



Prolonged air leak after video-assisted thoracic anatomical pulmonary resections: a clinical predicting model based on data from the Italian VATS group registry, a machine learning approach

Duilio Divisi¹, Marco Pipitone², Reinhold Perkmann², Luca Bertolaccini³, Carlo Curcio⁴, Francesco Baldinelli⁵, Roberto Crisci¹, Francesco Zaraca²; Italian VATS group*

¹Department of Life, Health and Environmental Sciences, Thoracic Surgery Unit, University of L'Aquila, L'Aquila, Italy; ²Department of Vascular and Thoracic Surgery, Central Hospital, Bolzano, Italy; ³Division of Thoracic Surgery, IEO European Institute of Oncology IRCCS, Milan, Italy; ⁴Department of Thoracic Surgery, Monaldi Hospital, Naples, Italy; ⁵First Service of Anesthesia and Intensive Care, Central Hospital, Bolzano, Italy
Contributions: (I) Conception and design: D Divisi, M Pipitone, L Bertolaccini, R Crisci, F Zaraca; (II) Administrative support: R Perkmann, C Curcio, R Crisci; (III) Provision of study materials or patients: Italian VATS group members; (IV) Collection and assembly of data: Italian VATS group members; (V) Data analysis and interpretation: D Divisi, M Pipitone, L Bertolaccini, R Crisci, F Zaraca; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Duilio Divisi, Prof, MD, PhD. Department of Life, Health and Environmental Sciences, University of L'Aquila, Palazzo Camponeschi, Piazza Santa Margherita 2, 67100 L'Aquila, Italy. Email: duilio.divisi@univaq.it; duilio.divisi@aslteramo.it.

Background: Prolonged air leak (PAL) is a frequent complication after lung resection surgery and has a high clinical and economic impact. A useful risk predictor model can help recognize those patients who might benefit from additional preventive procedures. Currently, no risk model has sufficient discriminatory capacity to be used in common clinical practice. The aim of this study is to identify predictive risk factors for PAL after video-assisted thoracoscopic surgery (VATS) anatomical resections in the Italian VATS group database and to evaluate their clinical and statistical performance.

Methods: We processed data collected in the second edition of the Italian VATS group registry. It includes patients that underwent a thoracoscopic anatomical resection for benign or malignant diseases, between November 2015 and December 2020. We used recursive feature elimination (RFE), using a backward selection process, to find the optimal combination of predictors. The study population was randomly split based on the outcome into a derivation (80%) and an internal validation cohort (20%). Discrimination of the model was measured using the area under the curve, or C-statistic. Calibration was displayed using a calibration plot and was measured using Emax and Eavg, the maximum and the average difference in predicted versus loess calibrated probabilities.

Results: A cohort of 6,236 patients was eligible for the study after application of the exclusion criteria. Five-day PAL rate in this patient cohort was 11.3%. For the construction of our predictive model, we used both preoperative and intraoperative variables, with a total of 320 variables. The presence of variables with missing values greater than 5% led to 120 remaining predictors. RFE algorithm recommended 8 features for the model that are relevant in predicting the target variable.

Conclusions: We confirmed significant prognostic risk factors for the prediction of PAL: decreased DLCO/VA ratio, longer duration of surgery, male sex, the need for adhesiolysis, COPD, and right side. We identified middle lobe resections and ground glass opacity as protective factors. After internal validation, a C statistic of 0.63 was revealed, which is too low to generate a reliable score in clinical practice.

Keywords: Prolonged air leak (PAL); risk factors; video-assisted thoracoscopic surgery lobectomy (VATS lobectomy); risk predictive model

* <https://cdn.amegroups.cn/static/public/jtd-21-1484-1.pdf>

Submitted Sep 09, 2021. Accepted for publication Jan 06, 2022. Published online Jan 13 2022.

doi: 10.21037/jtd-21-1484

View this article at: <https://dx.doi.org/10.21037/jtd-21-1484>

Introduction

Background

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 and the advent of enhanced recovery after surgery (ERAS). 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). Its incidence varies between 8% and 26% in patients who have undergone lobectomies (2). PAL has a high clinical and economic impact. The main consequences are longer hospital stays, higher incidence of atelectasis, pneumonia and infections of the pleural space and increased hospital readmissions, leading to higher hospital-related costs (3,4).

There are three potential ways to prevent PAL. First: recognizing patients at higher risk of PAL (5-8). The second way could be the adoption of a meticulous surgical technique as appropriate tissue manipulation and retraction, mobilization of all intrapleural adhesions, division of the inferior pulmonary ligament, routine pre-compression of staple lines, fissureless/fissurelast technique, and select use of surgical sealant (9,10). Finally, an objective intraoperative measurement of intraoperative air leak (IAL) is of importance (11,12).

A useful predictive risk model can help recognizing those patients who could benefit from supplemental preventive procedures. Many algorithms to risk-stratify patients undergoing lung resection for PAL have been reported from different study groups (5-8). Nevertheless, current risk models in the literature, even if they have many cases, a long period of enrollment and are scientifically relevant, do not have sufficient discriminatory capacity to be used in common clinical practice (13).

Objectives

Our aim is developing an accurate and yet simple model for predicting 5-day PAL using machine learning and

logistic regression. Machine learning is a method of data analysis that automates model building. It is seen as a part of artificial intelligence. We present the following article in accordance with the TRIPOD reporting checklist (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-21-1484/rc>).

Methods

Source of data

We processed data collected in the second edition of the Italian VATS group registry. It includes patients that underwent a thoracoscopic lobar or a sub-lobar resection for benign or malignant primary or secondary diseases, between November 2015 and December 2020. The VATS group registry is an online database, created in 2014 by Roberto Crisci, where data from 57 thoracic surgery units across Italy regarding video-assisted thoracoscopic surgery (VATS) anatomic lung resections are prospectively collected. Fifty-seven thoracic surgery units across Italy involved in the registry is on a voluntary basis. 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. 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. Collected variables in all these registries include information about the patient's medical history, characteristics, surgical procedures, and outcomes.

Participants

Our registry includes exclusively patients that underwent VATS lobar or sublobar resections for benign or malignant diseases. Therefore, are not included in the analyses, patients who underwent elective thoracotomy lobectomy, robotic resections, volume reduction, bullectomy or wedge resection. Exclusion criteria for our model included patients from units contributing an average of less than 10 procedures/year; patients with more than one lung surgery (only data of the first surgery is used); patients undergoing postoperative-

assisted mechanical ventilation; lung resection combined with diaphragm or chest wall resections; bronchial sleeve resections, and pneumonectomies.

Outcome

Our model was trained 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.

For the construction of our predictive model, we used both preoperative and intraoperative variables. All the analyzed variables were automatically pre-processed.

Sample size

The study population was randomly split based on the outcome into a derivation (80%) and an internal validation cohort (20%). Discrimination of the model was measured using the area under the curve, or C-statistic. Calibration was displayed using a calibration plot and was measured using Emax and Eavg, the maximum and the average difference in predicted versus loess calibrated probabilities.

Missing data

Prior to modeling, “near-zero-variance” predictors were identified and eliminated. When correlation between the predictors is higher than 0.5, the less statistic relevant features were removed. Continuous variables were not categorized. They have been scaled and centered to have a standard deviation of one and a mean of zero. During data entry, our database prevents the final validation of the case, if mandatory data are missing. Among the variables set as nonmandatory, we decided to exclude variables with a proportion of missing data >5%. We used K-nearest neighbors as method to impute remaining missing values.

Statistical analysis methods

The Model building algorithm applies to the derivation cohort through 4 distinct phases: data splitting; pre-processing; feature selection; variable importance estimation. We use recursive feature elimination (RFE), or shortly RFE algorithm, to select features that are relevant in predicting the target variable. RFE uses a backward selection process to find the optimal combination of

predictors. It builds a model based on all predictors and then calculates the importance of each predictor in the model and removes the ones with the least importance iteratively. We computed the predictor importance using a discrimination metric, C statistic. The trapezoidal rule is used to compute the area under the ROC curve. We used repeated 10-fold cross-validation with 5 repeats to factor in the variability caused by feature selection. The RFE function selects the least complex model within some percent tolerance (1.5%) of the largest performance value.

The final mode was than tested for their overall performance, discrimination, and calibration in the internal validation cohort. The 95% confidence intervals for these metrics were derived based on 500 bootstrap resampled replicates.

All analyses were conducted using R language for statistical computing software, 15 v.3.6.1 (July 5, 2019). For implementing the RFE method in R we used the caret package.

Ethical statement

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by national ethics board of the Italian VATS group and by the Institutional Research Review Board for data collection, transmission, storage, and analysis (Registration No. 81/2014/O/Oss), and individual consent for this retrospective analysis was waived.

Results

Participants

The Italian VATS group registry 2.0 recorded during the study period 6,348 patients who underwent a VATS anatomic lung resection. A cohort of 6,236 patients was eligible for the study after application of the exclusion criteria. The following patients/procedures were excluded: 23 patients from units contributing an average of less than 10 procedures/year; 7 procedures of patients with more than one lung surgery (only data of the first surgery is used); 30 patients undergoing postoperative-assisted mechanical ventilation; 52 patients because of lung resection combined with diaphragm or chest wall resections, bronchial sleeve resections, and pneumonectomies (*Figure 1*).

Five-day PAL rate in this patient cohort was 11.3%. The median age of the study participants was 69 years [the interquartile range (IQR) 63–75 years], 88.4%

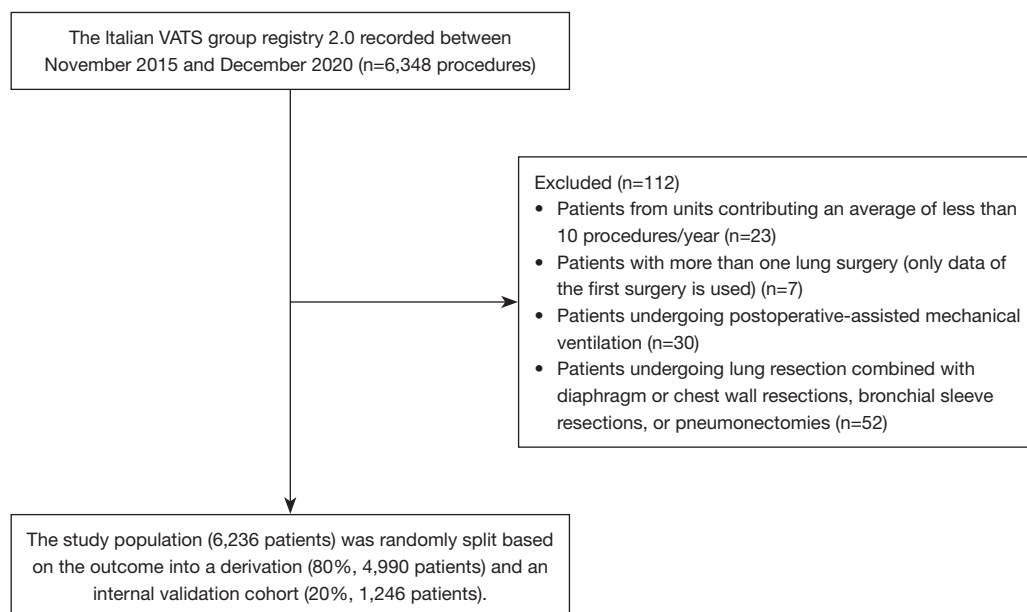


Figure 1 Patient's algorithm. VATS, video-assisted thoracoscopic surgery.

underwent a lobectomy, 9.5% a segmentectomy, and 2.1% a bilobectomy. No wedge resection, volume reduction or bulla resection were recorded in the registry. The baseline characteristics of patients in our cohort are summarized in *Table 1*. All surgeries were initially in video-assisted thoracoscopy and 537 patients (8.6%), were converted in open surgery. Adhesiolysis that is associated to high risk of parenchyma integrity impairment was performed in 1,688 cases (27.0%). The mean time with the chest tube was 4.6 days (IQR 3–5 days).

Model development

The study population was randomly split into a derivation (80%, n=4,990) and an internal validation cohort (20%, n=1,246). For the construction of our predictive model, we used both preoperative and intraoperative variables, with a total of 320 variables. The presence of variables with missing values greater than 5 percent led to 120 remaining predictors (*Table S1*).

RFE algorithm, using a backward selection process, recommended 8 features for the model that are relevant in predicting the target variable. The RFE function did not select the highest performance model but the least complex model within some percent tolerance (1.5%) of the largest performance value. We can also see the same results visually

in the following graphs (*Figure 2*). *Figure 2* graphically represents the performance profile of the different models across different subset sizes. The blue dot represents the optimal solution.

RFE algorithm allows to examine variable importance for the selected features. The bar graph in *Figure 3* shows which features are more important for predicting the target variable. Importance of a variable were scaled measurements to have a maximum value of 1 and derived from the area under the ROC curve.

Model specification

Figure 4 shows the forest plot of the logistic regression coefficients of the optimal model selected by RFE. The variables were arranged from top to bottom in order of importance and as is known, coefficient values greater than one indicated increased risk, while values less than one indicated protective effect. Six preoperative and two intraoperative variables were selected. Protective preoperative factors against the risk of PAL were middle lobe resection and ground glass opacity. The preoperative risk factors were decreased DLCO/VA ratio, male sex, chronic obstructive pulmonary disease (COPD), and right-side surgery. Intraoperative risk factors were prolonged duration of surgery and the need for adhesiolysis.

Table 1 Main characteristics of patients in our cohort

Variables	Italian VATS group (2.0)
Patients number	6,236
5-day PAL	702 [11.3]
Age, years	69 [63–75]
Gender (male)	3,725 [60]
BMI (kg/m ²)	24.5 [21.9–27.6]
Smoking history	
Never smoker	1,873 [30]
Past smoker	2,444 [39]
Current smoker	1,919 [31]
Zubrod Score (≥ 2)	1,870 [30]
Congestive heart failure	227 [4.0]
COPD	1,411 [23]
Diabetes	840 [13]
FEV1, % predicted (linear)	93 [80–106]
Pleural adhesions (present)	1,688 [27]
Open conversion	537 [8.6]
Pulmonary resection	
Bilobectomy	131 [2.1]
Lobectomy	5,513 [88.4]
Segmentectomy	592 [9.5]
Location	
Lower lobe	2,282 [36.6]
Middle lobe	470 [7.5]
Upper lobe	3,484 [55.9]
Side (right side)	3,702 [59.4]
Pulmonary pathology (malignant)	6,001 [96.2]

Results are expressed as counts and percentages [%] of patients for categorical variables, and as medians and interquartile ranges [IQR] 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.

Model performance

Both models were well calibrated as shown in the calibration plots (*Figure 5*), and according to the low values of Emax and Eavg (*Table 2*), C statistics was 0.68 (0.66–0.70)

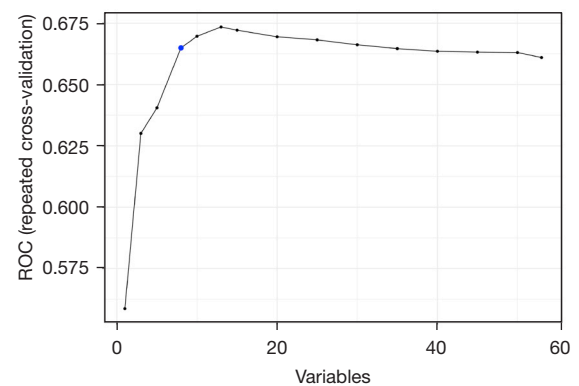


Figure 2 Graphical representation of the performance profile of the different models across different subset sizes. RFE algorithm allows to examine variable importance for the selected features. The blue dot represents the optimal solution. ROC, receiver operating characteristic; RFE, recursive feature elimination.

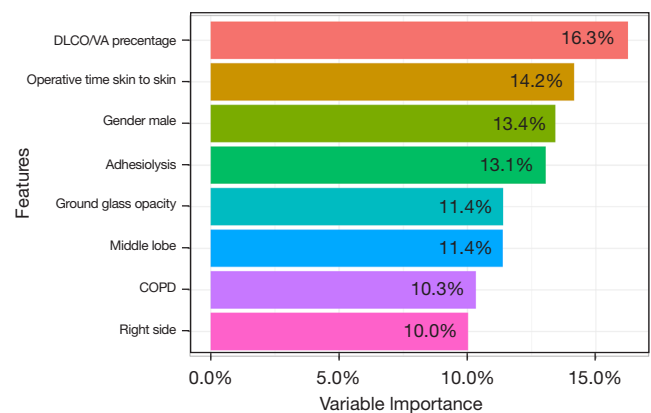


Figure 3 Measures of importance. When using machine learning models, it is important to understand which predictors are more influential on the outcome variable. The area under the ROC curve is conducted on each predictor and used as the measure of variable importance. All measures of importance are scaled to have a maximum value of 1 and expressed as percentages. DLCO/VA, diffusing capacity of the lung for carbon monoxide divided by alveolar volume; COPD, chronic obstructive pulmonary disease; ROC, receiver operating characteristic.

when the optimal model was tested in the training cohort, while it dropped to 0.63 (0.57–0.67) when tested in the internal validation cohort. The discriminative ability among the models is far from acceptable in clinical practice, as explained in our previous study (13).

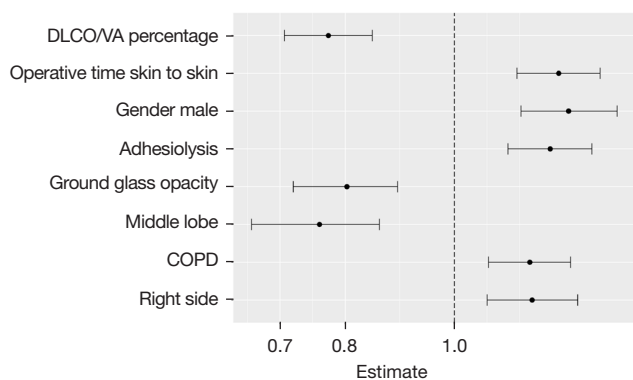


Figure 4 Forest plot of the logistic regression coefficients of the optimal model selected by RFE. The variables are sorted from top to bottom in order of importance and coefficient values greater than one indicate increased risk, while values less than one indicate a protective effect. DLCO/VA, diffusing capacity of the lung for carbon monoxide divided by alveolar volume; COPD, chronic obstructive pulmonary disease; RFE, recursive feature elimination.

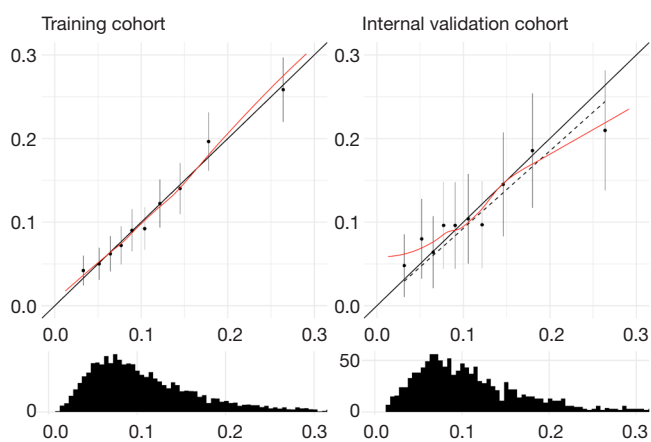


Figure 5 Calibration and distribution plot of our models. Black dots and their line ranges denote the observed probability for each decile of predicted risk, with their associated 95% CIs. 45-degree solid black line indicates perfect calibration. Dashed black line indicates the best-fitting straight line through the estimates (linear regression). Solid red line indicates the best-fitting curved line through the estimates (loess regression).

Discussion

Limitations

Our dataset comes from a more accurate edition of the Italian VATS group registry: 2.0. The completeness and integrity of any data collection is critical in these records.

Table 2 Performance of the original model

Parameters	Training cohort	Internal validation cohort
c-statistic	0.680 (0.66–0.70)	0.630 (0.57–0.67)
Brier score	0.095 (0.09–0.10)	0.098 (0.08–0.10)
E _{max}	0.050 (0.01–0.14)	0.142 (0.04–0.35)
E _{avg}	0.005 (0.002–0.009)	0.019 (0.01–0.03)

95% CI are indicated in the round brackets. E_{avg}, average difference in predicted and calibrated probabilities; E_{max}, maximum difference in predicted and calibrated probabilities.

In our database any patient was removed because of missing data. The Quality Committee of the Italian VATS registry removed approximately 5% of all registered procedures with incomplete or inaccurate data. The limitations of all datasets are that they are not designed for a specific research purpose. On the other hand, the benefits are more patient, and detection of the various practices used in the different centers.

A potential bias of our study is that 200 of the 320 critical variables were eliminated because of missingness. We did not include all of the variables that could predict postop of our leak because we elected to take a firm approach to missing data, and eliminate any variables that had more than 5% missing data.

Interpretation

PAL results in a longer duration of the chest tube which causes prolonged pain; limited ventilation leading to an increased risk of pneumonia; reduced mobility with a consequent increased risk of thromboembolism; need for pleurodesis, mechanical ventilation and sometimes reoperation; higher rate of readmission to hospital; prolonged hospital stays, and related higher overall costs (14). Its incidence, after the advent of MITS (minimally invasive thoracic surgery), was reduced to about 10% after anatomical resections (5,7). Our goal is to prevent air leaks because to “air” is to leak—to prevent is Devine (15). Recognizing patients at high risk of PAL could help surgeons during the operation. They may be more likely to add additional therapies such as sealants. Furthermore, in the presence of IAL on the ventilator test, which represents a reliable predictor of PAL, different thresholds could be adopted (11,12). In our previous study, we demonstrated that in unselected patients an IAL <100 mL/min., on the ventilator test, after a VATS lobectomy is self-limiting (16). Mr.

Brunelli and colleagues demonstrated that unselected patients submitted to lobectomy with an IAL >500 mL/min measured after resection would have an expected air leak duration of 15 days (11). Further studies are needed to provide evidence for the correlation between PAL and IAL, especially in high-risk patients.

The aim of this study was to provide an accurate and simple model for predicting 5-day PAL after anatomical VATS resections based on the Italian VATS group database.

The data collected by the register of the Italian VATS group represent the current picture of VATS lobectomy activities in Italy and are widely generalizable to the “real world”. Data come from 57 different hospitals in Italy. This heterogeneity represents, obviously, a potential bias. This problem is common to all national or international multicenter databases. Nevertheless, even if the ESTS (European Society of Thoracic Surgery) database currently collects data from more than 250 different European units, the updated version of the Eurolung risk models, recently published and elaborated from these data, achieved a worldwide appreciation (17). In order to obtain reliable data from a multicenter database, it is essential to have a strict protocol and control. Most of the Italian centers adopt the ERAS protocols (18). In fact, in 2017, the Italian VATS group database achieved the ESTS certification of data quality after an official inspection and audit. After that, the quality control of the Italian VATS group database has been regularly performed, and this guarantees the high quality of the recorded data. The enrollment of a cohort of 6,236 patients allowed the development of a more reliable predictive model.

According to our model for the prediction of PAL, the significant prognostic risk factors were male sex, COPD, decreased DLCO/VA ratio, right side, the longer duration of surgery, and the need for adhesiolysis. Protective factors were middle lobe and ground glass opacity resections.

Male sex has been found to be associated with PAL in many previous studies (6,7,9,19). Why male sex is associated with PAL is still unclear. COPD is probably the most consistent and validated risk factor contributing to the development of air leaks (6-9). On the same manner the PAL incidence is higher in patients with a decreased DLCO/VA ratio (9). In these patients a fragile lung parenchyma results in a high susceptibility to tearing during lung dissection or manipulation and an altered healing process especially in the suture lines. We have decided to report the DLCO/VA, also known as the transfer coefficient (KCO), because it represents a more accurate expression of the intrinsic

gas exchange function of the lung (20). DLCO/VA is not always linearly correlated with the predicted percent DLCO. They provide complementary information but may differ based on comorbidities. Right lobectomy and bi-lobectomy have also shown a higher incidence of PAL in other studies (9,21,22). The higher incidence of PAL on the right side could be secondary to the presence of 2 fissures, while for bi-lobectomy it is obviously related to the volume of lung removed and the ipsilateral residual space. The longer duration of surgery and the need for adhesiolysis are both validated risk factors and result in more extensive parenchymal manipulation and dissection favoring PAL (6,9,23). Resection of the middle lobe is probably a protective factor for its size and the small residual space. We cannot provide plausible explanations for the finding, not proven by other studies, that performing anatomical resections for ground glass opacities (defined as any lesions that has ground glass components) is protective for the onset of PAL. Larger studies are needed to confirm this finding. Other factors frequently associated with PAL in previous studies such as FEV₁, smoking history (current smoking and any past smoking history), BMI and steroid use were not confirmed in our analysis.

Implications

The analysis of the preoperative and intraoperative predictors for PAL in the 6,236 patients collected in the second edition of the Italian VATS group registry, after our internal validation, confirmed the finding of our previous study (10), where we evaluated both the clinical and statistical performances of the best 4 current PAL risk models (5-8). With a C statistic <0.65, predicting a PAL after VATS based on current predictive factors is possible but with a high rate of false positives and a low positive predictive value. Therefore, we decided not to generate a score for VATS anatomical resections, because the discriminative ability among our models is far from acceptable and its use in the clinical setting is currently limited. One of the possible reasons why our study failed to produce a reliable predictive model is that there is simply a high level of randomness to the occurrence of postoperative air leak that is irreducible. For example, tears in pulmonary staple lines or parenchyma may be largely random. As clinicians and scientists, we don't like to admit when we cannot predict, however it is useful to know when that is indeed the case.

Improving the success of the PAL prediction is strongly

in demand. One brilliant comment found a fitting analogy in forecasting the weather (24). Difficult to do due to a myriad of variables and able to modify human behavior but it can be achieved by exploiting experience and technology. While the focus of this investigation is on prediction, the actionability of prediction is another matter. In our view, the future perspective is the combination of identifying preoperative risk factors and identifying and quantifying intraoperative alveolar air leakage.

Conclusions

We confirmed significant prognostic risk factors for the prediction of PAL: decreased DLCO/VA ratio, longer duration of surgery, male sex, the need for adhesiolysis, COPD, and right side. We identified middle lobe resections and ground glass opacity as protective factors. After internal validation, a C statistic of 0.63 was revealed, which is too low to generate a reliable score in clinical practice.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Journal of Thoracic Disease*, for the series “Prolonged Air Leak after Lung Surgery: Prediction, Prevention and Management”. The article has undergone external peer review.

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-21-1484/rc>

Data Sharing Statement: Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-21-1484/dss>

Peer Review File: Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-21-1484/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-21-1484/coif>). The special series “Prolonged Air Leak after Lung Surgery: Prediction, Prevention and Management” was sponsored by Bard Limited. Bard Limited has no interference on the contents

of the special series. RC and FZ served as the unpaid Guest Editors of the series. LB serves as an unpaid editorial board member of the *Journal of Thoracic Disease*. 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 national ethics board of the Italian VATS group and by the Institutional Research Review Board for data collection, transmission, storage, and analysis (Registration No. 81/2014/O/Oss), and individual consent for this retrospective analysis was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Fernandez FG, Falcoz PE, Kozower BD, et al. The Society of Thoracic Surgeons and the European Society of Thoracic Surgeons general thoracic surgery databases: joint standardization of variable definitions and terminology. *Ann Thorac Surg* 2015;99:368-76.
2. Dugan KC, Laxmanan B, Murgu S, et al. Management of Persistent Air Leaks. *Chest* 2017;152:417-23.
3. Okereke I, Murthy SC, Alster JM, et al. Characterization and importance of air leak after lobectomy. *Ann Thorac Surg* 2005;79:1167-73.
4. Brunelli A, Xiume F, Al Refai M, et al. Air leaks after lobectomy increase the risk of empyema but not of cardio-pulmonary complications: a case-matched analysis. *Chest* 2006;130:1150-6.
5. 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* 2019;108:1478-83.
6. Rivera C, Bernard A, Falcoz PE, 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* 2011;92:1062-8; discussion 1068.
7. 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* 2017;153:957-65.
 8. Attaar A, Winger DG, Luketich JD, et al. A clinical prediction model for prolonged air leak after pulmonary resection. *J Thorac Cardiovasc Surg* 2017;153:690-699.e2.
 9. Geraci TC, Chang SH, Shah SK, et al. Postoperative Air Leaks After Lung Surgery: Predictors, Intraoperative Techniques, and Postoperative Management. *Thorac Surg Clin* 2021;31:161-9.
 10. McGuire AL, Yee J. Clinical outcomes of polymeric sealant use in pulmonary resection: a systematic review and meta-analysis of randomized controlled trials. *J Thorac Dis* 2018;10:S3728-39.
 11. 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* 2017;52:963-8.
 12. 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* 2017;3:179.
 13. Zaraca F, Pipitone M, Feil B, et al. Predicting a Prolonged Air Leak After Video-Assisted Thoracic Surgery, Is It Really Possible? *Semin Thorac Cardiovasc Surg* 2021;33:581-92.
 14. Attaar A, Luketich JD, Schuchert MJ, et al. Prolonged Air Leak After Pulmonary Resection Increases Risk of Noncardiac Complications, Readmission, and Delayed Hospital Discharge: A Propensity Score-adjusted Analysis. *Ann Surg* 2021;273:163-72.
 15. Cerfolio RJ. Commentary: To "Air" is to Leak-To Prevent is Devine. *Semin Thorac Cardiovasc Surg* 2021;33:595-6.
 16. 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* 2017;9:5230-8.
 17. Brunelli A, Cicconi S, Decaluwe H, et al. Parsimonious Eurolung risk models to predict cardiopulmonary morbidity and mortality following anatomic lung resections: an updated analysis from the European Society of Thoracic Surgeons database. *Eur J Cardiothorac Surg* 2020;57:455-61.
 18. Droghetti A. The ERAS project for VATS lobectomy-the Italian VATS Group. *J Thorac Dis* 2018;10:S490.
 19. Orsini B, Baste JM, Gossot D, et al. Index of prolonged air leak score validation in case of video-assisted thoracoscopic surgery anatomical lung resection: results of a nationwide study based on the French national thoracic database, EPITHOR. *Eur J Cardiothorac Surg* 2015;48:608-11.
 20. Kaminsky DA, Whitman T, Callas PW. DLCO versus DLCO/VA as predictors of pulmonary gas exchange. *Respir Med* 2007;101:989-94.
 21. Brunelli A, Varela G, Refai M, et al. A scoring system to predict the risk of prolonged air leak after lobectomy. *Ann Thorac Surg* 2010;90:204-9.
 22. Petrella F, Rizzo S, Radice D, et al. Predicting prolonged air leak after standard pulmonary lobectomy: computed tomography assessment and risk factors stratification. *Surgeon* 2011;9:72-7.
 23. Bernard A, Rivera C, Falcoz PE, et al. Application of model score of prolonged air leak in the French database. *Ann Thorac Surg* 2011;92:1548-50; author reply 1550.
 24. Demmy TL. Commentary: 20% Chance of Precipitation. *Semin Thorac Cardiovasc Surg* 2021;33:593-4.

Cite this article as: Divisi D, Pipitone M, Perkmann R, Bertolaccini L, Curcio C, Baldinelli F, Crisci R, Zaraca F; Italian VATS group. Prolonged air leak after video-assisted thoracic anatomical pulmonary resections: a clinical predicting model based on data from the Italian VATS group registry, a machine learning approach. *J Thorac Dis* 2023;15(2):849-857. doi: 10.21037/jtd-21-1484

Table S1 Predictors involved in the Model building algorithm

Group	Name
Admission	03_Admission.date, 04_Hospital.records.number, 05_Patient.ID, 06_Gender, 07_Patient.code, 08_Age.at.surgical.procedure.date, 09_Type, 10_Validation.date, 11_Experience.number
Comorbidities	03_Smoker, 08_Myocardial.infarction, 09_Congestive.heart.failure, 10_Peripheral.vascular.disease, 11_Cerebrovascular.disease, 12_Dementia, 13_COPD, 14_Connective.tissue.disease, 15_Peptic.ulcer.disease, 16_Diabetes.mellitus, 17_Moderate.to.severe.chronic.kidney.disease, 18_Hemiplegia, 19_Leukemia, 20_Malignant.lymphoma, 21_Solid.tumor, 22_Liver.disease, 23_AIDS, 24_Scoring..Age, 25_Final.score
ECOG	03_ECOG.score
Neoadjuvant.treatments	03_Done
PET	03_Done
Preoperative.diagnosis	03_Done
Pulmonary.function.evaluation	03_BMI..Height..cm., 04_BMI..Weight..kg., 05_BMI..value, 06_Spirometry..Done, 08_Spirometry..FEV1.perc, 10_Spirometry..FVC.perc, 12_Spirometry..DLCO.slash.VA.perc, 13_Hemogasanalysis..Done, 17_6.minutes.walking.test..Done, 23_Cardiac.stress..Done, 26_Lung.scintigraphy..Done
Surgical.procedure	03_Date, 04_Seniority..years.from.the.end.of.the.residency., 05_First.operator.experience.in.VATS.lobectomy, 06_Surgical.technique, 07_Number.of.incisions.apart.from.utility.thoracotomy, 08_Number.of.trocars, 09_Length.utility.thoracotomy..cm., 10_Degree.thoracoscope, 11_Length.thoracoscope..mm., 12_Operative.time..skin.to.skin.in.minutes., 13_Side, 14_Type.of.resection, 15_En.bloc, 16_Fissureless.lobectomy, 17_Lymph.node.dissection..Done, 18_Lymph.node.dissection..Type, 19_Lymph.node.dissection..Technique, 37_Number.of.N1.resected.lymph.nodes, 38_Number.of.N2.resected.lymph.nodes, 39_Number.of.N3.resected.lymph.nodes, 40_Total.resected.lymph.nodes, 41_Preoperative.histology, 42_Conversion, 60_Blood.loss, 61_Adhesiolysis, 62_Soft.tissue.retractors..Used, 64_Number.of.chest.tubes.used, 65_Chest.tube.1..FR.slash.CH., 66_Chest.tube.1.Under.water.seal.drain, 73_Pain.relief.techniques..Intercostal.block, 74_Pain.relief.techniques..Pericostal.cathethers, 75_Pain.relief.techniques..Peridural.cathethers, 76_Pain.relief.techniques..Elastomeric.pump.ev, 77_Pain.relief.techniques..PEC.block.1, 78_Pain.relief.techniques..PEC.block.2, 79_Pain.relief.techniques..Serratus.Plane.Block, 80_Pain.relief.techniques..Other, 83_Haemostasis.and.tissue.sealing..Used
Thorax.ct.scan	03_Done, 04_T..Value, 05_T..Size, 06_T..Lung, 07_T..Lobe, 09_N, 10_Station..N1.10, 11_Station..N1.11, 12_Station..N1.12_13_14, 13_Station..N2.1, 14_Station..N2.2, 15_Station..N2.3, 16_Station..N2.4, 17_Station..N2.5, 18_Station..N2.6, 19_Station..N2.7, 20_Station..N2.8, 21_Station..N2.9, 22_Station..N3.Contralateral, 23_Station..N3.Supraclavicular, 24_Station..N3.Other, 25_M, 26_M..Pleural.Effusion..citologically.proven., 27_M..Controlateral.Lung, 28_M..Thorax, 29_M..Brain, 30_M..Bone, 31_M..Extra.thoracic.nodes, 32_M..Kidney, 33_M..Adrenal.gland, 34_M..Bowel, 35_M..Liver, 36_M..Other, 37_TNM.Stage..8th.ed., 39_Ground.Glass.Opacity

PAL, prolonged air leak; BMI, body mass index; FEV1, forced expiratory volume in 1 second; COPD, chronic obstructive pulmonary disease; FVC, forced vital capacity; DLCO/VA, diffusing capacity of the lung for carbon monoxide divided by alveolar volume; ECOG, Eastern Cooperative Oncology Group; PET, positron emission tomography; CT, computer tomography.