Original Article

Predicting severe pain after major surgery: a secondary analysis of the Peri-operative Quality Improvement Programme (PQIP) dataset

R. A. Armstrong,[1,2] (iD) A. Fayaz,[3,4] G. L. P. Manning,[5] S. R. Moonesinghe,[6,7] (iD)
the Peri-operative Quality Improvement Programme (PQIP) delivery team and
C. M. Oliver,[6,8] for the PQIP collaborative

1 Academic Clinical Fellow, Department of Population Health Sciences, University of Bristol, Bristol, UK
2 Registrar, Department of Anaesthesia, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK
3 Consultant, 6 Honorary Consultant, Department of Anaesthesia and Peri-operative Medicine, University College London Hospital NHS Foundation Trust, London, UK
4 Honorary Associate Professor, 5 Registrar, Central London School of Anaesthesia, London, UK,
7 Professor, 8 Associate Professor, Centre for Peri-operative Medicine, Research Department for Targeted Intervention, Division of Surgery and Interventional Science, University College London, London, UK

Summary

Acute postoperative pain is common, distressing and associated with increased morbidity. Targeted interventions can prevent its development. We aimed to develop and internally validate a predictive tool to pre-emptively identify patients at risk of severe pain following major surgery. We analysed data from the UK Peri-operative Quality Improvement Programme to develop and validate a logistic regression model to predict severe pain on the first postoperative day using pre-operative variables. Secondary analyses included the use of peri-operative variables. Data from 17,079 patients undergoing major surgery were included. Severe pain was reported by 3140 (18.4%) patients; this was more prevalent in females, patients with cancer or insulin-dependent diabetes, current smokers and in those taking baseline opioids. Our final model included 25 pre-operative predictors with an optimism-corrected c-statistic of 0.66 and good calibration (mean absolute error 0.005, p = 0.35). Decision-curve analysis suggested an optimal cut-off value of 20–30% predicted risk to identify high-risk individuals. Potentially modifiable risk factors included smoking status and patient-reported measures of psychological well-being. Non-modifiable factors included demographic and surgical factors. Discrimination was improved by the addition of intra-operative variables (likelihood ratio χ^2 496.5, p < 0.001) but not by the addition of baseline opioid data. On internal validation, our pre-operative prediction model was well calibrated but discrimination was moderate. Performance was improved with the inclusion of peri-operative covariates suggesting pre-operative variables alone are not sufficient to adequately predict postoperative pain.

Introduction

Acute postoperative pain is common, with up to 47.2% of patients reporting severe pain within the first 24 h of surgery [1, 2]. It is associated with a negative patient experience and may also be associated with respiratory and/or cardiac complications, prolonged hospital stay, limited or delayed return to normal activity and the development of chronic postsurgical pain [2–4].

Identification of the patient at increased risk of problematic postoperative pain is challenging, with potential risk factors spanning patient, anaesthetic and surgical domains. Patient factors may include non-modifiable variables such as age and sex, as well as potentially modifiable characteristics such as anxiety, psychological distress and high levels of catastrophisation [5, 6]. Additionally differences in pain perception, pre-operative pain, pre-operative opioid usage and the presence of chronic pain have all been associated with problematic acute postoperative pain [5–7]. Anaesthetic and surgical factors may include the choice of analgesic regimen, inclusion of regional blocks, type and duration of surgery and the choice of surgical incision [5, 6].

We hypothesised that postoperative morbidity that is related to acute, or acute-on-chronic pain, could be mitigated by better identification and pre-emptive holistic management of `at-risk´ patients. The first part of this work is to attempt to identify patients at risk, ideally pre-operatively. Identification of patients at high risk of developing postoperative severe pain pre-operatively allows for better pre-operative decision-making, for counselling and for the introduction of evidence-based interventions. From a study perspective, it also has the advantage of preceding any anaesthetic or surgical intervention which might introduce unmeasured confounding. However, to determine the relative importance of predictors of severe pain, the effects of healthcare processes and peri-operative events should also be taken into account.

We undertook a secondary analysis of the Peri-operative Quality Improvement Programme (PQIP, www.pqip.org.uk) dataset examining potential risk factors for postoperative pain. We focused principally on pre-operative variables with additional consideration of intra-operative factors to ascertain their relative contributions to acute postoperative pain. Our aim was to develop and internally validate a model to predict risk for patients undergoing major surgery.

Methods

The Peri-operative Quality Improvement Programme is a prospective, multicentre, observational cohort study established in 2016 which collects data on adult patients (aged ≥ 18 y on date of surgery) undergoing major, planned non-cardiac surgery in UK National Health Service (NHS) hospitals [8, Moonesinghe et al., preprint, https://www.researchsquare.com/article/rs-708161/v1]. Case-mix, process and outcome data are collected on site and submitted electronically into the web-based study database. The full eligibility criteria and data specification are described elsewhere (https://pqip.org.uk). We performed a secondary analysis of anonymised patient-level data for procedures from December 2016 to June 2020 (cohort start and end dates determined by the study start date, and the date of data extraction). The Peri-operative Quality Improvement Programme has ethical approval from the Health Research Authority.

The Peri-operative Quality Improvement Programme contains several postoperative pain ratings. We selected severe pain at the site of surgery on day 1 after surgery as the primary outcome. This was measured by the Bauer Patient Satisfaction Score which asks patients to respond across 10 anaesthesia-related discomfort domains to the question: `At any stage after your operation have you had the following?´, with possible responses of `No´, `Yes, mild´, `Yes, moderate´ and `Yes, severe´. We categorised patient responses as a binary outcome: `Yes, severe´ vs. any of `No´, `Yes, mild´ or `Yes, moderate´ to reflect our clinical focus of identifying problematic postoperative pain.

The alternative postoperative pain measures recorded in PQIP include pain in recovery and pain on day 3 after surgery. The postoperative day 1 time-point was chosen in line with recommendations from the standardised endpoints in peri-operative medicine (StEP) initiative [9]. Pain in recovery reflects a different clinical context to later postoperative pain and is influenced by residual anaesthetic and analgesic drugs. Measures on postoperative day 3 may be biased by other factors such as complications or patients having been discharged and thus lost to follow-up.

Candidate explanatory variables were selected from those included in the PQIP dataset based on clinical rationale and previously published case series [2, 3, 5, 6, 10–12]. Given the primary focus on using only pre-operative data, we classified these variables according to whether they are available pre-operatively or only postoperatively (i.e. intra-operative and postoperative data). These are shown in Table 1. Primary model development used only pre-operative variables. Variables that were missing in > 10% of patients were omitted [13]. This included baseline opioid usage due to changes in the dataset over time. Given the potential clinical importance of this factor, we performed a secondary analysis in the subgroup of patients

Table 1 Peri-operative Quality Improvement Programme variables.

Pre-operative variables	Intra-/postoperative variables
Patient characteristics	**Type of anaesthesia**
Age	General
Sex	Spinal
ASA physical status	Epidural
Smoking history	Combined spinal and epidural
Current alcohol consumption	Regional block
Current occupation	Wound catheter**
Laboratory investigations	Local anaesthetic infiltration only
Serum sodium	Oral gabapentinoids**
Serum potassium	Intravenous paracetamol**
Serum urea	Intravenous NSAIDs**
Serum creatinine	Intravenous opioids**
White cell count	Intravenous ketamine**
Haemoglobin	Intravenous dexmedetomidine**
Past medical history	Intravenous lignocaine**
Respiratory history	General with TIVA**
Cerebrovascular disease	General with inhalational**
Cancer diagnosis	Intravenous analgesia
Dementia	Inhalational – desflurane**
Diabetes	Inhalational – isoflurane**
Opioid usage**	Inhalational – sevoflurane**
Baseline (pre-operative) PROMs	Inhalational – other**
Over the past 2 weeks has pain interfered with day-to-day activities	Inhalational – nitrous oxide**
Over the past 2 weeks have you felt worried or low because of pain	Intravenous propofol infusion**
QoR-15: Feeling of general wellbeing	Intravenous remifentanil infusion**
QoR-15: Moderate pain	General anaesthesia – other**
QoR-15: Severe pain	**Surgical details**
QoR-15: Feeling worried or anxious	Incision – thoracic
QoR-15: Feeling sad or depressed	Incision – upper abdominal
EQ5D: Usual activities	Incision – lower abdominal
EQ5D: Pain/discomfort	Incision – other
EQ5D: Anxiety/depression	Degree of peritoneal soiling
WHODAS 2.0: Past 30 days how many days totally unable to carry out usual activities because of health	Duration of surgery
Planned surgical procedure	**Recovery details**
Surgical speciality	Core temperature > 36° on arrival
Urgency of surgery	Abdominal drain present
Grading of surgery	Nasogastric present
Planned postoperative destination	Highest pain sore (severe vs none/mild/moderate)
Mode of surgery: open	**Day 1 postoperatively**
Mode of surgery: laparoscopic	Drinking
Mode of surgery: robotic-assisted	Eating
Mode of surgery: thoracoscopic	Begun to mobilise out of bed
How many operations in past 30 days	

Those marked with ** were only available for a subset of the inclusion period.
NSAIDs, non-steroidal anti-inflammatory drugs; TIVA, total intravenous anaesthesia; PROMs, patient-reported outcome measures; QoR-15, Quality of Recovery score; EQ5D, EuroQol 5-dimension; WHODAS 2.0, World Health Organization Disability Assessment Schedule.

in whom it was included. Secondary models were also developed with additional peri-operative (intra- and up to 24 h postoperative) variables. For the remaining variables, we performed complete-case analysis, that is, patients with missing values were removed. Full details on missing data are in online Supporting Information Tables S1 and S2.

Most laboratory test values (urea, creatinine, sodium, potassium, haemoglobin and white cell count) had extreme values at one or both ends of their distribution.

Consequently, the distributions were Winsorised at the 1st centile, 99th centile or both (resulting limits are shown in online Supporting Information Table S3).

For some continuous risk factors, the relationship with the primary outcome was linear. For most other variables it was possible to capture the non-linear relationship using restricted cubic splines (5 knots placed at 5th, 27.5th, 50th, 72.5th and 95th centiles) [14]. The WHO Disability Assessment Schedule 2.0 measure `days unable to carry out

usual activities´ was extremely skewed so values were log-transformed and the resulting relationship captured using restricted cubic splines with three knots (at 10th, 50th and 90th centiles). These non-linear relationships are summarised in online Supporting Information Figure S1. Categorical variables were regrouped (levels collapsed) where classes contained few individuals or events, as reported in the Results section.

As this was a secondary data analysis, sample size was determined by the eligible PQIP cohort. Adequacy of sample sizes and numbers of events were assessed according to recommendations for the development of prediction models for binary outcome data [15] (see online Supporting Information Appendix S1 for calculation).

A logistic regression model was trained on the entire sample using backward step-down selection. This method starts with a model fitted with all available predictors and then eliminates predictors in a stepwise fashion. The initial model and subsequent iterations are compared using the Akaike information criterion to determine the `best´ model. The Akaike information criterion balances the performance of the model (how well it fits the data) against its complexity (how many predictors it contains), penalising models with greater numbers of predictors.

For the primary analysis, pre-operative variables were limited to those which were present throughout the inclusion period (i.e. not including baseline opioid usage). Secondary analyses included: the subset of patients with baseline opioid data; surgical speciality subgroups; additional intra- and postoperative variables.

The output of a logistic regression model is the probability (log odds) of the outcome occurring for a given individual, based on the values of the predictors included in the model. Model performance was assessed in terms of discrimination using the c-statistic (equivalent to the area under the receiver operating characteristic curve in binary logistic regression). Discrimination refers to the ability of a model to distinguish those with and without the outcome, with the c-statistic representing the probability that a randomly selected subject experiencing a positive outcome will have a higher predicted probability of that outcome occurring than a randomly selected subject who did not experience a positive outcome. A model with a c-statistic of 0.5 is thus no better than chance, whilst 1.0 represents perfect performance; values of 0.5–0.7 are considered poor discrimination, ≥ 0.7 acceptable, ≥ 0.8 excellent and ≥ 0.9 outstanding [16].

Calibration (how well the model predictions fit the underlying data, across the full range of predicted values) was assessed visually by plotting the predicted values against observed values across the range of predictions, by calculation of mean prediction error and by the Hosmer-Lemeshow goodness-of-fit test (with a larger p value suggesting no significant difference between the predictions and the data, i.e. good fit).

Internal validation of discrimination ability and calibration was performed using bootstrap resampling to provide optimism-corrected results. This procedure aims to account for overfitting to the data used to develop the model (optimism) and give an unbiased estimate of how the model will perform on new, previously unseen data. In bootstrapping, a sample of the same size as the original sample is drawn, with replacement and the model is derived from that new bootstrap sample. This model is then applied to the original sample, and the difference in accuracy in the two samples (bootstrap and original) provides an estimate of the optimism (overfitting). This process is then repeated multiple times (we used 300 repeats) and the results averaged to obtain a final estimate of optimism. This value is then subtracted from the apparent accuracy of the initial model to get the optimism-corrected estimate [14].

Where the performance of models was compared, calibration was assessed as above. If both showed good calibration, discrimination (by c-statistic) was compared as were Brier scores (squared differences between actual binary outcomes and predictions). For nested models (i.e. where one model includes a subset of the predictors in another) the likelihood ratio test was used. This compares the goodness-of-fit of the models with the data and provides a chi-squared statistic (χ^2) and p value based on the null hypothesis that the models are the same fit to the data.

Decision-curve analysis was used to describe and compare the clinical implications of using each model at different thresholds. In decision-curve analysis, a model is considered to have clinical value if it has the highest net benefit across the whole range of thresholds for which a patient would be labelled as `high risk´ [17]. We also calculated the optimal threshold according to the Youden index, a means of summarising the receiver operating characteristic curve by giving equal weight to sensitivity and specificity [18].

All analyses were performed using R version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria) with the following packages loaded: cutpointr; dcurves; pmsampsize; readr; ResourceSelection; rms; tableone; tidyverse. The analysis code is available online [19]. Our findings are reported in accordance with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement [20].

Results

The dataset contained records for 27,843 patients, of which 20,125 (72.3%) included the primary outcome of severe pain on postoperative day 1. Reported reasons for missing outcome data included: 1272 (16.5%) declined; 301 (3.9%) already discharged; 1647 (21.3%) drowsy or asleep; nine (0.1%) language barrier; eight (0.1%) patients not available; and seven (0.1%) had died. The remaining were either `other´ (2549, 33.0%) or no reason given (1925, 24.9%) (online Supporting Information Table S4 shows comparison of patients with/without outcome data). To assess whether a lack of weekend data collection contributed to these cases, day of the week of surgery was examined; 16.7% of those with no reason and 30.5% of `other´ cases were performed on Fridays. A total of 3046 patients were removed due to missing data in other variables (full details in online Supporting Information Table S2). After the removal of patients with missing data, 17,079 were available for analysis based on pre-operative data (Fig. 1).

Patient characteristics and clinical features are shown in Table 2, stratified by the presence or absence of the primary outcome (with additional features in online Supporting Information Table S5).

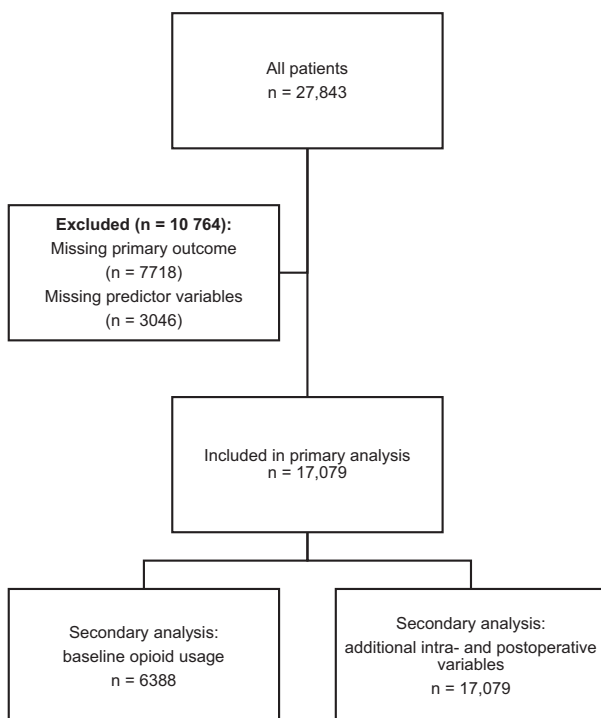The primary analysis led to the development of a model based on the pre-operative variables present throughout the inclusion period. The final model included 25 of 37 potential variables with those included and removed through backward selection as shown in Table 3. The full model specification is detailed in Table 4. For continuous predictors modelled through restricted cubic splines, the odds ratios cannot be interpreted individually. The relationship between these variables and the log odds of the outcome are shown in online Supporting Information Figure S2.

The c-statistic for apparent performance on the entire dataset was 0.68 (95%CI 0.67–0.69). After internal validation, the optimism-corrected performance was 0.66, suggesting a low degree of overfitting, and it was well calibrated across the range of predicted values (mean absolute error 0.005; $\chi^2 = 6.68$, p = 0.35) though tended to overpredict risk at the higher end of the range (Fig. 2a). Decision curve analysis revealed that a threshold of 50% predicted probability of severe pain is similar to treating nobody, and a threshold of 10% or lower is equivalent to treating everybody. The potential for net benefit with the model appears maximal with a threshold between 20% and 30% (Fig. 2b). The optimal cut-point according to the Youden index was slightly lower at 18%.

As described above, a subset of patients from 2019 onwards (n = 6388) had data on pre-operative opioid usage. The incidence of severe pain in this group was similar to the overall cohort (1149, 18.0%) but differed according to opioid status: 195/746 (26.1%) in those taking opioids vs. 954/5642 (16.9%) in those not taking opioids (p < 0.001).

A logistic regression model was developed on this subset of patients. However, the opioid variable was not included after backward selection suggesting it did not add to the predictive ability of the model (full details of included and deleted variables are in online Supporting Information Table S6). Whilst the apparent performance of this model was higher with a c-statistic of 0.69, internal validation revealed a higher degree of overfitting/optimism (as might be expected with the smaller sample size). Optimism-corrected performance was no better than the original model (c-statistic 0.66) and calibration was worse (mean absolute error 0.011; χ^2 8.23, p = 0.41).

The largest single surgical speciality was colorectal surgery with 63.1% of all patients having some type of abdominal surgery (Table 2). The predictive ability of the primary model according to surgical speciality was assessed using resampling to give optimism-corrected performance. Performance was broadly similar across specialities. The best discrimination was found in gynaecology patients, but calibration was poor (online Supporting Information Table S7).

Figure 1 Flow diagram of patients included in primary and secondary analyses.

Table 2 Patient characteristics stratified by presence or absence of primary outcome. Values are median (IQR [range]) or number (proportion).

	Overall n = 17,079	None/mild/moderate pain n = 13,939	Severe pain n = 3140
Age	66.1 (56.2–73.2 [18–95.6])	66.8 (57.1–73.5 [18–95.6])	62.9 (52.2–71.5 [18–93.8])
Female	7091 (41.5%)	5592 (40.1%)	1499 (47.7%)
ASA physical status			
1	1900 (11.1%)	1525 (10.9%)	375 (11.9%)
2	10,504 (61.5%)	8635 (61.9%)	1869 (59.5%)
3	4495 (26.3%)	3634 (26.1%)	861 (27.4%)
4 or 5	180 (1.1%)	145 (1.0%)	35 (1.1%)
Cancer diagnosis	5039 (29.5%)	4309 (30.9%)	730 (23.2%)
Patient has dementia	109 (0.6%)	89 (0.6%)	20 (0.6%)
Diabetes			
No	14,918 (87.3%)	12,244 (87.8%)	2674 (85.2%)
Insulin-dependent	526 (3.1%)	400 (2.9%)	126 (4.0%)
Non-insulin-dependent	1635 (9.6%)	1295 (9.3%)	340 (10.8%)
Smoking history			
Never smoked or no known	8891 (52.1%)	7365 (52.8%)	1526 (48.6%)
Current smoker	1692 (9.9%)	1282 (9.2%)	410 (13.1%)
Ex-smoker	6496 (38.0%)	5292 (38.0%)	1204 (38.3%)
Current alcohol consumption			
No alcohol	7216 (42.3%)	5729 (41.1%)	1487 (47.4%)
0–2 units/day	7372 (43.2%)	6112 (43.8%)	1260 (40.1%)
>2 units/day	2491 (14.6%)	2098 (15.1%)	393 (12.5%)
Over the past 2 weeks has pain been bad enough to interfere with your day-to-day activities? = Yes	4037 (23.6%)	2875 (20.6%)	1162 (37.0%)
Over the past 2 weeks have you felt worried or low because of pain? = Yes	3816 (22.3%)	2720 (19.5%)	1096 (34.9%)
Pre-operative QoR-15: having a feeling of general well-being	9 (6–10 [0–10])	9 (7–10 [0–10])	8 (5–10 [0–10])
Pre-operative QOR-15: moderate pain in the last 24 h	10 (6–10 [0–10])	10 (7–10 [0–10])	8 (4–10 [0–10])
Pre-operative QOR-15: severe pain in the last 24 h	10 (10–10 [0–10])	10 (10–10 [0–10])	10 (7–10 [0–10])
Pre-operative QOR-15: feeling worried or anxious	7 (5–10 [0–10])	7 (5–10 [0–10])	6 (3–9 [0–10])
Pre-operative QOR-15: feeling sad or depressed	10 (6–10 [0–10])	10 (7–10 [0–10])	9 (5–10 [0–10])
EQ5D: Usual activities			
No problems	11,559 (67.7%)	9812 (70.4%)	1747 (55.6%)
Slight problems	2670 (15.6%)	2087 (15.0%)	583 (18.6%)
Moderate problems	1808 (10.6%)	1316 (9.4%)	492 (15.7%)
Severe problems or unable	1042 (6.1%)	724 (5.2%)	318 (10.1%)
EQ5D pain/discomfort			
None	8977 (52.6%)	7774 (55.8%)	1203 (38.3%)
Slight	4098 (24.0%)	3342 (24.0%)	756 (24.1%)
Moderate	2839 (16.6%)	2056 (14.7%)	783 (24.9%)
Severe or extreme	1165 (6.8%)	767 (5.5%)	398 (12.7%)

(continued)

Table 2 (continued)

	Overall n = 17,079	None/mild/moderate pain n = 13,939	Severe pain n = 3140
EQ5D anxiety/depression			
None	6838 (40.0%)	5822 (41.8%)	1016 (32.4%)
Slight	6070 (35.5%)	5002 (35.9%)	1068 (34.0%)
Moderate	3141 (18.4%)	2408 (17.3%)	733 (23.3%)
Severe or extreme	1030 (6.0%)	707 (5.1%)	323 (10.3%)
WHODAS 2.0: in the past 30 days, for how many days were you totally unable to carry out your usual activities or work because of any health condition?	0 (0–0 [0–3.4])	0 (0–0 [0–3.4])	0 (0–2.1 [0–3.4])
Surgical speciality			
Abdominal – lower gastrointestinal	7810 (45.7%)	6425 (46.1%)	1385 (44.1%)
Abdominal – hepatobiliary	1349 (7.9%)	1110 (8.0%)	239 (7.6%)
Abdominal – other	516 (3.0%)	409 (2.9%)	107 (3.4%)
Abdominal – upper gastrointestinal	1096 (6.4%)	904 (6.5%)	192 (6.1%)
Burns and plastics	188 (1.1%)	175 (1.3%)	13 (0.4%)
Gynaecology	297 (1.7%)	265 (1.9%)	32 (1.0%)
Head and neck	282 (1.7%)	241 (1.7%)	41 (1.3%)
Orthopaedic	931 (5.5%)	700 (5.0%)	231 (7.4%)
Spinal	531 (3.1%)	408 (2.9%)	123 (3.9%)
Thoracic	1287 (7.5%)	998 (7.2%)	289 (9.2%)
Urology	2667 (15.6%)	2206 (15.8%)	461 (14.7%)
Vascular	125 (0.7%)	98 (0.7%)	27 (0.9%)
Open surgery	7780 (45.6%)	6276 (45.0%)	1504 (47.9%)
Laparoscopic surgery	8202 (48.0%)	6773 (48.6%)	1429 (45.5%)
Robotic-assisted	1487 (8.7%)	1301 (9.3%)	186 (5.9%)
Thoracoscopic surgery	751 (4.4%)	599 (4.3%)	152 (4.8%)
Baseline opioid usage	746 (11.7%)	551 (10.5%)	195 (17.0%)

QoR-15, Quality of Recovery score; EQ5D, EuroQol 5-dimension; WHODAS 2.0, World Health Organisation Disability Assessment Schedule.

The patients included in the primary analysis all had additional peri-operative (intra-operative and recovery) and postoperative (24 h) data (Table 1). A logistic regression model was developed with these additional predictors available for selection. The final model after backward selection included the same pre-operative variables as the primary model with the addition of several peri- and postoperative factors including both anaesthetic and surgical details (online Supporting Information Table S8). The full model specification (online Supporting Information Table S9) demonstrates that the additional predictors of increased risk were severe pain in recovery (OR (95%CI) 1.97 [1.60–2.41], p < 0.001) and the presence of a thoracic incision (OR 1.39 [1.10–1.74], p < 0.01). Additional predictors which reduced the risk of the primary outcome included: types of anaesthesia (presence compared with reference of absence: general OR 0.56 [0.45–0.70], p < 0.001; epidural OR 0.74 [0.66–0.84], p < 0.001; spinal OR 0.86 [0.77–0.95], p < 0.01; factors in recovery (absence of severe pain OR 0.68 [0.57–0.82], p < 0.001; presence of a nasogastric tube (OR 0.75 [0.65–0.87], p < 0.001); and factors at 24 h (mobilising OR 0.71 [0.64–0.79], p < 0.001; eating OR 0.76 [0.68–0.85], p < 0.001).

Comparing performance between this and the original, pre-operative only model, calibration was similar across the range of values (χ^2 11.16, p = 0.19, mean absolute error 0.006; online Supporting Information Figure S3a) and optimism-corrected discrimination performance improved with a c-statistic of 0.7, accompanied by a reduction in the Brier score (0.140 to 0.135). Comparing performance with the likelihood ratio test for nested models confirmed improved performance (LR χ^2 statistic 496.5, p < 0.001).

Table 3 Variables included and deleted through backward selection in primary model development.

Included variables	Deleted variables
Age	Serum sodium
Sex	Serum creatinine
Surgical speciality	Serum urea
Urgency of surgery	Current occupation
White cell count	EQ5D: usual activities
Respiratory history	ASA physical status
History of cerebrovascular disease	Dementia
Cancer diagnosis	Core question: over the past 2 weeks have you felt worried or low because of pain?
Diabetes	
Smoking history	≥2 operations in past 30 days
Current alcohol consumption	Pre-operative QoR-15: feeling sad or depressed
Planned postoperative destination	Serum potassium
Mode of surgery: open	Haemoglobin
Mode of surgery: laparoscopic	
Mode of surgery: robotic	
Mode of surgery: thoracoscopic	
Grade of surgery	
Core question: over the past 2 weeks has pain interfered with day-to-day activities?	
Pre-operative QoR-15: feeling of general well-being	
Pre-operative QoR-15: moderate pain	
Pre-operative QoR-15: severe pain	
Pre-operative QoR-15: feeling worried or anxious	
EQ5D: pain/discomfort	
EQ5D: anxiety/depression	
WHODAS 2.0: in the past 30 days, for how many days were you totally unable to carry out usual activities because of health?	

QoR-15, Quality of Recovery score; EQ5D, EuroQol 5-dimension; WHODAS 2.0, World Health Organisation Disability Assessment Schedule.

Decision-curve analysis showed increased net benefit across the range of probabilities (online Supporting Information Figure S3b), but this does not account for the difference in time-points at which decisions would be made.

Discussion

In this secondary analysis of data from the PQIP database, we have developed and internally validated a prediction model for severe pain on postoperative day 1 after major, non-cardiac surgery which utilises only pre-operative patient data. Model performance was limited, but several potential contributory factors were identified. This is the first attempt to systematically develop a peri-operative pain prediction model using such a large, high-quality dataset in a mixed surgical population.

Severe pain occurred in 18.4% of patients. Whilst in keeping with the ranges reported in previous case series [1, 3] this represents a significant minority of patients, with potential implications for patient outcomes given the associations between postoperative pain and morbidity [2]. We focused on pre-operative variables to increase the clinical utility of our model [10] with the ultimate aim that prediction would enable clinicians to take preventative

rather than reactive approaches. Many of the key predictors were non-modifiable patient and surgical aspects, for example female sex and younger age, both of which have been reported in previous studies [11, 12, 21]; a past medical history of diabetes; and thoracic surgery. A potentially modifiable lifestyle factor identified was current smoking status, which has previously been cited [12]. We did not replicate the previously reported finding of baseline opioid usage as a risk factor [11, 12].

We found baseline patient-reported outcome measures to be relatively strong predictors, particularly those around psychological symptoms of anxiety/ depression but also reported pain/discomfort. Anxiety and depressive symptoms, as well as higher levels of pre-operative pain, have previously been recognised as contributors to the risk of problematic postoperative pain [11, 12, 21]. Alongside these are complex factors we were not able to assess, such as pain catastrophising, feelings of helplessness and higher than expected pain [11, 21]. The importance of functional domains has also previously been highlighted [22]. Whilst our model did include a WHO Disability Assessment Schedule 2.0 measure of ability to undertake normal activities, the EQ5D domain of `usual

Table 4 Full specification of primary model after backward selection. Continuous variables modelled with restricted cubic splines have multiple odds ratios (OR), with each component marked (' to '''); reference categories for categorical variables shown.

	OR	95%CI lower	95%CI upper	p value
Intercept	0.64	0.27	1.52	0.314
Age	0.99	0.98	1.00	0.003
Age'	1.00	0.97	1.03	0.910
Age''	0.95	0.68	1.33	0.779
Age'''	1.15	0.47	2.85	0.766
Sex (reference = male)	-			
Female	1.16	1.06	1.27	0.001
Speciality (reference = abdominal – lower gastrointestinal)	-			
Abdominal – hepatobiliary	1.11	0.93	1.33	0.241
Abdominal – other	0.86	0.68	1.09	0.212
Abdominal – upper gastrointestinal	1.12	0.93	1.35	0.220
Burns and plastics	0.28	0.16	0.50	0.001
Gynaecology	0.40	0.27	0.59	0.001
Head and neck	0.65	0.45	0.92	0.014
Orthopaedic	0.82	0.67	1.00	0.046
Spinal	0.63	0.49	0.80	0.001
Thoracic	1.33	1.06	1.67	0.013
Urology	1.10	0.95	1.28	0.203
Vascular	0.91	0.57	1.43	0.671
Urgency (reference = elective)	-			
Expedited	0.84	0.73	0.97	0.019
White cell count	1.08	0.95	1.22	0.249
White cell count'	0.60	0.21	1.75	0.351
White cell count''	4.34	0.06	322.25	0.501
White cell count'''	0.32	0.00	45.45	0.651
Respiratory history (reference = None)	-			
Dyspnoea limiting exertion or at rest	1.23	0.98	1.54	0.073
Dyspnoea on exertion	1.14	1.00	1.29	0.050
Cerebrovascular disease	1.19	0.97	1.46	0.092
Cancer diagnosis	0.85	0.77	0.94	0.002
Diabetes (reference = None)	-			
Insulin-dependent	1.43	1.15	1.77	0.001
Non-insulin-dependent	1.33	1.17	1.52	0.001
Smoking history (reference = never smoked or not known)	-			
Current smoker	1.18	1.03	1.35	0.016
Ex-smoker	1.15	1.05	1.26	0.002
Alcohol consumption (reference = none)	-			
0–2 units/day	0.91	0.84	1.00	0.047
>2 units/day	0.88	0.77	1.00	0.051
Planned postoperative destination (reference = ward care)	-			
Level 1	0.82	0.72	0.94	0.004
Level 2 or level 1.5 (enhanced care)	0.80	0.72	0.88	0.001
Level 3	0.95	0.77	1.18	0.643

Table 4 (continued)

	OR	95%CI lower	95%CI upper	p value
Open surgery	0.73	0.59	0.91	0.005
Laparoscopic surgery	0.68	0.55	0.84	0.001
Robotic-assisted	0.63	0.49	0.81	0.001
Thoracoscopic surgery	0.51	0.36	0.70	0.001
Grade of surgery (reference = major)	-			
Xmajor	0.84	0.71	1.00	0.044
Complex	0.72	0.61	0.85	0.001
Over the past 2 weeks has pain been bad enough to interfere with your day-to-day activities? = Yes	1.13	0.99	1.28	0.062
Pre-operative QOR-15: having a feeling of general well-being	1.07	1.02	1.12	0.006
Pre-operative QOR-15: having a feeling of general well-being'	0.90	0.83	0.97	0.009
Pre-operative QOR-15: having a feeling of general well-being''	2.13	0.82	5.54	0.120
Pre-operative QOR-15: moderate pain	0.98	0.96	1.00	0.028
Pre-operative QOR-15: severe pain	0.98	0.96	1.00	0.035
EQ5D: Pain/discomfort (reference = none)	-			
Slight problems	1.20	1.07	1.34	0.001
Moderate problems	1.06	0.87	1.29	0.573
Severe problems or unable	1.00	0.72	1.38	0.992
WHODAS 2.0: in the past 30 days, for how many days were you totally unable to carry out your usual activities or work because of any health condition?	1.15	1.06	1.26	0.001
WHODAS 2.0: in the past 30 days, for how many days were you totally unable to carry out your usual activities or work because of any health condition?´	0.87	0.75	1.00	0.057
Pre-operative QOR-15: feeling worried or anxious	0.97	0.96	0.99	0.001
EQ5D: Anxiety/depression (reference = none)	-			
Slight	0.99	0.89	1.10	0.795
Moderate	1.13	0.94	1.34	0.189
Severe or extreme	1.35	1.01	1.80	0.041

QoR-15, Quality of Recovery score; EQ5D, EuroQol 5-dimension; WHODAS 2.0, World Health Organisation Disability Assessment Schedule.

activities´ was not included. These patient-reported symptoms represent potential areas of modifiable benefit if a holistic approach is taken to peri-operative assessment and pre-optimisation before major surgery.

The factors most strongly associated with reduced risk of pain were all surgical factors: speciality, mode and grade of surgery. Whilst these speciality-specific factors might not be directly modifiable by the anaesthetist, they could contribute to a shared decision-making approach with both surgeons and patients, in which their potential impact on postoperative course could be explored [10].

Whilst predictions based purely on pre-operative data might have optimal utility in clinical practice [10], the discriminatory performance of the model was limited, though a c-statistic of 0.66 is similar to other reported predictive models [3, 11]. One of the largest existing analyses is a model developed using data from approximately 50,000 patients in the international PAIN OUT registry [11]. However, this model included patients from 2011 to 2015, and the express aim was to produce a simple scoring tool, patient reported outcome measures were only recorded postoperatively rather than at baseline. Additionally, the most important predictor for severe postoperative pain found was the country in which the surgery took place [11], our analysis has the advantage of including patients from only one healthcare system.

Taken together, our primary model and existing analyses suggest that pre-operative data alone are not sufficient to accurately predict which patients will go on to experience problematic acute postoperative pain. We
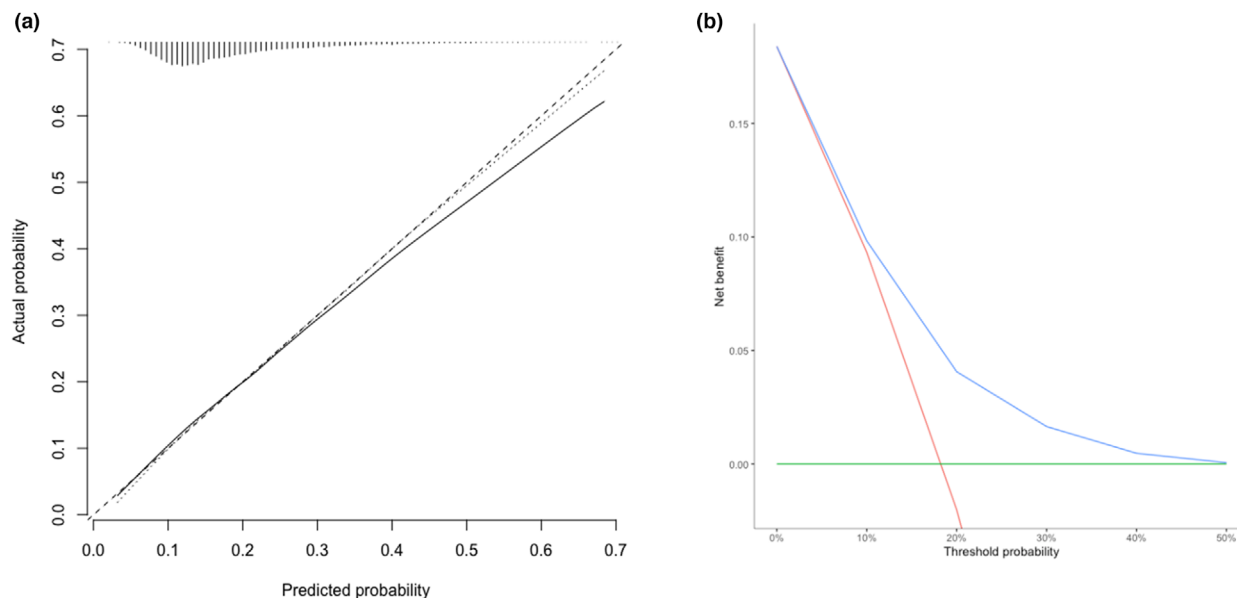
Figure 2 (a) Calibration plot showing apparent (. . . .), bias-corrected (—) and ideal (- - -) performance and (b) decision-curve analysis showing net benefit of treating all patients (–), no patients (–) or based on primary model optimism-corrected predicted probability (–).

explored whether the addition of peri- and postoperative variables would improve performance as these additional factors might reveal more modifiable areas for interventions to benefit patients. Our analyses demonstrated that adding peri- and postoperative variables did increase the predictive ability of the resulting models. Anaesthetic factors related to reduced risk included the use of epidurals and spinals. Surgical factors associated with increased risk of pain included thoracic incisions (in keeping with the increased risk with thoracic surgery), whilst the presence of a nasogastric tube in recovery was associated with reduced risk of severe pain. Patients who did not receive general anaesthesia largely underwent orthopaedic or urological surgery, whilst patients with nasogastric tubes were predominantly in the general surgery group. The model performed similarly in these groups compared with other surgical specialities. Additionally, severity of pain in recovery strongly predicted pain at 24 h postoperatively. On postoperative day 1, important predictors of reduced pain were two aspects of `DrEaMing´ (drinking, eating, mobilising), recognised as a surrogate indicator of good recovery [23], namely mobilising and eating. These findings highlight potential aspects of care for improvement and intervention throughout the peri-operative journey, for example aiming for good pain control in the immediate postoperative period and adherence to the principles of enhanced recovery after surgery to facilitate early mobilisation and resumption of oral intake.

Whilst these additional variables improved model performance, they are not available to the anaesthetist and patient before surgery, thus potentially resulting in a missed opportunity to change patient trajectory in the early peri-operative period. However, these factors could be used pre-operatively in a similar fashion to the National Emergency Laparotomy Audit (NELA) risk prediction model in which anticipated values are used [24], with the additional option of recalculating risk at the end of a surgical procedure if significant changes have occurred.

The clinical utility and feasibility of such a model is yet to be explored, though it would suit implementation similar to the widely adopted NELA score for emergency laparotomy in which the predicted probability of an outcome is used to classify patients as `high-risk´ and the package of care guided by this determination [25]. The calibration of our primary model was excellent and patient risk could be stratified based on the model prediction, for example using a cut-off of 20–30% based on decision-curve analysis. This would aid informed consent for surgery and anaesthesia in the context of shared decision-making by advising patients of potential outcomes after surgery. Those patients identified as `high-risk´ might then receive an adjusted anaesthetic and/or analgesic plan, or the use of a score at the point of leaving recovery could be used to highlight patients to inpatient pain services for ongoing care in the early postoperative period. Whilst other pain prediction tools have deliberately focused on simple additive tools for

ease of use, accepting lower performance [11], a model such as that developed here would be used through digital/online implementation, in line with other widely used risk prediction models [26].

This analysis does have some limitations. The case-mix was representative of major surgical activity in the NHS, but as a result is dominated by colorectal and other abdominal surgery, with around half of cases laparoscopic and/or robotic-assisted. Whilst we did assess performance across other specialities, the reduced risk of pain associated with laparoscopic surgery may have contributed to a lower overall rate of severe pain compared with other case series. The primary outcome was missing in approximately a quarter of patients and those patients differed in some predictor variables, for example more underwent open procedures and were planned for level 3 postoperative care. Whilst these factors may help to explain why outcome data were missing, they also have potential implications for the generalisability of our findings. As with all secondary analyses of observational data, we are limited by the variables included in the PQIP dataset so potentially important predictors were unavailable. For example, additional baseline risk factors such as ethnicity, pre-existing chronic pain, pain at the site of incision, detailed analgesic usage or previous negative experiences of pain were not collected. Baseline opioid usage and details on type of anaesthesia were not available for the whole cohort due to changes in the dataset over time and so reduced the available sample size. Several variables relating to anaesthesia and analgesia delivered to patients during surgery are of clinical significance, for example specific analgesic drugs used, but were missing for the majority of patients and so were excluded. These factors will be important in any future work focused on the impact of peri-operative processes on postoperative pain. Finally, we were focused on the specific issue of acute postoperative pain and so did not explore the performance of our models at predicting longer-term outcomes such as chronic postsurgical pain.

We have shown that data collected as part of a prospective cohort study can be used to develop a tool for predicting severe pain on postoperative day 1 with limited performance, but which serves to highlight several potentially important and modifiable contributing factors. However, it is likely that several important patient and process measures are missing from the available dataset and that pre-operative factors alone are not sufficient for accurate prediction beyond risk-stratification. Further work will be required to explore additional factors which might improve predictive performance and assess how the tool might be applied in clinical practice.

Acknowledgements

References

1. Walker EMK, Bell M, Cook TM, Grocott MPW, Moonesinghe SR. Patient reported outcome of adult perioperative anaesthesia in the United Kingdom: a cross-sectional observational study. *British Journal of Anaesthesia* 2016; **117**: 766e.
2. Gerbershagen HJ, Aduckathil S, van Wijck AJM, Peelen LM, Kalkman CJ, Meissner W. Pain intensity on the first day after surgery: a prospective cohort study comparing 179 surgical procedures. *Anesthesiology* 2013; **118**: 934–44.
3. Janssen KJM, Kalkman CJ, Grobbee DE, Bonsel GJ, Moons KGM, Vergouwe Y. The risk of severe postoperative pain: modification and validation of a clinical prediction rule. *Anesthesia and Analgesia* 2008; **107**: 1330–9.
4. Kehlet H, Jensen TS, Woolf CJ. Persistent postsurgical pain: risk factors and prevention. *Lancet* 2006; **367**: 1618–25.
5. Caumo W, Schmidt AP, Schneider CN, et al. Preoperative predictors of moderate to intense acute postoperative pain in patients undergoing abdominal surgery. *Acta Anaesthesiologica Scandinavica* 2002; **46**: 1265–71.
6. Ip HYV, Abrishami A, Peng PWH, Wong J, Chung F. Predictors of postoperative pain and analgesic consumption: a qualitative systematic review. *Anesthesiology* 2009; **111**: 657–77.
7. Chapman CR, Davis J, Donaldson GW, Naylor J, Winchester D. Postoperative pain trajectories in chronic pain patients undergoing surgery: the effects of chronic opioid pharmacotherapy on acute pain. *Journal of Pain* 2011; **12**: 1240–6.
8. Gilhooly D, Moonesinghe S. The Perioperative Quality Improvement Programme: improving outcomes. *British Journal of Hospital Medicine* 2018; **79**: 117–7.
9. Myles PS, Boney O, Botti M, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: patient comfort. *British Journal of Anaesthesia* 2018; **120**: 705–11.
10. Yang MMH, Riva-Cambrin J. Prediction tools for postoperative pain. *Pain Reports* 2021; **6**: e875.
11. Schnabel A, Yahiaoui-Doktor M, Meissner W, Zahn PK, Pogatzki-Zahn EM. Predicting poor postoperative acute pain outcome in adults: an international, multicentre database analysis of risk factors in 50,005 patients. *Pain Reports* 2020; **5**: e831.

12. Yang MMH, Hartley RL, Leung AA, et al. Preoperative predictors of poor acute postoperative pain control: a systematic review and meta-analysis. *British Medical Journal Open* 2019; **9**: e025091.

13. Sainani KL. Dealing with missing data. *Physical Medicine and Rehabilitation* 2015; **7**: 990–4.

14. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Cham: Springer International Publishing, 2015.

15. 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.

16. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*, 3rd edn. Chichester: Wiley, 2013.

17. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making* 2006; **26**: 565–74.

18. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical Journal. Biometrische Zeitschrift* 2008; **50**: 419–30.

19. Armstrong RA. raarmstrong/pqip-pain-public. v1.0. 2022. https://zenodo.org/record/7113266 (accessed 26/09/2022).

20. Collins G, Reitsma J, Altman D, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. https://www.equator-network.org/reporting-guidelines/tripod-statement/ (accessed 13/12/2021).

21. van Boekel RLM, Bronkhorst EM, Vloet L, Steegers MAM, Vissers KCP. Identification of preoperative predictors for acute postsurgical pain and for pain at three months after surgery: a prospective observational study. *Scientific Reports* 2021; **11**: 16459.

22. Pogatzki-Zahn EM, Liedgens H, Hummelshoj L, et al. Developing consensus on core outcome domains for assessing effectiveness in perioperative pain management: results of the PROMPT/IMI-PainCare Delphi Meeting. *Pain* 2021; **162**: 2717–36.

23. Levy N, Mills P, Mythen M. Is the pursuit of DREAMing (drinking, eating and mobilising) the ultimate goal of anaesthesia? *Anaesthesia* 2016; **71**: 1008–12.

24. NELA Risk Calculator. https://data.nela.org.uk/riskcalculator (accessed 27/01/2022).

25. National Emergency Laparotomy Audit. Standards Documents. https://www.nela.org.uk/Standards-Documents#pt (accessed 13/12/2021).

26. Surgical Outcome Risk Tool. http://www.sortsurgery.com/SORT2_home (accessed 07/07/2021).

Supporting Information

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

Appendix S1. Sample size calculation.

Appendix S2. Members of Peri-operative Quality Improvement Programme (PQIP) delivery team and collaborative.

Figure S1. Non-linear relationships between predictors and outcome.

Figure S2. Relationship between variables modelled through restricted cubic splines and the log odds of the primary outcome.

Figure S3. (a) Calibration plot and (b) decision curve for secondary model developed using pre- and peri-operative variables.

Table S1. Variables removed as > 10% missing.

Table S2. Missing data in individual patients by variable.

Table S3. Limits to continuous variables after Winsorisation.

Table S4. Characteristics of patients with/without outcome data.

Table S5. Additional patient characteristics and clinical features stratified by presence or absence of primary outcome.

Table S6. Included and deleted variables after backward selection for subset of patients with baseline opioid data.

Table S7. Model performance by surgical speciality.

Table S8. Variables included and deleted through backward selection in secondary model development.

Table S9. Full secondary model specification using pre, intra and postoperative variables after backward selection.