Risk stratification in high-risk patients undergoing lung resection

Marcus Taylor[1], Stuart W. Grant[2], Glen P. Martin[3]

[1]Department of Cardiothoracic Surgery, Manchester University Hospital Foundation Trust, Wythenshawe Hospital, Manchester, UK; [2]Division of Cardiovascular Sciences, University of Manchester, ERC, Manchester University Hospitals Foundation Trust, Manchester, UK; [3]Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, Manchester Academic Heath Science Centre, University of Manchester, Manchester, UK

Contributions: (I) Conception and design: M Taylor; (II) Administrative support: M Taylor; (III) Provision of study materials or patients: M Taylor; (IV) Collection and assembly of data: M Taylor; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Marcus Taylor. Department of Cardiothoracic Surgery, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, M23 9LT, UK. Email: Marcus.taylor1@nhs.net.

Background: Although guidelines advocate using clinical prediction models (CPMs) to estimate mortality after lung resection, analysis of model performance in patients with pre-defined high-risk characteristics is lacking. The aim was to validate three existing models in subgroups of patients with high-risk characteristics.

Methods: The Thoracoscore, RESECT-90 and Safi models were validated using a database of 3,426 patients undergoing lung resection between 2012 and 2019 in a single UK centre. Models were validated for the outcome for which they were developed. Validation was performed for the overall cohort and for five high-risk subgroups [age ≥80 years, predicted post-operative (PPO) forced expiratory volume in 1 second (FEV1) <40%, PPO diffusion capacity of the lung for carbon monoxide (DLCO) <40%, Performance Status (PS) ≥2 and extended resection (bilobectomy, sleeve resection, pneumonectomy)]. The area under the receiver operating characteristic curve (AUC) and observed to expected (O:E) ratio were used to assess model discrimination and calibration.

Results: Overall in-hospital, peri-operative and 90-day mortality rates were 1.8% (n=63), 2.2% (n=76) and 3.3% (n=113) respectively. In total, 28.7% (n=982) of patients were included in at least one of the high-risk categories. Only the RESECT-90 model demonstrated acceptable performance for the overall cohort. Overall calibration was acceptable in four, four and two of the subgroups for the Thoracoscore, RESECT-90 and Safi models, respectively.

Conclusions: Discriminatory ability of models is diminished when applied to homogeneous subgroups. In this study, only the RESECT-90 model emerged as potentially useful for estimating mortality in high-risk patients being considered for lung resection but further validation work is required.

Keywords: Lung resection; 90-day mortality; RESECT-90 model; risk models; high-risk surgery

Received: 16 May 2022; Accepted: 18 July 2022; Published: 30 July 2022.
doi: 10.21037/shc-22-27
View this article at: https://dx.doi.org/10.21037/shc-22-27

Introduction

Existing guidelines highlight the need for a robust pre-operative assessment process for all patients being considered for lung resection (1). This is of particular importance in the contemporary era as alternatives to lung resection such as stereotactic ablative body radiotherapy (SABR) have demonstrated promising results for early-stage lung cancer in patients who are either anatomically unsuitable for resection or at a prohibitively high risk of peri-operative complications (2).

Multiple clinical prediction models (CPMs) have been developed for predicting the risk of short-term mortality after lung resection (3,4). A large number (>50) of variables have been included in these models but there are several key risk factors that are present in the majority of existing models. These include advanced age, poor functional status, abnormal pulmonary function tests, abnormal physiology and more extensive resection (3). Patients with these risk factors would generally be identified as high-risk by clinicians and the majority of CPMs.

CPMs are usually developed from overall cohorts of patients across a wide spectrum of risk. Robust assessment of CPM performance may include evaluation in clinically relevant subgroups of the overall population (5). These clinically relevant subgroups may include low or high-risk groups. CPM model performance at the extremes of risk can be inadequate and this may not always be identified if only overall model validation metrics are used (5).

Given evolving treatment alternatives, there remains a pressing need for the ability to accurately predict peri-operative mortality for patients with high-risk features being considered for lung resection. Therefore, the aim of this study was to analyse outcomes and assess risk model performance in a contemporary cohort of patients deemed to be high risk, according to the presence of one or more risk factors which have previously emerged as being associated with adverse peri-operative outcomes in patients undergoing lung resection. We present the following article in accordance with the STROBE reporting checklist (https://shc.amegroups.com/article/view/10.21037/shc-22-27/rc).

Methods

Patients

All consecutive patients who underwent lung resection for primary non-small cell lung cancer (NSCLC) between January 2012 and December 2019 at Manchester University NHS Foundation Trust were included. All cases of NSCLC were confirmed pathologically, and post-operative staging was assigned based on the post-operative histological analysis according to the 8th edition of the Tumour Node Metastasis Classification for Lung Cancer (6).

Data

Our data collection methods have been detailed in previous publications (4). Variables with more than 15% of data missing were excluded. Missing categorical data were imputed based on an assumption that missingness was equal to absent, whilst missing continuous data was replaced with either the mean (for normally distributed data) or median (for non-normally distributed data) value. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). All data were cleaned and stored in the Northwest Clinical Outcomes Research Registry (NCORR) (IRAS 260294). The NCORR database has full ethical approval from the regional Research Ethics Committee of the Health Research Authority. This project was approved by the NCORR steering committee and individual patient consent was waived.

Defining the high-risk patient

Deeming a patient to be high-risk is a subjective process. Nevertheless, there are studies and guidelines in the literature which provide some consensus regarding those patients likely to have a higher peri-operative risk of mortality and morbidity (1,7). Consequently, for this study we have identified several different subgroups of high-risk patients based upon their fulfilment of any of the following criteria:

- Age ≥80 years;
- Predicted post-operative (PPO) forced expiratory volume in 1 second (FEV1) <40%;
- PPO diffusion capacity of the lung for carbon monoxide (DLCO) <40%;
- Extended resection (bilobectomy, sleeve lobectomy, pneumonectomy);
- Performance Status (PS) score ≥2.

Data and outcomes

Continuous variables were presented as mean and standard deviation (SD), or median and interquartile range (IQR) for normal and non-normally distributed variables, respectively. Discrete variables were presented as percentages. Normality of distribution was assessed visually using histograms and statistically using the Kolmogorov-Smirnov test.

The primary outcomes were peri-operative, in-hospital and 90-day mortality. Peri-operative mortality is a composite endpoint comprising both in-hospital and 30-day mortality, whilst 90-day mortality is becoming increasingly recognised as the most important measure of short-term mortality after lung resection (8). These three primary outcomes are the three endpoints against which the risk models validated in this study were developed to predict.

Risk model performance

The three risk models analysed in this study were the Thoracoscore (9), the RESECT-90 score (10) and the Safi model (11). Model performance was assessed for the overall cohort (comprising all high risk and non-high-risk patients) and also separately for each of the five high-risk subgroups (as defined above). Model performance was assessed using measures of discrimination and calibration. Model discrimination was assessed by calculating the area under the receiver operating characteristic curve (AUC). An AUC of >0.7 was deemed to represent acceptable discrimination with values >0.8 deemed to represent excellent discrimination.

Model calibration was measured using observed to expected (O:E) ratios. An O:E ratio above 1 represents that model systematically under-estimates risk (and vice versa) (12). Validating each high-risk subgroup separately complies with the principles of strong calibration, as outlined by Van Calster *et al.*, whereby analysing the correlation between predicted risks and observed event rates for multiple covariate patterns represents the strongest measure of model calibration (5).

Given that the majority of data included in this study (all patients undergoing surgery between 2012 and 2018) were part of the dataset used to develop the RESECT-90 model, bootstrapping (1,000 iterations) was performed to adjust for in-sample optimism with regards to performance of the RESECT-90 model in this study.

All tests were 2-sided and statistical significance was defined as P value <0.05. All statistical analysis was undertaken using SPSS version 28 (SPSS, Inc., Chicago, IL, USA).

Results

Patient characteristics

During the study period 3,426 patients underwent surgery, of whom 28.7% (n=982) were included in at least one of the high-risk categories and were therefore defined as high risk. The overall mean age was 68.3 years (±9.0 years) and 47.3% (n=1,621) were male. Complete patient characteristics are shown in *Table 1*. The overall in-hospital, peri-operative and 90-day mortality rates were 1.8% (n=63), 2.2% (n=76) and 3.3% (n=113) respectively. No patients were lost to follow-up.

Risk model performance: Thoracoscore model and peri-operative mortality

Discrimination and calibration of the Thoracoscore model for the overall cohort were both inadequate (AUC 0.65, 95% CI: 0.58–0.72 and O:E ratio 0.66, P<0.001). Whilst discrimination was inadequate for all five subgroups (range, 0.55–0.67), calibration was acceptable for four of the five subgroups.

Risk model performance: RESECT-90 model and 90-day mortality

Discrimination and calibration of the RESECT-90 model for the overall cohort were both acceptable (AUC 0.72, 95% CI: 0.67–0.77 and O:E ratio 1.01, P=0.886). Discrimination was acceptable for two of the five subgroups and calibration was acceptable for four of the five subgroups.

Risk model performance: Safi model and in-hospital mortality

Discrimination and calibration of the Safi model for the overall cohort were both inadequate (AUC 0.65, 95% CI: 0.58–0.71 and O:E ratio 0.38, P<0.001). Discrimination was inadequate for four of the five subgroups and calibration was acceptable for two of the five subgroups.

The results of the model validation are displayed in *Table 2*.

Discussion

This study has analysed outcomes for patients deemed to be high-risk according to the presence of one or more risk factors which have previously emerged as being associated with adverse peri-operative outcomes. With regards to model validation, only the RESECT-90 model demonstrated acceptable model performance when applied to the cohort as a whole.

When analysing subgroups of a population in which a model was developed for, a lower AUC value is generally to be expected. This is because the act of creating a subgroup leads to a more homogenous cohort, potentially removing important discriminatory risk factors leading to reduced discrimination overall. For that reason, when assessing model performance in subgroups, model calibration performance may be more clinically relevant. With the exception of a single subgroup (extended resection), the

Table 1 Patient characteristics

Variable	Value	Missing data (%)
Age (years) (mean ± SD)	68.3 (±9.0)	0
Male sex	47.3% (n=1,621)	0
ASA score (median, IQR)	3.0 (2.0–3.0)	1.1
PS score (median, IQR)	1.0 (0–1.0)	2.2
% Predicted FEV1 (mean ± SD)	86.5% (±21.0%)	6.7
% Predicted DLCO (mean ± SD)	72.4% (±16.5%)	14.3
BMI (kg/m^2) (mean ± SD)	26.6 (±4.9)	10.9
Creatinine (µmol/L) (median, IQR)	72.0 (64.0–82.6)	11.4
Anaemia	23.4% (n=803)	11.9
Diabetes mellitus	13.3% (n=456)	2.0
Hypercholesterolaemia	17.4% (n=596)	2.0
Hypertension	37.5% (n=1,286)	2.0
Smoking	80.8% (n=2,767)	2.0
Arrhythmia	6.2% (n=212)	4.6
Ischaemic heart disease	14.3% (n=490)	4.6
COPD	32.5% (n=1,113)	2.5
Cerebrovascular disease	7.4% (n=253)	3.2
Peripheral vascular disease	6.4% (n=220)	1.1
Right-sided resection	61.4% (n=2,105)	0
Resected segments (mean ± SD)	4.0 (±1.8)	0
Thoracotomy	81.2% (n=2,782)	0
Extent of resection		
Complex lobectomy	8.0% (n=273)	0
Pneumonectomy	5.2% (n=179)	0
Risk scores		
Thoracoscore (mean ± SD)	2.8% (±2.0%)	0
RESECT-90 (mean ± SD)	3.4% (±3.4%)	0
High-risk characteristics		
Age ≥80 years	8.6% (n=296)	0
PPO FEV1 <40%	6.0% (n=204)	6.7
PPO DLCO <40%	12.0% (n=410)	14.3
BMI <18.5 kg/m^2	3.1% (n=107)	10.9
Extended resection	13.2% (n=452)	0
PS ≥2	8.0% (n=273)	2.2

SD, standard deviation; ASA, American Society of Anesthesiologists; IQR, interquartile range; PS, performance status; FEV1, forced expiratory volume in 1 second; DLCO, diffusion capacity of the lung for carbon monoxide; BMI, body mass index; Anaemia, anaemia defined as haemoglobin <120 g/L for women and <130 g/L for men as per World Health Organisation classifications; COPD, chronic obstructive pulmonary disease; Complex lobectomy, bilobectomy or sleeve lobectomy or chest wall resection; PPO, predicted post-operative; Extended resection, complex lobectomy or pneumonectomy.

Table 2 Risk model validation

Model	Discrimination			Calibration	
	AUC	95% CI, lower	95% CI, higher	O:E ratio	P value
Thoracoscore					
Overall cohort	0.65	0.58	0.72	0.66	<0.001
Age ≥80 years	0.54	0.37	0.71	1.41	0.075
PPO FEV1 <40%	0.56	0.33	0.78	0.50	<0.001
PPO DLCO <40%	0.55	0.44	0.66	1.00	0.928
Extended resection	0.67	0.56	0.77	0.83	0.252
PS score 2	0.58	0.40	0.76	0.85	0.377
RESECT-90					
Overall cohort	0.72	0.67	0.77	1.01	0.886
Age ≥80 years	0.59	0.41	0.78	1.08	0.570
PPO FEV1 <40%	0.58	0.40	0.76	0.92	0.572
PPO DLCO <40%	0.69	0.59	0.80	0.89	0.307
Extended resection	0.64	0.53	0.74	1.29	0.031
PS score 2	0.79	0.67	0.91	1.09	0.571
Safi					
Overall cohort	0.65	0.58	0.71	0.38	<0.001
Age ≥80 years	0.49	0.34	0.64	0.51	<0.001
PPO FEV1 <40%	0.59	0.34	0.83	0.61	0.016
PPO DLCO <40%	0.73	0.62	0.85	0.81	0.199
Extended resection	0.50	0.38	0.61	0.86	0.388
PS score 2	0.55	0.38	0.73	0.63	0.007

AUC, area under the curve; CI, confidence intervals; O:E, observed to expected; PPO, predicted post-operative; FEV1, forced expiratory volume in 1 second; DLCO, diffusion capacity of the lung for carbon monoxide; PS, Performance Status.

RESECT-90 model demonstrated acceptable calibration for all high-risk cohorts analysed. The Thoracoscore and Safi models demonstrated acceptable calibration in four and two of the subgroups, respectively.

The Thoracoscore was published in 2007 and developed from 15,183 patients undergoing all forms of thoracic surgery in multiple centres across France between 2002 and 2005. The primary endpoint was a composite endpoint comprising in-hospital and 30-day mortality. It includes several variables including age, PS score and undergoing pneumonectomy but does not include any measures of pulmonary function. It is advocated in guidelines for lung resection published by both the British Thoracic Society (1) and the National Institute for Clinical Excellence, however several studies demonstrating inadequate model performance outside of the initial patient cohort have since been published (4,13).

Given that the Thoracoscore model was developed to be applied to patients undergoing all forms of thoracic surgery, it is perhaps not surprising that model performance was inadequate for this cohort comprised solely of high-risk subgroups of patients undergoing lung resection. These results are however in keeping with other studies suggesting that the model is not suitable for use as a risk stratification tool in contemporary thoracic surgery practice.

The RESECT-90 model was developed from 6,600 patients undergoing lung resection in two UK centres between 2012 and 2018 with an outcome metric of 90-day mortality. Internal validation of the model demonstrated acceptable model performance (10) however no external validations of the model have yet been published. The model is comprised of twelve variables including age, PS score, DLCO and number of resected bronchopulmonary segments, all expressed as continuous variables. In this study, the model demonstrated acceptable performance for the cohort as a whole. Whilst its subgroup discriminatory ability was somewhat varied (acceptable for two of the five subgroups), calibration was adequate for four of the five high-risk subgroups, suggesting that the model has potential to be a useful risk stratification tool in high-risk patient cohorts.

Pneumonectomy is recognised as a procedure carrying particularly high peri-operative mortality, with contemporary studies reporting a 90-day mortality rate of around 10% (14). Hence, the Safi model, designed specifically to predict in-hospital mortality for patients undergoing pneumonectomy, can be considered to be the only existing model, to the best of our knowledge, developed solely for high-risk patients (11). Comprising five variables (age, alcohol use, pre-operative white cell count, coronary artery disease and undergoing pneumonectomy as a palliative procedure), internal validation demonstrated acceptable model performance. However, there are a number of concerns with regards to model development [limited number of deaths, an outdated method of internal validation (split sample approach) and the inclusion of variables not clinically relevant in the contemporary era (pneumonectomy is now rarely performed as a palliative procedure)] for the Safi model. These factors, when coupled with the model's poor performance in both a recent external validation study (14) and also in this study, mean that the Safi model cannot be recommended as a risk stratification tool for patients undergoing either pneumonectomy or other forms of lung resection. A study validating the performance of the Thoracoscore model in a cohort of patients undergoing pneumonectomy found the model's performance to be similarly inadequate (15).

This study has a number of limitations. The standard of a retrospective study is defined by the quality and missingness of data available for analysis. Although below the threshold for exclusion, the relatively high rate of missing data for DLCO, a key variable, is a drawback of this work. Defining the high-risk cohort, whilst based on recognised risk factors, remains a subjective exercise which may not meet with universal agreement. Nevertheless, our results have been derived from a relatively large and contemporary dataset reflective of thoracic surgery activity in the United Kingdom.

Conclusions

An important proportion of patients undergoing lung resection have characteristics which are clinically recognised as being associated with increased peri-operative risks. This study has shown that, from the three existing CPMs validated, only the RESECT-90 model can be considered a potentially useful tool when attempting to risk stratify patients within these groups. Given the increased availability and efficacy of non-surgical treatments for early-stage lung cancer, further work is required in larger cohorts to improve the ability of clinicians to robustly risk stratify high-risk lung cancer patients who are potentially suitable for lung resection.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Shanghai Chest* for the series "Thoracic Surgery in High Risk Patients". The article has undergone external peer review.

Reporting Checklist: The authors have completed the STROBE reporting checklist. Available at https://shc.amegroups.com/article/view/10.21037/shc-22-27/rc

Data Sharing Statement: Available at https://shc.amegroups.com/article/view/10.21037/shc-22-27/dss

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://shc.amegroups.com/article/view/10.21037/shc-22-27/coif). The series "Thoracic Surgery in High Risk Patients" was commissioned by the editorial office without any funding or sponsorship. MT served as the unpaid Guest Editor of the series and serves as an unpaid editorial board member of *Shanghai Chest*. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the NCORR steering committee (NCORR database IRAS 260294) and individual consent for this retrospective analysis was waived.

References

1. Lim E, Baldwin D, Beckles M, et al. Guidelines on the radical management of patients with lung cancer. Thorax 2010;65 Suppl 3:iii1-27.
2. Bahig H, Chen H, Louie AV. Surgery versus SABR for early stage non-small cell lung cancer: the moving target of equipoise. J Thorac Dis 2017;9:953-6.
3. Taylor M, Hashmi SF, Martin GP, et al. A systematic review of risk prediction models for perioperative mortality after thoracic surgery. Interact Cardiovasc Thorac Surg 2021;32:333-42.
4. Taylor M, Szafron B, Martin GP, et al. External validation of six existing multivariable clinical prediction models for short-term mortality in patients undergoing lung resection. Eur J Cardiothorac Surg 2021;59:1030-6.
5. Van Calster B, Nieboer D, Vergouwe Y, et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016;74:167-76.
6. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J Thorac Oncol 2016;11:39-51.
7. Salati M, Brunelli A. Risk Stratification in Lung Resection. Curr Surg Rep 2016;4:37.
8. Taylor M, Grant SW, West D, et al. 90-day mortality: re-defining the peri-operative period after lung resection. Clinical Lung Cancer 2021;22:e642-5.
9. Falcoz PE, Conti M, Brouchet L, et al. The Thoracic Surgery Scoring System (Thoracoscore): risk model for in-hospital death in 15,183 patients requiring thoracic surgery. J Thorac Cardiovasc Surg 2007;133:325-32.
10. Taylor M, Martin GP, Abah U, et al. Development and internal validation of a clinical prediction model for 90-day mortality after lung resection: the RESECT-90 score. Interact Cardiovasc Thorac Surg 2021;33:921-7.
11. Grant SW, Collins GS, Nashef SAM. Statistical Primer: developing and validating a risk prediction model. Eur J Cardiothorac Surg 2018;54:203-8.
12. NICE Guideline [NG122] Lung Cancer: Diagnosis and Management; 2019.
13. Brunswicker A, Taylor M, Grant SW, et al. Pneumonectomy for primary lung cancer: contemporary outcomes, risk factors and model validation. Interact Cardiovasc Thorac Surg 2022;34:1054-61.
14. Safi S, Benner A, Walloschek J, et al. Development and validation of a risk score for predicting death after pneumonectomy. PLoS One 2015;10:e0121295.
15. Qadri SS, Jarvis M, Ariyaratnam P, et al. Could Thoracoscore predict postoperative mortality in patients undergoing pneumonectomy? Eur J Cardiothorac Surg 2014;45:864-9.

doi: 10.21037/shc-22-27
Cite this article as: Taylor M, Grant SW, Martin GP. Risk stratification in high-risk patients undergoing lung resection. Shanghai Chest 2022;6:21.

