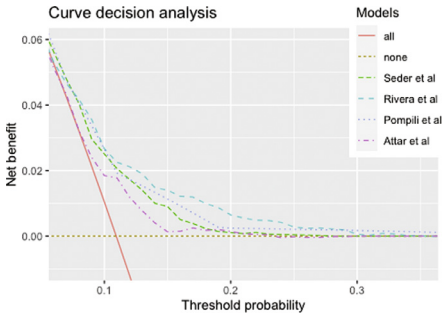
Predicting a Prolonged Air Leak After Video-Assisted Thoracic Surgery, Is It Really Possible?

Francesco Zaraca, MD,[*] Marco Pipitone, MD,[*] Birgit Feil, MD,[*] Reinhold Perkmann, MD,[*] Luca Bertolaccini, MD, FCCP,[†] Carlo Curcio, MD,[‡] and Roberto Crisci, MD,[§] the Italian VATS group [#]

Validation of predictive risk models for prolonged air leak (PAL) is essential to understand if they can help to reduce its incidence and complications. This study aimed to evaluate both the clinical and statistical performances of 4 existing models. We selected 4 predictive PAL risk models based on their scientific relevance. We referred to these models as Chicago, Bordeaux, Leeds and Pittsburgh model, respectively, according to the affiliation place of the first author. These predicting risk models were retrospectively applied to patients recorded on the second edition of the Italian Video-Assisted Thoracoscopic Surgery Group registry. Predictions for each patient were calculated based on the logistic regression coefficient values provided in the original manuscripts. All models were tested for their overall performance, discrimination, and calibration. We recalibrated the original models with the re-estimation of the model intercept and slope. We used curve decision analysis to describe and compare the clinical effects of the studied risk models. Better statistical metrics characterize the models developed on larger populations (Chicago and Bordeaux models). However, no model has a valid benefit for threshold probability greater than 0.30. The Net benefit of the most performing model (Bordeaux model) at the threshold probability of 0.11 is 23 of 1000 patients, burdened by 333 false positive cases. One of 1000 is the Net benefit at the threshold probability of 0.3. The use of PAL scores based on preoperative predictive factors cannot be currently used in a clinical setting because of a high false positive rate and low positive predictive value.

Semin Thoracic Surg 33:581–592

Keywords: Video-assisted thoracic surgery, VATS lobectomy, Prolonged air leak, Risk factors



Decision curve analysis of the scores of the Chicago, Bordeaux, Leeds, and Pittsburgh models for diagnosing 5-day prolonged air leak (PAL). The 4 curves are compared to the curves of treating none and all patients. It shows that all models have almost no net benefit as soon as the threshold probability of 0.30 is exceeded. Predicting a prolonged air leak after Video-Assisted Thoracoscopic Surgery (VATS) based on preoperative predictive factors is possible but with a high rate of false positives.

Central Message

The use of PAL scores based on preoperative predictive factors is currently burdened by a high false positive rate and low positive predictive value.

Abbreviations: PAL, Prolonged air leak; VATS, Video-Assisted Thoracoscopic Surgery; ESTS, European Society of Thoracic Surgeons; FEV1, forced expiratory volume in 1 second; BMI, body mass index; %FEV1, percent forced expiratory volume in 1 second; ERAS, Enhanced Recovery After Surgery; MITS, Minimally Invasive Thoracic Surgery; GOLD, Global strategy for the diagnosis, management, and prevention of chronic Obstructive Lung Disease; COPD, Chronic Obstructive Pulmonary Disease; IAL, intraoperative air leak; LOS, Length of Stay; E_{max}, the average absolute difference in predicted vs. calibrated probabilities; E_{avg}, the maximum absolute difference in predicted vs calibrated probabilities

*Department of Vascular and Thoracic Surgery, Regional Hospital, Bolzano, Italy
[†]Division of Thoracic Surgery, IEO European Institute of Oncology IRCCS, Milan, Italy
[‡]Department of Thoracic Surgery, Monaldi Hospital, Naples, Italy
[§]Division of Thoracic Surgery, Thoracic Surgery Unit, University of L'Aquila, "G. Mazzini" Hospital, Teramo, Italy

Disclosure: The authors report no funding for the article, no proprietary or commercial interest in any product mentioned or concept discussed in this manuscript.

[#]The members of the Italian VATS group are listed in Appendix 1 at the end of the article.

Address reprint requests to Francesco Zaraca, MD, PhD, Privat-Dozent (UniKlinik Marburg, DE), Department of Vascular and Thoracic Surgery, Regional Hospital Bolzano, L. Böhler Street 5, 39100 Bolzano Italy. E-mail: francesco.zaraca@sabes.it

OBJECTIVES

Prolonged air leak (PAL) is a frequent complication after lung resection surgery, but its causes and management are still not clearly defined. The definition of PAL is an air leak that lasts beyond the average postoperative hospitalization time, which has been reduced by the development of minimally invasive techniques. The air leak duration for defining PAL varies between the different series published and ranges, most of the time, from the fifth to the seventh postoperative day.[1,2] PAL increases the risk of complications and has an impact on hospital stay and, therefore, on costs. Fundamental to prevent and manage PAL is the analysis of risk factors, a proper surgical technique, and a correct management of the chest tube. Therefore, a useful predictive risk model can help recognizing those patients who could benefit from supplemental preventive procedures. Among the recently published models, we have selected four predictive PAL risk models, based on the number of cases involved, enrollment period, and scientific relevance: models published by Rivera et al,[1] Pompili et al,[2] Seder et al,[3] and Attaar et al.[4] In this manuscript, we referred to the models as Bordeaux, Leeds, Chicago, and Pittsburgh model, respectively, according to the affiliation place of the first author.

This study aimed to evaluate the performance of these 4 models using a multicenter national-wide registry, the Italian Video-Assisted Thoracoscopic Surgery (VATS) group registry.

Perspective Statement

A predictive risk model for prolonged air leak (PAL) after VATS could help reducing its incidence. We demonstrated that the use of PAL predictive scores based on preoperative predictors adds a minimal net benefit compared to the strategy of treating all or no patients, but at the price of high false-positive rate. The low decision analysis performances of predictive scores question their value in a clinical setting.

METHODS

Source of Data

Four predicting models for PAL risk were retrospectively applied to patients recorded on the second edition of the Italian VATS Group registry.[5] Our cohort included patients that underwent a lobar or a sub-lobar resection between November 2015 and December 2019. The VATS group registry is an online database, created in 2014 by Roberto Crisci, prospectively collecting the data regarding VATS anatomic lung resections. Fifty-seven thoracic surgery units across Italy are involved in the registry voluntarily. The local ethical committee of each thoracic surgery unit reviewed and approved the submission of data to the database and written informed consent was obtained from each participant.

The Chicago group[3] analyzed data from the Society of Thoracic Surgeons General Thoracic Surgery database, collected between 2009 and 2016, and their sample included 52,198 procedures. The Bordeaux group[1] processed data collected in the French national thoracic database, Epithor (French Society of Thoracic and Cardiovascular Surgery), between 2004 and 2008, with 24,113 procedures. The Leeds model[2] used data registered in the European Society of Thoracic Surgeons database, from July 2007 to August 2015, with 5069 procedures. The patients in the Italian VATS group validation cohort were not included in the European Society of Thoracic Surgeons Registry, used to create the Leeds model. The Pittsburgh model[4] sample included patients operated at 8 hospital sites, from January 2009 to June 2014, with 2317 procedures. Collected variables in all these registries include information about the patient's medical history, characteristics, surgical procedures, and outcomes.

Participants: Eligibility and Exclusion Criteria

The eligibility criteria for the Chicago model included all patients who underwent elective thoracoscopic or thoracotomic lobectomy, bilobectomy, segmentectomy, or wedge resection for lung cancer. The Bordeaux model comprised wedge and volume reduction, besides lobectomy, bilobectomy, and segmentectomy. Both the surgical approaches were evaluated: VATS and thoracotomy. The Leeds model and our sample included exclusively patients that underwent VATS lobar or sublobar resections for benign or malignant diseases (Supplemental Table 1).

The Pittsburgh sample, compared to the Bordeaux sample, did not include volume reduction and bullectomy. Our patients, therefore, represents a subgroup of the Chicago, Bordeaux, and Pittsburgh models sample. Exclusion criteria of the Chicago model were age <18 years, pneumonectomy, chest wall or diaphragm resection, sleeve lobectomy or bilateral procedures. Pneumonectomies and explorative thoracotomies were excluded in the Bordeaux model. Exclusion criteria for the Leeds model were lung resection combined with diaphragm or chest wall resections; bronchial sleeve resections; pneumonectomies; sublobar resections; patients undergoing postoperative-assisted mechanical ventilation and patients from units contributing <20 procedures. In the Pittsburgh model were excluded patients with more than one surgery with only data of the most recent surgery; patients that underwent bronchial sleeve lobectomy, patients who died before the discharge, pneumonectomies, and extended chest wall/diaphragm resections.

The exclusion criteria were applied to our series according to the model we had to validate.

Outcome, Predictors, and Missing Data

We validate each model, adopting the 5-day PAL definition, that is, an air leak that lasts beyond the fifth postoperative day. The Italian VATS group registry allows to distinguish the daily amount of fluid and air leaks from thoracic drainage. All four original models share standard developing methods that include backward or forward stepwise logistic regression,

Table 1. Logistic Regression Coefficients of Predictors for Prolonged Air Leak and Their Associated Score Values, as Provided in the Original Manuscripts

Variable	Categories	Chicago		Bordeaux		Leeds		Pittsburgh	
		Coef.	Score	Coef.	Score	Coef.	Score	Coef.	Score
Intercept		*	-	−1.3406	-	−2.7	-	−3.137	-
Gender	Male sex	0.330	4	0.39	4	0.44	1		
BMI (Kg/m^2)	Linear			−0.109	24-BMI			−0.064	14.6-(0.36*BMI)
	<25	0.668	7						
	<18.5					0.96	2		
Smoker								0.445	2.50
Zubrod Score	>2							0.398	2.25
Dyspnea score	Linear			0.187					
FEV1, % predicted	≥60 and <80				6	0.4	1	0.377	2
	≤70	0.358	5						
	<60					0.4	1	0.801	4.5
	<50				10				
	Unmeasured							−0.967	5.5, if measured
Pleural adhesions	Present			0.366	4				
Pulmonary resection	Wedge resection			ref				−0.851	
	Lobectomy	0.668	6	0.717	7			ref	5.5
	Segmentectomy			0.717	7			ref	5.5
	Bilobectomy	0.668	6	1.06	13			0.575	8.6
	Bulla resection			0.2	2			-	
	Pulmonary volume reduction			1.43	14			-	
Location	Lower or middle lobe			ref					
	Upper lobe			0.425	4				
	Right upper lobe	0.336	3						
Side	Right side							0.044	
Thoracotomy (vs VATS)								−0.278	
Reoperation								0.572	3.25
Preoperative hospitalization ≥1d								−1.063	6.25
Interaction terms	Right-sided thoracotomy							1.068	6.26
	Wedge resection by thoracotomy							0.696	4.75

*Value not available in the original manuscript. PAL, Prolonged air leak; BMI, body mass index; FEV1, forced expiratory volume in 1 second; VATS, Video-Assisted Thoracoscopic Surgery.

bootstrap resampling or cross-validation for internal validation, and temporal validation as external validation.

Selected predictors, as well their associated coefficients and scores are shown in Table 1.

In our registry the dyspnea score, defined according to the Medical Research Council, is not available. Although dyspnea may be caused by deconditioning and other factors that could affect air leak beside spirometry, we decided to adopt the forced expiratory volume in 1 second (FEV1) value as a surrogate predictor. According to the Global strategy for the diagnosis, management, and prevention of chronic Obstructive Lung Disease classification of Chronic Obstructive Pulmonary Disease,[6] we categorized it into three groups: FEV1 more than 80%, considered no dyspnea, 0 points; FEV1 between 50% and 80% regarded as moderate dyspnea with 3 points, and a FEV1 less than 50% considered severe with 5 points.

Our register has undergone regular Data Quality Check of the data under the supervision of the Quality Committee, a biostatistician, and the engineers involved in maintenance of the database. Moreover, the database prevents the final validation of the case, if mandatory data are missing. Only after integration of the missing information the case is considered part of the register.[5] Among the variables set as nonmandatory, we decided to exclude variables with a proportion of missing data >10%. Missing data were imputed by choosing the most common value among the nonmissing values.

Statistical Analysis Methods

Predictions for each patient were calculated based on the logistic regression coefficient values provided in the original manuscripts and reported in Table 1. All models were tested for their overall performance, discrimination, and calibration.

The models' overall performance was evaluated using the Brier score. Discrimination was measured using the C statistic. Calibration was displayed using a calibration plot and was measured using E_{max} and E_{avg}, the maximum and the average difference in predicted versus loess calibrated probabilities. The 95% confidence intervals for these metrics were derived based on a 1000 bootstrap resampled replicates.[7] We recalibrated the original models following the procedure explained in Vergouwe.[8] We estimated 2 additional logistic regression models: recalibration in the large (re-estimation of model intercept); and logistic recalibration (re-estimation of intercept and slope). We did not estimate new coefficients for the whole set of variables used in the models, so-named "model revision" (re-estimation of all factors). A revised model is just like a newly developed model, and prior information from the derivation sample is disregarded.[9,10] The performance of the updated models was assessed using the metrics as mentioned above (Supplemental Fig. 1).

Score values for each patient were calculated and calibrated probabilities for each score value were inferred through a weighted loess regression. We used decision curve analysis and Net benefit to describe and compare the clinical effects of the four 5-day PAL risk models using the scores and their calibrated probabilities.[11–14] In the Net benefit, the threshold probability value identifies both the probability above which the patients are considered at high risk, and the ratio between true/false positives at which the advantages of a treatment are balanced by the disadvantages. The Net benefit is the number of patients beyond this ratio.

Since the incidence of PAL would probably be around 10%, a sample size of over 3000 patients should be suitable for statistical analyses. The study was approved by the Institutional Review Board of the Italian VATS group and by the Institutional Research Reviewer Board for data collection, transmission, storage and analysis (81/2014/O/Oss). All analyses were conducted using R language for statistical computing software,15 v.3.6.1 (July 5, 2019).[15]

RESULTS

Patients

The Italian VATS group registry recorded 3965 patients who underwent a VATS anatomic lung resection. Five-day PAL rate in this patient cohort was 11.0%. The baseline characteristics of patients both in our cohort and in the validated models' groups are summarized in Table 2. The median age of the study participants was 69 years (the interquartile range [IQR] 62–75), 88% underwent a lobectomy, 9.1% a segmentectomy, and 2.4% a bilobectomy. No wedge resection, volume reduction or bulla resection were recorded in the registry. We did not remove any patient because of missing data. In the FEV1 variable, 94 measurements were missing (2.4%) and were replaced by the median value of the category. No other values were missing. All surgeries were initially in video-assisted thoracoscopy and 342 patients (8.6%), were converted in open surgery. Adhesiolysis that is associated to high risk of parenchyma integrity impairment was performed in 1046 cases (26.4%). The median predicted FEV1 was 94% (IQR 80–107) and 633 patients (22%) had a predicted FEV1 lower than 80%. The median time with the chest tube was 4 days (IQR 3–5). The number of participants and the prevalence of PAL differ between models' cohorts. We had a 5-day PAL in up to 11% of cases, in contrast with the 10,4%, 9.9%, and 8.6% of the Chicago, Leeds, and Pittsburgh sample, respectively. Seven-day PAL occurred in 308 patients (7.7%) in our cohort, whereas it involved 6.9% of patients in the Bordeaux cohort.

Model Performance and Updating

The Chicago model was assessed after a recalibration in the large (re-estimation of model intercept), since the original intercept coefficient value was not published. The model is well calibrated in terms of slope, as shown by the minimum variation of the slope at logistic recalibration, the low values of E_{max} and E_{avg} and the nonsignificant variation of the model fit (LRT χ^2 P value) at recalibration. Model discrimination does not differ from the results obtained by the authors during internal validation, C statistics of 0.63 (0.60–0.65). As shown in Figure 1, the Bordeaux model regularly underestimates the probability of PAL events. The model was developed for 7-day postoperative PAL. After the re-estimation of model intercept, the Bordeaux model is well calibrated in terms of slope. Although developed on a different outcome, this model has the greatest discriminative ability among the models studied, C statistics 0.65(0.62–0.68). Leeds and Pittsburgh models perform worse in terms of calibration and discriminative ability, when applied to our sample. The Brier score fluctuates in all models between 0.1 and 0.09 without a significant difference among the models (Table 3 and Fig. 1).

Curve Decision Analysis

Table 4 shows how many patients out of 1000 would be recognized as being at high risk and how many of them really experienced a PAL postoperatively, based on different thresholds probabilities and different predicting model scores. Among the tested models, Bordeaux has the best net benefit for all the threshold probabilities greater than the prevalence of PAL (Fig. 2). Furthermore, the curve decision analysis shows that all models have almost no net benefit as soon as the threshold probability of 0.30 is exceeded. The net benefit of each predictive model, regarding threshold probabilities lower than the prevalence of disease, decreases when related to the hypothesis of treating all patients (Table 4, Fig 2).

For example, the Bordeaux score is able to identify 81 of 1000 patients with a risk of PAL greater than 0.2 in our cohort. Of the identified patients, 21 will develop PAL (true positives) and 60 will be false positives. Six patients exceed the neutral true positive threshold dictated by the proportion 60* (1:4) = 15, where 1:4 represents the odd ratio for this threshold probability. The effectiveness of this probability score, net of the ratio of false positive patients considered acceptable, is for

Table 2. Characteristics of Patients in the Different Cohorts

Variable	Categories	Italian VATS Group (2.0)	Chicago	Bordeaux	Leeds	Pittsburgh
Patients number		3965	52,198	24,113	5069	2317
5-day PAL		436 (11.0)	5453 (10.4)		504 (9.9)	200 (8.6)
7-day PAL		308 (7.7)		1655 (6.9)		
Age		69 (62–75)	68 (55–81)	59.2 (46–72)	64 (57–71)	65 (53–77)
Gender	Male sex	2379 (60)	23,759 (46)	16,717 (69)	2862 (56)	1021 (44)
BMI (Kg/m^2)		25.8 (23.3–28.3)	26.9 (19.9–33.9)	22.9 (18.5–26.4)	25.5 (22.5–28)	28 (21–35)
Smoking history	Never smoker	1204 (30)	6783 (13)			684 (30)
	Past smoker	1533 (38)	32,775 (63)			1180 (50)
	Current smoker	1228 (31)	12,640 (24)			453 (20)
Zubrod Score	≥2	1289 (32)	2203 (4.2)			390 (17)
Congestive heart failure		144 (3.6)				76 (4)
COPD		928 (23)				771 (33)
Diabetes		541 (13)	10,026 (19)		188 (3.7)	419 (18)
FEV1, % predicted	Linear	94 (80–107)	81 (55–107)		84.4 (72.9–96)	83 (61–105)
	≥60 and <80	735 (19.0)			1838 (36)	285(12)
	<60	160 (4.1)				561 (24)
	Missing	94 (2.4)				297 (13)
Pleural adhesions	Present	1046 (16)		5383 (22)		
Surgical approach	VATS vs Thoracotomy	3965 (100)	34,185 (65)	4803 (20)	5069 (100)	1714 (74)
Pulmonary resection	Bilobectomy	95 (2.4)	1661(3)	931 (3.9)	-	37 (1)
	Lobectomy	3509 (88)	39,660 (76)	13,100 (54)	5069 (100)	1500 (65)
	Segmentectomy	361 (9.1)	2978 (6)	955 (3.9)	-	
	Wedge resection	-	7899 (15)	7653 (32)	-	780 (34)
	Bullectomy or LVRS	-	-	1473 (6)	-	-
Location	Lower lobe	1396 (35)	17,535(33)	7039 (29)		
	Middle lobe	310 (7.8)	3307 (6)	1063 (4)		
	Upper lobe	2164 (55)	31,305 (60)	11,877 (50)	3215 (63)	
Side	Right side	2351 (59)	31,121 (60)	14,055 (58)	3035 (60)	1377 (59)
Pulmonary pathology	Malignant	3806 (96)		14,096 (58)		1950 (84)

Results are expressed as counts and percentages of patients for categorical variables, and as medians and interquartile ranges for numeric variables. PAL, Prolonged air leak; BMI, body mass index; FEV1, forced expiratory volume in 1 second; COPD, Chronic obstructive pulmonary disease, LVRS, Lung volume reduction surgery.

a threshold probability of 0.2 of 6 of 1000 patients (Net benefit). In order to make the Bordeaux score more effective in identifying high-risk patients, lower true/false positive ratios must be accepted. The Net benefit at the threshold probability of 0.11 is 23 of 1000 patients, but it will be burdened by 333 false positive cases.

The other three models have slightly lower Net benefits (Table 4).

DISCUSSION

Prolonged air leak is one of the most common complications after pulmonary resection. Although it is a condition of significant impact on hospital stay it is poorly characterized. It worsens outcomes increasing morbidity, mortality, and, consequently, hospital-related costs.[16,17] With the development of less invasive techniques and fast-track discharge pathways,[18,19] proper prevention, and management of PAL are essential. A strategy for PAL control should include prevention through meticulous surgical procedure, standardized chest tube management and could benefit from efficient validation of the preexisting predicting risk models to select high-risk patients.

Although many air leaks resolve spontaneously within 48 hours, some persist for days, and the incidence of 5 days PAL varies between 7% and 12%.[17,20,21] Some patients are at a higher risk of PAL, and a meticulous surgical technique is crucial in reducing its incidence. The traditional rules, effective also in the era of Minimally Invasive Thoracic Surgery, are mobilization of all intrapleural adhesions, division of the inferior pulmonary ligament, routine precompression of staple

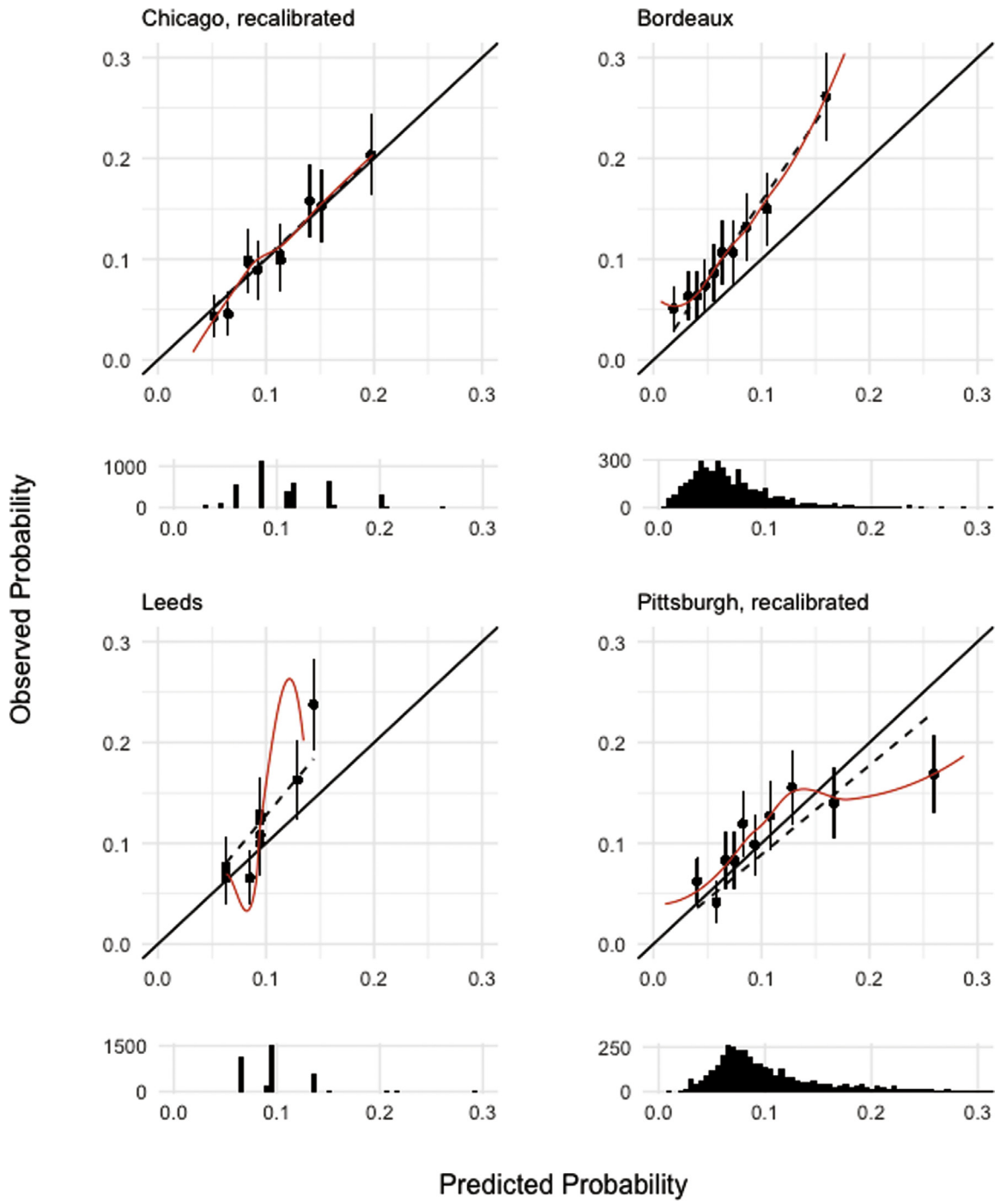


Figure 1. Calibration and distribution plot of the studied models (Chicago, Bordeaux, Leeds, and Pittsburgh). The Chicago and Pittsburgh models were plotted after re-estimation of model intercept. Black dots and their line ranges denote the observed probability for each deciles of predicted risk, with their associated 95% confidence intervals (CIs). 45-degree solid black line indicates perfect calibration. Dashed black line indicates the best-fitting straight line through the estimates (linear regression). Solid red line indicates the best-fitting curved line through the estimates (loess regression). Color version of figure is available online.

lines, and the fissureless/fissurelast technique. At the end of the anatomic resection, intraoperative assessment of intraoperative air leak (IAL) is of importance in order to prevent PAL. According to 2 recent studies, the presence of IAL at the ventilator test represents a reliable predictor for PAL.[22,23] These studies overcome the need to use the Macchiarini's classification with submersion test,[24] which is, according to our opinion, too subjective. The use of lung sealants after anatomic lung resections is still matter of debate. Our recent randomized controlled trial demonstrated that the use of sealant in patients with a

Table 3. Performance of the Original and Updated Models

		Original Model	Calibration in the Large	Logistic Recalibration
Chicago Model	Intercept	-	−3.39	−3.51
	Slope	-	1	1.09
	Residual deviance	-	2625.27	2624.75
	Df	-	3932	3931
	LRT Chi-square *P* value	-	<0.001	0.47
	Brier score	-	0.095 (0.09−0.1)	0.095 (0.09−0.1)
	E_{max}	-	0.027 (0.01−0.08)	0.023 (0.01−0.08)
	E_{avg}	-	0.006 (0−0.02)	0.006 (0−0.02)
	c-statistic	-	0.63 (0.6−0.65)	0.63 (0.6−0.65)
Bordeaux Model	Intercept	0	0.53	0.3
	Slope	1	1	0.91
	Residual deviance	2695.53	2603.31	2602.13
	Df	3940	3939	3938
	LRT Chi-square *P* value	-	<0.001	0.27
	Brier score	0.096 (0.09−0.1)	0.094 (0.09−0.1)	0.094 (0.09−0.1)
	E_{max}	0.238 (0.08−0.46)	0.06 (0.02−0.26)	0.113 (0.02−0.28)
	E_{avg}	0.042 (0.03−0.05)	0.006 (0−0.01)	0.007 (0−0.01)
	c-statistic	0.65 (0.63−0.68)	0.65 (0.62−0.68)	0.65 (0.62−0.68)
Leeds Model	Intercept	0	0.24	1.28
	Slope	1	1	1.47
	Residual deviance	2431.51	2413.12	2405.25
	Df	3498	3497	3496
	LRT Chi-square *P* value	-	<0.001	0.005
	Brier score	0.099 (0.09−0.11)	0.099 (0.09−0.11)	0.098 (0.09−0.11)
	E_{max}	0.376 (0.11−0.63)	0.32 (0.06−0.57)	0.171 (0.05−0.41)
	E_{avg}	0.025 (0.02−0.04)	0.012 (0.01−0.02)	0.005 (0−0.02)
	c-statistic	0.62 (0.59−0.65)	0.62 (0.59−0.65)	0.62 (0.59−0.65)
Pittsburgh model	Intercept	0	3.22	1.03
	Slope	1	1	0.58
	Residual deviance	4431.61	2614.91	2589.78
	Df	3854	3853	3852
	LRT Chi-square *P*value	-	<0.001	<0.001
	Brier score	0.107 (0.1−0.12)	0.96 (0.09−0.1)	0.95 (0.09−0.1)
	E_{max}	0.265 (0.14−0.59)	0.299 (0.13−0.46)	0.072 (0.02−0.2)
	E_{avg}	0.102 (0.09−0.11)	0.023 (0.01−0.03)	0.01 (0−0.02)
	c-statistic	0.61 (0.58−0.64)	0.61 (0.58−0.64)	0.61 (0.58−0.64)

95% confidence intervals are indicated in the round brackets. E_{avg}, average difference in predicted and calibrated probabilities; E_{max}, maximum difference in predicted and calibrated probabilities; LRT, likelihood ratio test; df, degrees of freedom.

moderate IAL between 100 and 400 mL/min reduces the incidence of PAL, length of stay (LOS), and hospital costs.[25] Other retrospective studies confirmed our results.[26,27]

At present, we have no evidence that a specific drainage system (digital or traditional) or aspiration regimen in PAL patients reduces drainage length or LOS.[28−33] The Enhanced Recovery After Surgery studies demonstrated that the adoption of a standardized protocol for drainage removal reduces LOS.[18,28] Different but similar protocols are adopted worldwide.[34,35] Air leak management with digital drainage allows us to observe and record the air leak trend closely and carefully. It helps to remove the tube immediately after air leak resolution.[28]

Identifying high-risk patients through predictive models could help to reduce PAL incidence, adopting aggressive preventive measures. It can help surgeons identify selected patients in whom the following procedures could be routinely performed: buttressing of suture lines, creation of an apical pleural tent at the time of upper lobectomy, creation of pneumoperitoneum at the time of lower lobe resection or injection of the phrenic nerve with a local anesthetic.

Furthermore, for these high-risk patients with a detected PAL who are unresponsive to conservative measures, surgeons could have a low threshold to take other minimally invasive measures to resolve air loss such as chemical pleurodesis, insertion of endobronchial valves or Heimlich valve. Finally, a reliable predicting model could be a useful tool in preoperative counseling, especially in centers where it is common practice to send patients home very early (ie, on postoperative day 1) because it may help to predict and prepare patients who will be discharged home with a digital air leak system.

Table 4. Clinical and Net Benefit Outcomes Per 1000 Patients, Basing Treatment on Different High-Risk Thresholds

				N° People at Risk		N° of PAL Events			
	Score Value	Threshold Probability	Odd Ratio	High Risk	Low Risk	Identified	Not Identified	Events Beyond Chance[†]	Net Benefit[*]
CHICAGO Model	-Inf	Treat all	1:Inf	1000	0	109	0	0	0.000
	6	0.05	1:19	840	160	103	6	11	0.003
	11	0.1	1:9	530	470	76	34	18	0.015
	13	0.11	1:8.1	307	693	52	57	18	0.021
	14	0.12	1:7.3	303	697	52	57	19	0.017
	14	0.13	1:6.7	303	697	52	57	19	0.014
	15	0.14	1:6.1	275	725	47	62	17	0.010
	16	0.15	1:5.7	215	785	40	69	17	0.009
	20	0.2	1:4	25	975	6	103	3	0.001
	25	0.3	1:2.3	0	1000	0	109	0	0.000
	25	0.4	1:1.5	0	1000	0	109	0	0.000
BORDEAUX Model	-Inf	Treat all	1:Inf	1000	0	109	0	0	0.000
	4	0.05	1:19	892	108	104	5	6	0.000
	13	0.1	1:9	397	603	64	46	21	0.016
	13	0.11	1:8.1	397	603	64	46	21	0.023
	14	0.12	1:7.3	338	662	59	50	22	0.021
	15	0.13	1:6.7	273	727	52	57	22	0.019
	16	0.14	1:6.1	223	777	44	65	20	0.015
	17	0.15	1:5.7	173	827	38	72	19	0.014
	20	0.2	1:4	81	919	21	88	12	0.006
	25	0.3	1:2.3	16	984	5	104	3	0.001
	29	0.4	1:1.5	3	997	2	108	2	0.001
LEEDS Model	-Inf	0	1:Inf	1000	0	114	0	0	0.000
	-Inf	0.05	1:19	1000	0	114	0	0	0.000
	0	0.1	1:9	676	324	91	23	14	0.011
	0	0.11	1:8.1	676	324	91	23	14	0.015
	1	0.12	1:7.3	186	814	37	77	16	0.017
	1	0.13	1:6.7	186	814	37	77	16	0.015
	1	0.14	1:6.1	186	814	37	77	16	0.013
	1	0.15	1:5.7	186	814	37	77	16	0.011
	2	0.2	1:4	7	993	3	111	2	0.002
	2	0.3	1:2.3	7	993	3	111	2	0.002
	3	0.4	1:1.5	2	998	1	113	1	0.001
PITTSBURGH Model	-Inf	0	1:Inf	1000	0	108	0	0	0.000
	11	0.05	1:19	992	8	108	0	1	0.000
	21	0.1	1:9	589	411	76	32	12	0.010
	23	0.11	1:8.1	394	606	59	49	16	0.018
	24	0.12	1:7.3	305	695	47	61	14	0.012
	25	0.13	1:6.7	236	764	37	71	12	0.008
	26	0.14	1:6.1	177	823	28	80	9	0.004
	28	0.15	1:5.7	103	897	17	91	6	0.001
	33	0.2	1:4	14	986	4	104	3	0.001
	40	0.3	1:2.3	0	1000	0	108	0	0.000
	40	0.4	1:1.5	0	1000	0	108	0	0.000

"Score value" is the threshold score that identifies the threshold probability in our patient sample.

*Net benefit with the positive Treat all Net Benefit subtracted.

[†]"Event beyond chance" identifies the number of true positive events beyond the expected events due to the prevalence of the disease.

This study provides information about the performance of four different prognostic models developed for the evaluation of PAL risk.[1–4] The variables used are listed in Table 1. The Chicago, Bordeaux, Leeds, and Pittsburgh models were validated using a multicenter national-wide registry, the Italian VATS group database. These four models were, therefore, assessed by investigators not involved in the model's development and applied to a new cohort of patients from a different geographic region. Performance of the Chicago, Bordeaux, Leeds, and Pittsburgh models was assessed in terms of

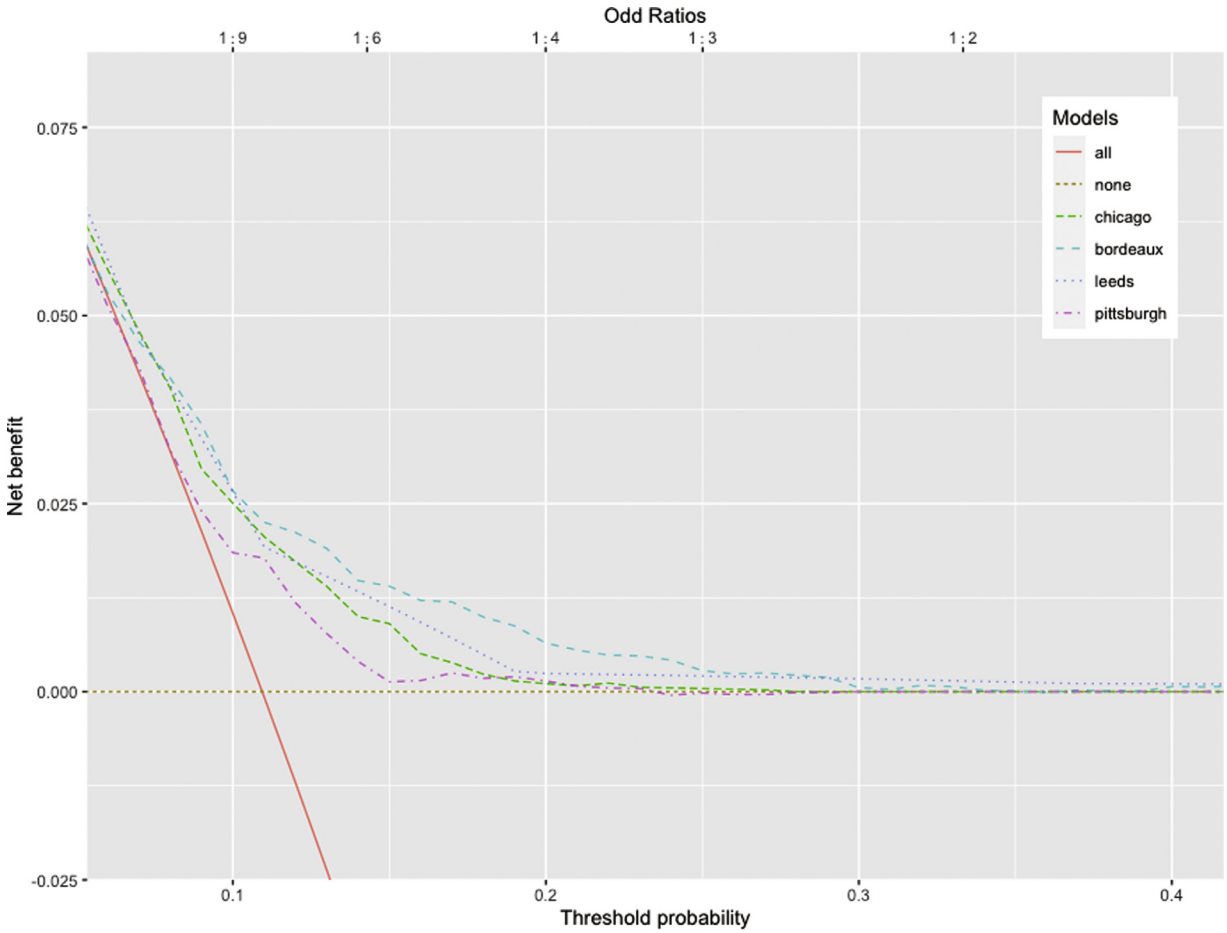


Figure 2. Decision curve analysis of the scores of the Chicago, Bordeaux, Leeds, and Pittsburgh models for diagnosing 5-day PAL. The 4 curves are compared to the curves of treating none and all patients.

discrimination, calibration, and overall performance. The linear predictor for each patient of the studied models was calculated based on the logistic regression coefficient values provided in the original manuscripts. The models' performances were modestly improved by recalibration (Table 3). Statistical performances reported in the original validation studies were attenuated in the VATS group registry cohort. Chicago model discrimination in our sample slightly differs from the results obtained by the authors during cross-validated internal validation, C statistics of 0.63 (0.60−0.65) versus 0.644. The Bordeaux model C-index was 0.69 (95% confidence interval, 0.66−0.72) using 7-day PAL as outcome in the external validation performed by the authors, not so far from our result with a 5-day PAL outcome, 0.65 (0.63−0.68). In the original manuscript of Pompili et al, no C index value was provided, then we cannot compare it with our results, 0.62 (0.59−0.65). The discriminatory accuracy of the Pittsburgh model was initially reported as 0.73, but in our cohort, it was 0.61 (0.58−0.63).

However, even if validated and recalibrated to the studied sample, can the currently published PAL scores have a benefit in clinical practice? Traditional statistical metrics, such as discrimination and calibration, are not directly informative about the clinical value of a model. Although the scores reliably identify patients with increased risk, the predicted risk of PAL deviates little from the prevalence of disease for most patients. A similar degree of uncertainty is mainly related to the poor predictive ability of preoperative predictors. The Net benefit analysis has shown that these scores are effective only at the cost of a large number of false positives. The best performing PAL score at the 1:2 true/false positive ratio has a net benefit of 1 of 1000 patient. Although the application of a score has no contraindications or costs, it is time consuming. Its use is not reasonable given the low reliability. To make these scores more effective and reliable in identifying high-risk patients, lower true / false positive ratios must be accepted. Is it reasonable to use a treatment on high-risk patients knowing that the ratio between true and false positives is 1:5 or lower? No surgical procedure is complication-free and cost-free, and furthermore, no surgical technique is effective in 100% of cases. On the other hand, the use of a drug (eg, statins) with minimal side effects to prevent cardiovascular events is justified, even on a large population of healthy patients.

In conclusion, according to our decision analysis, the use of PAL scores based only on preoperative predictive factors is not reliable and could lead to uselessly treating a large number of patients with negligible clinical benefits and increased health care costs.

LIMITATIONS

The comprehensiveness and integrity of any data collection are fundamental when the information is used for validation analysis. Our dataset comes from a multicenter nationwide registry. This is an advantage because patients recruited in only one institution are not representative of the various practices used to manage air leaks. Furthermore, data coming from an organizational registry are, most of the time, current and generalizable. On the other hand, the national databases were not designed for a specific research purpose. Some explanatory variables of interest such as method of data collections of air leak, suction regimen, and type of drainage system were not available. Furthermore, the dyspnea score evaluation was not available and was replaced in calculations with %FEV1 value ranges. We did not remove any patient because of missing data, since the Quality Committee of the Italian VATS registry regularly removes procedures with incomplete or inaccurate data. The amount of procedures removed is approximately 5% of all registered procedures.[5] This could be a potential source of bias that could slightly affect the performance of our external validations of the models analyzed. The median %FEV1 (94%) was higher in our database than in other registries. Some predictors were entirely subjective; for example, the definition of "pleural adhesions and adhesiolysis."

Our cohort selection as well as the Leeds sample, represented a subgroup of the Bordeaux and Pittsburgh models sample. Rivera,[1] in his analyses, included volume reduction and bullectomy. He also included procedures performed with traditional thoracotomy, like Attar.[4] However, the analyzed scores have been developed to be adopted on a population that includes our sample.

CONCLUSIONS

After our external validation, the Chicago and Bordeaux models represent the best 2 current predictive risk models for PAL after VATS anatomical resections, but with a C statistic of

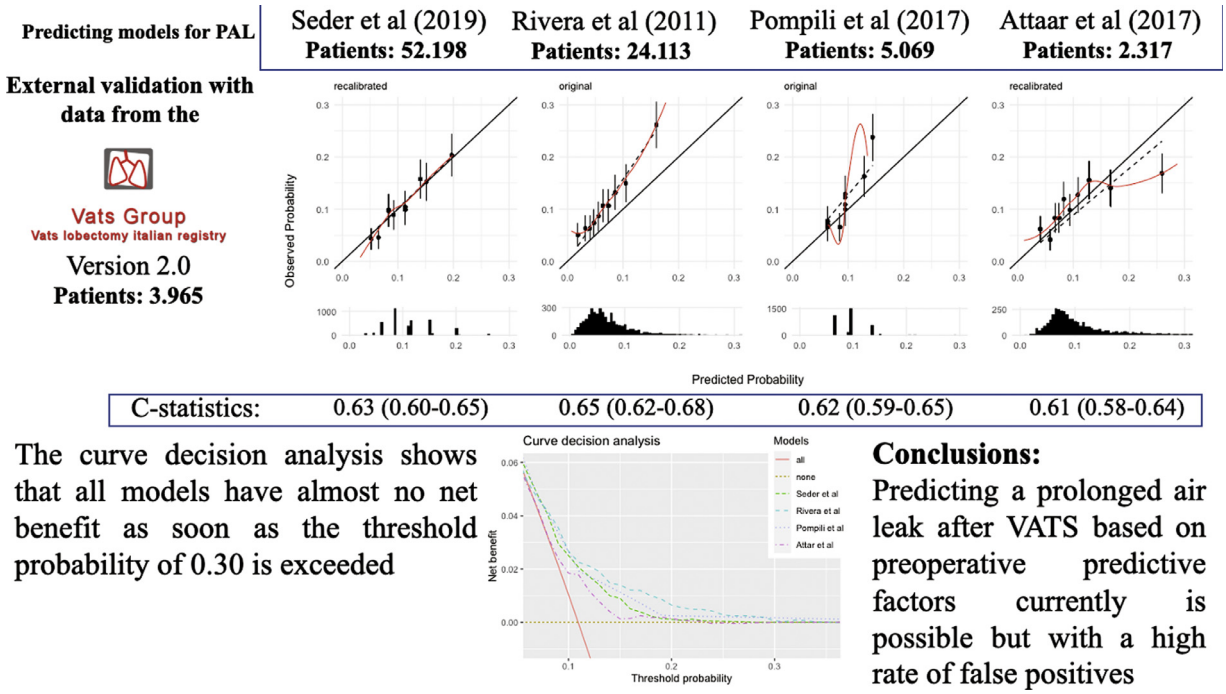


Figure 3. Graphical abstract: among the recently published models, 4 selected predictive prolonged air leak risk models were retrospectively applied to patients recorded on Italian VATS Group registry. Predictions for each patient were calculated based on the logistic regression coefficient values provided in the original manuscripts. All models were tested for their overall performance, discrimination, and calibration. We recalibrate the original models with the re-estimation of the model intercept and slope. We used decision curve analysis and Net benefit to describe and compare the clinical effects of the four 5-day PAL risk models using the scores and their calibrated probabilities.
After our external validation, the Chicago and Bordeaux models represent the best two current predictive risk models for PAL after VATS anatomical resections, with a C statistic of no more than 0.65. Furthermore, according to our decision analysis, all predictive models are burdened by a high number of false positive cases.
Predicting a prolonged air leak after VATS based on preoperative predictive factors is possible but with a high rate of false positives.

no more than 0.65. Predicting a prolonged air leak after VATS based on preoperative predictive factors is possible but with a high rate of false positives (Fig. 3). Further research to identify additional intraoperative risk factors, as well as the development, validation, and refinement of new risk prediction models are required to improve routine clinical practice.

Acknowledgments

For our statistical procedures, we were inspired by the online tutorial based on the HOMR model at the website (https://darrendahly.github.io/post/homr/) and by Zhang et al's article entitled "Decision curve analysis: a technical note".[13]

SUPPLEMENTARY MATERIAL

Scanning this QR code will take you to the article title page to access supplementary information.

APPENDIX 1. ITALIAN VATS GROUP MEMBERS

Mancuso M. — Pernazza F.	Alessandria
Refai M.	Ancona
Srella F. — Argnani D.	ASL Romagna
Marulli G. — De Palma A.	Bari Policlinico
Bortolotti L. — Rizzardi G.	Bergamo Humanitas Gavazzeni
Solli PG. — Dolci GP.	Bologna Osp. "S. Orsola"
Perkmann R. — Zaraca F.	Bolzano Osp. Di Bolzano
Benvenuti M. — Gavezzoli D.	Brescia
Cherchi R. — Ferrari P.	Cagliari
Mucilli F. — Camplese P.	Chieti Osp. "S. Maria Annunziata"
Melloni G. — Mazza F.	Cuneo
Cavallesco G. — Maniscalco P.	Ferrara Policlinico Universitario
Voltolini L. — Gonfiotti A.	Firenze Osp "Careggi"
Sollitto F. — Ardò N.	Foggia
Pariscenti GL. — Risso C.	Genova S. Martino
Surrente C. - Lopez C.	Lecce Osp "Fazzi"
Droghetti A. — Giovanardi M.	Mantova Osp C. Poma
Breda C. — Lo Giudice F.	Mestre
Alloisio M. — Bottoni E.	Milano Humanitas
Spaggiari L. — Gasparri R.	Milano IEO
Torre M. — Rinaldo A.	Milano Osp. "Niguarda"
Nosotti M. — Tosi D.	Milano Policlinico Universitario
Negri GP. — Bandiera A.	Milano S. Raffaele
Baisi A. — Raveglia F.	Milano S. Paolo
Stefani A. — Natali P.	Modena
Scarci M. — Pirondini E.	Monza
Curcio C. — Amore D.	Napoli Monaldi
Rena O.	Novara Osp. "Maggiore della Carità"
Nicotra S. — Dell'Amore A.	Padova
Bertani A. — Tancredi G.	Palermo ISMETT
Ampollini L. — Carbognani P.	Parma Policlinico Universitario
Puma F. — Vinci D.	Perugia Policlinico Universitario
Cardillo G. — Carleo F.	Roma Forlanini
Margaritora S. — Meacci E.	Roma Gemelli
Luzzi L. — Ghisalberti M.	Siena Policlinico Universitario
Crisci R. — Divisi D.	Teramo Osp "Mazzini"
Lausi P. — Guerrera F.	Torino "Molinette"
Fontana D. — Della Beffa V.	Torino "S. Giovanni Bosco"
Morelli A. — Londero F.	Udine Osp "S. Maria della Misericordia"
Imperatori A. — Rotolo N.	Varese Osp. Di Circolo
Terzi A. — Viti A.	Verona Negrar
Infante M. — Benato C.	Verona Osp. Borgo Trento

(*continued*)

REFERENCES

1. Rivera C, Bernard A, Falcoz P-E, et al: Characterization and prediction of prolonged air leak after pulmonary resection: A nationwide study setting up the index of prolonged air leak. Ann Thorac Surg 92:1062–1068, 2011
2. Pompili C, Falcoz PE, Salati M, et al: A risk score to predict the incidence of prolonged air leak after video-assisted thoracoscopic lobectomy: An analysis from the European Society of Thoracic Surgeons database. J Thorac Cardiovasc Surg 153:957–965, 2017
3. Seder CW, Basu S, Ramsay T, et al: A prolonged air leak score for lung cancer resection: An analysis of the Society of Thoracic Surgeons General Thoracic Surgery Database. Ann Thorac Surg 108:1478–1483, 2019
4. Attaar A, Winger DG, Luketich JD, et al: A clinical prediction model for prolonged air leak after pulmonary resection. J Thorac Cardiovasc Surg 153:690–699.e2, 2017
5. Solli P, Bertolaccini L, Droghetti A, et al: 2016 annual report from the Italian VATS Group. Future Oncol. 14:23–28, 2018
6. Vestbo J, Hurd SS, Agustí AG, et al: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 187:347–365, 2013
7. Curtin D, Dahly DL, Smeden M, et al: Predicting 1−year mortality in older hospitalized patients: External validation of the HOMR model. J Am Geriatr Soc 67:1478–1483, 2019
8. Vergouwe Y, Nieboer D, Oostenbrink R, et al: A closed testing procedure to select an appropriate method for updating prediction models: Method selection to update a prediction model. Stat Med 36:4529–4539, 2017
9. Moons KGM, Altman DG, Reitsma JB, et al: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 162:W1, 2015
10. Steyerberg EW, Vickers AJ, Cook NR, et al: Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology 21:128–138, 2010
11. Vickers AJ, Van Calster B, Steyerberg EW: Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ 352:i6, 2016
12. Vickers AJ, van Calster B, Steyerberg EW: A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res 3:18, 2019
13. Zhang Z, Rousson V, Lee W-C, et al: Decision curve analysis: A technical note. Ann Transl Med 6:308, 2018

14. Collins GS, Altman DG: Predicting the 10-year risk of cardiovascular disease in the United Kingdom: Independent and external validation of an updated version of QRISK2. BMJ 344:e4181, 2012
15. Wickham H, Grolemund G: R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. First edition. Sebastopol, CA, O'Reilly; 2016, pp 492
16. Brunelli A, Xiume F, Al Refai M, et al: Air leaks after lobectomy increase the risk of empyema but not of cardiopulmonary complications. Chest 130:1150–1156, 2006
17. Burt BM, Shrager JB: Prevention and management of postoperative air leaks. Ann Cardiothorac Surg 3:216–218, 2014
18. Batchelor TJP, Rasburn NJ, Abdelnour-Berchtold E, et al: Guidelines for enhanced recovery after lung surgery: Recommendations of the Enhanced Recovery After Surgery (ERAS®) Society and the European Society of Thoracic Surgeons (ESTS). Eur J Cardiothorac Surg 55:91–115, 2019
19. Rogers LJ, Bleetman D, Messenger DE, et al: The impact of enhanced recovery after surgery (ERAS) protocol compliance on morbidity from resection for primary lung cancer. J Thorac Cardiovasc Surg 155:1843–1852, 2018
20. Mueller MR, Marzluf BA: The anticipation and management of air leaks and residual spaces post lung resection. J Thorac Dis 6:271–284, 2014
21. Varela G, Jiménez MF, Novoa N, et al: Estimating hospital costs attributable to prolonged air leak in pulmonary lobectomy. Eur J Cardiothorac Surg 27:329–333, 2005
22. Brunelli A, Salati M, Pompili C, et al: Intraoperative air leak measured after lobectomy is associated with postoperative duration of air leak. Eur J Cardiothorac Surg 52:963–968, 2017
23. Zaraca F, Vaccarili M, Zaccagna G, et al: Can a standardised ventilation mechanical test for quantitative intraoperative air leak grading reduce the length of hospital stay after video-assisted thoracoscopic surgery lobectomy? J Vis Surg 3:179–188, 2017
24. Macchiarini P, Wain J, Almy S, Dartevelle P: Experimental and clinical evaluation of a new synthetic, absorbable sealant to reduce air leaks in thoracic operations. J Thorac Cardiovasc Surg 117:751–758, 1999
25. Zaraca F, Vaccarili M, Zaccagna G, et al: Cost-effectiveness analysis of sealant impact in management of moderate intraoperative alveolar air leaks during video-assisted thoracoscopic surgery lobectomy: A multicentre randomised controlled trial. J Thorac Dis 9:5230–5238, 2017
26. Park BJ, Snider JM, Bates NR, et al: Prospective evaluation of biodegradable polymeric sealant for intraoperative air leaks. J Cardiothorac Surg 11:168, 2016
27. Mortman KD, Corral M, Zhang X, et al: Length of stay and hospitalization costs for patients undergoing lung surgery with Progel pleural air leak sealant. J Med Econ 21:1016–1022, 2018
28. Mayor JM, Lazarus DR, Casal RF, et al: Air leak management program with digital drainage reduces length of stay after lobectomy. Ann Thorac Surg 106:1647–1653, 2018
29. Miller DL, Helms GA, Mayfield WR: Digital drainage system reduces hospitalization after video-assisted thoracoscopic surgery lung resection. Ann Thorac Surg 102:955–961, 2016
30. Lijkendijk M, Licht PB, Neckelmann K: Electronic versus traditional chest tube drainage following lobectomy: A randomized trial. Eur J Cardiothorac Surg 48:893–898, 2015
31. Zhou J, Lyu M, Chen N, et al: Digital chest drainage is better than traditional chest drainage following pulmonary surgery: A meta-analysis. Eur J Cardiothorac Surg 54:635–643, 2018
32. Holbek BL, Christensen M, Hansen HJ, et al: The effects of low suction on digital drainage devices after lobectomy using video-assisted thoracoscopic surgery: A randomized controlled trial. Eur J Cardiothorac Surg 55:673–681, 2019
33. Gocyk W, Kużdżał J, Włodarczyk J, et al: Comparison of suction versus nonsuction drainage after lung resections: A prospective randomized trial. Ann Thorac Surg 102:1119–1124, 2016
34. Nakanishi R, Fujino Y, Kato M, et al: Early chest tube removal after thoracoscopic lobectomy with the aid of an additional thin tube: A prospective multi-institutional study. Gen Thorac Cardiovasc Surg 66:723–730, 2018
35. Xie H-Y, Xu K, Tang J-X, et al: A prospective randomized, controlled trial deems a drainage of 300 ml/day safe before removal of the last chest drain after video-assisted thoracoscopic surgery lobectomy. Interact Cardiovasc Thorac Surg 21:200–205, 2015