA, American Society of Anaesthesiologists ; ECG, Electrocardiogram

Hospital-level mortality rates

There were five hospitals with crude mortality rates above the upper 99.8% control limit (Figure 5). No hospital had an unadjusted rate below the lower 99.8% limit.

After adjustment, the 30-day mortality rates for all hospitals were within the 99.8% limits (Figure 6), however there were 12 hospitals above and seven hospitals below the 95% limits, which is more than would be expected by chance alone.

Figure 5
Funnel plot of unadjusted 30-day postoperative mortality rates

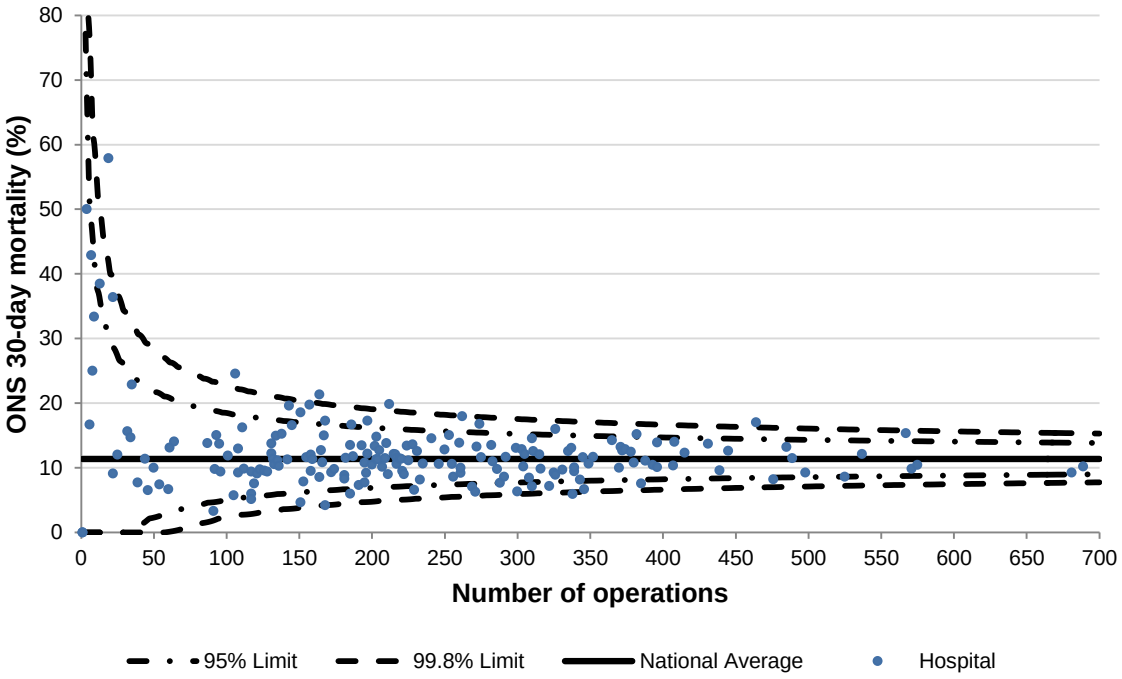
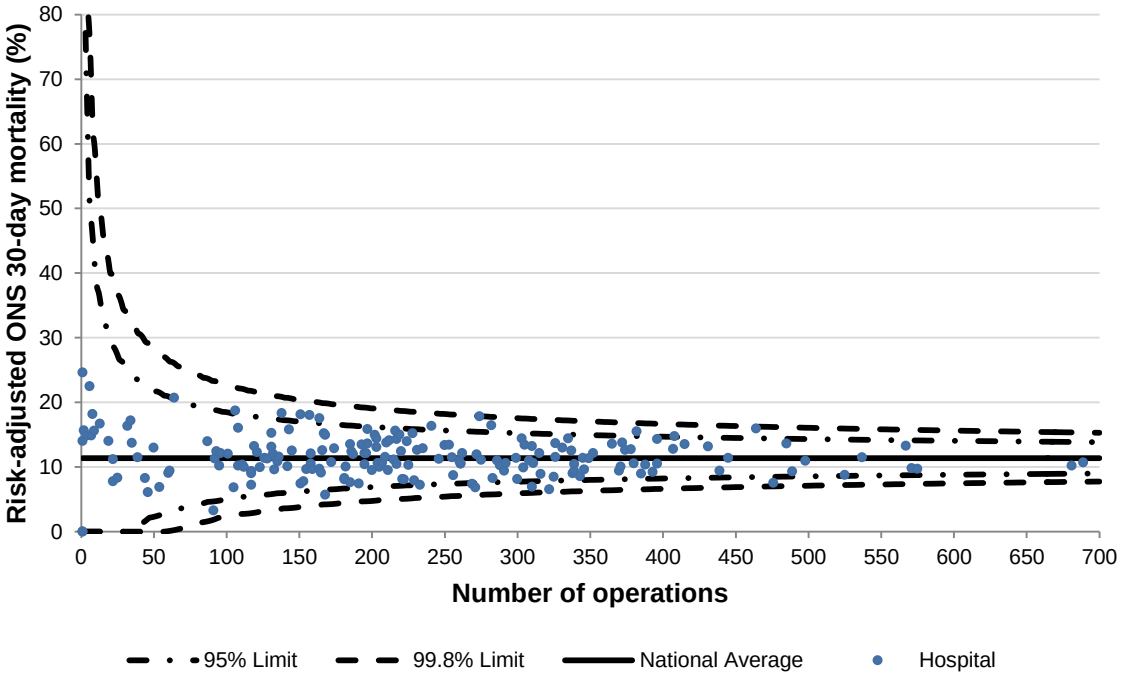
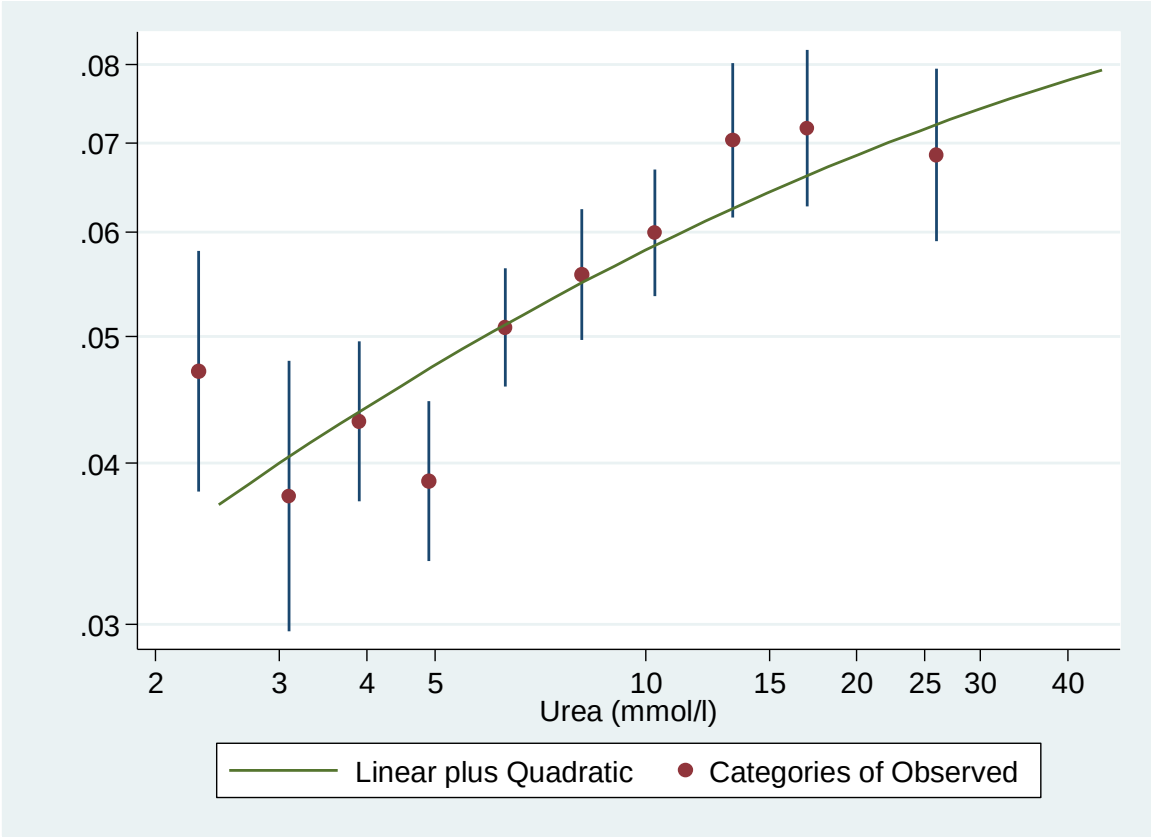


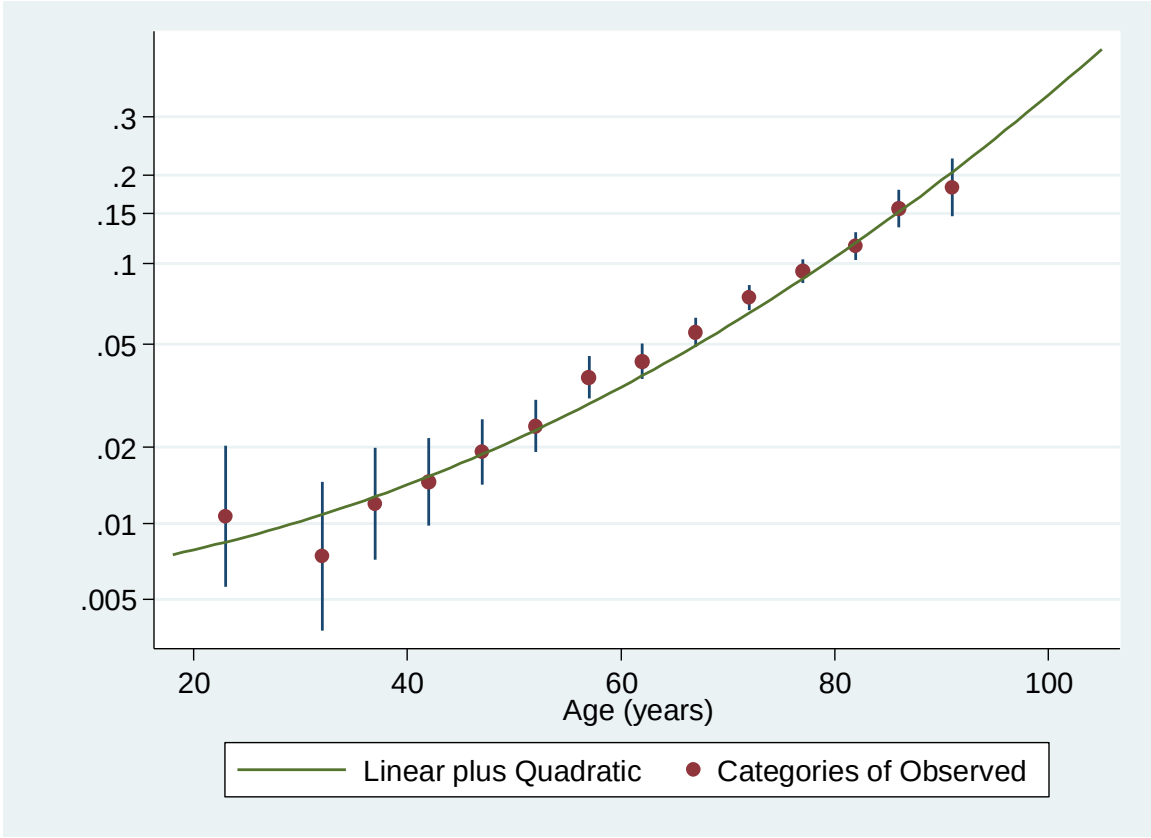
Figure 6
Funnel plot of adjusted 30-day postoperative mortality rates

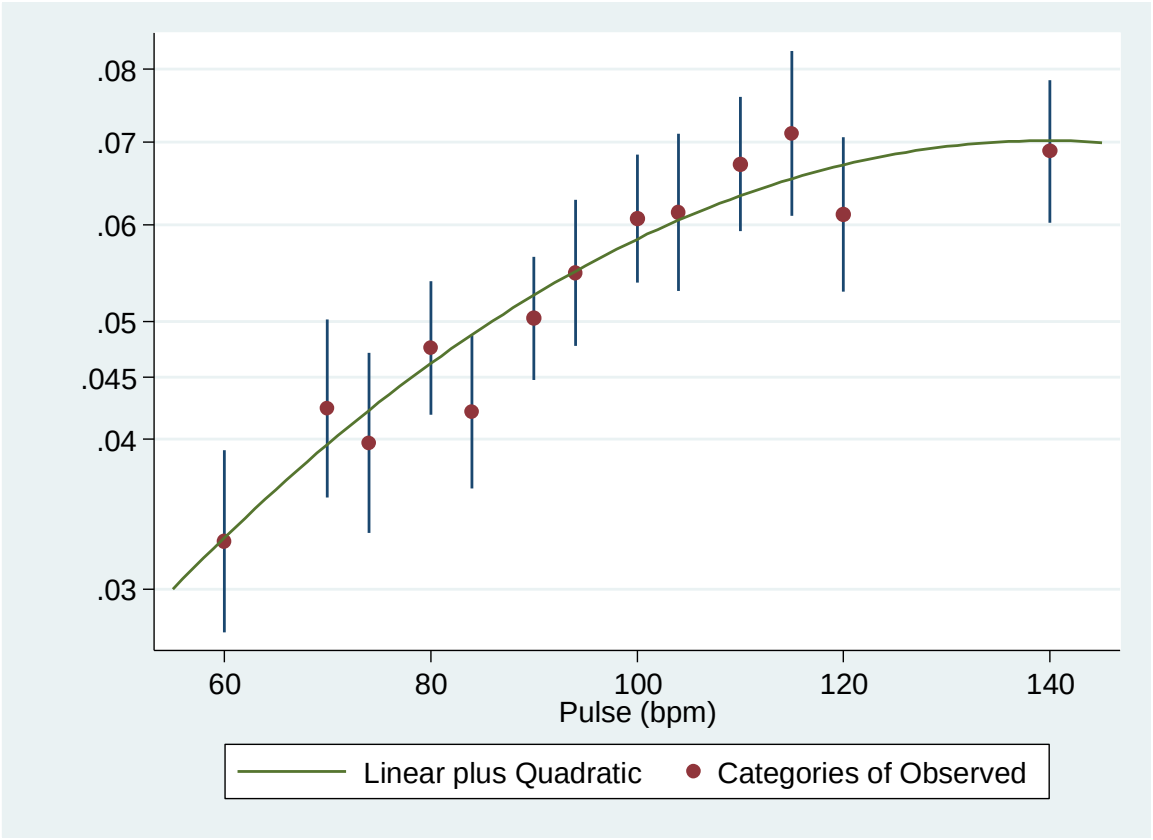


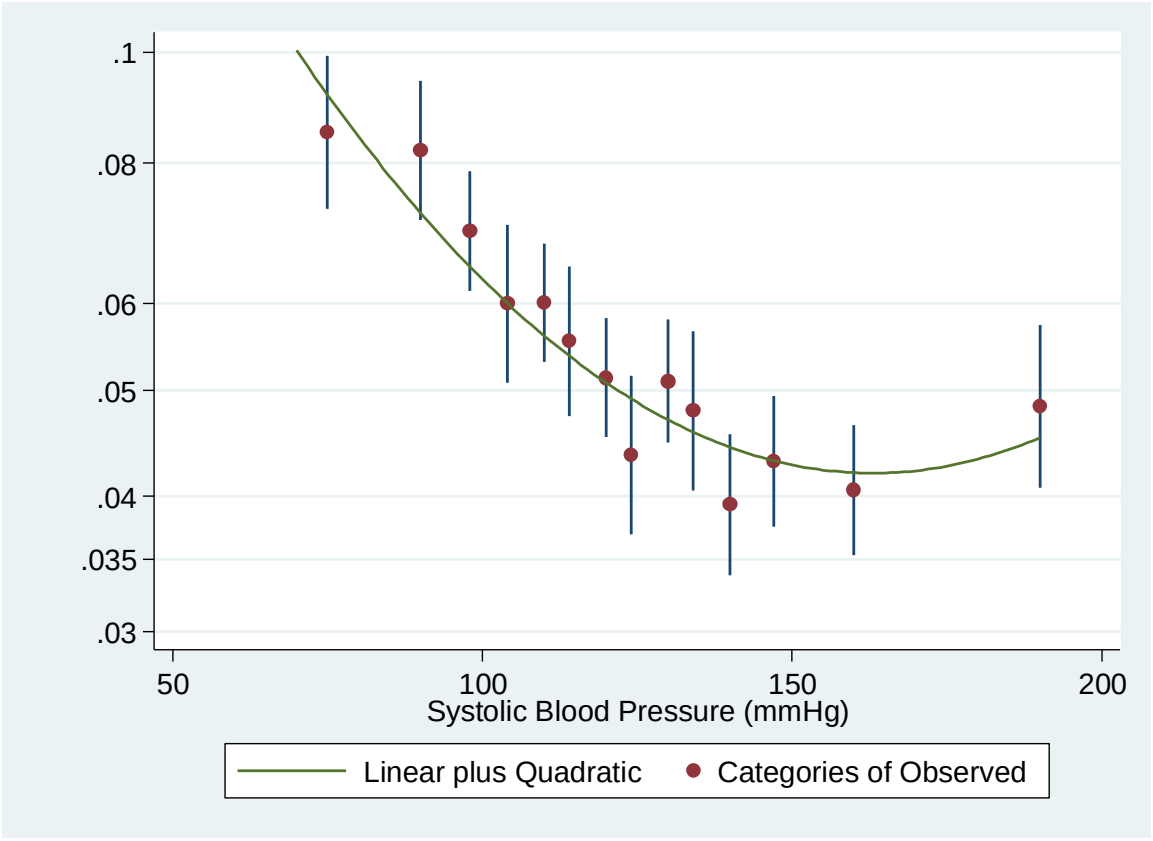
Appendix 1

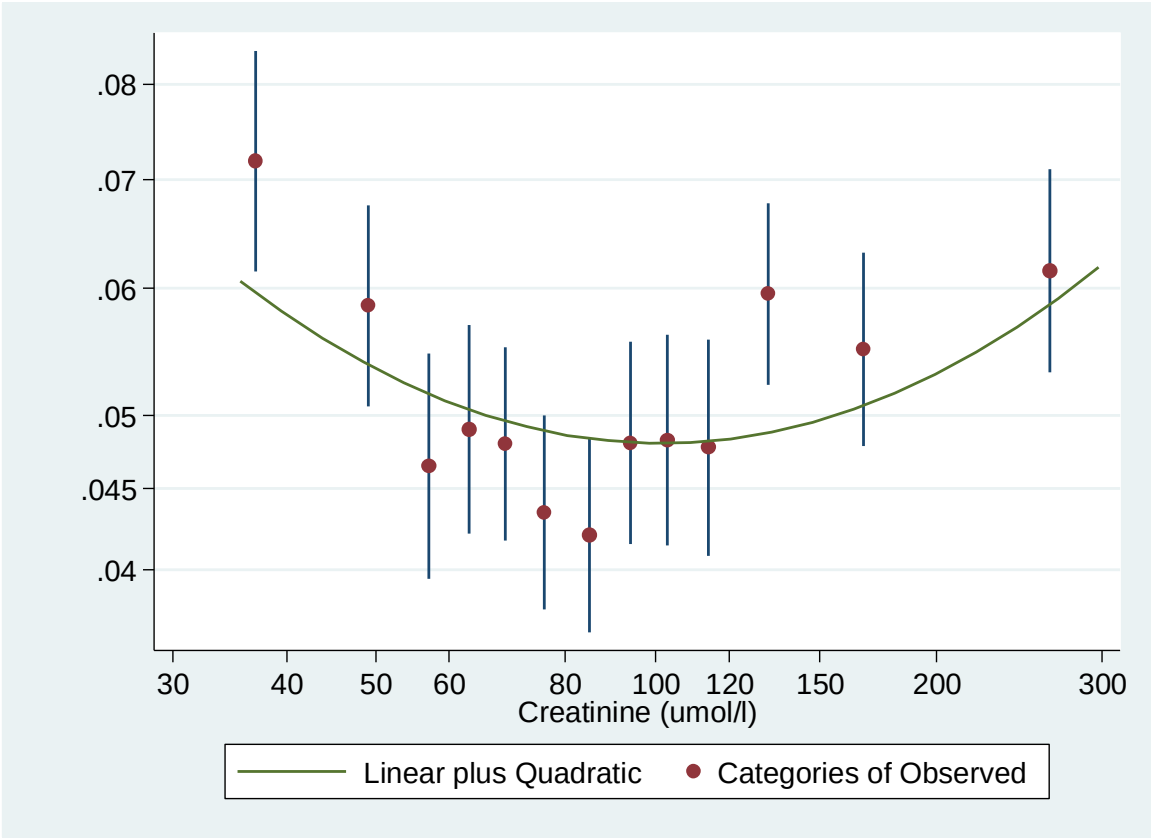
Graphs showing the non-linear relationships between 30-day mortality and continuous variables

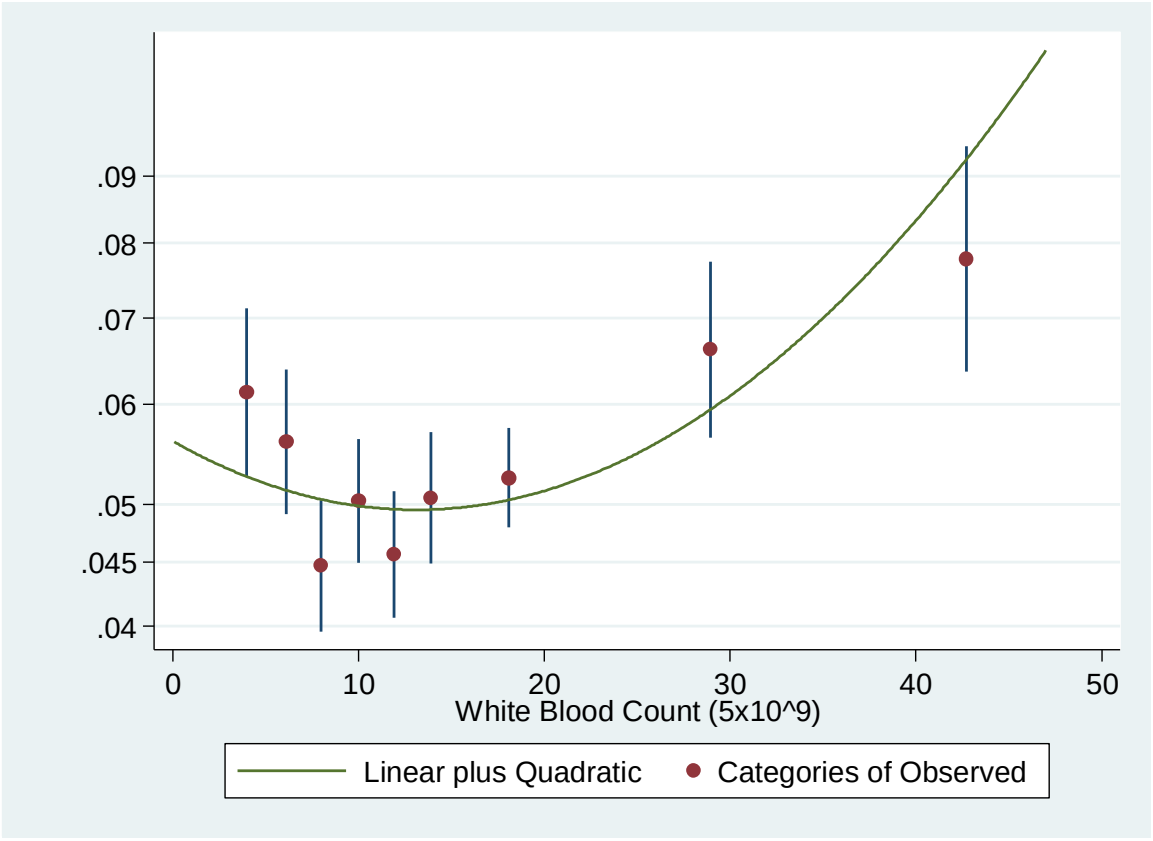


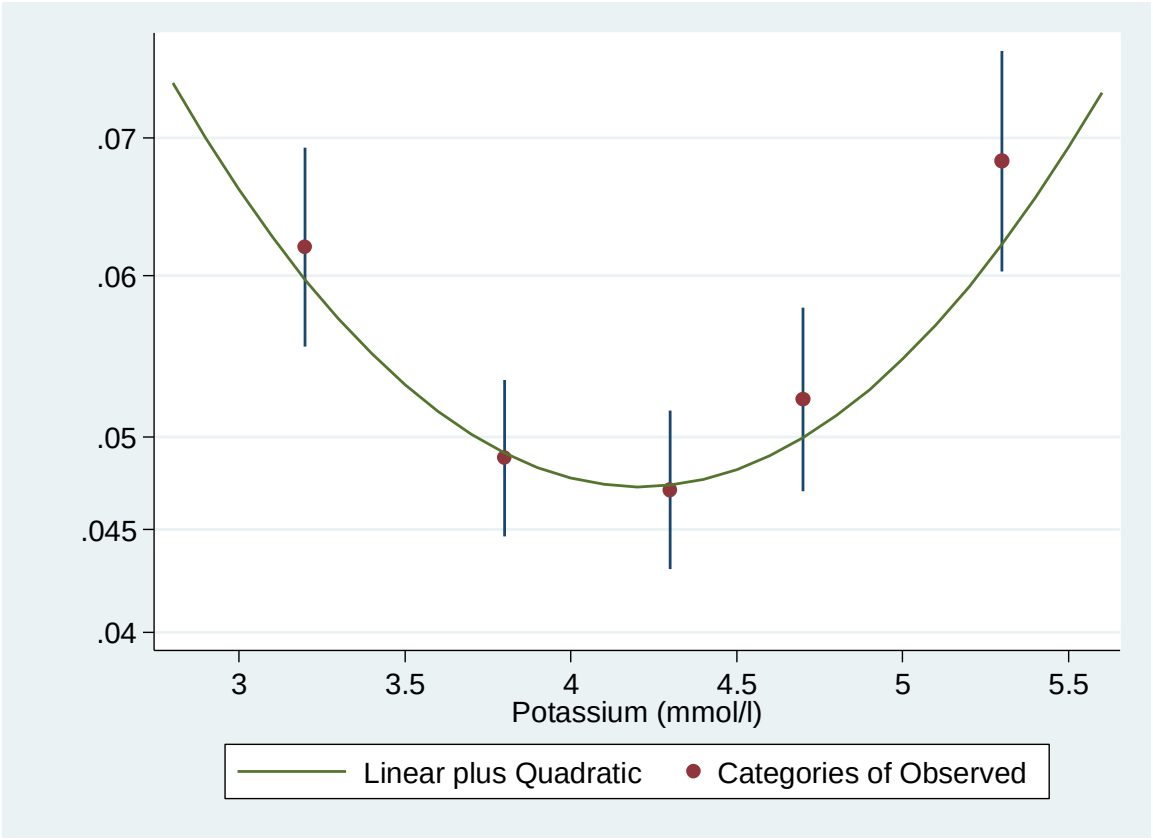


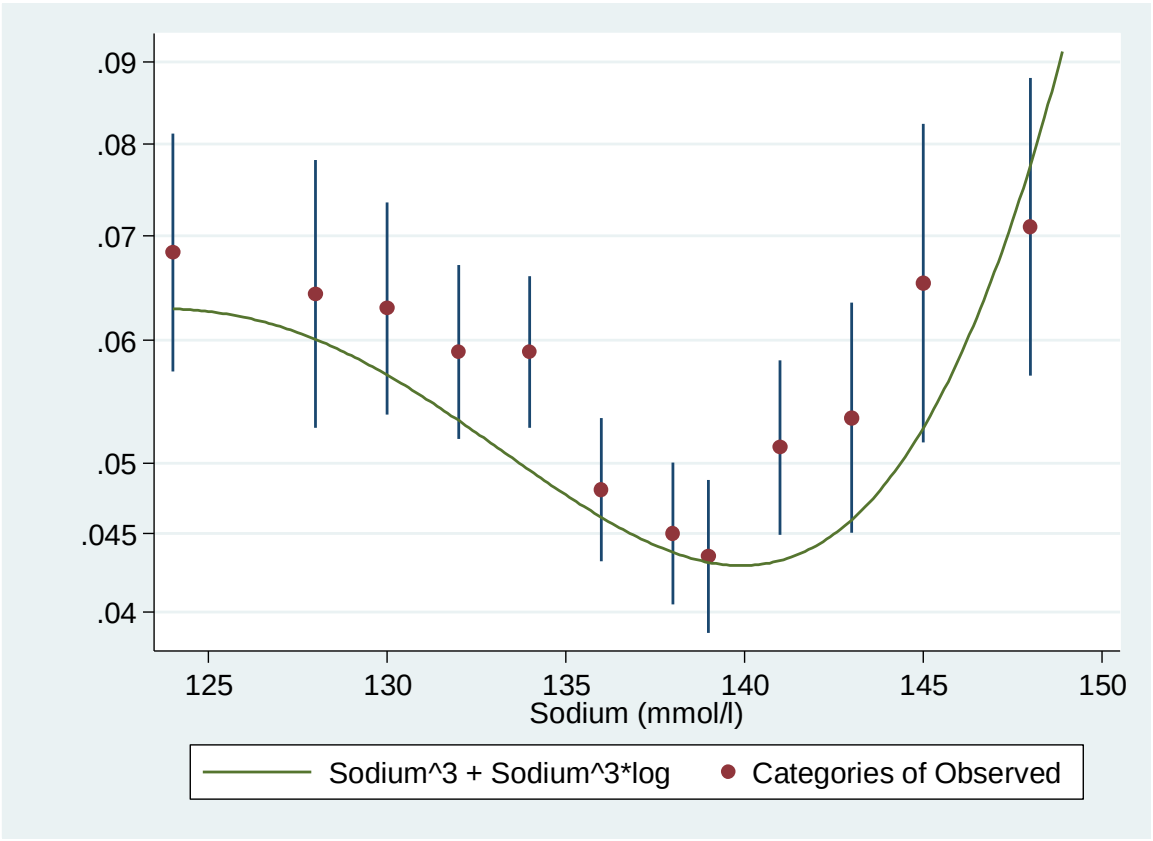

.08
.07
.06
.05
.045
.04
.03
60
80
100
120
140
Pulse (bpm)
Linear plus Quadratic
Categories of Observed

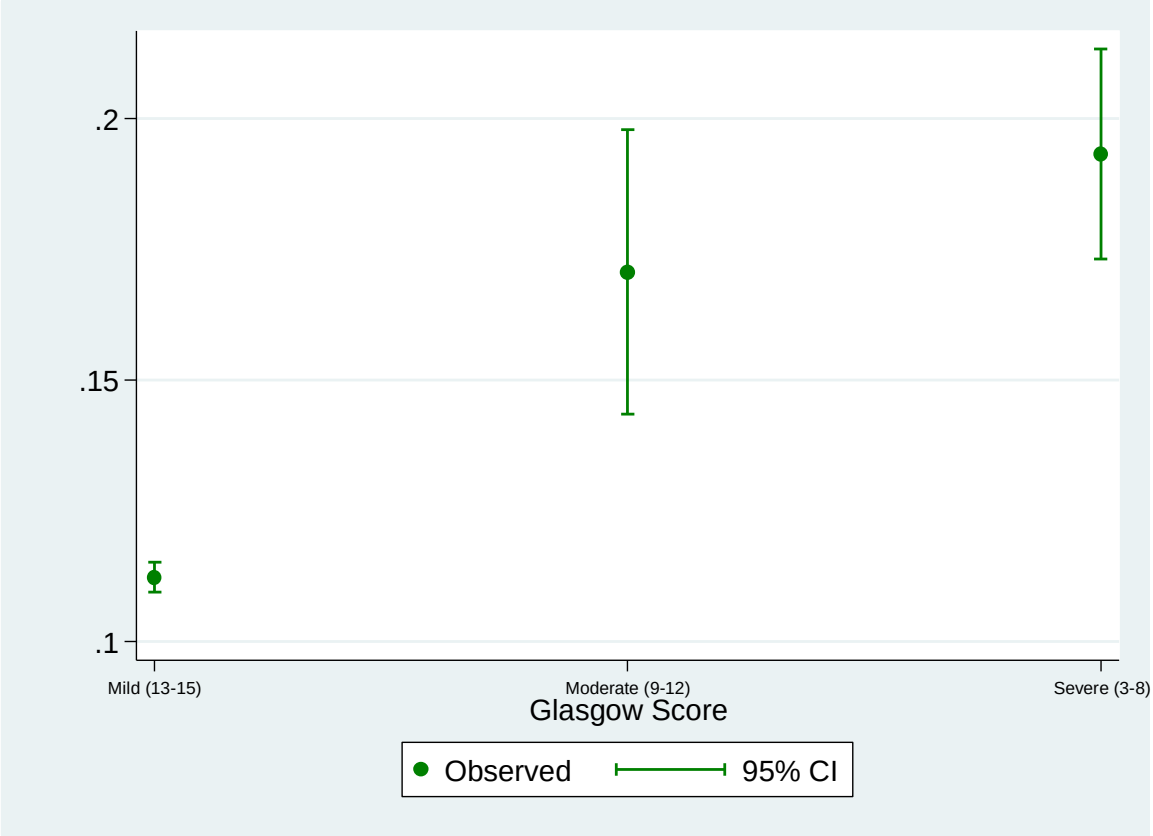

.1
.08
.06
.05
.04
.035
.03
50
100
150
200
Systolic Blood Pressure (mmHg)
Linear plus Quadratic
Categories of Observed


.08
.07
.06
.05
.045
.04
30
40
50
60
80
100
120
150
200
300
Creatinine (umol/l)
Linear plus Quadratic
Categories of Observed


.09
.08
.07
.06
.05
.045
.04
0
10
20
30
40
50
White Blood Count (5x10^9)
Linear plus Quadratic
Categories of Observed


.07
.06
.05
.045
.04
3
3.5
4
4.5
5
5.5
Potassium (mmol/l)
Linear plus Quadratic
Categories of Observed


.09
.08
.07
.06
.05
.045
.04
125
130
135
140
145
150
Sodium (mmol/l)
Sodium^3 + Sodium^3*log
Categories of Observed

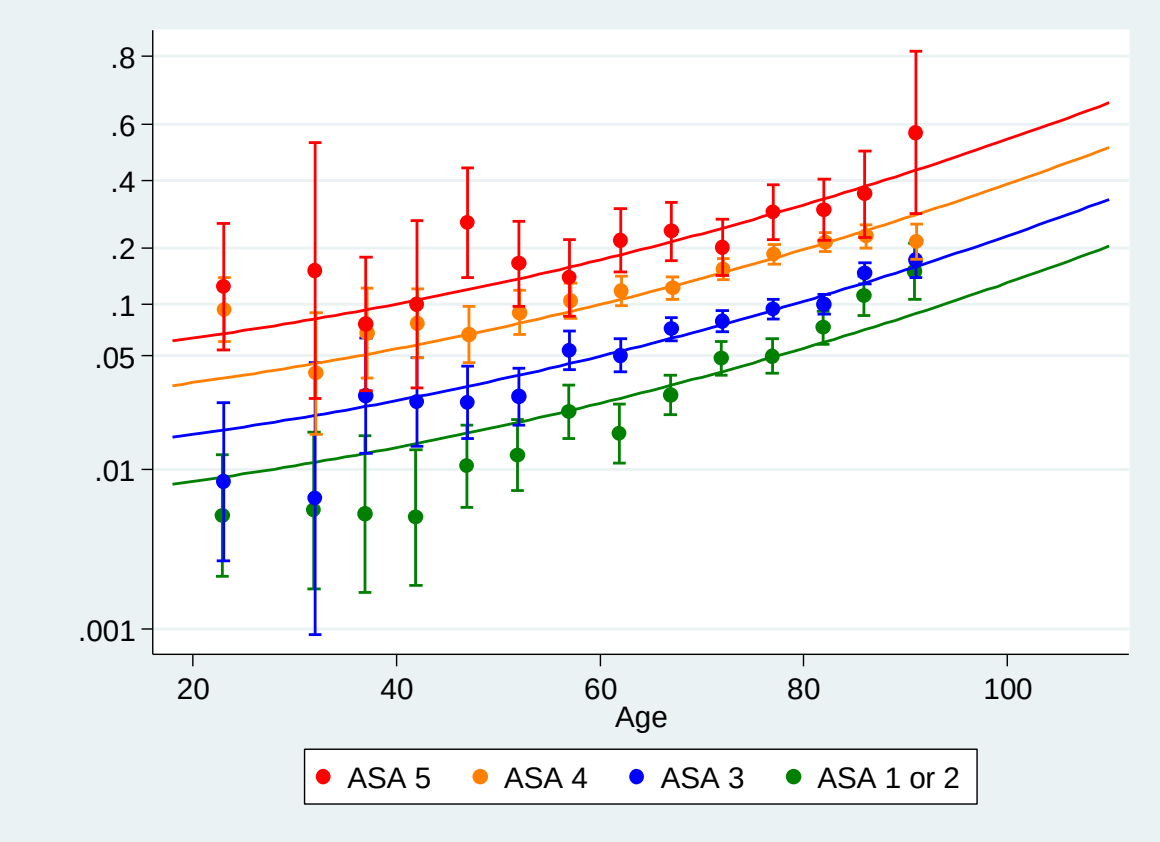

.2
.15
.1
Mild (13-15)
Moderate (9-12)
Severe (3-8)
Glasgow Score
Observed
95% CI

Appendix 2

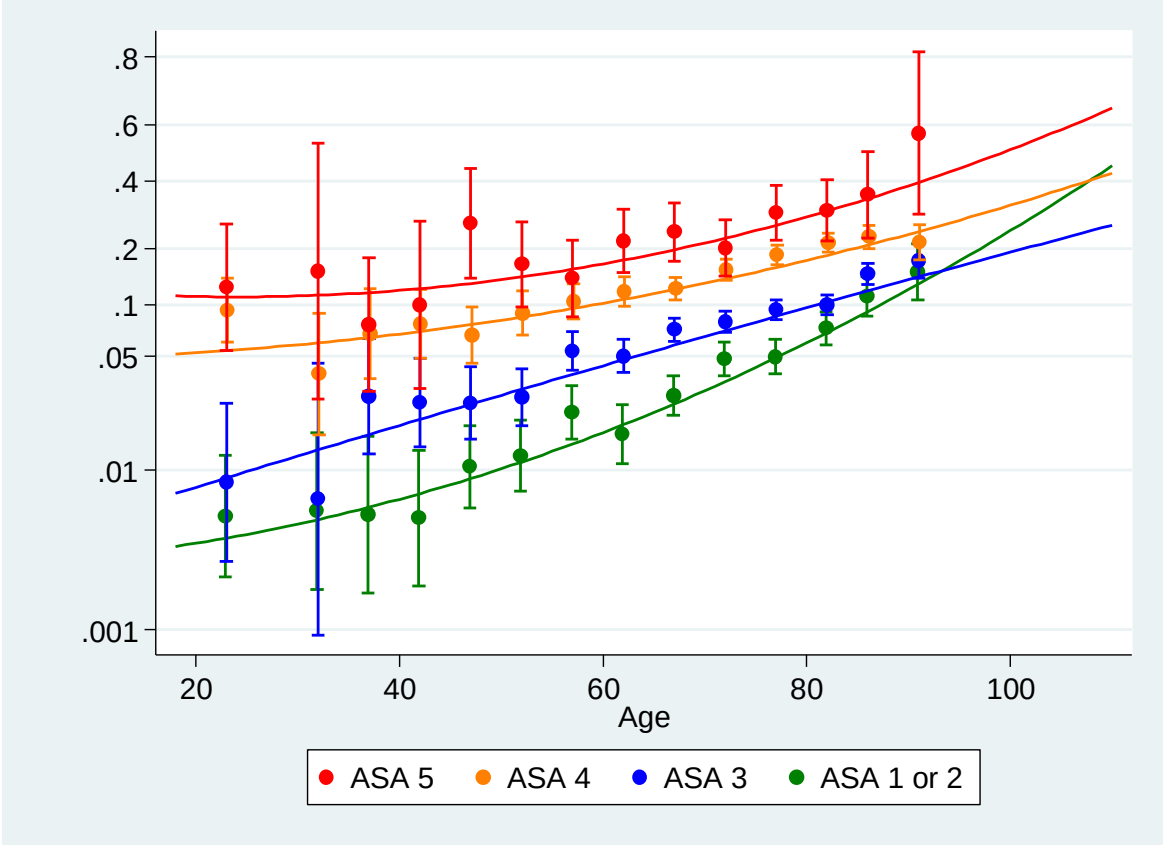
Interaction terms included in the NELA risk model

Age x ASA interaction

The simple model with the same shape relationship between age and mortality in each ASA group. Model fit is worse for young patients (mortality under-estimated for ASA 1 to 3 and over-estimated for ASA 5). Whilst it appears to fit for ASA grades 3, 4 and 5, it appears to fit less well for patients in ASA grades 1 and 2.

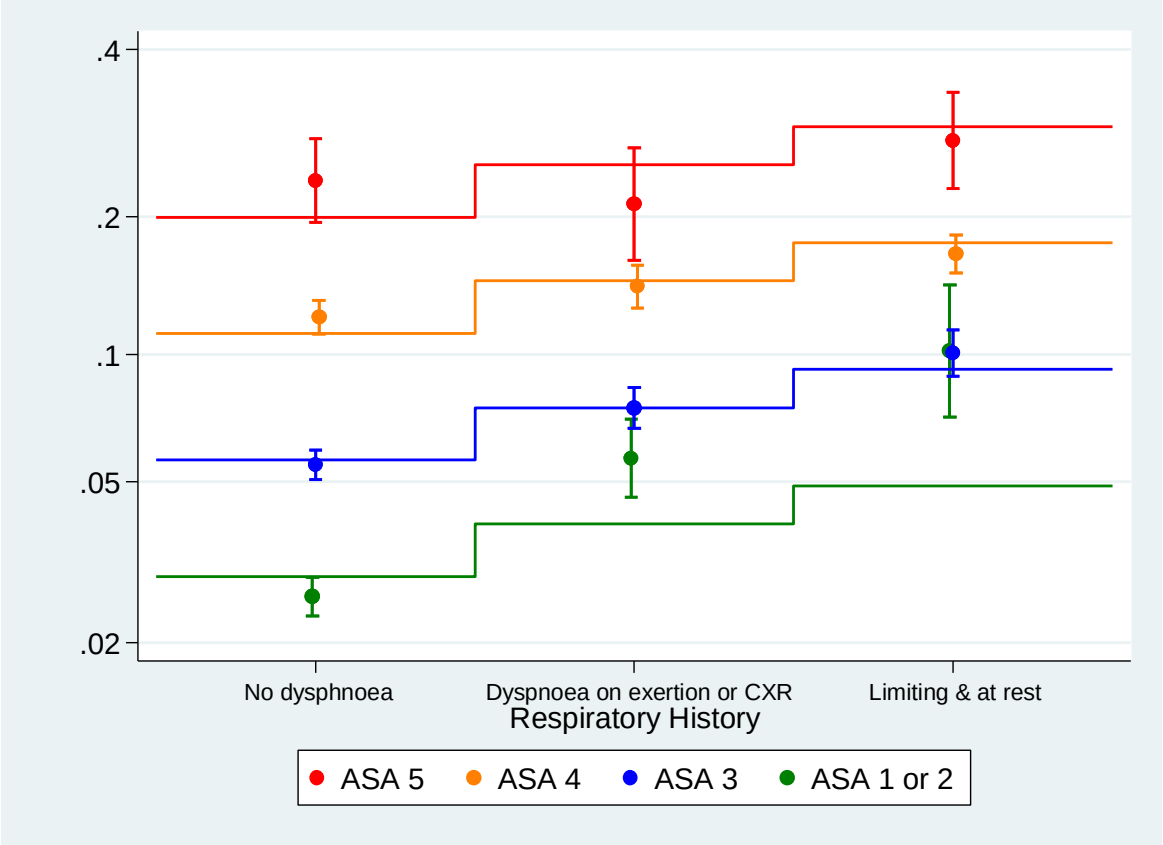


Including an interaction term allows the model to have a different shaped relationship between age and mortality for each ASA group (i.e. an interaction between age and ASA).



ASA x respiratory history interaction

The simple model forces the effect of ASA on mortality to be the same across categories of respiratory history. This approach produces a reasonable fit to the data (dots with error bars) except for patients with ASA 1 or 2.



Including an interaction term in the model allows the effect of ASA on mortality to be varied across categories of respiratory history.