Anaesthesia 2022, 77, 1356–1367 doi:10.1111/anae.15858

Original Article

Development and internal validation of a model for postoperative morbidity in adults undergoing major elective colorectal surgery: the peri-operative quality improvement programme (PQIP) colorectal risk model

J. Bedford,[1,2] **P. Martin,**[3] **S. Crowe,**[4] **D. Wagstaff,**[1,2] **C. Santos,**[5] **G. Singleton,**[6] **R. Baumber,**[7] **C. Vindrola-Padros,**[8] **R. Vohra,**[9] **M. Swart,**[10] **C. M. Oliver,**[1,11] **J. Dorey,**[12] **I. Leeman,**[12] and **S. R. Moonesinghe,**[13,14] **for the peri-operative quality improvement programme (PQIP) collaborative***

1 Consultant, 13 Professor, UCLH Surgical Outcomes Research Centre, Department of Anaesthesia and Peri-operative Medicine, University College London Hospitals NHS Foundation Trust, London, UK
2 PhD student, 6 MPhil student, 11 Associate Professor, 14 Professor, Centre for Peri-operative Medicine, 8 Senior Research Fellow, Research Department for Targeted Intervention, UCL Division of Surgery and Interventional Science, London, UK
3 Lecturer, Department of Applied Health Research, University College London, UK
4 Professor, Clinical Operational Research Unit, University College London, UK
5 Data analyst, Health Services Research Centre, National Institute for Academic Anaesthesia, Royal College of Anaesthetists, London, UK
7 Consultant, Department of Anaesthesia, Royal National Orthopaedic Hospital NHS Trust, London, UK
9 Consultant, Department of Surgery, Nottingham University Hospitals NHS Trust, Nottingham, UK
10 Consultant, Department of Anaesthesia, Torbay and South Devon NHS Trust, Devon, UK
12 Representative, Lay Committee, Royal College of Anaesthetists and Lay representatives PQIP project team

Summary

Over 1.5 million major surgical procedures take place in the UK NHS each year and approximately 25% of patients develop at least one complication. The most widely used risk-adjustment model for postoperative morbidity in the UK is the physiological and operative severity score for the enumeration of mortality and morbidity. However, this model was derived more than 30 years ago and now overestimates the risk of morbidity. In addition, contemporary definitions of some model predictors are markedly different compared with when the tool was developed. A second model used in clinical practice is the American College of Surgeons National Surgical Quality Improvement Programme risk model; this provides a risk estimate for a range of postoperative complications. This model, widely used in North America, is not open source and therefore cannot be applied to patient populations in other settings. Data from a prospective multicentre clinical dataset of 118 NHS hospitals (the peri-operative quality improvement programme) were used to develop a bespoke risk-adjustment model for postoperative morbidity. Patients aged ≥ 18 years who underwent colorectal surgery were eligible for inclusion. Postoperative morbidity was defined using the postoperative morbidity survey at postoperative day 7. Thirty-one candidate variables were considered for inclusion in the model. Death or morbidity occurred by postoperative day 7 in 3098 out of 11,646 patients (26.6%). Twelve variables were incorporated into the final model, including (among others): Rockwood clinical frailty scale; body mass index; and index of multiple deprivation quintile. The C-statistic was 0.672 (95%CI 0.660–0.684), with a bootstrap optimism corrected C-statistic of 0.666 at internal validation. The model demonstrated good calibration across the range of morbidity estimates with a mean slope gradient of predicted risk of 0.959 (95%CI 0.894–1.024) with an index-corrected intercept of −0.038

(95%CI −0.112–0.036) at internal validation. Our model provides parsimonious case-mix adjustment to quantify risk of morbidity on postoperative day 7 for a UK population of patients undergoing major colorectal surgery. Despite the C-statistic of < 0.7, our model outperformed existing risk-models in widespread use. We therefore recommend application in case-mix adjustment, where incorporation into a continuous monitoring tool such as the variable life adjusted display or exponentially-weighted moving average-chart could support high-level monitoring and quality improvement of risk-adjusted outcome at the population level.

Correspondence to: S. R. Moonesinghe
Email: ramani.moonesinghe@nhs.net
Accepted: 28 July 2022
Keywords: colorectal surgery; monitoring; morbidity; quality improvement; risk-adjustment
*See online Supporting Information Appendix S1
This article is accompanied by an editorial by T. G. Coulson et al., *Anaesthesia* 2022; **77**: 1332–5.
Twitter: @jbedford1984; @duncanwagstaff; @rachelbaumber; @CeciliaVindrola; @CMOliver_; @rmoonesinghe

Introduction

Over 1.5 million major surgical procedures take place in the UK NHS every year and approximately 25% of patients develop a complication following major surgery [1, 2]. Colorectal surgery, including for colorectal cancer, accounts for a substantial proportion of major surgery and complications, with almost 2 million new colorectal cancer diagnoses worldwide per year and 1 million deaths [3]. Postoperative morbidity is predictive of reduced long-term survival across a range of specialties, including colorectal surgery [4, 5].

Two frequently cited risk-adjustment models of postoperative morbidity are the physiological and operative severity score for the enumeration of mortality and morbidity (POSSUM) model and the American College of Surgeons national surgical quality improvement programme (ACS-NSQIP) risk prediction calculator [6, 7]. A third model that could be considered for use in major colorectal surgery populations is the surgical outcome risk tool (SORT) for morbidity, a modification of the validated SORT mortality tool [8, 9]. The POSSUM and SORT morbidity models were derived from small, single-centre cohorts. Both included patients undergoing non-colorectal procedures. The morbidity risk estimate for POSSUM has not been updated since its first publication in 1991, and now overestimates morbidity in contemporary practice [10, 11]. Although the ACS-NSQIP risk prediction calculator demonstrates good discrimination, many of the 21 variables required for risk calculation are not routinely collected in NHS clinical datasets [6]. As model coefficients have not been published, it is difficult to validate this tool outside the ACS-NSQIP programme and performance in a UK population is unknown.

Given the limitations of existing models, we sought to develop and internally validate a bespoke model in a UK population of patients undergoing major elective colorectal surgery. Our aim was to support risk adjustment and monitoring of morbidity outcomes across a range of institutions by using the model to inform continuous display methods, such as the variable life-adjusted display and exponentially-weighted moving average charts [12, 13]. These tools can be used by institutions to benchmark practice and drive quality improvement programmes to enhance outcomes for patients [14].

Methods

This study was approved by the Health Research Authority and analysis approved following discussion with the PQIP National Project Team and Chief Investigator. In the period under investigation, PQIP was recruiting consenting patients in 112 NHS hospitals in England and six NHS hospitals in Wales. Seventy-nine of the 118 NHS hospitals had recruited 50 or more patients.

We report data handling and analysis in line with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement [15]. Data were collected prospectively as part of the national PQIP study by local site investigators, typically research nurses and clinicians. Patients aged ≥18 years, undergoing colorectal surgery between 15 December 2016 and 31 March 2020 were eligible for inclusion within this analysis (online Supporting Information, Appendix S2). Cases after

31 March 2020 were excluded to limit the impact of the COVID-19 pandemic on model development. Patients undergoing concurrent procedures were included in the analysis if the primary operation was recorded as colorectal.

Risk-adjustment variables were defined a priori. Morbidity was defined using the validated and reliable postoperative morbidity survey (POMS) at postoperative day 7 [16]. As the national median length of stay following major colorectal cancer resection in the UK is 7 days (IQR 5–11) [17], postoperative length of stay >7 days due to morbidity identifies patients with substantial postoperative morbidity. Our primary outcome, treated as a binary variable, was defined as `morbidity present´ if a patient remained in hospital at postoperative day 7 with any POMS-defined morbidity in the preceding 24 h. Patients discharged before postoperative day 7 were assumed to be morbidity free. Our secondary outcome was the presence of any POMS-major defined morbidity at day 7 (online Supporting Information, Table S1 [8]). Patients who died before postoperative day 7 were assigned to have morbidity. Cases with missing outcome data were excluded from analysis.

Implausible values were removed and treated as missing (online Supporting Information, Figure S1). Index of multiple deprivation scores for England and Wales local super output areas were generated for each patient record using published methods [18] and converted into UK quintiles using open source government data [19–21]. Postcodes, adjusted index of multiple deprivation scores and UK index of multiple deprivation quintiles are available here: https://github.com/jbedford1984/PQIP-CR-model-development. Due to its inclusion in the PQIP dataset in January 2019, Rockwood clinical frailty scale scores were only available for a subgroup of the dataset [22]. However, we included the clinical frailty scale in our analysis, imputing missing items, due to evidence supporting its association with postoperative morbidity [22, 23]. The effect of the high proportion of missing data (65%) of this is explored in one of our sensitivity analyses.

Continuous variables were centred and modelled as restricted cubic splines (online Supporting Information, Figure S2). Categorical variable groups containing <0.5% of cases were combined with similar categories where clinically appropriate or excluded if not. Multiple imputation with chained equations was used to impute missing predictor data (10 complete datasets) [24]. Data were assumed to be missing at random. The imputation model included all predictor and outcome variables under consideration and additional variables to support the missing at random assumption (online Supporting Information, Appendix S3).

Thirty-one candidate predictors, deemed non-modifiable at the point of admission for surgery, were considered for inclusion in the model (online Supporting Information, Table S2). Intra- and postoperative variables influenced by quality of care provided, such as intra-operative blood loss, were excluded. To minimise the risk of overfitting, a sample size of 2632 patients was required to fit a model with 31 candidate predictors, providing 21 events per predictor [25].

We identified predictors of POMS-defined postoperative morbidity by fitting backwards-stepwise logistic regression models, selected on Akaike Information Criterion, across 500 bootstrap re-samples of each of the 10 imputed datasets (total 5000 models). Variables selected into ≥ 70% of the 5000 bootstrap models were included in our model, the PQIP colorectal risk model, hereafter named PQIP-CR. We performed variable selection rather than including all 31 predictors to develop a parsimonious risk-adjustment model to facilitate implementation in clinical practice. Model coefficients and performance estimates were pooled using Rubin's rules. We assessed discrimination with the C-statistic. Calibration was assessed visually with the bootstrap bias corrected calibration curve (500 bootstrap replicates of each imputed dataset) and through calculation of the slope and intercept estimates at internal validation. Overall model fit was evaluated with the Brier score.

Linear shrinkage factors were estimated to calibrate the PQIP-CR model when predicting our secondary outcome POMS-major by fitting the log odds estimated by the PQIP-CR model as the predictor variable. Internal validation was performed using bootstrap correction of optimism, with 1000 bootstrap re-samples of each imputed dataset [26]. We also calculated the sensitivity, specificity, positive predictive value and negative predictive value of our model at multiple thresholds of estimated risk. We compared performance of PQIP-CR to that of the POSSUM and SORT morbidity models [7, 8]. Slope and intercept values for the existing models were freely estimated using published variables and categorisations, with POMS-defined morbidity as the outcome measure. We were not able to compare performance of the ACS-NSQIP model as some variables required for risk calculation were not available in our dataset. Planned sensitivity analyses are described in the online Supporting Information (Appendices S4–S7).

Statistical analyses were performed using R version 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria) and online Supporting Information (Appendix S8) shows packages used for data management and visualisation.

Results

A total of 11,646 cases contributed to model development and internal validation (99.6% of potential eligible cases; Fig. 1). Of these, 6241 (53.6%) patients were discharged before postoperative day 7 and therefore assumed to have no POMS-defined morbidity at this time-point. Table 1 shows patient characteristics for our cohort. There were 3098 (26.6%) patients with POMS defined morbidity at postoperative day 7 (including 33 patients who died within 7 days of surgery, Table 2). POMS-major defined morbidity occurred in 2067 (17.7%) patients. Median postoperative length of stay for the cohort was 6 days (IQR 4–9). Overall data completeness for predictor variables was excellent (95.8%). The proportion of missing data for each predictor variable is shown in online Supporting Information (Table S3).

Twelve variables were selected into ≥70% of the 5000 backwards-selection bootstrap models, resulting in the PQIP-CR model containing the following predictors: mode of surgery; clinical frailty scale; severity of surgery based on the Clinical Coding and Schedule Development Group classification; ASA physical status; BMI; serum sodium; index of multiple deprivation quintile; white cell count; serum urea; age; and serum creatinine (online Supporting Information, Table S4). Table 3 shows the odds ratios and 95%CIs for each categorical predictor in the model. For continuous variables, which were modelled as restricted

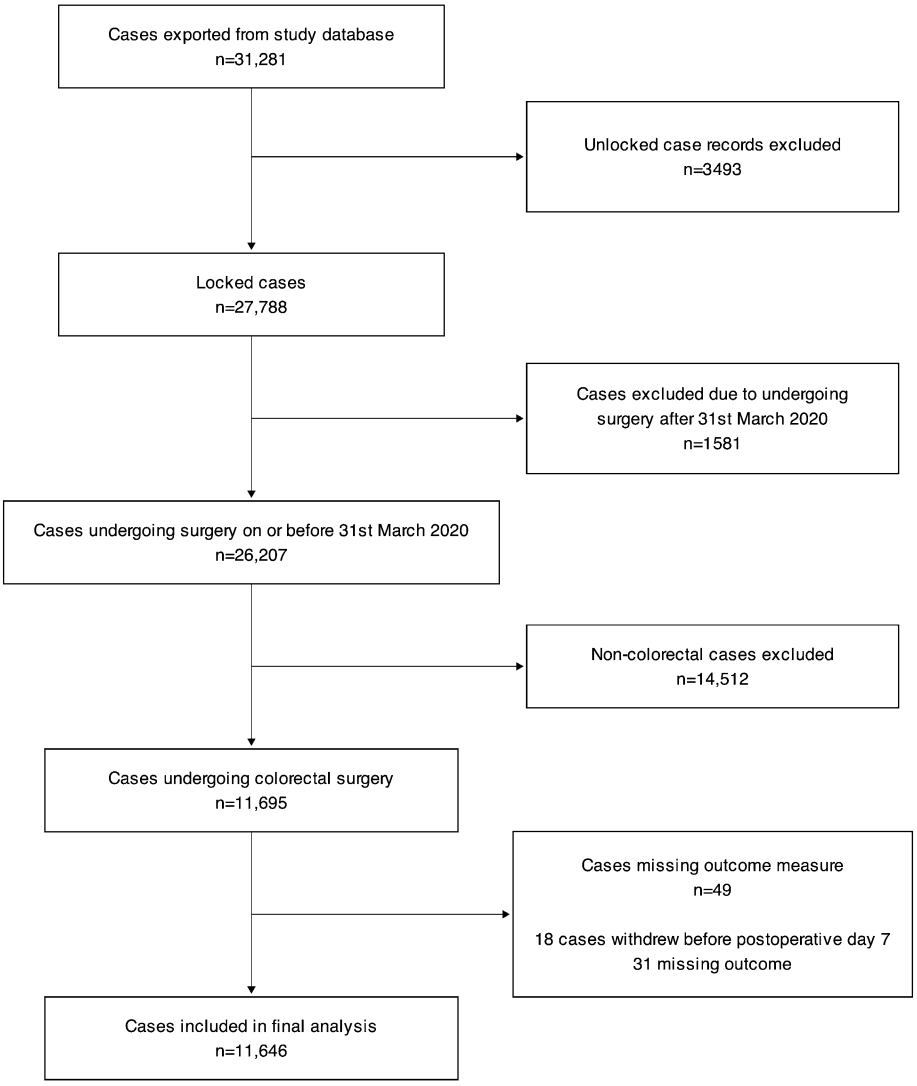

Figure 1 Study flow diagram showing study sample construction.

Table 1 Patient characteristics and comorbidities for the PQIP-colorectal model derivation cohort. Values are number (proportion), mean (SD) or median [IQR (range]).

Total	11,646 (100%)
Age; y	64 (14)
Sex	
Female	5082 (43.6%)
ASA	
1	236 (10.9%)
2	1740 (62.1%)
3	1066 (25.8%)
4/5	131 (1.1%)
Severity of surgery*	
Minor/intermediate/major	195 (1.7%)
Xmajor	6019 (51.7%)
Complex	5432 (46.6%)
NCEPOD categorisation	
Elective	10,550 (90.6%)
Expedited	1096 (9.4%)
Mode of surgery	
Laparoscopic	7367 (63.3%)
Open	4052 (34.8%)
Robotic	227 (1.9%)
Rockwood clinical frailty scale	
1	2169 (18.6%)
2	4113 (35.3%)
3	3572 (30.7%)
4	977 (8.4%)
5	485 (4.2%)
6–9	330 (2.8%)
Index of multiple deprivation quintile	
1 – most deprived	1716 (14.7%)
2	2078 (17.8%)
3	2365 (20.3%)
4	2691 (23.1%)
5 – least deprived	2796 (24.0%)
BMI; $kg.m^2$	27.6 (5.4)
Urea; $mmol.l^{-1}$	5.1 (4.2–6.2 [1.5–33.0])
Creatinine; $\mu mol.l^{-1}$	75 (65–88 [26–898])
White cell count; $\times 10^9.l^{-1}$	7.2 (5.9–8.8 [0.1–31.4])
Diabetes	
No	10,154 (87.2%)
Yes, HbA1c $\leq$ 69 $mmol.mol^{-1}$	1192 (10.2%)
Yes, HbA1c >69 $mmol.mol^{-1}$	300 (2.6%)
Smoking history	
Non-smoker	6136 (52.7%)
Ex-smoker (> 6 months)	3818 (32.8%)
Ex-smoker (< 6 months)	490 (4.2%)
Current smoker	1202 (10.3%)
ECG findings	
Normal	8927 (76.7%)
AF rate 60–90	483 (4.1%)
AF rate > 90/any other abnormal rhythm	1337 (11.5%)
Not done	901 (7.7%)
NYHA classification	
1	9592 (82.4%)
2	1723 (14.8%)
3/I4	331 (2.8%)
Diagnosis of cancer in last 5 years	
Yes	8342 (71.6%)
Respiratory history findings	
Normal	9968 (85.6%)
Dyspnoea on exertion	1310 (11.2%)
Dyspnoea on minimal exertion/at rest/chest X-ray	368 (3.2%)
Cardiovascular history findings	
No failure	8712 (74.8%)
Diuretic, digoxin, antihypertensive therapy	2686 (23.1%)
Peripheral oedema, warfarin therapy, raised JVP or cardiomegaly on chest X-ray	248 (2.1%)

Xmajor, extra major; AF, atrial fibrillation; NYHA, New York Heart Association; JVP, jugular venous pulse.
*Clinical Coding and Schedule Development Group classification.

cubic splines, we have shown point estimates of odds ratios at various points on the continuous scale. Online Supporting Information (Figure S2) shows the relationship between continuous variables selected and postoperative morbidity.

The C-statistic of the PQIP-CR model was 0.672 (95%CI 0.660–0.684), online Supporting Information (Figure S3) shows the receiver operating characteristic curve with 95% CIs. At internal validation, the optimism-corrected C-statistic was 0.666 (95%CI 0.654–0.678). A C-statistic of 0.5 indicates a model is no better at predicting an outcome than random chance, ≥ 0.7 indicates acceptable discrimination and ≥ 0.8 is considered strong discrimination. The model demonstrated good apparent and bias corrected calibration across the range of morbidity predictions (Fig. 2), with a maximum difference between predicted and observed morbidity of 2.4 percentage points across deciles of predicted risk (predicted morbidity 33.8%, observed morbidity 36.2%). Calibration was maintained after

Table 2 Incidence of postoperative morbidity and complication outcomes in our cohort. Values are number (proportion).

Total number of patients in cohort	11,646
Patients remaining in hospital at postoperative day 7	5405 (46.4%)
Patients with POMS-defined morbidity at postoperative day 7 (including death ≤ postoperative day 7), of which:	3098 (26.6%)
Pulmonary	575 (4.9%)
Cardiovascular	259 (2.2%)
Infectious	1553 (13.3%)
Neurological	196 (1.7%)
Pain	749 (6.4%)
Renal	1114 (9.6%)
Haematological	94 (0.8%)
Gastrointestinal	1785 (15.3%)
Wound	488 (4.2%)
Patients with POMS-major defined morbidity at postoperative day 7 (including death ≤ postoperative day 7), of which:	2067 (17.7%)
Pulmonary major	575 (4.9%)
Cardiovascular major	259 (2.2%)
Infectious major	1483 (12.7%)
Neurological major	196 (1.7%)
Pain major	77 (0.7%)
Renal major	159 (1.4%)
Haematological major	94 (0.8%)
Wound major	488 (4.2%)
Patients who died within 7 days of surgery	33 (0.3%)
Patients with Clavien-Dindo grade 2 or above complication at any time-point during hospital admission	2997 (25.7%)

POMS, postoperative morbidity survey. POMS-major is a subset of the POMS criteria (online Supporting Information, Table S1). The Clavien-Dindo grading of surgical complications is shown in online Supporting Information (Table S7).

bootstrap correction of bias, with observed risk of POMS and POMS-major morbidity falling within the 95%CIs of predicted PQIP-CR risk across the full range of predictions. At internal validation, the mean slope gradient of predicted risk for the model was 0.959 (95%CI 0.894–1.024) with an index-corrected intercept of −0.038 (95%CI −0.112–0.036). The Brier score was 0.181. Online Supporting Information (Appendix S9) details how to calculate the predicted risk of morbidity from our model. Online Supporting Information (Table S5) shows the sensitivity, specificity, positive predictive value and negative predictive value of the PQIP-CR model at various thresholds of estimated risk.

The C-statistic of the PQIP-CR model when predicting POMS-major morbidity was 0.669 (95%CI 0.655–0.682), with an optimism-corrected C-statistic of 0.668 (95%CI 0.655–0.682). Online Supporting Information (Figure S3) shows the receiver operating characteristic curve with 95% CIs. Bias corrected calibration was good (Fig. 2) across the range of predictions with a slope estimate of 1.001 (95%CI 0.918–1.084) and intercept estimate of 0.002 (95%CI −0.127–0.130) at internal validation. Online Supporting Information (Appendix S10) shows how to calculate the estimated risk of POMS-major defined morbidity using the estimated linear shrinkage factors. Online Supporting Information (Table S6) shows the sensitivity, specificity, positive predictive value and negative predictive value of the PQIP-CR model when predicting POMS-major defined morbidity at various thresholds of estimated risk.

When estimating the risk of Clavien–Dindo [27] grade 2 and above complications the C-statistic was 0.635 (95%CI 0.623–0.647), with performance maintained at internal validation (online Supporting Information, Table S7 and Appendix S4). Calibration was good with a slope estimate of 1.001 (95%CI 0.910–1.093) and an intercept estimate of 0.001 (95%CI −0.099–0.103). Online Supporting Information (Appendix S4) shows the estimated shrinkage factors, how to apply them and the calibration curves following application for Clavien–Dindo defined complications.

Data were available to make morbidity risk estimates for 9173 cases using the POSSUM model and 11,640 for the SORT morbidity model. Fewer estimates were made with the POSSUM model due to missing intra-operative variables required for risk estimation Discrimination of PQIP-CR was higher despite recalibration of the two existing models in our cohort ($p < 0.001$, Table 4). Calibration of PQIP-CR was also superior, providing well calibrated estimates across a wider range of risk predictions.

There were 5241 cases included in our second sensitivity analysis (online Supporting Information, Appendix S5). Missing data for the clinical frailty scale variable reduced from 64.8% in the main cohort to 22.9% (n = 1202). The rate of missing data was similar when comparing patients aged <65 years and those aged ≥65 years. There was no significant difference when comparing clinical frailty scale odds ratios between PQIP-CR and those in the sensitivity analysis. Our third sensitivity analysis assessed the ability of a model to estimate the risk of a more homogenous outcome, gastrointestinal morbidity, defined by POMS (online Supporting Information, Appendix S6). The C-statistic of this model was 0.626 (95%CI 0.611–0.641). Our multilevel sensitivity

Table 3 Estimated odds ratios and coefficients for the PQIP-CR model.

Variable	Odds ratio (95%CI)	Coefficient	Standard error
Mode of surgery			
Laparoscopic	1	Reference	
Open	2.06 (1.88–2.24)	0.719	0.045
Robotic	1.41 (1.04–1.92)	0.346	0.156
Rockwood clinical frailty scale			
1	1	Reference	
2	1.29 (1.11–1.50)	0.255	0.077
3	1.68 (1.40–2.01)	0.517	0.092
4	2.02 (1.53–2.68)	0.705	0.138
5	2.24 (1.64–3.06)	0.809	0.155
6–9	2.48 (1.68–3.66)	0.908	0.192
Sex			
Female	1	Reference	
Male	1.36 (1.23–1.51)	0.310	0.052
Severity of surgery#			
Minor/intermediate/major	1	Reference	
Xmajor	1.19 (0.83–1.70)	0.175	0.183
Complex	1.77 (1.24–2.53)	0.570	0.182
ASA			
1	1	Reference	
2	1.17 (1.00–1.38)	0.160	0.082
3	1.52 (1.27–1.83)	0.421	0.093
4/5	1.56 (1.02–2.38)	0.444	0.215
BMI; kg.m^2			
18	1.09 (0.88–1.34)	*−0.022	0.019
25	1	*0.175	0.074
30	1.21 (1.10–1.34)	*−0.521	0.216
40	1.34 (1.15–1.56)		
Sodium; mmol.l^{-1}			
120	1.84 (1.03–3.28)	*−0.029	0.018
130	1.37 (1.08–1.73)	*−0.012	0.041
140	1	*0.346	0.314
150	1.44 (0.87–2.38)		
Index of multiple deprivation quintile			
5 – least deprived	1	Reference	
4	1.03 (0.91–1.17)	0.031	0.066
3	0.99 (0.86–1.13)	−0.014	0.068
2	1.19 (1.04–1.37)	0.176	0.069
1 – most deprived	1.17 (1.02–1.35)	0.160	0.073
White cell count; × 10^9.l^{-1}			
3.0	1.00 (0.78–1.29)	*−0.010	0.044
7.0	1	*0.132	0.201
11.0	1.15 (1.04–1.28)	*−0.293	0.521
15.0	1.32 (1.21–1.56)		

(continued)

Table 3 (*continued*)

Variable	Odds ratio (95%CI)	Coefficient	Standard error
Urea; mmol.l^{-1}			
3.0	1.17 (0.99–1.37)	*−0.587	0.220
8.0	1	*0.897	0.583
15.0	1.02 (0.76–1.37)	*−2.968	2.964
20.0	1.04 (0.68–1.59)		
Age; y			
30	1	*0.002	0.003
40	1.02 (0.96–1.08)	*0.003	0.004
50	1.04 (0.91–1.18)		
60	1.07 (0.90–1.27)		
70	1.13 (1.01–1.47)		
80	1.22 (1.06–1.66)		
Serum creatinine; μmol.l^{-1}			
30	1.70 (1.13–2.58)	*−0.787	0.311
60	1	*2.697	0.937
90	1.12 (0.98–1.29)	*−9.184	3.418
120	1.12 (0.95–1.33)		

[#]Clinical Coding and Schedule Development Group classification. Xmajor, extra major; Point estimates of odds ratios are shown for continuous variables. These do not represent categorisations and cannot be used to calculate risk estimates.
*Coefficients estimated in model fit. PQIP-CR model performance: Pooled R^2 = 0.102, Likelihood ratio χ^2 statistic 846.22, df 34, $p < 0.001$; Brier score 0.181.

analysis had a variance of 0.136 for the random effect of hospital site, and a median odds ratio of 1.42 for the random intercept terms [28]. Online Supporting Information (Appendix S7) shows the variation in intercept estimates by hospital site. The estimated odds ratios for the fixed effects in the model are also shown alongside the median odds ratio for the hospital level variable.

Discussion

Our model, PQIP-CR, demonstrates good calibration to risk-adjust postoperative day 7 morbidity defined by the POMS in the setting of elective major colorectal surgery with discrimination performance superior to published morbidity risk models [7, 8]. Importantly, except for ASA-PS and clinical frailty scale grading, all variables included in PQIP-CR are objective and reproducible and the model does not require adjustment for the quality-of-care patients receive intra- and postoperatively. The multicentre, prospective cohort we studied is the largest used for morbidity model development outside of the USA and our parsimonious model has potential to be generalisable to other populations [29]. We recommend that this model be used to prospectively monitor outcomes using methods such as variable life-adjusted display or exponentially-weighted moving average charts [12, 13] at the population level. Given the poor performance of POSSUM [7] and SORT morbidity [8] models in our cohort, we do not recommend their use in patients undergoing elective or expedited colorectal surgery. We encourage the peri-operative community to undertake further work to refine, improve and validate our model internationally in major colorectal surgical cohorts.

Our model provided greater discrimination than existing POSSUM and SORT morbidity models ($p < 0.001$). Discrimination of these models was lower in our contemporary cohort than previously published (Table 4). Differences in case-mix and surgical complexity between our cohort and those used to develop the SORT and POSSUM morbidity models may contribute to their poorer performance. The SORT morbidity model was derived from a single-centre cohort of patients undergoing generally less complex, predominantly orthopaedic procedures [8]. The POSSUM model was developed in a single-centre cohort of patients undergoing elective and emergency general surgery in 1988–1989. Its poor discrimination in our cohort may reflect changes in surgical practice over the past 30 years. Changes in the diagnosis, management and prognosis of respiratory and cardiac disease may result in POSSUM model variables no longer capturing the impact of the chronic disease on outcome after surgery.

The ACS-NSQIP risk calculator, developed from a large cohort of over 1.4 million patients, is one of the most

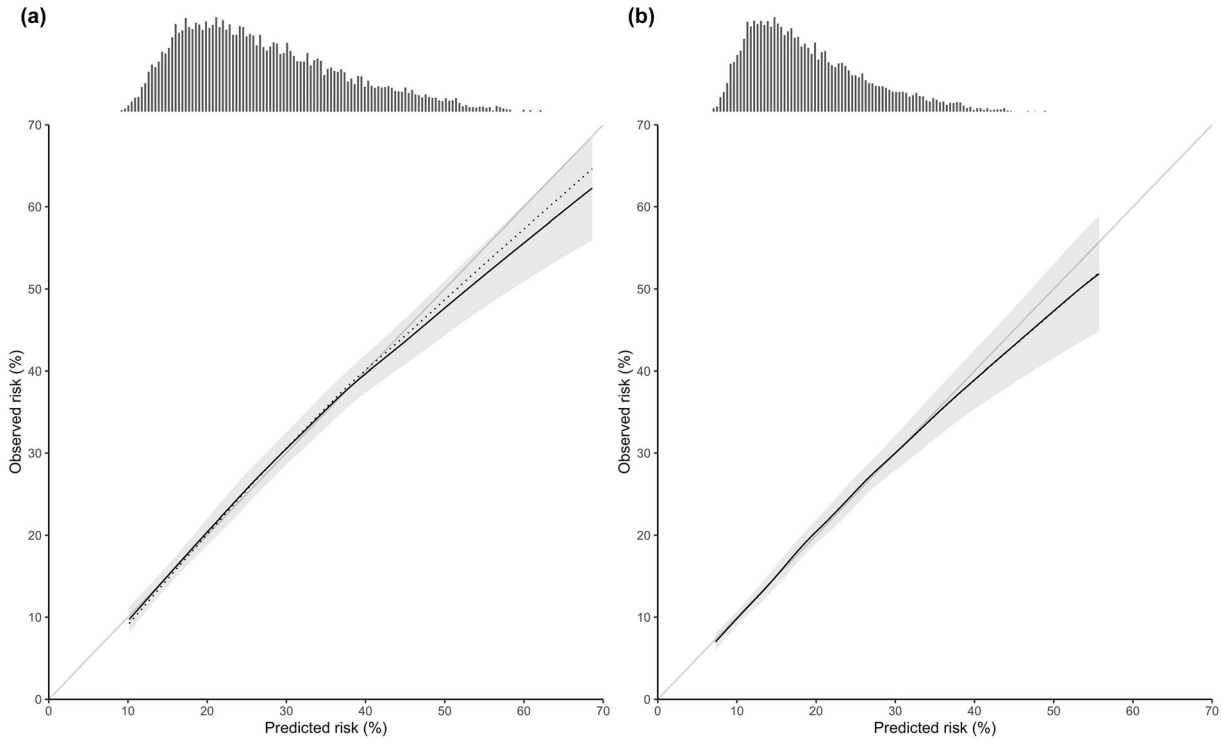

Figure 2 Bias corrected calibration curve comparing observed morbidity defined by (a) POMS and (b) POMS-major at postoperative day 7 against predicted risk for PQIP-CR model. Dotted black lines represent apparent calibration; Solid black lines represent bootstrap bias corrected calibration; Grey shaded area represents 95%CI for bootstrap bias corrected calibration curves. The histogram above each curve shows the distribution of PQIP-CR predicted risk.Notes: PQIP-CR, Peri-operative Quality Improvement Programme colorectal risk model; POMS, Postoperative Morbidity Survey; POMS-major; a subclassification of the Postoperative Morbidity Survey (see online Supporting Information, Table S1).

Table 4 Discrimination of the PQIP model compared with existing morbidity adjustment models.

Model	C-statistic reported in original study (95%CI)	Original study sample size (n)	C-statistic at validation in PQIP cohort (95%CI)
PQIP-CR	0.672 (0.660–0.684)	11,646	0.666 (0.654–0.678)
POSSUM [7]	Not stated	1372	0.591 (0.577–0.603)*
SORT morbidity [8]	0.72 (0.67–0.77)	1583	0.602 (0.590–0.614)#

PQIP-CR, PQIP-CR, Peri-operative Quality Improvement Programme colorectal risk model; POSSUM, physiological and operative severity score for the enumeration of mortality and morbidity; SORT morbidity, surgical outcome risk tool morbidity.
*POSSUM estimates were calculated for 9173/11,646 (78.8%) due to missing predictor variables required for POSSUM calculation.
#SORT morbidity estimates were calculated for 11,640/11,646 (99.9%) in the PQIP cohort due to missing age data.

frequently cited peri-operative risk models internationally [6]. Discrimination of the ACS-NSQIP colorectal risk calculator at development was higher than that of PQIP-CR when predicting risk of any postoperative morbidity (C-statistic 0.727 vs. 0.672 (95%CI 0.660–0.684), respectively). When predicting specific postoperative morbidity, the C-statistic of the ACS-NSQIP colorectal model ranged from 0.671 (risk of surgical site infection) to 0.844 (cardiac complications). We were unable to assess ACS-NSQIP model performance in our cohort as variables required for risk prediction were not available in our dataset. Twenty-one predictors are required for risk estimation using the calculator, nine more than those required by the PQIP-CR model. Importantly, ACS-NSQIP colorectal model coefficients have not been published and performance has not been validated in large cohorts outside the programme in the USA.

A major strength of our study was excellent data quality and completeness, with overall >95% completion of predictor variables and only 0.4% missing outcome data

causing case exclusion. The large, multicentre cohort included patients undergoing surgery in over 100 NHS hospitals across England and Wales. Two main limitations of our analysis are important to highlight. First, although sites are asked to recruit either all eligible patients or a random sample, patient consent is still required; this may introduce sampling bias. The large number of institutions contributing to the dataset should minimise this risk. In addition, the > 25% morbidity rate and median postoperative length of stay of 6 days (IQR 4–9 days) also suggest that we did not recruit a low-risk cohort. All OPCS procedure codes eligible for recruitment to the PQIP study (online Supporting Information, Appendix S2) were classified in a `restrictive´ category in a recent ecological study using hospital episode data, a category which was associated with a 90-day postoperative mortality of 2.8% [2]. Second, there was potential for measurement error in our POMS outcome. Our clinical dataset was collected by local clinical teams, commonly including research nurses, anaesthetists and surgical colleagues. Measurement error is likely to be higher for morbidity than mortality, potentially meaning that morbidity is harder to predict using statistical models. This hypothesis has previously been suggested to explain the generally lower discrimination of morbidity models compared with mortality models [6–9, 11]. These are undoubted generic challenges in developing models to predict or risk adjust for postoperative morbidity, but by using the POMS we have used an objective measure that has been formally validated and is widely deployed both in research and audit/quality improvement [16]. Model discrimination did not improve when we assessed the effect using a more homogenous outcome measure, gastrointestinal morbidity; suggesting the limited discrimination of PQIP-CR is not related to use of a composite measure. In our multilevel sensitivity analysis, the median odds ratio of 1.42 demonstrates moderate unexplained variation in outcome at the hospital level. Comparison of the median odds ratio with estimates for patient-level variables suggests that only severe frailty, surgical complexity and high ASA-PS grade (≥3) had stronger associations with the primary outcome than residual unexplained hospital level variation. Therefore, in our cohort, this residual unexplained hospital level variation appears to play an important role in patient outcome and may reduce the performance of our model (online Supporting Information, Appendix S7). Such variation may result from differences in the quality of care between centres in our cohort, or from between-hospital differences in aspects of case-mix that we did not measure and therefore did not adjust for in PQIP-CR. Hospital level structures and processes which may influence morbidity rates include the clinical expertise of the surgical and peri-operative team; the use of and adherence to enhanced recovery pathways; the availability of prehabilitation programmes; and the availability and utilisation of higher-level care areas such as high dependency and intensive care units postoperatively. Exploration of the association between health system factors and patient outcome may identify additional peri-operative process measures to improve model performance in the future.

We acknowledge a possible disadvantage of including socio-economic status in a risk-adjustment model, if in practice this results in an institutionalised acceptance that patients with lower status have worse outcomes. However, our work in patients undergoing emergency colorectal surgery has demonstrated that differences in clinical outcome by deprivation status are not obviously attributable to differences in standards of inpatient care [30]. These differences, therefore, are more likely to be due to the wider social determinants of health, on which there is a substantial literature [31]. We feel inclusion of index of multiple deprivation data as a predictor variable offers the potential for improved management of these patients.

Postoperative morbidity is associated with increased hospital length of stay and healthcare costs following emergency and elective colorectal surgery and is an important short-term postoperative outcome to risk-adjust and monitor [32]. Despite the broad nature of morbidity included in our composite POMS and POMS-major outcomes, the incorporation of the PQIP-CR model into a continuous monitoring tool, such as the variable life-adjusted display or exponentially-weighted moving average chart, will provide a high-level overview of risk-adjusted outcome after major colorectal surgery. Such methods have been incorporated into the Intensive Care National Audit and Research Centre (ICNARC) reporting system and the National Emergency Laparotomy Audit where exponentially-weighted moving average charts function as mortality monitoring tools that alert clinicians to poor outcome trends in their institutions earlier than data presented as annual standardised mortality ratios. This allows timely investigation of local data and intervention to improve quality of care and patient outcomes. The organ-specific criteria routinely recorded as part of the POMS data collection process allows clinicians to identify domain specific postoperative morbidity and implement timely, targeted quality improvement.

Incorporation of peri-operative outcome variables into electronic healthcare records provides the opportunity to create large-scale datasets that will support procedure specific risk model development. Progress in creating

specific, objective and well-defined standardised peri-operative outcome measures may further improve discrimination of morbidity risk-adjustment models [33]. Additional predictor variables that accurately capture hospital level structures and processes such as availability and utilisation of prehabilitation programmes or the routine use of enhanced recovery pathways may support improved risk prediction. The increased availability of peri-operative data through programmes such as PQIP and ACS-NSQIP allow more frequent updating of risk-adjustment models than was previously possible. As such, risk models should no longer be considered a fixed entity, remaining unchanged for decades. Instead, they should be dynamically updated and refined at regular intervals following the example of the ICNARC programme, ensuring risk estimates accurately inform clinical practice [34].

Acknowledgements

This study was supported by The Health Foundation as part of the Innovating for Improvement funding stream. Information on external PQIP data access requests and project support are available at: https://pqip.org.uk/pages/datarequest. SM, JB and SC received Health Foundation grant funding under the `Innovating for Improvement´ stream to support pilot development of a variable life-adjusted display morbidity monitoring tool. The peri-operative quality improvement programme is jointly funded by The Health Foundation and the Royal College of Anaesthetists and is delivered by the Health Services Research Centre, National Institute of Academic Anaesthesia, Royal College of Anaesthetists. No other competing interests declared.

References

1. Pearse RM, Clavien PA, Demartines N, et al. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. *British Journal of Anaesthesia* 2016; **117**: 601–9.
2. Abbott TEF, Fowler AJ, Dobbs TD, Harrison EM, Gillies MA, Pearse RM. Frequency of surgical treatment and related hospital procedures in the UK: a national ecological study using hospital episode statistics. *British Journal of Anaesthesia* 2017; **119**: 249–57.
3. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Translational Oncology* 2021; **14**: 101174.
4. Khuri SF, Henderson WG, DePalma RG, Mosca C, Healey NA, Kumbhani DJ. Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. *Annals of Surgery* 2005; **242**: 326–43.
5. Moonesinghe SR, Harris S, Mythen MG, Rowan KM, Haddad FS, Emberton M, Grocott MPW. Survival after postoperative morbidity: a longitudinal observational cohort study. *British Journal of Anaesthesia* 2014; **113**: 977–84.
6. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, Cohen ME. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *Journal of the American College of Surgeons* 2013; **217**: 833–842.e3.
7. Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *British Journal of Surgery* 1991; **78**: 355–60.
8. Wong DJN, Oliver CM, Moonesinghe SR. Predicting postoperative morbidity in adult elective surgical patients using the surgical outcome risk tool (SORT). *British Journal of Anaesthesia* 2017; **119**: 95–105.
9. 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.
10. Bagnall NM, Pring ET, Malietzis G, Athanasiou T, Faiz OD, Kennedy RH, Jenkins JT. Perioperative risk prediction in the era of enhanced recovery: a comparison of POSSUM, ACPGBI, and E-PASS scoring systems in major surgical procedures of the colorectal surgeon. *International Journal of Colorectal Disease* 2018; **33**: 1627–34.
11. Moonesinghe SR, Mythen MG, Das P, Rowan KM, Grocott MP. Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery: qualitative systematic review. *Anesthesiology* 2013; **119**: 959–81.
12. Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet* 1997; **350**: 1128–30.
13. Cook DA, Coory M, Webster RA. Exponentially weighted moving average charts to compare observed and expected values for monitoring risk-adjusted hospital indicators. *British Medical Journal Quality and Safety* 2011; **20**: 469–74.
14. Duckett SJ, Coory M, Sketcher-Baker K. Identifying variations in quality of care in Queensland hospitals. *Medical Journal of Australia* 2007; **187**: 571–5.
15. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Annals of Internal Medicine* 2015; **162**: 55–63.
16. Grocott MPW, Browne JP, Van der Meulen J, et al. The postoperative morbidity survey was validated and used to describe morbidity after major surgery. *Journal of Clinical Epidemiology* 2007; **60**: 919–28.
17. Healthcare Quality Improvement Partnership. *National Bowel Cancer Audit Annual Report 2019 v2.0.*, 2020.
18. Abel GA, Barclay ME, Payne RA. Adjusted indices of multiple deprivation to enable comparisons within and between constituent countries of the UK including an illustration using mortality rates. *British Medical Journal Open* 2016; **6**: 1–6.
19. GOV.UK. English indices of deprivation. 2019. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019 (accessed 31/08/2021).
20. GOV.WALES. Welsh Index of Multiple Deprivation (full Index update with ranks). 2019. https://gov.wales/welsh-index-multiple-deprivation-full-index-update-ranks-2019 (accessed 31/08/2021).
21. Office for National Statistics. Postcode to output area to lower layer super output area to middle layer super output area to local Authority District (August 2019) Lookup in the UK. 2019. https://geoportal.statistics.gov.uk/datasets/06938ffe68de49de98709b0c2ea7c21a/about (accessed 31/08/2021).
22. Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I, Mitnitski A. A global clinical measure of fitness and frailty in elderly people. *Canadian Medical Association Journal* 2005; **173**: 489–95.
23. Sandini M, Pinotti E, Persico I, Picone D, Bellelli G, Gianotti L. Systematic review and meta-analysis of frailty as a predictor of morbidity and mortality after major abdominal surgery. *British Journal of Surgery Open* 2017; **1**: 128–37.

24. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software* 2011; **45**: 1–67.
25. Riley RD, Snell KIE, Ensor J, Burke DL, Harrell Jr FE, Moons KGM, Collins GS. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Statistics in Medicine* 2019; **38**: 1276–96.
26. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. New York: Springer, 2001.
27. Dindo D, Demartines N, Clavien P-A. Classification of surgical complications. *Annals of Surgery* 2004; **240**: 205–13.
28. Merlo J, Chaix B, Ohlsson H, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health* 2006; **60**: 290–7.
29. Souwer ETD, Bastiaannet E, Steyerberg EW, Dekker JWT, van den Bos F, Portielje JEA. Risk prediction models for postoperative outcomes of colorectal cancer surgery in the older population - a systematic review. *Journal of Geriatric Oncology* 2020; **11**: 1217–28.
30. Poulton TE, Moonesinghe R, Raine R, Martin P. Socioeconomic deprivation and mortality after emergency laparotomy: an observational epidemiological study. *British Journal of Anaesthesia* 2020; **124**: 73–83.
31. Marmot M, Wilkinson R. *Social Determinants of Health*. 2nd edn. Oxford: Oxford University Press, 2005.
32. Ma C, Ma C, Teriaky A, et al. Morbidity and mortality after surgery for nonmalignant colorectal polyps: a 10-year nationwide analysis. *American Journal of Gastroenterology* 2019; **114**: 1802–10.
33. Boney O, Moonesinghe SR, Myles PS, et al. Core outcome measures for perioperative and Anaesthetic care (COMPAC): a modified Delphi process to develop a core outcome set for trials in perioperative care and anaesthesia. *British Journal of Anaesthesia* 2022; **128**: 174–85.
34. Harrison DA, Ferrando-Vivas P, Shahin J, Rowan KM. Ensuring comparisons of health-care providers are fair: development and validation of risk prediction models for critically ill patients. *Health Services and Delivery Research* 2015; **3**: 1–132.

Supporting Information

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

Figure S1. Data cleaning – removal of implausible values for continuous variables.

Figure S2. Non-linear relationship between morbidity and continuous variables included in the PQIP-CR model.

Figure S3. Receiver operating characteristic curves for the PQIP-CR model predicting (a) POMS defined morbidity and (b) POMS-major defined morbidity.

Table S1. Definition and incidence of morbidity defined by the Postoperative Morbidity Survey in our cohort.

Table S2. Thirty-one predictors considered for inclusion in the PQIP-CR model and their association with unadjusted postoperative morbidity and postoperative length of stay.

Table S3. Number and proportion of missing predictor variable data.

Table S4. Frequency each candidate variable was selected into each backwards step-wise model across the 5000 bootstrap samples. Variables selected into > 70% of bootstrap models were selected into the PQIP-CR model.

Table S5. PQIP-CR model performance at specified cut off thresholds of predicted POMS-defined morbidity.

Table S6. PQIP-CR model performance (after linear shrinkage) at specified cut off thresholds of predicted POMS-major defined morbidity.

Table S7. Definition and incidence of the Clavien–Dindo grading of surgical complications in our cohort.

Appendix S1. Additional authors and collaborators.

Appendix S2. Colorectal procedures and Office of Population Censuses and Surveys (OPCS) codes eligible for patient recruitment to PQIP at the time of model development.

Appendix S3. Variables used in multiple imputation process.

Appendix S4. Sensitivity analysis – Linear shrinkage factors for predicting risk of Clavien–Dindo defined complications using PQIP-CR model.

Appendix S5. Sensitivity analysis – Effect of missing data in Rockwood Clinical Frailty Scale variable on model fitting process.

Appendix S6. Sensitivity analysis – Gastrointestinal POMS-defined morbidity at postoperative day 7.

Appendix S7. Sensitivity analysis – multilevel model incorporating random intercept term for hospital site.

Appendix S8. R packages used in data management and analysis.

Appendix S9. Model equations for calculating risk of postoperative morbidity using PQIP-CR.

Appendix S10. Linear shrinkage factors for predicting the risk of POMS-major defined morbidity at postoperative day 7 using PQIP-CR model.

