Evaluation of prognostic risk models for postoperative pulmonary complications in adult patients undergoing major abdominal surgery

STARSurg Collaborative and TASMAN Collaborative; Li, Elizabeth

DOI:
10.1016/S2589-7500(22)00069-3

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
STARSurg Collaborative and TASMAN Collaborative & Li, E 2022, 'Evaluation of prognostic risk models for postoperative pulmonary complications in adult patients undergoing major abdominal surgery: a systematic review and international external validation cohort study', *The Lancet Digital Health*, vol. 4, no. 7, pp. e520-e531. https://doi.org/10.1016/S2589-7500(22)00069-3

Evaluation of prognostic risk models for postoperative pulmonary complications in adult patients undergoing major abdominal surgery: a systematic review and international external validation cohort study

*STARSurg Collaborative and TASMAN Collaborative**

Lancet Digit Health 2022; 4: e520–31

*All contributing authors are shown in the appendix (pp 14–25)

Correspondence to:
Mr James Glasbey, Academic Department of Surgery, University of Birmingham, Birmingham B15 2TH, UK
j.glasbey@bham.ac.uk

See Online for appendix

Summary

Background Stratifying risk of postoperative pulmonary complications after major abdominal surgery allows clinicians to modify risk through targeted interventions and enhanced monitoring. In this study, we aimed to identify and validate prognostic models against a new consensus definition of postoperative pulmonary complications.

Methods We did a systematic review and international external validation cohort study. The systematic review was done in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. We searched MEDLINE and Embase on March 1, 2020, for articles published in English that reported on risk prediction models for postoperative pulmonary complications following abdominal surgery. External validation of existing models was done within a prospective international cohort study of adult patients (≥18 years) undergoing major abdominal surgery. Data were collected between Jan 1, 2019, and April 30, 2019, in the UK, Ireland, and Australia. Discriminative ability and prognostic accuracy summary statistics were compared between models for the 30-day postoperative pulmonary complication rate as defined by the Standardised Endpoints in Perioperative Medicine Core Outcome Measures in Perioperative and Anaesthetic Care (StEP-COMPAC). Model performance was compared using the area under the receiver operating characteristic curve (AUROCC).

Findings In total, we identified 2903 records from our literature search; of which, 2514 (86·6%) unique records were screened, 121 (4·8%) of 2514 full texts were assessed for eligibility, and 29 unique prognostic models were identified. Nine (31·0%) of 29 models had score development reported only, 19 (65·5%) had undergone internal validation, and only four (13·8%) had been externally validated. Data to validate six eligible models were collected in the international external validation cohort study. Data from 11 591 patients were available, with an overall postoperative pulmonary complication rate of 7·8% (n=903). None of the six models showed good discrimination (defined as AUROCC ≥0·70) for identifying postoperative pulmonary complications, with the Assess Respiratory Risk in Surgical Patients in Catalonia score showing the best discrimination (AUROCC 0·700 [95% CI 0·683–0·717]).

Interpretation In the pre-COVID-19 pandemic data, variability in the risk of pulmonary complications (StEP-COMPAC definition) following major abdominal surgery was poorly described by existing prognostication tools. To improve surgical safety during the COVID-19 pandemic recovery and beyond, novel risk stratification tools are required.

Funding British Journal of Surgery Society.

Introduction

Postoperative pulmonary complications account for a substantial proportion of surgical complications following major abdominal surgery, affecting 3–30% of individuals depending on the definition and patient population under investigation.[1–3] Before the COVID-19 pandemic, evidence existed to support the association between postoperative pulmonary complications and a moderate risk of death and critical care usage.[4–7]

Accurate preoperative stratification of patients' risk of pulmonary complications has several potential advantages. Firstly, evidence-based but resource-intensive perioperative prophylaxis can be targeted to those at highest risk of postoperative pulmonary complications. Examples include lung-protective intraoperative ventilation, prophylactic respiratory physiotherapy, and epidural analgesia.[8] Secondly, accurate preoperative stratification can inform shared decision making with patients and can determine optimal use of enhanced care areas with additional monitoring or elevated nursing ratios. Finally, accurate preoperative stratification can be used to enrich patient populations for clinical trials that evaluate novel therapies.

Several prognostic tools have been proposed to inform clinical decision making. However, these tools are yet to be used routinely in clinical practice possibly due to inconsistencies in outcome definitions,[3] paucity of

Research in context

Evidence before this study

We searched PROSPERO and Open Science Framework on June 1, 2021, for registered systematic reviews. We used the search terms "surgery", "pulmonary complication", and "postoperative pulmonary complication". We also searched MEDLINE and Embase, from database inception to March 1, 2020, with languages restricted to English. One previous systematic review was identified reporting prognostic models for postoperative pulmonary complications (not performed in line with the EQUATOR guidelines), and none were registered. 29 unique prognostic models were identified; however, very few had been validated externally (four [13·8%]) or in prospective international data (two [6·9%]).

Added value of this study

In this study, we present a comprehensive systematic review of prediction models for postoperative pulmonary complications in abdominal surgery. We showed that substantial heterogeneity exists in definitions used for postoperative pulmonary complications, and substantial flaws exist in the reporting and methodological approaches to the prognostic models developed. These factors limit comparison between models and the generalisability of study results. We also present external validation of six of the models identified in a large, prospective, international cohort study completed before the COVID-19 pandemic. This cohort study doubles the number of scores that have been externally validated to date and allows direct comparison of performance in an independent dataset. No models showed good discrimination on external validation using the Standardised Endpoints in Perioperative Medicine Core Outcome Measures in Perioperative and Anaesthetic Care (StEP-COMPAC) postoperative pulmonary complication definition (for model discrimination, an area under the receiver operating characteristic curve of 0·60–0·69 was considered moderate; 0·70–0·79 was considered good; and ≥0·80 was considered excellent), and data were insufficient to explore calibration of five of the models. Where calibration was performed, systematic over-estimation of risk was observed.

Implications of all the available evidence

Postoperative pulmonary complications remain an urgent clinical concern as surgery continues through and beyond the COVID-19 pandemic. Accurate stratification of patients at high risk of postoperative pulmonary complications following major abdominal surgery is required to provide appropriate clinical advice regarding risk of complications, or target enhanced monitoring and perioperative interventions to mitigate modifiable risk factors. However, the prediction of postoperative pulmonary complications (StEP-COMPAC definition) for patients following major abdominal surgery is not sufficiently described by existing risk prediction tools. As surgery is upscaled to meet the growing demands on health-care systems worldwide, an urgent clinical need exists to better differentiate between modifiable risk and fixed risk of pulmonary complications and develop validated prediction tools that are clinically relevant to a global context.

external validation,[9] lack of generalisability, and the complexity of implementing stratified care in clinical pathways.[10,11] The recent standardised, consensus-driven definitions of postoperative pulmonary complications by the Standardised Endpoints in Perioperative Medicine Core Outcome Measures in Perioperative and Anaesthetic Care (StEP-COMPAC) group[3] has superseded previous definitions and provides an opportunity to allow fair comparison between different models developed to date. The so-called transportability of these prognostic tools[12] has yet to be evaluated to determine whether these tools can be generalised to predict the occurrence of post-operative pulmonary complications in a broad major abdominal surgical population. Therefore, in this study, we aimed to identify existing prognostic models from the literature, and externally validate selected scores in an international, prospective, pre-pandemic cohort study to predict postoperative pulmonary complications according to the StEP-COMPAC definition.

Methods

Systematic review

Search strategy and selection criteria

For the **predefined protocol** see https://osf.io/ceypm

We did a systematic review according to a predefined protocol, which was registered on Open Science Framework, and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.[13] We developed a search strategy to identify risk prognostication models that predict postoperative pulmonary complications following abdominal surgery (appendix p 13). We did a comprehensive search of MEDLINE and Embase, on March 1, 2020, for articles published in English. This search was supplemented through hand-searching citation and reference lists from relevant articles. No date restrictions were applied.

Studies were eligible for inclusion if they developed or externally validated a model that sought to predict risk of pulmonary complications after abdominal surgery using preoperative or operative characteristics, or both. Models that included patients having non-abdominal surgical procedures were eligible if these models also included abdominal surgical procedures. The exclusion criteria were development or validation done in non-adult patients (<18 years) in whom development and performance were not separate for adult and children, inclusion of postoperative parameters in the risk score, a score non-specific to pulmonary complications (ie, all complications) in which pulmonary and other complications were not reported separately, procedure-specific risk scores, and those that did not provide sufficient description of risk

parameters to allow replication of the algorithm in the external validation study.

Data analysis

Following the removal of duplicate publications, titles and abstracts were screened and full texts of relevant publications were reviewed. Data fields of interest relating to study characteristics (year of publication, setting, sample size, inclusion criteria, technique for model development), pulmonary complications (definition used, number of cases), and the risk score (validation status, clinical parameters included, and any metrics reported regarding prognostic accuracy and model performance) were extracted from eligible papers.

Study screening and data extraction were completed independently by seven reviewers (SKK, WAC, VM, OK, SB, SQS, and MK), with any disagreements resolved through a consensus-based approach. Finally, quality assessment of eligible studies was done using the reporting guidelines of the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)[14] and the Prediction model Risk Of Bias ASsessment Tool (PROBAST).[15] Narrative synthesis of results was done, with the data extracted and summarised using frequencies and percentages for dichotomous variables. No meta-analysis was planned or performed.

External validation cohort study

Study design and participants

We did an international external validation cohort study of the risk scores identified from the systematic review in accordance with the TRIPOD reporting guidelines.[14] The international external validation cohort study was a preplanned secondary analysis of the REspiratory COmplications after abdomiNal surgery (RECON) study, done according to a prespecified protocol.[16] RECON was a prospective, multicentre, international study investigating pulmonary complications after major abdominal surgery. Eligible patients were identified by local collaborators at each participating hospital during two data collection periods: Jan 1, 2019, to April 30, 2019 (UK and Ireland), and Sept 1, 2019, to Oct 31, 2019 (Australia). Consecutive adult patients undergoing a broad range of major abdominal surgeries were eligible across four surgical disciplines (abdominal visceral resection, reversal of stoma, open vascular surgery, anterior abdominal wall hernia repair, or transplant surgery through any operative approach). Planned day case procedures and abdominal surgeries without visceral resection were excluded. This process collected routine anonymised data with no change to clinical care pathways, and so was considered an audit in the UK and Ireland but required ethical approval in Australia. Confirmation of appropriate local or national regulatory approval was required before patient enrolment. The international external validation cohort study was registered according to the appropriate local or national approval pathways in each participating country, and data was submitted using a secure Research Electronic Data Capture server.[17]

Data collection and procedures

Data relating to baseline demographics, perioperative care, and 30-day outcomes for each patient were collected using a prespecified case report form.[16] The Assess Respiratory Risk in Surgical Patients in Catalonia (ARISCAT) score[18] was identified a priori by the external advisory group for validation based on the likelihood of clinical adoption, methodological rigor (ie, sufficient internal validation during development, one or more existing external validation studies), and accessibility of covariable data in routine practice. Other models identified from the systematic review were externally validated if the corresponding clinical parameters were available in the RECON dataset.

Variables from the RECON dataset that were identified for use in the external validation process were age (years), sex (male or female), body-mass index (kg/m^2), history of cardiac disease (formal diagnosis of myocardial infarction, angina, congestive cardiac failure, or hypertension recorded on the patient health-care record), history of chronic respiratory disease (formal diagnosis of asthma, chronic obstructive pulmonary disease, bronchiectasis, pulmonary fibrosis, lung cancer, or obstructive sleep apnoea recorded on the patient health-care record), history of neurological disease (formal diagnosis of stroke, dementia, multiple sclerosis, Parkinson's disease, or neurological tumour recorded on the patient health-care record), the American Society of Anesthesiologists' physical status classification (grades 1–5), smoking status within the last 6 weeks (yes or no), preoperative (ie, ≤1 month) respiratory tract infection (yes or no), preoperative oxygen saturation (>96%, 91–95%, or ≤90% as measured by pulse oximetry on 21% fractional concentration of oxygen in inspired air or room air), last preoperative albumin concentration (g/L), last preoperative haemoglobin concentration (g/L), operative indication (benign or malignant), operative urgency (elective [intervention planned in advance of admission] or emergency [intervention planned after admission]), operative approach (open or minimally invasive [laparoscopic or robotic techniques with no conversion to open]), highest level of operative incision (highest incision performed at the thoracicoabdominal, upper abdominal, or lower abdominal level), and operative duration (min). Full data were extracted from routinely collected patient health records, with no changes to clinical care pathways.

Outcome definition

Postoperative pulmonary complications were defined in this study according to the StEP-COMPAC definition.[3]

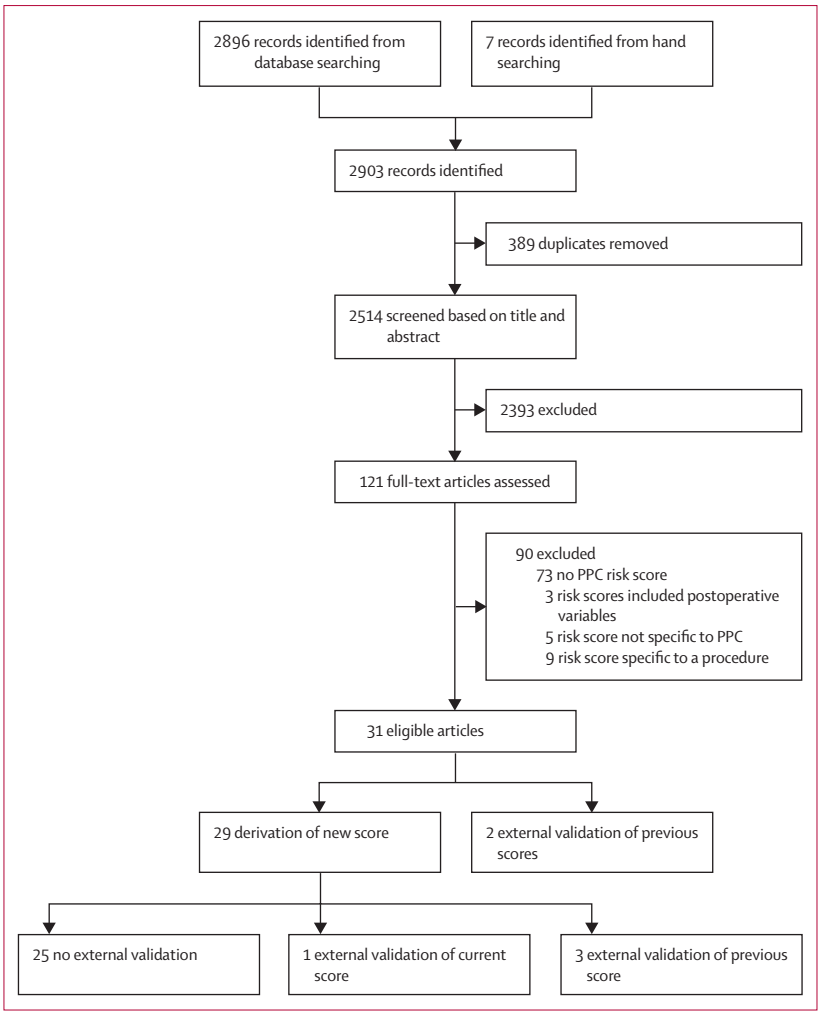

Figure 1: **PRISMA flow diagram**
PRISMA=Preferred Reporting Items for Systematic Reviews and Meta-Analyses. PPC=postoperative pulmonary complications.

Core outcome sets are an agreed set of outcomes that all clinical trials should report as a minimum in a specific area, and so improve transparency, consistency, and reproducibility of our study design and analysis.[19] This definition is globally accepted as the benchmark for perioperative outcome selection, guided by inter-disciplinary leaders in each field of interest.[20] This consensus composite outcome definition of postoperative pulmonary complications includes four respiratory pathologies that share a common pathophysiological mechanism of pulmonary collapse and airway contamination (atelectasis [radiological evidence], pneumonia [US Centers for Disease Control definition], acute respiratory distress syndrome [Berlin consensus definition], and pulmonary aspiration [clear clinical history with radiological evidence]).[3] This definition excludes other complications that do not share these mechanisms but have previously been considered as postoperative pulmonary complications (eg, pulmonary embolism, pleural effusion, pneumothorax, cardiogenic pulmonary oedema, and bronchospasm) or other causes of respiratory failure requiring mechanical ventilation.[3] All postoperative pulmonary complication events were recorded at 7 days and 30 days after surgery. The StEP-COMPAC composite postoperative pulmonary complication outcome was used for all external validation models, irrespective of those used in the original derivation models. Data collection teams were unaware of the predictors that would be included in this external validation, therefore no blinding to outcomes or other predictors was deemed necessary.

Statistical analysis

Risk score models identified from the systematic review were externally validated if the corresponding clinical parameters were available in the RECON dataset.[16] Where appropriate, clinical parameters that were aligned but not exactly equivalent in RECON were rationalised on a pragmatic basis. Definitions of each parameter were harmonised to facilitate validation. Scores are referenced in the text by the acronym or name of the score where this is available, or the name of the first author in the development report. Multivariable logistic regression was done for each individual model, with pairwise exclusion done for patients that had incomplete data for clinical parameters required.

Model performance was compared using the area under the receiver operating characteristic curve (AUROCC) and prognostic accuracy summary statistics (sensitivity, specificity, positive predictive value, and negative predictive value). Coefficients, intercepts, and cutoff values were used in accordance with the most recent paper reporting development or validation, or both. The cutoff with the highest AUROCC value was chosen if a cutoff point had not been previously reported. An AUROCC of 0·60–0·69 was considered to be moderate, 0·70–0·79 considered good, and 0·80 or more considered excellent model discrimination.[21] Calibration was assessed through visual inspection and the calibration intercept (calibration-in-the-large) and slope[22] (an intercept of 0 and slope of 1 indicates perfect calibration). No recalibration was planned or performed to allow determination of whether these models were transportable in their original iteration[12] (eg, continue to produce accurate predictions in a related but different population). A sensitivity analysis was done using 7-day postoperative pulmonary complications as the outcome to determine whether model discrimination remained consistent.

We did all statistical analyses using R Studio (version 3.6.1) with the following packages: tidyverse, finalfit, pROC, and collaboratoR.[23]

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

	Country	Study type	Outcomes	Operative population	Number of events	Event rate	Validation
Brooks-Brunn (1997)[33]	USA	Prospective, multicentre	Postoperative pulmonary complications (pneumonia [alternate]; atelectasis [alternate])	n=400; elective non-cardiac (abdominal)	90	22·5%	Internal (subgroup)
Brooks-Brunn (1998)[24]	USA	Prospective, multicentre	Postoperative pulmonary complications (pneumonia [alternate]; atelectasis [alternate])	n=276; elective non-cardiac (abdominal)	73	26·4%	None
Arozullah et al (2000)[34]	USA	Retrospective, multicentre (NSQIP)	Postoperative respiratory failure (unanticipated reintubation [undefined]; failure to wean [<48 h])	n=81 719; non-cardiac	2746	3·4%	Internal (subgroup)
Fuso et al (2000)[25]	Italy	Retrospective, single centre	Postoperative pulmonary complications (pneumonia [undefined]; atelectasis [undefined])	n=480; elective non-cardiac (abdominal)	88	18·3%	None
Arozullah et al (2001)[35]	USA	Retrospective, multicentre (NSQIP)	Postoperative pulmonary complication (pneumonia [StEP-COMPAC])	n=160 805; non-cardiac	2466	1·5%	Internal (subgroup)
McAlister et al (2003)[36]	Canada	Prospective, single centre	Postoperative respiratory failure (unanticipated reintubation [undefined]) or postoperative pulmonary complications (pneumonia [alternate]; atelectasis [alternate]; pleural effusion [alternate]; pneumothorax [alternate])	n=272; elective non-cardiac	22	8·1%	Internal (cross-validation)
McAlister et al (2005)[26]	Canada	Prospective, single centre	Postoperative respiratory failure (unanticipated reintubation [undefined]) or postoperative pulmonary complications (pneumonia [alternate]; atelectasis [alternate]; pleural effusion [alternate]; pneumothorax [alternate])	n=1055; elective non-cardiac	28	2·7%	None
Johnson et al (2007)[37]	USA	Retrospective, multicentre (NSQIP)	Postoperative respiratory failure (unanticipated reintubation [undefined]; failure to wean [<48 h])	n=90 055; non-cardiac	2691*	3·0%	Internal (subgroup)
Scholes et al (2009)[27]	Australia	Prospective, multicentre	Postoperative pulmonary complications (pneumonia [alternate]; atelectasis [alternate])	n=268; elective non-cardiac (upper abdominal)	35	13·1%	None
Canet et al (2010)[18]	Spain	Prospective, multicentre	Postoperative pulmonary complications (traditional [alternate])	n=2464; non-cardiac and cardiac	123	5·0%	Internal (bootstrap) and external[48,51,52,53]
Smith et al (2010)[28]	USA	Retrospective, single centre	Postoperative pulmonary complications (traditional [alternate])	n=359; non-cardiac (general)	25	7·0%	None
Kor et al (2011)[39]	USA	Prospective, single centre	Postoperative pulmonary complication (acute respiratory distress syndrome [alternate])	n=4363; elective non-cardiac and cardiac	113	2·6%	Internal (cross-validation) and external[45]
Gupta et al (2011)[38]	USA	Retrospective, multicentre (NSQIP)	Postoperative respiratory failure (unanticipated reintubation [undefined]; failure to wean [<48 h])	n=211 410; non-cardiac and cardiac	6531	3·1%	Internal (subgroup)
Ramachandran et al (2011)[40]	USA	Retrospective, multicentre (NSQIP)	Postoperative respiratory failure (unanticipated reintubation [30 days])	n=222 094; elective non-cardiac	1853	0·8%	Internal (subgroup)
Hua et al (2012)[41]	USA	Retrospective, multicentre (NSQIP)	Postoperative respiratory failure (unanticipated reintubation [30 days])	n=231 548; non-cardiac and cardiac	5028	2·2%	Internal (subgroup)
Blum et al (2013)[29]	USA	Retrospective, single centre	Postoperative pulmonary complication (acute respiratory distress syndrome [alternate])	n=50 367; non-cardiac	93	0·2%	None
Gupta et al (2013)[43]	USA	Retrospective, multicentre (NSQIP)	Postoperative pulmonary complication (pneumonia [alternate])	n=211 410; non-cardiac and cardiac	3825	1·8%	Internal (subgroup)
Brueckmann et al (2013)[42]	USA	Retrospective, single centre	Postoperative respiratory failure (unanticipated reintubation [72 h])	n=16 885; non-cardiac and cardiac	65	0·4%	Internal (subgroup) and external[51]
Jeong et al (2014)[44]	South Korea	Prospective, single centre	Postoperative pulmonary complication (traditional [alternate])	n=2059; non-cardiac and cardiac	140	6·8%	Internal (bootstrap)
Kor et al (2014)[45]	USA, Turkey	Prospective, multicentre	Postoperative pulmonary complication (acute respiratory distress syndrome [alternate])	n=1562; elective non-cardiac and cardiac	117	7·5%	Internal (bootstrap)
Canet et al (2015)[46]	Europe	Prospective, multicentre	Postoperative respiratory failure (hypoxaemia [5 days])	n=5384; non-cardiac and cardiac	224	4·2%	Internal (bootstrap)
Yang et al (2015)[30]	USA	Retrospective, multicentre (NSQIP)	Postoperative respiratory failure (unanticipated reintubation [2 days]; failure to wean [2 days]) or postoperative pulmonary complication (pneumonia [alternate])	n=165 196; elective non-cardiac (general)	9596	5·8%	None
Johnson et al (2017)[47]	USA	Retrospective, multicentre (NSQIP)	Postoperative respiratory failure (unanticipated reintubation [30 days]; failure to wean [<48 h])	n=151 700; elective non-cardiac	2900	1·9%	Internal (subgroup)

(Table 1 continues on next page)

	Country	Study type	Outcomes	Operative population	Number of events	Event rate	Validation
(Contined from previous page)							
Neto et al (2018)[48]	Global	Prospective, multicentre	Postoperative respiratory failure (hypoxaemia [5 days]; unanticipated reintubation [5 days]; failure to wean [5 days]) or postoperative pulmonary complications (pneumonia [alternate]; acute respiratory distress syndrome [alternate]; pneumothorax [alternate])	n=3919; non-cardiac and cardiac	419	10·7%	Internal (subgroup)
Lukannek et al (2019)[51]	USA	Retrospective, single centre	Postoperative respiratory failure (unanticipated reintubation [<72 h])	n=90 893; non-cardiac	699	0·8%	External[51]
Russotto et al (2019)[31]	Europe	Prospective, multicentre	Postoperative pulmonary complication (pneumonia [alternate])	n=5094; non-cardiac and cardiac	120	2·4%	None
Kawasaki et al (2019)[49]	Japan	Prospective, single centre	Postoperative pulmonary complication (pneumonia [alternate])	n=1050; non-cardiac (abdominal)	56	5·3%	Internal (bootstrap)
Takesue et al (2019)[50]	Japan	Retrospective, multicentre	Postoperative pulmonary complication (pneumonia [StEP-COMPAC])	n=247 604; non-cardiac (general)	7196*	2·9%	Internal (subgroup)
Baba et al (2020)[32]	Japan	Retrospective, single centre	Postoperative pulmonary complication (pneumonia [alternate])	n=1016; elective non-cardiac (general)	67	6·6%	None

NSQIP=National Surgical Quality Improvement Program. StEP-COMPAC=Standardised Endpoints in Perioperative Medicine Core Outcome Measures in Perioperative and Anaesthetic Care. *Event rate in derivation sample was not stated in the paper, and was thus calculated based on the event rate in combined derivation and internal validation cohorts.

Table 1: **Characteristics of included studies describing an original risk score**

Results

In total, we identified 2903 records from our literature search; of which, 2514 (86·6%) unique records were screened based on the title and abstract, and 121 (4·8%) of 2514 full texts were assessed for eligibility (figure 1). We identified 29 unique risk scores to predict the risk of pulmonary complications after abdominal surgery: nine (31·0%) had score development reported only,[24–32] 19 (65·5%) had undergone internal validation,[18,33–50] and only four (13·8%) had been externally validated (table 1).[18,39,42,51]

Most scores were developed for patients either undergoing procedures from all surgical specialties (non-cardiac and cardiac; 11 [37·9%] of 29) or all forms of non-cardiac surgery (eight [27·6%]; table 1). Several scores were developed only for patients undergoing elective procedures (12 [41·3%]; table 1).[24–27,30,32,33,36,39,40,45,47] Although 18 (62·1%) studies were multicentre, only 12 (41·4%) were developed or validated in prospective cohort data. Furthermore, nine (31·0%) were derived from the subsets of the same registry (National Surgical Quality Improvement Program).[30,34,35,37,38,40,41,43,47] All scores developed shared a common approach of using logistic regression to derive their models.

Adherence to the TRIPOD reporting guidelines[14] was typically mixed, with no clear pattern of improvement since TRIPOD was published in 2015 (appendix p 1). Consistently poor reporting remained for methods of blinding of outcome or predictors, how missing data were handled, and methods and results of model updating. Similarly, all scores (derived or validated), which were quality assessed using the PROBAST tool,[15] were rated to be at an overall high risk of bias (32 [88·9%] of 36) or unclear risk of bias (four [11·1%]; appendix p 2)—namely, due to the signalling questions regarding the outcome or analysis, or both (appendix p 4).

A wide variation was reported in the definition of outcomes in included studies. Studies defined outcomes according to either the aetiology of postoperative pulmonary complications (16 [55·2%] of 29), severity of pulmonary complications (eg, postoperative respiratory failure; nine [31·0%]), or a composite outcome of both (four [13·8%]; table 1). Within these classifications, a substantial variation existed in the included postoperative pulmonary complications and prognostic criteria used. For example, although 17 studies had a primary outcome including pneumonia, only two (11·8%) adhered to the US Centre for Disease Control definition.[3] No studies used the StEP-COMPAC consensus criteria.

A wide variation in the pulmonary complication rates was reported between studies (range 0·2%[29] to 24·6%[24]; figure 2). Definition of outcome appeared to influence the rate observed; studies only reporting acute respiratory distress syndrome or respiratory failure had a corresponding lower event rate.

The 29 risk prediction models were based on 53 unique clinical parameters, with substantial heterogeneity in the number of variables (median 7 [IQR 5–8]) and types of variables included within the predictive models (figure 3; appendix p 5). Common variables identified as predictive included the type of surgery and surgical location (21 [72·4%] of 29), a diagnosis of chronic respiratory disease (17 [58·6%]), patient age (16 [55·2%]), and whether the procedure was listed as an emergency (14 [48·3%]).

Of the 29 eligible risk prediction models, all models demonstrated good or excellent discrimination for

postoperative pulmonary complications in their development study (median AUROCC 0·815 [IQR 0·780–0·854]; appendix pp 6–7). However, in four risk prediction models that had previously undergone external validation,[18,39,42,51] the AUROCC remained more than 0·7 for one score only (ie, ARISCAT[18]).

Measures of prognostic accuracy were reported for only eight (27·6%) of 29 original models during development, with a substantial variation in the sensitivity (IQR 0·500–0·790) and specificity (0·749–0·902) at the reported cutoff point. Although 17 (58·6%) of 29 models reported an evaluation of calibration during development, just one model (ie, SLIP[39]) had calibration reported within an external validation cohort.[45]

In the external validation cohort study, 11 591 patients undergoing major abdominal surgery from 150 centres across the UK, Ireland, and Australia were included (appendix p 8). Overall, 903 (7·8%) of 11 591 patients reported a 30-day postoperative pulmonary complication rate, including atelectasis (526 [4·5%]), pneumonia (493 [4·3%]), pulmonary aspiration (39 [0·3%]), and acute respiratory distress syndrome (16 [0·1%]). 721 (79·8%) of 903 postoperative pulmonary complications occurred within the first 7 days after surgery. A further 451 (3·9%) of 11 591 reported other postoperative pulmonary complications that did not meet the StEP-COMPAC definition, and so were not included as an outcome event in the external validation cohort. Substantial differences were observed between the surgical populations in the derivation cohorts and the RECON cohort (appendix pp 9–11)—notably, three (50%) of six studies that included patients undergoing non-abdominal procedures[18,42,44] (30–60% of the overall samples) and two studies requiring emergency procedures.[24,33]

Of the 17 variables that were identified in five or more models, all variables except for functional status and respiratory symptoms at the time of surgery were represented in the RECON dataset. Six original risk score models[18,24,33,42,44,49] were able to be externally validated with the data collected in the RECON cohort study based on the prerequisite clinical parameters being present (table 1; appendix pp 8–11). Of these models, only two scores (ARISCAT[18] and SPORC[42]) had been previously externally validated (table 2).

None of the six models demonstrated good discrimination for identifying postoperative pulmonary complications according to the StEP-COMPAC definition in this external validation cohort (figure 4A; table 2). Performance ranged from an AUROCC of 0·574 to 0·700, with the ARISCAT[18] score showing the best discrimination (AUROCC 0·700 [95% CI 0·683–0·717]). All models performed significantly worse than their derivation or internal validation results (figure 4B). Furthermore, model discrimination remained consistent on sensitivity analysis using the 7-day postoperative pulmonary complication outcome (appendix pp 3, 12).

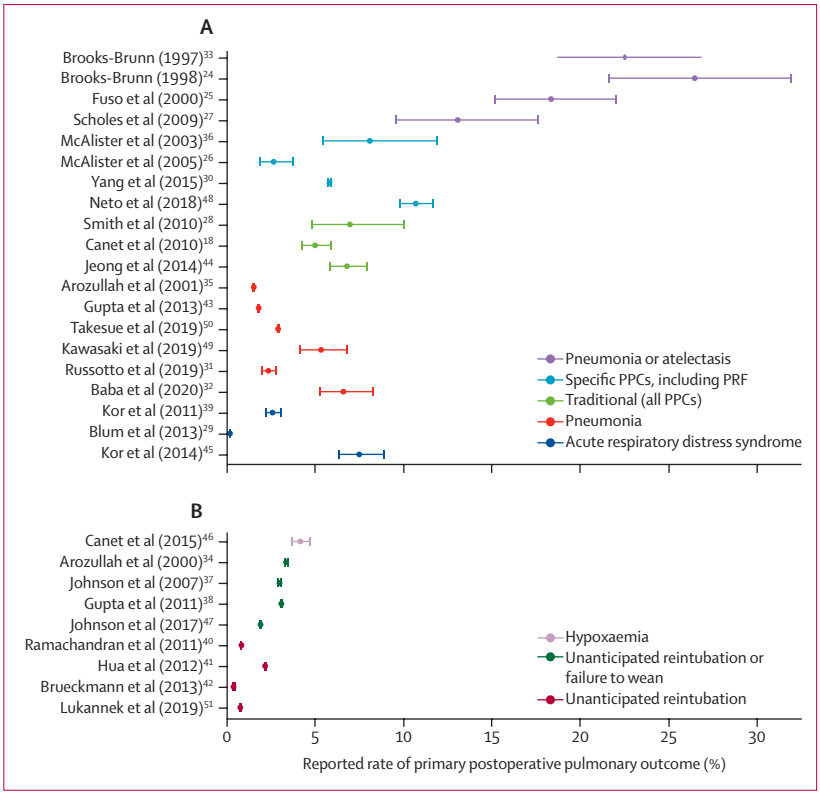

Figure 2: **Rate of primary pulmonary outcome reported in included studies**
Error bars show 95% CIs. (A) PPCs. (B) PRF only. PPCs=postoperative pulmonary complications. PRF=postoperative respiratory failure.

The intercept for the original models, which is required to calculate calibration of the models in the RECON dataset, was only reported in one study.[33] This model showed poor calibration within the RECON validation dataset with substantial overestimation of the risk of postoperative pulmonary complications (calibration-in-the-large –1·867; calibration slope 0·377).

Discussion

Mitigation of postoperative pulmonary complications is a priority in surgery, with these complications representing a leading source of morbidity and mortality.[4–7] This study found that before the COVID-19 pandemic 29 original models had been developed for the prediction of postoperative pulmonary complications that were relevant to patients undergoing abdominal surgery. The clinical relevance of most existing models is questionable at present because of a paucity of external validation, the heterogeneity and number of variables required for use, and substantial variation in the populations and outcomes under investigation. Of the six models[18,24,33,42,44,49] able to be externally validated in the largest prospective dataset of postoperative pulmonary complications done before the COVID-19 pandemic, none were significantly higher than the a priori threshold for good discrimination in the cohort (ie, AUROCC ≥0·70).

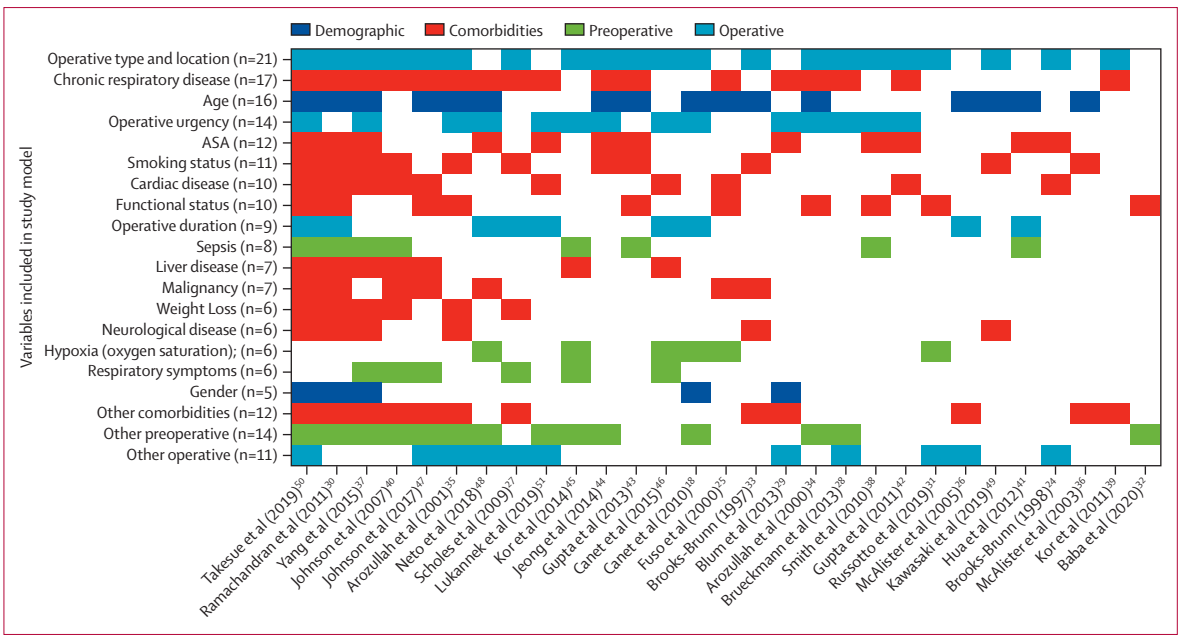

Figure 3: **Heatmap of variables included in all models identified from the systematic review**
A full list of variables included in less than five models is shown in the appendix (p 5). ASA=American Society of Anesthesiologists.

Despite being one of the most common and clinically significant postoperative complications in modern surgical practice, the prediction of multifactorial outcomes such as postoperative pulmonary complications remain challenging tasks.[54] As this study demonstrates, a large number of demographic, comorbidity, preoperative, and intraoperative factors exist, which have been previously identified as predictive of postoperative pulmonary complications. Despite the potential for bias and methodological concerns in derivation of the models,[9] the most frequently highlighted variables represent those that are most likely to have the greatest clinical significance. These represent important factors to investigate in any future development of risk prediction models, although the potential for unaccounted confounders or collinearity must be considered. Furthermore, the challenges in prediction are further compounded by the paucity of agreement regarding what the target outcome of these models should entail. Substantial heterogeneity exists in the definition of postoperative pulmonary complications observed in the included studies, reflecting the variation observed in the literature.[3] This heterogeneity limits the inter-study comparability in the event rates reported and subsequent model performance on external validation. The consensus-based global guidelines for standardisation of postoperative pulmonary complications[3] that we have used in this study provide a reproducible framework that can be used by the perioperative community for future comparison of prognostic models, clinical trials to which our risk stratification tool might be applied, and routine practice.

Furthermore, the paucity of external validation studies and poor quality of reporting constitute major barriers for the safe use of developed risk prediction models in clinical practice.[55] This study identified that only one (17%) of six original risk scores of postoperative pulmonary complication and only three [19%] of 16 of those available for over 5 years have undergone external validation to date, which is reflective of a wider issue with replication of risk prediction models.[56] Because of the nature of predictive models and the pragmatic design of our international multicentre study, only those with corresponding data for all prognostic variables present within dataset could be externally validated. This effect should be considered a major limitation of this analysis. Using the RECON dataset, the number of models externally validated has doubled from four to eight, yet 16 (55%) of 29 risk models remain unvalidated. We prioritised the ARISCAT model[18] a priori based on study quality and likelihood of clinical adoption. However, despite RECON including 11 (73%) of 15 most commonly identified variables, in addition to other less commonly identified variables, the capability to externally validate further models was restricted by the heterogeneity and number of variables identified in the models. This limitation could have implications for the clinical utility of these alternative models, with even greater challenges implementing complex datapoints and scoring systems within existing care pathways.[57,58]

As frequently observed in other contexts,[56] the discrimination on external validation was significantly poorer than in the derivation cohorts. Furthermore, these models either did not report information needed to assess

	AUROCC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Brooks-Brunn (1997)[33]	0·607 (0·587–0·628)	0·664 (0·629–0·698)	0·525 (0·514–0·536)	0·110 (0·100–0·119)	0·947 (0·940–0·953)
Brooks-Brunn (1998)[24]	0·660 (0·642–0·679)	0·745 (0·717–0·774)	0·531 (0·521–0·540)	0·118 (0·110–0·127)	0·961 (0·956–0·966)
Canet et al (2010)[18]	0·700 (0·683–0·717)	0·753 (0·725–0·781)	0·564 (0·555–0·574)	0·127 (0·118–0·136)	0·964 (0·960–0·969)
Brueckmann et al (2013)[42]	0·574 (0·556–0·593)	0·241 (0·214–0·269)	0·875 (0·869–0·881)	0·140 (0·123–0·158)	0·932 (0·927–0·937)
Jeong et al (2014)[44]	0·688 (0·671–0·706)	0·714 (0·685–0·744)	0·578 (0·568–0·587)	0·125 (0·116–0·134)	0·960 (0·955–0·965)
Kawasaki et al (2019)[49]	0·662 (0·645–0·678)	0·800 (0·773–0·826)	0·453 (0·444–0·463)	0·110 (0·102–0·118)	0·964 (0·959–0·969)

AUROCC=area under the receiver operating characteristic curve. PPV=positive predictive value. NPV=negative predictive value. RECON=REspiratory COmplications after abdomiNal surgery.

Table 2: **Discriminatory performance of six risk stratification scores in the external validation cohort (RECON) to predict pulmonary complications in the 30-day postoperative period**

calibration or demonstrated poor calibration[33] within the RECON cohort, and so could not be reliably used to assess risk at an individual patient level without recalibration.[22] Poor calibration and prognostic accuracy might reflect selection biases or overfitting within derivation cohorts, as well as differences in eligibility criteria and outcome definitions from the RECON cohort. For example, half the externally validated studies broadly matched the RECON population by including only patients undergoing abdominal surgery,[24,33,49] whereas the other half included patients undergoing other non-abdominal procedures (including cardiac surgery), who comprised 30–60% of the overall derivation samples.[18,42,44] Similarly, two studies excluded patients requiring emergency procedures,[24,33] which might be of particular importance as emergency surgery was frequently identified as a key prognostic factor. These differences from the original population might partly explain the differences in event rate between studies and reduced prognostic accuracy observed on external validation (ie, these models might not be readily transportable to patients undergoing emergency surgery). The other important consideration would be that all models were being assessed in the prediction of the StEP-COMPAC definition of postoperative pulmonary complications,[3] rather than the definition originally used in the development of this study. The StEP-COMPAC definition was selected because this definition has reached consensus as a new global standard, therefore ensuring the analyses in our study are relevant to the future of perioperative research and practice. However, this decision means that poor model performance in our validation study does not inherently reflect poorly on the model as originally described. In addition, this choice of outcome does not account for the severity of postoperative pulmonary complications, and this fact might explain the poor performance of models such as SPORC,[42] which focuses on the prediction of unplanned mechanical ventilation irrespective of aetiology. In light of these differences from the original derivation cohorts in the RECON cohort, this external validation analysis should be viewed as only testing the transportability (ie, generalisability) of these models, rather than producing consistent results when tested under the same conditions as their original derivation cohorts. Nonetheless, our results still indicate that none of the current models were suitable for the reliable prediction of postoperative pulmonary complications in a broad cohort of patients undergoing major abdominal surgery.

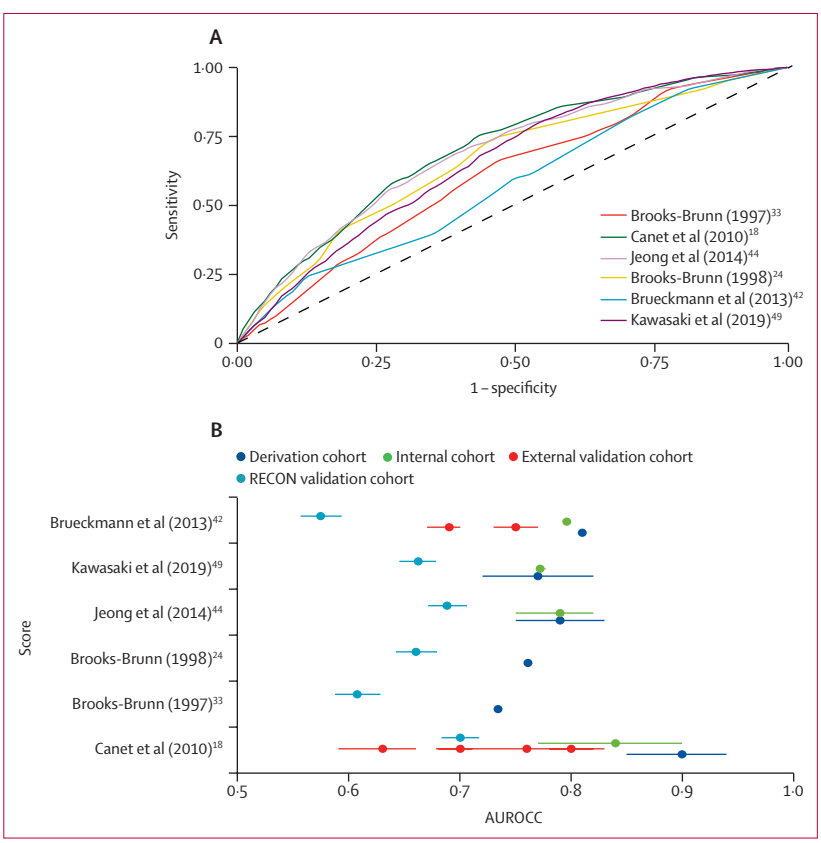

Figure 4: **Discriminatory performance of six risk stratification scores in the external validation cohort to predict pulmonary complications in the 30-day postoperative period**

Error bars show 95% CIs. (A) Receiver operating characteristic curves. (B) Discrimination compared with previous literature. AUROCC=area under the receiver operating characteristic curve.

This systematic review presents the most comprehensive analysis to date of risk scoring systems for the prediction of pulmonary complications after abdominal surgery. Furthermore, we report the simultaneous external

validation of six prognostic models identified, allowing direct contrast of their performance in a broad cohort of patients undergoing major abdominal surgery. The external validation is done on the largest prospective international cohort study on the topic of postoperative pulmonary complications before the COVID-19 pandemic, and uses the consensus-based StEP-COMPAC definition of postoperative pulmonary complications to facilitate future comparisons.[3]

This study, however, has several important limitations. First, although a systematic methodology was adopted, the search was limited to prognostic scores for the prediction of postoperative pulmonary complications in a broad adult surgical population, and thus did not include procedure-specific models because of anticipated limited generalisability to a broader population. Additionally, the search was limited to English language papers only. Because of these search limitations, the scores reported here might not encompass every score developed globally. Furthermore, this systematic review was limited by the quality of reporting in the included studies, and it should be noted that the majority were published before the TRIPOD statement,[14] with few reporting measures of prognostic accuracy and model calibration to facilitate comparison. Secondly, RECON did not collect data for all possible risk factors identified in the prognostic scores; for example, functional status of patients or respiratory symptoms at the time of surgery. Data collected were restricted to what was routinely recorded and practical to be collected across international health systems, informed by prognostic models published at the time of design (ie, published before January, 2019). This methodology has been previously validated for both case ascertainment and data accuracy across a wide range of settings.[59–63] More than half of the models identified remain unvalidated, including updated versions of the scores such as the SPORC-2 score,[51] and our results are not generalisable to models not evaluated within this study. However, these models often had other methodological or practical challenges that might limit their capacity for clinical adoption, such as complex or poorly defined covariables or insufficient internal validation. Furthermore, pragmatic decisions were made to equate risk factors when the data collected in RECON did not exactly match the original study, and these decisions might have contributed to the reduced discrimination observed on external validation. For example, high Brinkman index corresponding to current smoking status, or history of angina equated to a broader history of cardiac disease. Thirdly, none of the datasets or models incorporated perioperative SARS-CoV-2 infection, and this infection represents an important covariable that remains unaccounted for.[64] However, even during periods with high community infection rates, the overall perioperative SARS-CoV-2 infection rates have remained low (<5%),[64,65] and so the absolute risk of pulmonary sequelae of COVID-19 and thus the effect on the discriminative ability of these prognostic models is likely to be minimal. Fourthly, RECON collected routine data with no change to clinical care pathways. Given that community health-care attendances are not always accessible from hospital, the potential for missed postoperative pulmonary complications after discharge exists. However, the StEP-COMPAC definition requires radiological evidence of postoperative pulmonary complications for diagnosis; we expect all postoperative pulmonary complications that meet our study definition to have been identified. Furthermore, sensitivity analyses showed that the predictive ability remained consistent whether for early 7-day follow-up or late 30-day follow-up for postoperative pulmonary complications. Finally, RECON was limited to only high-income countries (UK, Ireland, and Australia), as were patients included in the included model development studies. More data are required from low-income and middle-income countries to understand whether these models are valid across low-resource health-care settings.

The risk of postoperative pulmonary complications should be minimised for all patients where modifiable; however, without the capability to accurately identify those at the highest risk, provision of appropriate clinical advice regarding the risk of complications or target-enhanced monitoring and perioperative interventions is challenging.[8] The ARISCAT model[18] remains the most validated in the literature and showed the highest discrimination in the RECON dataset, with potential clinical utility in identifying patients unlikely to develop postoperative pulmonary complications according to the StEP-COMPAC definition. However, because of the paucity of calibration observed in this study or previous external validation studies, no scores can be currently recommended for routine clinical use in a broad cohort of patients undergoing major abdominal surgery. As surgery is upscaled to meet the growing demands on health-care systems worldwide,[66] an urgent clinical need exists to better differentiate between modifiable risk and fixed risk of pulmonary complications and develop validated prediction tools that are clinically relevant to a global context.

Contributors

The writing group (OK, VM, SB, WAC, MK, SQS, SRK, SKK, KAM, and JCG), external advisory group (AB, BC-B, ME, EMH, ML, DN, TP, PP, RP, TR, NS, and RV), and data analysis group (KAM, OK, SRK, and JCG) contributed to writing, data interpretation, and critical revision of the manuscript. KAM, OK, and JCG designed the study. SKK, WAC, VM, OK, SB, SQS, and MK did the literature search, acquired data, screened records, extracted data, and assessed risk of bias with oversight from KAM and JCG. The steering committee, external advisory group, and regional leads contributed to conception, protocol development, study delivery, and management of the RECON study. The collaborators contributed to data collection and study governance across included sites. All members of the writing group had full access to the data in the study, and the data analysis group (KAM, OK, SRK, and JCG) verified the underlying data in the study. The senior authors (JCG and KAM) and the writing committee had final responsibility for the decision to submit for publication. Detailed role descriptions of all contributing authors are shown in the appendix (pp 14–25).

Declaration of interests
RP has previously received research grants from Intersurgical. All other authors declare no competing interests.

Data sharing
Data sharing requests will be considered by the management group upon written request to the corresponding author. If agreed, deidentified participant data will be made available, subject to a data sharing agreement.

Acknowledgments
This study was funded through support from the British Journal of Surgery Society. This study has been presented at four meetings to date: Association of Surgeons in Training Surgical Summit, National Research Collaborative Meeting 2020, Surgical Research Society 2021 Virtual Annual Meeting, and Surgical Research Society of Southern Africa 48th Surgical Research Society Meeting.

References

1 Miskovic A, Lumb AB. Postoperative pulmonary complications. *Br J Anaesth* 2017; **118:** 317–34.
2 Kirmeier E, Eriksson LI, Lewald H, et al. Post-anaesthesia pulmonary complications after use of muscle relaxants (POPULAR): a multicentre, prospective observational study. *Lancet Respir Med* 2019; **7:** 129–40.
3 Abbott T, Fowler AJ, Pelosi P, et al. A systematic review and consensus definitions for standardised end-points in perioperative medicine: pulmonary complications. *Br J Anaesth* 2018; **120:** 1066–79.
4 Patel K, Hadian F, Ali A, et al. Postoperative pulmonary complications following major elective abdominal surgery: a cohort study. *Perioper Med* 2016; **5:** 10.
5 Khuri SF, Henderson WG, DePalma RG, et al. Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. *Ann Surg* 2005; **242:** 326–43.
6 Fernandez-Bustamante A, Frendl G, Sprung J, et al. Postoperative pulmonary complications, early mortality, and hospital stay following noncardiothoracic surgery: a multicenter study by the perioperative research network investigators. *JAMA Surgery* 2017; **152:** 157–66.
7 STARSurg Collaborative and COVIDSurg Collaborative. Death following pulmonary complications of surgery before and during the SARS-CoV-2 pandemic: a comparative analysis of two prospective international cohort studies. *Br J Surg* 2021; **108:** 1448–64.
8 Odor PM, Bampoe S, Gilhooly D, Creagh-Brown B, Moonesinghe SR. Perioperative interventions for prevention of postoperative pulmonary complications: systematic review and meta-analysis. *BMJ* 2020; **368:** m540.
9 Nijbroek SG, Schultz MJ, Hemmes SNT. Prediction of postoperative pulmonary complications. *Curr Opin Anaesthesiol* 2019; **32:** 443–51.
10 Peden CJ, Stephens T, Martin G, et al. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery (EPOCH): a stepped-wedge cluster-randomised trial. *Lancet* 2019; **393:** 2213–21.
11 Dutoit L, Kluyts HL, Gobin V, et al. The African surgical outcomes-2 (ASOS-2) pilot trial, a mixed-methods implementation study. *South Afr J Anaesth Analg* 2019; **25:** 14–31.
12 Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999; **130:** 515–24.
13 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; **339:** b2535.
14 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015; **13:** 1.
15 Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019; **170:** 51–58.
16 STARSurg Collaborative. REspiratory COmplications after abdomiNal surgery (RECON): study protocol for a multi-centre, observational, prospective, international audit of postoperative pulmonary complications after major abdominal surgery. *Br J Anaesth* 2020; **124:** e13–16.
17 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42:** 377–81.
18 Canet J, Gallart L, Gomar C, et al. Prediction of postoperative pulmonary complications in a population-based surgical cohort. *Anesthesiology* 2010; **113:** 1338–50.
19 Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. *Trials* 2017; **18:** 280.
20 Myles PS, Grocott MPW, Boney O, et al. Standardizing end points in perioperative trials: towards a core and extended outcome set. *Br J Anaesth* 2016; **116:** 586–89.
21 Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med* 2003; **29:** 1043–51.
22 Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019; **17:** 230.
23 McLean K, Ots R, Drake TM, Harrison EM. CollaboratoR: scalable multi-centre research using R and REDCap. 2020. https://github.com/kamclean/collaborator (accessed Dec 18, 2020).
24 Brooks-Brunn JA. Validation of a predictive model for postoperative pulmonary complications. *Heart Lung* 1998; **27:** 151–58.
25 Fuso L, Cisternino L, Di Napoli A, et al. Role of spirometric and arterial gas data in predicting pulmonary complications after abdominal surgery. *Respir Med* 2000; **94:** 1171–76.
26 McAlister FA, Bertsch K, Man J, Bradley J, Jacka M. Incidence of and risk factors for pulmonary complications after nonthoracic surgery. *Am J Respir Crit Care Med* 2005; **171:** 514–17.
27 Scholes RL, Browning L, Sztendur EM, Denehy L. Duration of anaesthesia, type of surgery, respiratory co-morbidity, predicted VO2max and smoking predict postoperative pulmonary complications after upper abdominal surgery: an observational study. *Aust J Physiother* 2009; **55:** 191–98.
28 Smith PR, Baig MA, Brito V, Bader F, Bergman MI, Alfonso A. Postoperative pulmonary complications after laparotomy. *Respiration* 2010; **80:** 269–74.
29 Blum JM, Stentz MJ, Dechert R, et al. Preoperative and intraoperative predictors of postoperative acute respiratory distress syndrome in a general surgical population. *Anesthesiology* 2013; **118:** 19–29.
30 Yang CK, Teng A, Lee DY, Rose K. Pulmonary complications after major abdominal surgery: National Surgical Quality Improvement Program analysis. *J Surg Res* 2015; **198:** 441–49.
31 Russotto V, Sabaté S, Canet J. Development of a prediction model for postoperative pneumonia: a multicentre prospective observational study. *Eur J Anaesthesiol* 2019; **36:** 93–104.
32 Baba H, Tokai R, Hirano K, et al. Risk factors for postoperative pneumonia after general and digestive surgery: a retrospective single-center study. *Surg Today* 2020; **50:** 460–68.
33 Brooks-Brunn JA. Predictors of postoperative pulmonary complications following abdominal surgery. *Chest* 1997; **111:** 564–71.
34 Arozullah AM, Daley J, Henderson WG, Khuri SF. Multifactorial risk index for predicting postoperative respiratory failure in men after major noncardiac surgery. The National Veterans Administration Surgical Quality Improvement Program. *Ann Surg* 2000; **232:** 242–53.
35 Arozullah AM, Khuri SF, Henderson WG, Daley J. Development and validation of a multifactorial risk index for predicting postoperative pneumonia after major noncardiac surgery. *Ann Intern Med* 2001; **135:** 847–57.
36 McAlister FA, Khan NA, Straus SE, et al. Accuracy of the preoperative assessment in predicting pulmonary risk after nonthoracic surgery. *Am J Respir Crit Care Med* 2003; **167:** 741–44.
37 Johnson RG, Arozullah AM, Neumayer L, Henderson WG, Hosokawa P, Khuri SF. Multivariable predictors of postoperative respiratory failure after general and vascular surgery: results from the patient safety in surgery study. *J Am Coll Surg* 2007; **204:** 1188–98.
38 Gupta H, Gupta PK, Fang X, et al. Development and validation of a risk calculator predicting postoperative respiratory failure. *Chest* 2011; **140:** 1207–15.
39 Kor DJ, Warner DO, Alsara A, et al. Derivation and diagnostic accuracy of the surgical lung injury prediction model. *Anesthesiology* 2011; **115:** 117–28.

40 Ramachandran SK, Nafiu OO, Ghaferi A, Tremper KK, Shanks A, Kheterpal S. Independent predictors and outcomes of unanticipated early postoperative tracheal intubation after nonemergent, noncardiac surgery. *Anesthesiology* 2011; **115:** 44–53.

41 Hua M, Brady JE, Li G. A scoring system to predict unplanned intubation in patients having undergone major surgical procedures. *Anesth Analg* 2012; **115:** 88–94.

42 Brueckmann B, Villa-Uribe JL, Bateman BT, et al. Development and validation of a score for prediction of postoperative respiratory complications. *Anesthesiology* 2013; **118:** 1276–85.

43 Gupta H, Gupta PK, Schuller D, et al. Development and validation of a risk calculator for predicting postoperative pneumonia. *Mayo Clin Proc* 2013; **88:** 1241–49.

44 Jeong BH, Shin B, Eom JS, et al. Development of a prediction rule for estimating postoperative pulmonary complications. *PLoS One* 2014; **9:** e113656.

45 Kor DJ, Lingineni RK, Gajic O, et al. Predicting risk of postoperative lung injury in high-risk surgical patients: a multicenter cohort study. *Anesthesiology* 2014; **120:** 1168–81.

46 Canet J, Sabaté S, Mazo V, et al. Development and validation of a score to predict postoperative respiratory failure in a multicentre European cohort: a prospective, observational study. *Eur J Anaesthesiol* 2015; **32:** 458–70.

47 Johnson AP, Altmark RE, Weinstein MS, Pitt HA, Yeo CJ, Cowan SW. Predicting the risk of postoperative respiratory failure in elective abdominal and vascular operations using the National Surgical Quality Improvement Program (NSQIP) participant use data file. *Ann Surg* 2017; **266:** 968–74.

48 Neto AS, da Costa LGV, Hemmes SNT, et al. The LAS VEGAS risk score for prediction of postoperative pulmonary complications: an observational study. *Eur J Anaesthesiol* 2018; **35:** 691–701.

49 Kawasaki K, Yamamoto M, Suka Y, et al. Development and validation of a nomogram predicting postoperative pneumonia after major abdominal surgery. *Surg Today* 2019; **49:** 769–77.

50 Takesue Y, Miyata H, Gotoh M, et al. Risk calculator for predicting postoperative pneumonia after gastroenterological surgery based on a national Japanese database. *Ann Gastroenterol Surg* 2019; **3:** 405–15.

51 Lukannek C, Shaefi S, Platzbecker K, et al. The development and validation of the Score for the Prediction of Postoperative Respiratory Complications (SPORC-2) to predict the requirement for early postoperative tracheal re-intubation: a hospital registry study. *Anaesthesia* 2019; **74:** 1165–74.

52 Mazo V, Sabaté S, Canet J, et al. Prospective external validation of a predictive score for postoperative pulmonary complications. *Anesthesiology* 2014; **121:** 219–31.

53 Jin Y, Xie G, Wang H, et al. Incidence and risk factors of postoperative pulmonary complications in noncardiac Chinese patients: a multicenter observational study in university hospitals. *Biomed Res Int* 2015; **2015:** 265165.

54 Shah N, Hamilton M. Clinical review: can we predict which patients are at risk of complications following surgery? *Crit Care* 2013; **17:** 226.

55 Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014; **14:** 40.

56 Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015; **68:** 25–34.

57 Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. *BJOG* 2017; **124:** 423–32.

58 Chu KM, Weiser TG. Real-world implementation challenges in low-resource settings. *Lancet Glob Health* 2021; **9:** e1341–42.

59 Student Audit and Research in Surgery (STARSurg) Collaborative. Impact of post-operative non-steroidal anti-inflammatory drugs on adverse events after gastrointestinal surgery. *Br J Surg* 2014; **101:** 1413–23.

60 Student Audit and Research in Surgery (STARSurg) Collaborative. Multicentre prospective cohort study of body mass index and postoperative complications following gastrointestinal surgery. *Br J Surg* 2016; **103:** 1157–72.

61 Student Audit and Research in Surgery (STARSurg) Collaborative. Prognostic model to predict postoperative acute kidney injury in patients undergoing major gastrointestinal surgery based on a national prospective observational cohort study. *BJS Open* 2018; **2:** 400–10.

62 EuroSurg Collaborative. Safety and efficacy of non-steroidal anti-inflammatory drugs to reduce ileus after colorectal surgery. *Br J Surg* 2020; **107:** e161–69.

63 GlobalSurg Collaborative and National Institute for Health Research Global Health Research Unit on Global Surgery. Global variation in postoperative mortality and complications after cancer surgery: a multicentre, prospective cohort study in 82 countries. *Lancet* 2021; **397:** 387–97.

64 COVIDSurg Collaborative. Mortality and pulmonary complications in patients undergoing surgery with perioperative SARS-CoV-2 infection: an international cohort study. *Lancet* 2020; **396:** 27–38.

65 COVIDSurg Collaborative. Effect of COVID-19 pandemic lockdowns on planned cancer surgery for 15 tumour types in 61 countries: an international, prospective, cohort study. *Lancet Oncol* 2021; **22:** 1507–17.

66 COVIDSurg Collaborative. Elective surgery cancellations due to the COVID-19 pandemic: global predictive modelling to inform surgical recovery plans. *Br J Surg* 2020; **107:** 1440–49.