Aus der Klinik und Poliklinik für
Herzchirurgie, herzchirurgische Intensivmedizin und Thoraxchirurgie
der Universität zu Köln
Direktor: Universitätsprofessor Dr. med. Th. Wahlers

Eine vergleichende Studie von vier Thorax-Mortalitäts-Scores

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Medizinischen Fakultät
der Universität zu Köln

vorgelegt von
Andrés Omar Amorín Estremadoyro
aus Lima, Peru

promoviert am 14. Juni 2024

Gedruckt mit Genehmigung der Medizinischen Fakultät der Universität zu Köln

2024

Dekan: Universitätsprofessor Dr. med. G. R. Fink

1. Gutachter: Professor Dr. med. K. Hekmat

2. Gutachter: Privatdozent Dr. med. K. Frank

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Dissertationsschrift ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskriptes habe ich Unterstützungsleistungen von Prof. Dr. med. Khosro Hekmat. Weitere Personen waren an der Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe einer Promotionsberaterin/eines Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertationsschrift stehen. Die Dissertationsschrift wurde von mir bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die in dieser Arbeit verwendeten Datensätze wurden von mir ohne unzulässige Hilfe Dritter gesichtet, zur Datenanalyse extrahiert und anschließend anonymisiert statistisch ausgewertet. Alle patientenbezogenen Daten stammen aus den entsprechenden digitalisierten Patientenakten im Krankenhausinformationssystem Orbis® (ORBIS AG, Deutschland) sowie im Quantitative Imaging in Medicine and Surgery (QIMS) (terraconnect GmbH & Co. KG ·Deutschland) des Kölner Herzzentrums und der CardioClinic Kalk.

Erklärung zur guten wissenschaftlichen Praxis:

Ich erkläre hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten (Amtliche Mitteilung der Universität zu Köln AM 132/2020) der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.

Köln, den 25.09.2023

Andrés Amorín

Andrés Omar Amorín Estremadoyro

Acknowledgments

First, I would like to thank my doctoral supervisor Prof. Dr. med. Khosro Hekmat for the granting of the topic as well as to Prof. Dr. med Anton Sabashnikov and Dr. med. Matthias Heldwein for the support, stimulation, and advice during the preparation of this work.

A special thanks goes to my parents, who supported my siblings and me in every way. Finally, my heartfelt thanks go to Laura for her full support, and without her, I would never have been able to create this work.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AUC	Area under the curve
ASA	American society of anesthesiology
BMI	Body mass index
CHD	Coronary heart disease
CI	Confidence interval
CT	Computed tomography
CVD	Cardiovascular disease
DLCO	Diffusing capacity for carbon monoxide
FEV_1	Forced expiratory volume in the first second
HL	Hosmer-Lemeshow test
MDT	Multidisciplinary team
NSCLC	Non-small cell lung cancer
OCC	Overall correct classification
OR	Odds ratio
$ppoFEV_1$	Predicted postoperative FEV_1
ROC	Receiver-operating-characteristics
SD	Standard deviation
UICC	Union for International Cancer Control
VATS	Video-assisted thoracic surgery
RATS	Robotic-assisted thoracic surgery

1 Summary

Der Prozentsatz der Patienten im resezierbaren Stadium bei der initialen Diagnose von nicht-kleinzelligem Lungenkrebs (NSCLC) nimmt zu. Der Anteil der Patienten, die zwar eine Operation durchführen könnten, jedoch aufgrund ihres schlechten Allgemeinzustands nicht für eine Lungenresektion geeignet sind, steigt im Anschluss. In diesem Kontext gewinnen Risikovorhersagemodelle zunehmend an Bedeutung für die geeignete Auswahl von Patienten mit vorhergesagter adäquater kurzfristiger Überlebensrate. Hier haben wir die vier bekanntesten Bewertungsmodelle Thoracoscore, Epithor, Eurolung 2 und den Eurolung 2b hinsichtlich ihrer diskriminierenden Fähigkeit zur Vorhersage der 30-Tage-Mortalität für unsere Kohorte validiert und verglichen.

Die vier Bewertungssysteme wurden unter Verwendung von Kalibrierungs- und Diskriminierungsstatistiken analysiert. Wir verglichen die Fläche unter der Kurve (AUC) der Receiver-Operating-Characteristic (ROC) mittels DeLong-Methode und berechneten Gesamtkorrektklassifikationswerte (OCC).

In unserer Analyse wurden insgesamt 624 aufeinanderfolgende Patienten eingeschlossen, die sich zwischen 2012 und 2018 an unserer Einrichtung einer Operation wegen NSCLC unterzogen. Wir beobachteten eine 30-Tage-Mortalität von 2,2% (14 Patienten). Die AUC für Eurolung 2 und Eurolung 2b (0,82) waren größer als die der anderen Bewertungssysteme wie Epithor (0,71) und Thoracoscore (0,65). Die DeLong-Analyse zeigte eine signifikante Überlegenheit von Eurolung 2 und Eurolung 2b gegenüber dem Thoracoscore (p=0,04).

Eurolung 2 und der Eurolung 2b waren die bevorzugten Bewertungssysteme für die Risikostratifizierung der 30-Tage-Mortalität im Vergleich zu Thoracoscore und Epithor. Bewertungssysteme sind ein wertvolles Werkzeug, um Patienten während der präoperativen Bewertung für weitere funktionelle Tests zu screenen und den Entscheidungsfindungsprozess zu unterstützen. Dennoch sollte kein Bewertungssystem allein darüber entscheiden, einen Patienten von einer Operation auszuschließen.

2 Introduction

2.1 Pulmonary Malignancies Worldwide

Pulmonary malignancies have consistently ranked as the second most common type of tumor worldwide, contributing to significant mortality rates. The prevalence and prognosis of these malignancies often depend on their histopathological classification and stage [1].

2.2 Non-Small Cell Lung Cancer (NSCLC)

Non-small cell lung cancer (NSCLC) stands out with a relatively better prognosis and survival rates compared to other forms of pulmonary malignancies, such as small cell lung cancer. Early-stage NSCLC, when diagnosed, offers the opportunity for curative anatomical lung resection, which remains the gold standard treatment [2]. Advances in diagnostic tools, particularly the widespread use of high-resolution computed tomography (CT), have led to an increase in the number of patients diagnosed at early, curatively treatable stages (UICC-Stage I-IIIA) [3-5].

2.3 Challenges in Operability

However, an aging population with significant comorbidities poses challenges to the operability of patients. This demographic shift has resulted in a growing number of patients with resectable UICC stages for NSCLC who are not suitable candidates for lung resection due to poor general health conditions [5-7]. Research by Wang et al. and Lin et al. has highlighted the rising incidence of early-stage NSCLC among patients aged 65 and older with comorbidities and impaired lung function [8, 9]. To bridge the gap between the potential for curative resection in NSCLC patients and the high perioperative mortality among high-risk individuals, risk prediction models for preoperative risk stratification have gained clinical significance in improving post-surgery survival.

2.4 Evolution of Risk Stratification Models

Over time, various risk stratification strategies have been proposed, ranging from relatively straightforward calculations like the predicted postoperative FEV1 (ppoFEV1) [10], to more complex modified versions of ppoFEV1 [11], and highly intricate logarithmic scoring systems. Ultimately, four globally recognized models have emerged as the most widely established: the Thoracoscore by Falcoz et al. 2007 [12], the Epithor by Bernard et al. 2011 [13], Eurolung 2 by Brunelli et al. [14], and the simplified Eurolung 2 by Brunelli et al. [15] (referred to as Eurolung 2b for simplification in this work). These scoring systems were designed with the overarching goal of reliably predicting in-hospital mortality in lung cancer patients following pulmonary resection.

2.5 Objectives of the Study

The primary objective of this work is to assess and compare the performance of these four scoring models in predicting 30-day mortality within a large cohort of patients from our institution. Additionally, we aim to identify independent factors associated with 30-day mortality across different scoring systems.

3 Material and Methods

3.1 Patient Selection

Perioperative data were collected from our institutional patients' database. Patients with reasonable suspicion of lung cancer were staged and diagnosed according to European lung cancer guidelines. Furthermore, pulmonary function testing was obligatory, including measurements of the diffusing capacity for carbon monoxide (DLCO) and the forced expiratory volume in the first second (FEV_1). All patients were discussed in the multidisciplinary team (MDT). The surgical approach for anatomic resection was in most cases, uniportal video-assisted thoracic surgery (VATS). Furthermore, we used a muscle-sparing thoracotomy with an incision of 4-7 cm as a surgical approach. Surgery was performed by experienced senior surgeons only. 30-day mortality was measured from the time of surgery to the date of death.

3.2 Statistical Analyzes

Statistical analyses were performed with SPSS software version 23 (SPSS Inc., Chicago, IL). Continuous scale data are presented as mean ± standard deviation (SD) and were analyzed using the two-tailed Student's t-test for independent samples. The Kolmogorov-Smirnov test showed a normal distribution of the continuous data. A p-value of <0.05 was considered significant. Calibration was performed using the Hosmer-Lemeshow (HL) test (goodness-of-fit-test) to ensure the absence of a significant discrepancy between predicted and observed mortality. Calibration was considered good when there was a low $\chi 2$ value and a high p-value (>0.05).

Discrimination (the ability of a scoring model to differentiate between survival and death) was evaluated with receiver-operating-characteristic (ROC) curves; the area under the curve (AUC) indicates the discriminative ability of the scores, hence the ability to discriminate survivors from non-survivors. AUCs enable direct comparison of different scoring systems.

AUC of 0.5 (a diagonal line) is equivalent to random chance, AUC >0.7 indicates a moderate prognostic model, and AUC >0.8 (a bulbous curve) indicates a good prognostic model [16]. To compare the ROC curves of the scoring systems, we performed an analysis according to the method of DeLong et al. [17]. The overall correct classification (OCC) (the ratio of the number of correctly predicted survivors and non-survivors to the total number of patients) values of the scores were calculated. The risk of mortality is given as odds ratios for all scorers with 95%-confidence intervals.

Furthermore, we analyzed whether parameters within the different scoring systems were significant factors for 30-day mortality in our cohort. If univariate analyses showed significant differences between patients who died in the first 30 days after surgery compared to patients who survived, multinomial regression analysis followed for further evaluation of the results. To analyze which parameters of the individual scores were independent factors for 30-day mortality in our cohort. Multinomial differences were described by odds ratio (OR) and 95% confidence interval (CI). Categorical variables were analyzed using Pearson’s chi-squared or Fisher exact test. Continuous parameters were expressed as mean ± standard deviation (SD) of the mean and were analyzed via unpaired Student t-test. A p-value <0.05 was considered statistically significant. The variables included in these four scores are shown in Table 1.

3.3 Ethics Statement

This investigation was approved by the institutional review board of our university hospital of Cologne and conformed to the principles outlined in the Declaration of Helsinki. Due to its retrospective study, the need for written informed consent was waived.

4 Results

4.1 Patients’ Characteristics

Table 2 represents patients’ perioperative clinical characteristics according to the score models. These variables provide information on demographic characteristics, clinical features, lung function, comorbidities, and surgical details of the study participants. Data from 625 consecutive patients who underwent anatomical pulmonary resections between 2012 and 2018 due to NSCLC at our institution were included in the present analysis.

4.1.1 Demographic Characteristics

The mean age of the participants in the study 64.0 ± 11.0 years, whereas 59.5% were male, indicating a slightly higher representation of males in the study compared to females. The mean body mass index (BMI) of the participants was 25.9 ± 4.9. A BMI between 18.5 and 24.9 was considered normal weight with the majority of included patients falling within the normal weight range or being slightly overweight. In the patient cohort analyzed, 8.0% of the participants had an ASA (American Society of Anesthesiologists) score greater than 3. The ASA score assesses the physical health and fitness of a patient before surgery, with a higher score indicating more systemic disease or functional impairment.

4.1.2 Clinical Features and Health Status

In addition, 9.1% of this cohort showed a performance status classification greater than 3. Performance status is a measure of a patient's overall functional ability and is commonly used in oncology to assess patient's ability to carry out daily activities and self-care. In terms of symptomatic dyspnea, 10.1% of the patients had a dyspnea score greater than 2, meaning greater sensation of breathlessness or difficulty in breathing, as a higher score indicates more severe symptoms.

As referred in Table 2, the mean forced expiratory volume in one second (FEV_1) indicated as percentage of predicted value for the participants was 79.3 ± 18.1%. FEV_1 is a lung function parameter, representing the actual value of the participants relative to the value that would be expected for their age, sex, and height. The mean predicted postoperative forced expiratory volume in one second ($ppoFEV_1$) for the participants was 1.79 ± 0.6 L. The value of $ppoFEV_1$ was an estimate of the expected FEV_1 in lung function test after surgery, considering the individual's preoperative lung function and the planned extent of surgery.

4.1.3 Oncological Characteristics

From the oncological perspective, no participants demonstrated past medical history of any malignancy, indicating that none of the participants in the study had been previously diagnosed with cancer. Regarding the actual diagnosis, 25.9% of the patients were classified with a tumor stage higher than IIB. Tumor staging is a system used to describe the size and spread of cancer, with higher stages indicating more advanced disease.

4.1.4 Comorbidities

From the entire cohort, 35% reached a comorbidity score greater than 3. The comorbidity score assesses the presence and severity of other medical conditions in addition to the primary diagnosis in our study. Complementary to their comorbidity score, 21.8% of the participants demonstrated the presence of coronary heart disease (CHD), whereas 3.0% of the participants manifested other cardiovascular diseases (CVD), a broad term that encompasses various conditions affecting the heart and blood vessels.

4.1.5 Surgical Details

The following percentages indicate the distribution of participants based on the surgical approach used. All surgical procedures were performed in an elective manner with no urgent or emergent cases. Regarding the operation side, 59.7% were performed on the right side, while 40.3% were performed on the left side of the thorax. Moreover, 6.6% of the participants underwent pneumonectomy. The other 14.7% of patients underwent extended resections which typically refers to the removal of additional adjacent tissues or structures along with the primary surgical target. Whereas 28% of the patients were operated on through thoracotomy, surgical access on 72% of participants was approached via video-assisted thoracic surgery (VATS), a minimally invasive surgical approach that uses small incisions and a camera for visualization. Overall, 30-day mortality after surgery was 2.2% (14 patients).

4.2 Analysis of our Cohort

Parameters of the different scoring systems were analyzed in terms of occurrence frequency in patients who died within 30 days compared to patients who survived the first 30 days after surgery. Parameters of the scoring systems are shown in Table 1 whereas patients' characteristics according to the score models are shown in Table 2.

In the 30-day mortality group, significantly more patients were men (85.7% vs. 58.9%; $p=0.01$). Male gender was also an independent factor for 30-day mortality (OR 9.1, CI 1.12-70.4; $p=0.03$). Furthermore, FEV_1 (77.3 ± 18.2 L vs. 83.3 ± 22.6 L; $p=0.01$) and the predicted postoperative FEV_1 ($ppoFEV_1$) (1.52 ± 0.52 L vs. 1.80 ± 0.6 L; $p=0.05$) differed significantly between groups in the univariate analysis.

However, these two parameters showed no independent effect on 30-day mortality in our cohort. Patients who underwent pneumonectomy died significantly more often in the first 30-days after surgery (21.4% vs. 6.2%; $p=0.001$). Pneumonectomy was an independent factor for 30-day mortality in the multivariate analysis (OR 8.58, CI 2.82-27.8; $p<0.001$).

Table 1: Summary of variables included in the different scoring systems

	Thoracoscore	*Epithor*	*Eurolung 2*	*Eurolung 2b*
Preoperative parameters:				
Age	X	X	X	X
Gender	X	X	X	X
BMI		X	X	X
ASA Score	X	X		
Performance status classification	X	X		
Dyspnea score	X			
FEV_1%		X		
$ppoFEV_1$			X	X
Comorbidities	X	X		
CHD			X	
CVD			X	
Benign/malign	X			
Tumor stage		X		
Surgical approach:				
Priority of surgery (elective/urgent)	X			
Operation side		X		
Pneumonectomy (other)	X	X	X	X
Extended resection (sleeve)		X	X	
Thoracotomy/VATS			X	X

Abbreviations: ASA classification: American Society of Anesthesiology classification; BMI: Body mass index; CHD: coronary heart disease; CVD: cardiovascular disease; FEV_1: forced expiratory volume in the first second; $ppoFEV_1$: predicted postoperative FEV_1; VATS: Video-assisted thoracic surgery

Table 2: Mortality scores in our cohort

	Total cohort n=624	30-days mortality (n=14)	30-days survival (n=610)	univariate p-value	multivariate p-value
Preoperative parameters:					
Age	64.0 ± 11.0	67.7 ± 15.9	63.9 ± 10.8	0.08	
Male gender (%)	59.5	85.7	58.9	**0.01**	**0.03**
BMI	25.9 ± 4.9	26.1 ± 2.7	25.8 ± 4.9	0.72	
ASA score >3 (%)	8.0	9.2	7.9	0.17	
Performance status classification >3 (%)	9.1	14.3	7.1	0.14	
Dyspnea score >2 (%)	10.1	13.1	9.3	0.12	
FEV_1 %	79.3 ± 18.1	77.3 ± 18.2	83.3 ± 22.6	**0.01**	0.3
$ppoFEV_1$ l	1.79 ± 0.6	1.52 ± 0.52	1.80 ± 0.6	**0.05**	0.4
Comorbidity score >3 (%)	35	37	35	0.17	
CHD (%)	21.8	21.6	21.9	0.63	
CVD (%)	3.0	3.0	7.1	0.36	
Benign/malign (%)	0/100	0/100	0/100	/	
Tumor stage >IIB (%)	25.9	26.1	25.8	0.81	
Surgical approach:					
Priority of surgery (elective/urgent) (%)	100/0	100/0	100/0	/	
Operation side (right/left) (%)	59.7/40.3	57.1/42.9	59.7/40.3	0.76	
Pneumonectomy (other) (%)	6.6	21.4	6.2	**<0.001**	**<0.001**
Extended resection (%)	14.7	21.8	14.6	0.46	
Thoracotomy/VATS (%)	28/72	36/64	28/72	0.23	

Abbreviations: ASA classification: American Society of Anesthesiology classification; BMI: body mass index; CHD: coronary heart disease; CVD: cardiovascular disease; FEV_1: forced expiratory volume in the first second; $ppoFEV_1$: predicted postoperative FEV_1; VATS: video-assisted thoracic surgery

4.3 Comparison of the Mortality Scores

Table 3 summarizes the OCC, calibration, and discrimination of all four score systems. There were no significant differences between expected and observed mortality for *Thoracoscore*, *Epithor*, *Eurolung 2*, and *Eurolung 2b* using the HL-test. Figures 1-4 show the ROC-Curves of all scoring models. The AUC for *Eurolung 2* and *Eurolung 2b* (0.82) were greater than those of the other scoring systems (Figures 3-4). The AUC of *Epithor* was 0.71 and of 0.65, respectively (Table 3, Figures 1-2). The DeLong analysis showed a significant superiority of *Eurolung 2* and *Eurolung 2b* over the *Thoracoscore* (p=0.04).

The DeLong analysis of the other ROC curves showed no significant differences. The OCC are presented in Table 3. All scoring systems reached the same OCC of 97.8%. Furthermore, there were no significant differences between expected and observed mortality using the HL-test. The risk for mortality prediction was expressed by the log-rank test. There were no significant differences in the prediction of mortality between the scoring systems.

Table 3: Summary of variables included in the different scoring systems

Scoring model	Logistic regression		OCC	HL-test		ROC-Analysis	
	OR	CI 95%	%	x^1	p-value	AUC	CI 95%
Thoracoscore	1.1	1.0-1.2	97.8	14.3	0.67	0.65	0.51-0.79
Epithor	1.06	1.03-1.09	97.8	4.7	0.78	0.71	0.57-0.81
Eurolung 2	1.2	1.1-1.3	97.8	4.4	0.82	0.82	0.67-0.96
Eurolung 2b	1.4	1.2-1.5	97.8	12.5	0.13	0.82	0.6-0.85

Abbreviations: AUC: area under the curve; CI: confidence interval; HL: Hosmer-Lemeshow test; OR: Odds ratio, OCC: overall correct classification; ROC: receiver-operating-characteristic

5 Discussion

5.1 Prognostic Variation in Scoring Methods

The overall correct classification obtained for all the selected scoring methods in this retrospective validation, as well as in previous studies of *Thoracoscore*, *Epithor*, *Eurolung 2*, and simplified *Eurolung 2b* reached a high value of 97.8% [12–15]. Furthermore, there was no significant discrepancy on the Hosmer-Lemeshow test between expected and observed mortality in the patient collective analyzed. However, following detailed analysis of each predictive model, it became evident that various predictive methods performed differently in terms of their prognostic efficacy (Figures 1-4). Whereas *Thoracoscore*'s AUC showed a value of 0.65 and the *Epithor*'s reached 0.78, thus corresponding to a reasonable prognostic model, *Eurolung 2* and *2b* accounted for AUC of 0.82, corresponding to an effective prognostic model based on our observation.

5.2 Rejecting Patients with High Mortality Scores

There is a growing trend towards rising proportion of patients with surgically resectable early-stage NSCLC, however not being suitable for surgery due to their overall poor medical condition or presence of significant comorbidities [5-7]. In other words, whereas there are no issues in terms of technical aspects of surgical resection, the risk/benefit ratio could be significantly compromised, if no thoughtful selection is performed. Therefore, it is crucial to use appropriate risk stratification tools, such as mortality assessment estimation. In our study *Eurolung 2* and *2b* revealed an AUC of 0.82, providing more powerful strategies for prognostic purposes after surgery than *Thoracoscore* and *Epithor* (Figures 3, 4) corroborating previous studies [14, 15]. As expected per these predictive models, patients with high scores had a higher risk of dying within 30 days after surgery. Whereas scoring systems can provide valuable assessment orientation in terms of risk/benefit ratio of lung resection, they should be used with caution as patient rejection cannot be solely rely on prediction models. This is due to the fact that other clinical factors that cannot be assessed by scoring models, such as frailty, comorbidities, previous surgery and other individual factors. In addition, none of the four assessment methods undergoing analysis offer a cutoff value for denial of surgery and no other scoring system has been claimed to be capable of this judgement [15].

5.3 Reliability of 30-day mortality prediction following Test Calibration

Fortunately, in-hospital mortality following pulmonary resection remains fairly low. The 30-day mortality rate in our cohort accounted for 2.2% (14 patients from the entire cohort included). This corroborates previously published data by Brunelli et al. on accuracy of thoracic mortality score of *Eurolung 2* and *Eurolung 2 simplified* with mortality rates of 2.7% and 2.2%, respectively [14, 15]. Low perioperative mortality rate of 2.2% was also shown by Falcoz et al. in their research on efficacy of *Thoracoscore* risk model [12]. Furthermore, Pompili et al found comparable low 30-day mortality rates in their validation study on *Eurolung 2* [18].

The calibration of scoring systems is generally more challenging the more frequently the endpoint under examination occurs, and even more if it occurs very infrequently, as in the case of mortality following lung resections. These issues in prediction are not associated with a single scoring system, or the variables included; it is an intrinsic fault of any attempt to predict low-rate events [19].

5.4 The Formula for our Cohort Scores

In general, a well-functioning and reliable scoring system should be on one hand simple in use and on the other hand reproducible for a large number of patients [16]. The four scoring methods analyzed in our study varied greatly in multiple aspects. Whereas each predicting score model was uniformly reproducible, the simplicity as well as the accuracy of the tests fluctuated substantially (Tables 1, 3, and Figures 1-4).

5.4.1 Thoracoscore

The oldest scoring system used in this investigation for in-hospital mortality following lung resection is the *Thoracoscore* by Falcoz et al., which was published in 2007 [12]. Employing this score system is uncomplicated. The least accurate of the included systems, the *Thoracoscore* displayed an AUC of 0.65 (Figure 1). *Thoracoscore* underperformed significantly compared to *Eurolung 2* and *Eurolung 2b*, according to the DeLong analysis (p=0.04).

Only NSCLC patients who experienced lobectomy, pneumonectomy, or sleeve resection were included in our cohort. *Thoracoscore*, on the other hand, takes into account the malignancy parameter, which is amplified by a factor of 1.2423. For this mentioned scoring system the operation urgency is also included in this prediction model. This criterion is not applicable to our cohort as only NSCLC patients who received elective surgery were included in our study (Table 1). In our population, pneumonectomy and male gender were independent risk factors for 30-day mortality (Table 2).

Of all the scoring methods, *Thoracoscore* scores male gender with a multiplier of 0.4505, which is low compared to other scoring systems. On the other hand, pneumonectomy was the highest multiplier in *Thoracoscore* among other prediction score systems accounting for 1.2176. As demonstrated by Qadri et al. in their study, *Thoracoscore* was not sufficient enough to accurately predict the probability of in-hospital death in pneumonectomy patients [20]. De Loucou et al. presented a modified version of *Thoracoscore* in 2020 that appeared to modestly enhance its performance [21].

Figure 1. AUC for *Thoracoscore* prognostic system. Qadri et al., 2014, European Journal of Cardio-Thoracic Surgery, 45, p. 867 (Adapted Graphic).

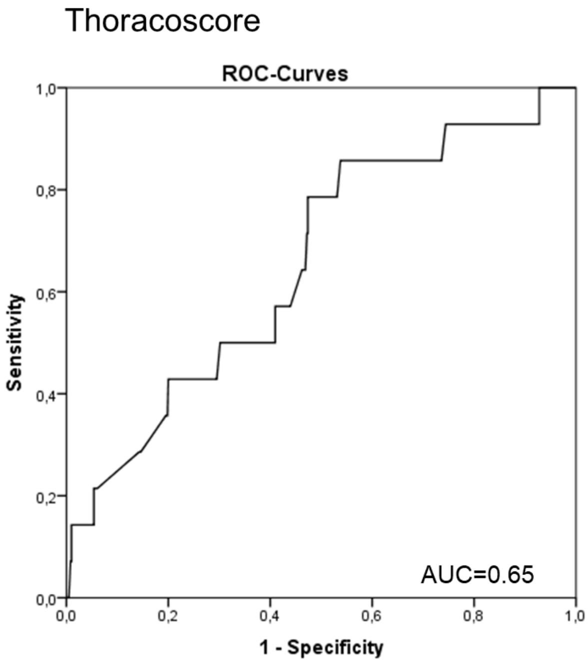
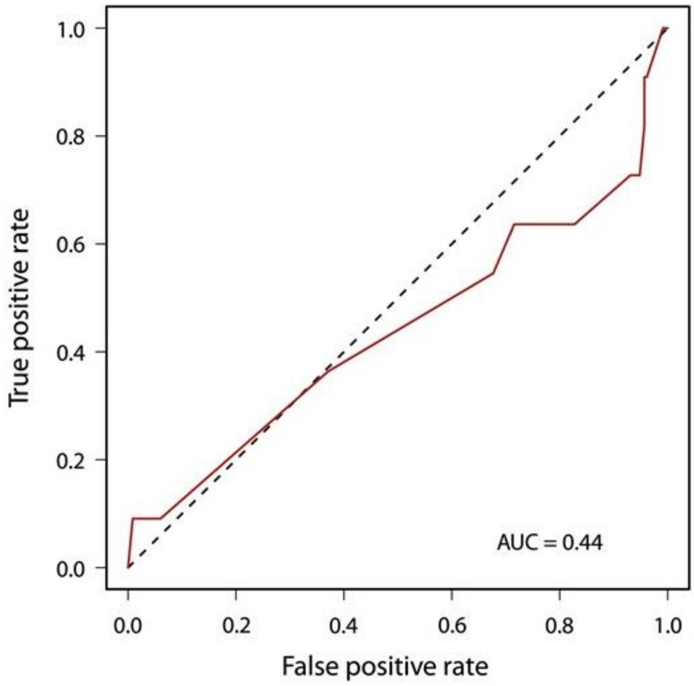


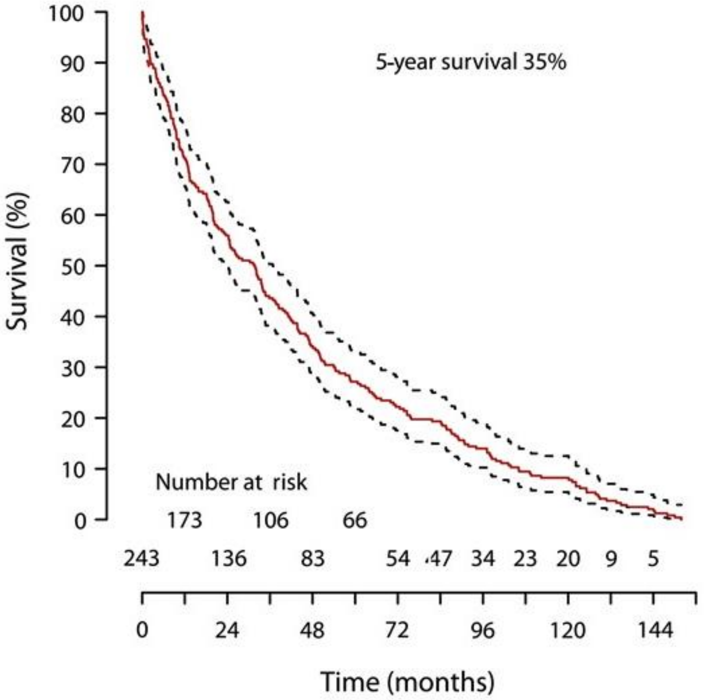
Figure 2 exhibits that the discriminatory ability of *Thoracoscore* in the group of patients in the study by Syed et al. In-hospital mortality was not satisfactorily measured by the C statistics/ROC curve (area under the receiver operator characteristic curve (AUC) accounted for 0.44) [20].

Figure 2. ROC for in-hospital mortality by *Thoracoscore*. Qadri et al., 2014, European Journal of Cardio-Thoracic Surgery, 45, p. 867 (Adapted Graphic).



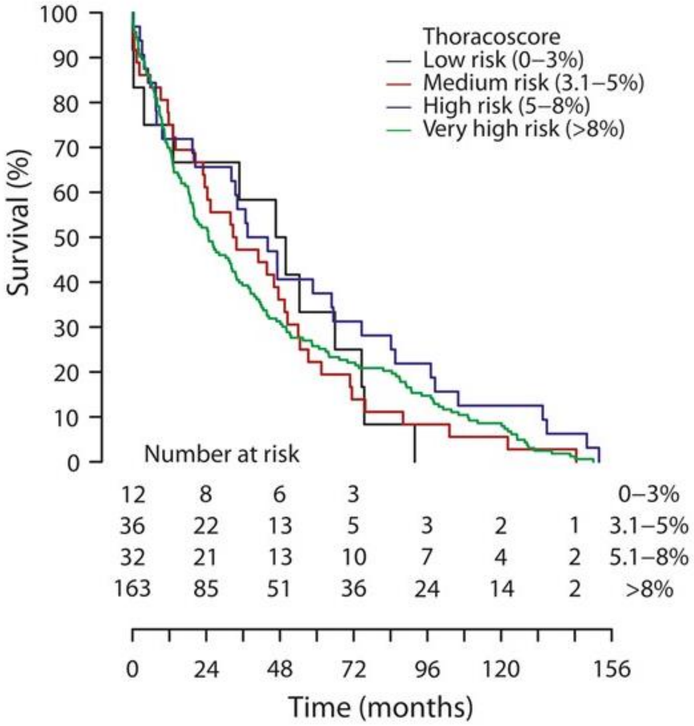
Survival analysis by Kaplan–Meier in the analysis of Syad et al. showed that 30-day, 1-year, 2-year and 3-year observed mortalities after pneumonectomy were 5.3%, 29%, 43% and 56%, respectively. Overall 5-year survival of all patients was 35±6% of years (Figure 3).

Figure 3. Kaplan-Meier survival analysis after pneumonectomy. Qadri et al., 2014, European Journal of Cardio-Thoracic Surgery, 45, p. 867 (Adapted Graphic).



In the same study, survival comparison of incremental risk groups showed poor 5-year survival (25%) in the very high-risk group (Figure 4).

Figure 4. Kaplan–Meier survival curves of all risk groups. Qadri et al., 2014, European Journal of Cardio-Thoracic Surgery, 45, p. 867 (Adapted Graphic).



5.4.2 Epithor

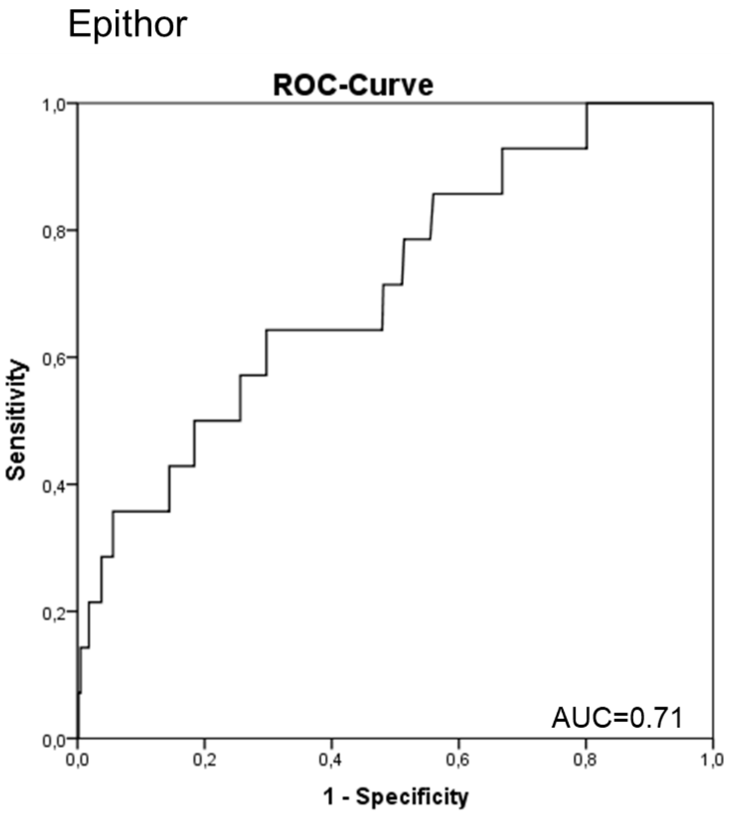
The *Epithor* assessment method was developed and presented by Bernard et al. in 2011 [13]. In our research group, *Epithor* demonstrated an AUC of 0.71 (Figure 5). According to the DeLong analysis, there were no significant differences compared to *Thoracoscore*, *Eurolung 2*, and *Eurolung 2b.* When compared to the other scoring systems, *Epithor* has by far the highest number of assessment variables. Additionally, different scoring methods, such as the performance score, are used in both *Epithor* and *Thoracoscore*. Moreover, a sub score with nine comorbidities is included in the *Epithor* prediction model.

Compared to other systems the *Epithor* is rather complicated in use. It is not clear if this contributes to the *Epithor* score being the least used or cited test in the literature. The *Epithor* was not even mentioned by Taylor et al. in their comparison study on external validation of six existing multivariable clinical prediction models for short-term mortality in patients undergoing lung resection [22]. No mobile application or website makes it simple to enter the score, in contrast to the other scores. Only the *Epithor* score takes into account both the tumor stage and the operation side. Also, this score benefits the operations on the left side. According to previous research, surgery on the left side has a survival advantage for pneumonectomy [23, 24].

The left upper lobe resection, on the other hand, is renowned for its technical complexity. For left upper lobe resection, prolonged air-leak has been frequently reported; in some cases, this resulted in redo surgical procedures with per se higher mortality risk [25]. Additionally, it has been noted that this resection considerably increases the risk of perioperative stroke [26].

Tumor stages III and IV had been evaluated individually by the *Epithor* score. Roughly 26% of the members of the group we studied were in UICC stages III or IV. Stage IV patients are considered with a 0.5 multiplier. Every stage IV patient who is at an oligo metastatic stage, at least in our department, is thoroughly reviewed in the MDT. In this regard, no stage IV patient in our group deceased within 30 days of surgery. Therefore, this component might be overstated.

Figure 5. Evaluation of Epithor Score in UICC Stages III and IV for Surgical Outcomes in Cancer Patients. Bernard et al., 2011, The Journal of Thoracic and Cardiovascular Surgery, 141, p. 449 (Adapted Graphic).



Alain et al. described in their *Epithor* risk stratification model three interactions among the variables surgery side and pneumonectomy, FEV_1 and pneumonectomy, and FEV_1 and extended resection. The predicted logit at a certain value of FEV_1 differed according to the type of pulmonary resection. In patients who had undergone pneumonectomy, the predicted logit varied to some extent according to the value of FEV_1. This was not the case for patients who had undergone limited resection or lobectomy (Figure 6).

The predicted logit varied according to the value of FEV_1 in patients who had undergone extended resection, whereas it decreased linearly among patients who had undergone simple pulmonary resection (Figure 7).

Patients who had undergone right-sided pneumonectomy had an adjusted odds ratio of 2.9 (95% CI, 1.44-5.88), and patients with left-sided pneumonectomy had an adjusted odds ratio of 1.78 (95% CI, 0.87-3.645; Figure 8) [13].

Figure 6. Interaction of pneumonectomy and forced expiratory volume (FEV_1; as a percentage). CI, Confidence interval. Bernard et al., 2011, The Journal of Thoracic and Cardiovascular Surgery, 141, p. 449 (Adapted Graphic).

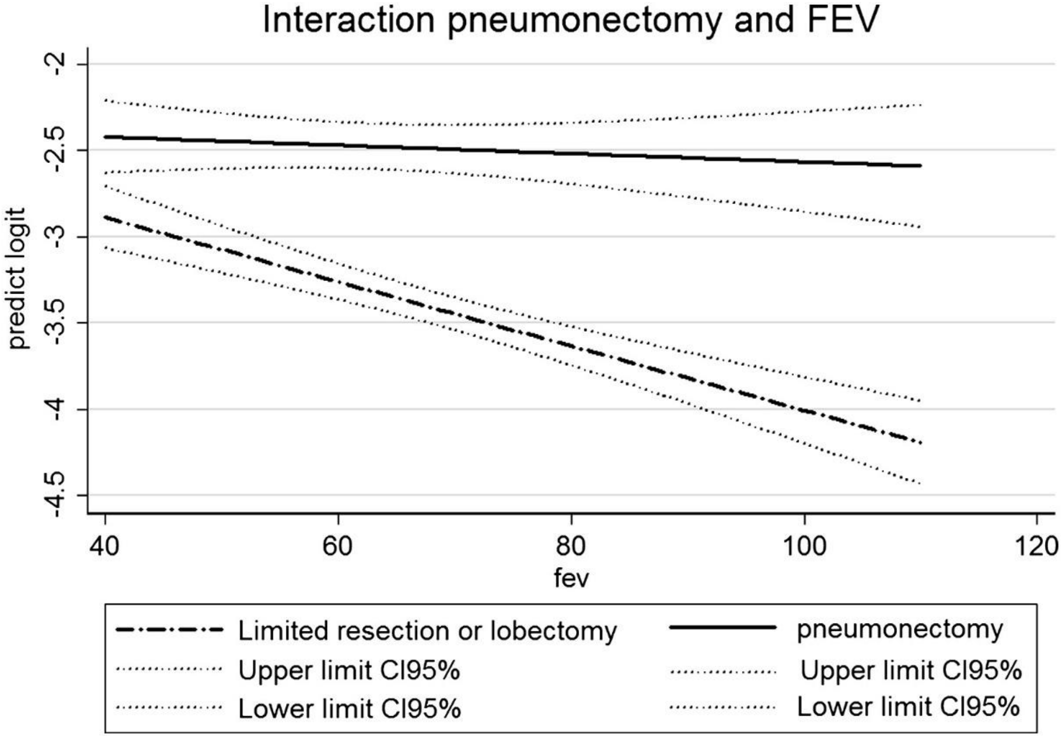


Figure 7. Interaction of extended resection and forced expiratory volume (FEV_1; as a percentage). CI, Confidence interval. Bernard et al., 2011, The Journal of Thoracic and Cardiovascular Surgery, 141, p. 449 (Adapted Graphic).

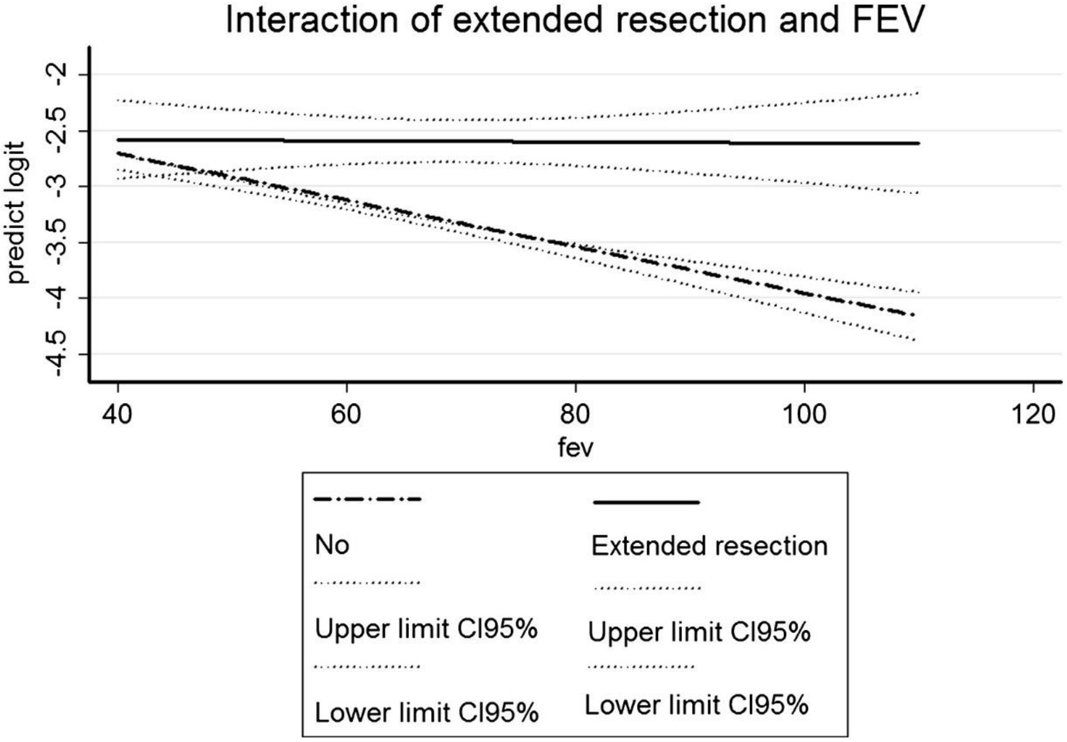
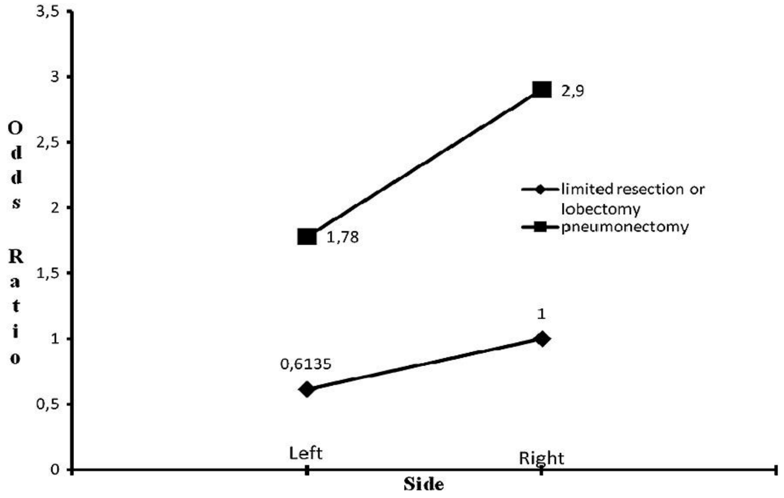


Figure 8. Interaction of side and type of lung resection. Bernard et al., 2011, The Journal of Thoracic and Cardiovascular Surgery, 141, p. 449 (Adapted Graphic).



5.4.3 Eurolung 2 and 2b

Eurolung 2 and 2b were both developed by Brunelli in 2017 and 2020, respectively [14, 15]. The AUC for *Eurolung 2* and *Eurolung 2b* (0.82 each) were greater than those for the other scoring systems (Figures 9 and 10). Variables age and pneumonectomy were associated with highest multipliers in the scores.

Figure 9. Performance Comparison of Eurolung 2 and Other Scoring Systems (AUC = 0.82). Brunelli et al., 2017, European Journal of Cardio-Thoracic Surgery, 51, p. 490 (Adapted Graphic).

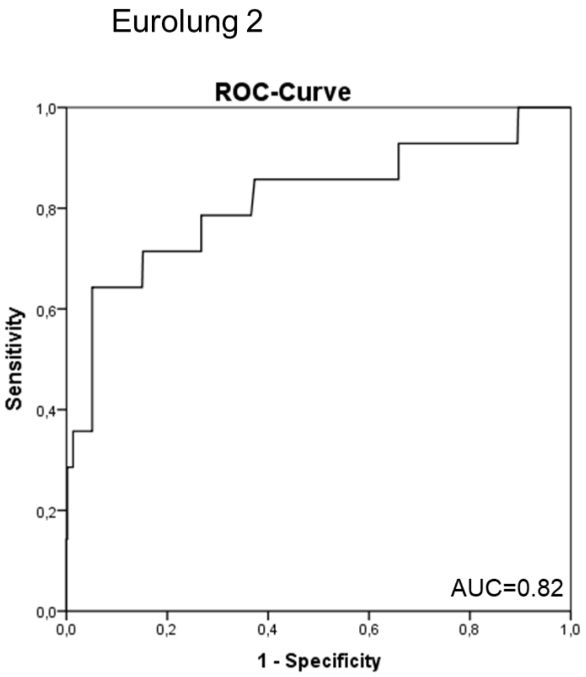
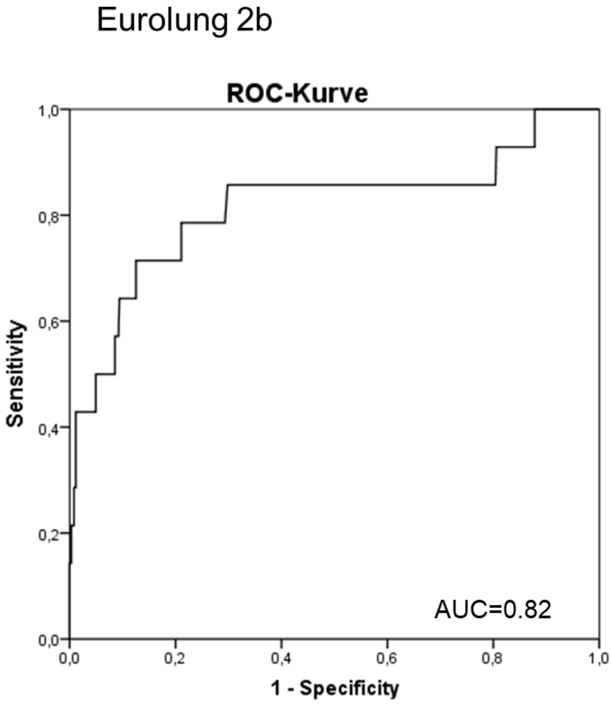


Figure 10. Performance Comparison of Eurolung 2 and Other Scoring Systems (AUC = 0.82). Brunelli et al., 2017, European Journal of Cardio-Thoracic Surgery, 51, p. 490 (Adapted Graphic).



These two parameters were also independent factors for 30-day mortality in our cohort analyzed (Table 2). Interestingly, the AUC and the OCC of the two scores were the same (Figures 3 and 4, Table 3).

In *Eurolung 2b* Brunelli et al. did not take the parameters coronary heart disease (CHD) and cardiovascular disease (CVD) into account. In our study the distribution of CHD was very similar, whereas the prevalence of CVD in general was relatively low (Table 2). That could explain why those factors did not influence the OCC of the two tests.

Furthermore, Brunelli et al. excluded extended resection in the modification that was also shown to be of low significance in our cohort. Thoracotomy was shown in both scoring systems to be one of the strongest multipliers influencing short-term survival.

Therefore, VATS appeared to be the preferred approach for better outcomes when technically feasible. However, no subgroup analysis in terms of uniportal VATS, triportal VATS or even RATS (Robotic-assisted thoracic surgery) was performed, whereas the 30-day mortality of these three procedures might differ [27, 28]. The absence of this stratification may be only a minor downside as both scores were shown to be very reliable prognostic models in general.

Finally, both tests are easy to use. For both scores, there are specific smartphone applications available.

Figure 11 shows the predicted morbidity plotted against the observed morbidity with the patients ordered by increasing risk of morbidity according to EuroLung1. The two lines almost overlap, indicating the high precision of the model [14].

Figure 11. (A) Lowess smoothing plots of the predicted and observed cardiopulmonary morbidity rates with the patients ordered by increasing risk of cardiopulmonary morbidity according to EuroLung1. (B) Lowess smoothing plots of the predicted and observed mortality rates with the patients ordered by increasing risk mortality according to EuroLung2. Brunelli et al., 2017, European Journal of Cardio-Thoracic Surgery, 51, p. 490 (Adapted Graphic).

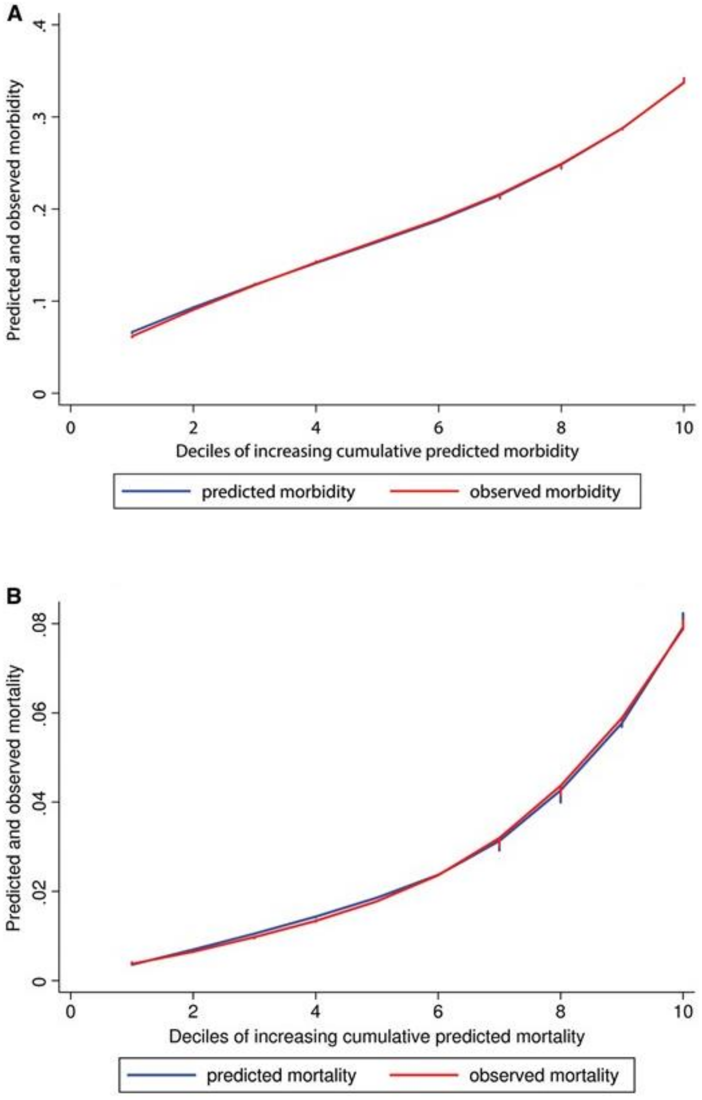


Figure 12 and 13 show the Locally Weighted Scatterplot Smoothing plots of the observed and predicted morbidity and mortality, respectively of the full and parsimonious Eurolung risk models with the patients ordered by deciles of predicted morbidity (according to the full model). The 3 curves are almost overlapped [15].

Figure 12. Locally Weighted Scatterplot Smoothing plots of the observed and predicted morbidity of the full and parsimonious models with the patients ordered by deciles of predicted morbidity (according to the full model). Brunelli et al., 2020, European Journal of Cardio-Thoracic Surgery, 57, p. 455 (Adapted Graphic).

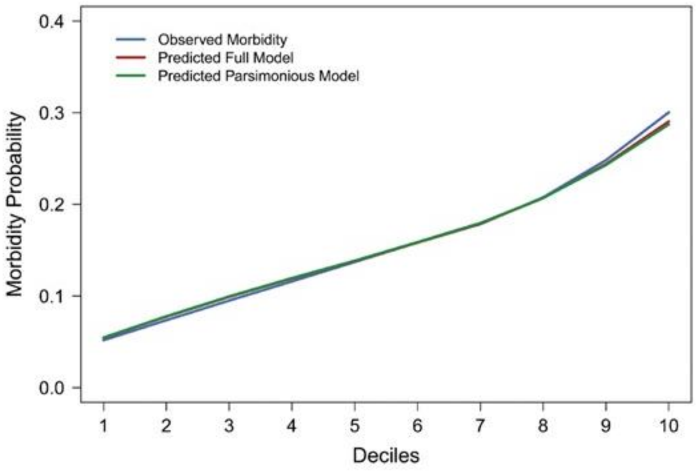
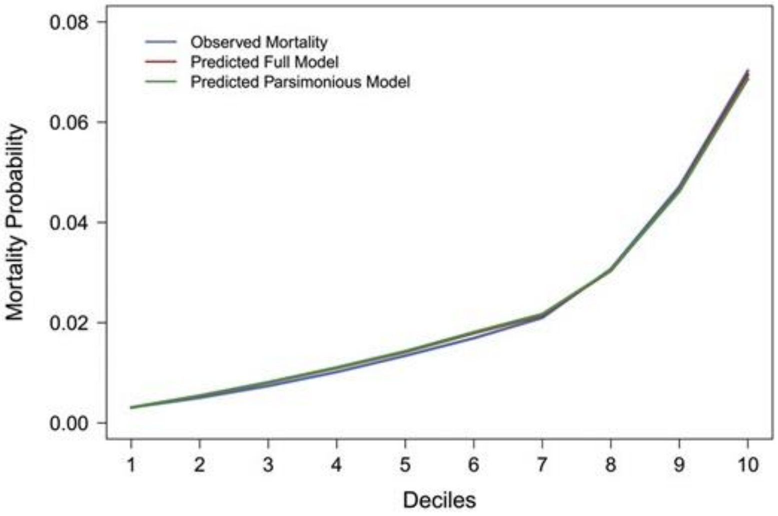


Figure 13. Locally Weighted Scatterplot Smoothing plots of the observed and predicted mortality of the full and parsimonious models with the patients ordered by deciles of predicted mortality (according to the full model). Brunelli et al., 2020, European Journal of Cardio-Thoracic Surgery, 57, p. 455 (Adapted Graphic).



5.5 Limitations

This study represents a retrospective analysis conducted at a single institution, and it is essential to emphasize that all data included in this analysis were sourced from an unselected patient population, reflecting the diversity encountered in everyday clinical practice. However, it is crucial to acknowledge the limitations associated with a single-center analysis, as this design inherently carries the potential for referral or selection bias. While the data collected offer valuable insights into the real-world experiences of patients, it is important to remain cautious when generalizing the findings to broader populations, as the influence of specific institutional practices or patient demographics cannot be ruled out.

Another noteworthy aspect is the absence of a statistical power estimation in this study. Power estimation is a critical step in research design, as it helps determine the minimum sample size required to detect significant effects accurately. In the absence of such estimation, there is a risk that the study may not have had sufficient statistical power to detect meaningful differences or associations. As a result, it is prudent to interpret the results with caution, recognizing the possibility of type II errors (i.e., failing to detect true effects due to inadequate sample size).

Additionally, it is important to acknowledge that the cohort analyzed in this study was relatively small. While the insights gained from this cohort are valuable, larger sample sizes are typically preferred in research to enhance the reliability and generalizability of findings. A larger cohort can help mitigate the effects of random variation and increase the precision of estimates.

In light of these limitations, it is evident that further research efforts are warranted. Specifically, larger prospective randomized studies with long-term follow-up periods are necessary to build upon these preliminary results. Such studies would not only help corroborate the current findings but also provide more robust and evidence-based data that can better inform clinical decision-making. In summary, while this retrospective analysis offers valuable insights, it serves as a foundation for future research endeavors aimed at achieving a deeper and more reliable understanding of the subject matter.

5.6 Conclusions for the daily practice

The current analysis has shed light on the performance of different scoring systems in risk-stratifying patients undergoing non-small cell lung cancer (NSCLC) surgery. Specifically, both the *Eurolung 2* scoring system and the simplified *Eurolung 2b*, as proposed by Brunelli et al., demonstrated their superiority in predicting the risk of 30-day mortality when compared to the Thoracoscore and *Epithor* models. It is paramount to delve into the reasons underlying the comparative inferiority of *Thoracoscore* and *Epithor* in this context.

One possible explanation for the observed inferiority of *Thoracoscore* could be its limited discriminatory ability. The ability of a scoring system to distinguish between patients at different levels of risk is a fundamental aspect of its utility. If *Thoracoscore* demonstrated reduced discriminatory power in this study, it implies that it may not effectively differentiate between patients with varying risks of 30-day mortality after NSCLC surgery. This limitation raises concerns about its reliability for risk assessment in this specific patient population.

Conversely, the limitations of *Epithor* may stem from the presence of redundant variables within the scoring system. Redundant variables, or factors that overlap in their predictive value, can lead to inaccuracies and complexities in risk prediction models. The redundancy within *Epithor* might have contributed to its suboptimal performance when compared to the more efficient *Eurolung 2* and *Eurolung 2b*.

It is worth emphasizing that while scoring systems provide valuable insights into risk assessment, their application should always be exercised thoughtfully and with caution. Medical decision-making should not be solely reliant on prediction models, as individual patient characteristics, clinical judgment, and other factors play crucial roles in determining treatment strategies and surgical operability.

Nevertheless, scoring systems, in general, hold great value in the realm of preoperative evaluation. They serve as valuable tools to screen patients for further functional testing, aiding clinicians in identifying those who may benefit from additional assessments or interventions before surgery. Furthermore, scoring systems contribute to the process of patient selection and prognosis assessment, assisting healthcare professionals in making informed decisions regarding the most suitable treatment approaches.

In conclusion, this analysis underscores the importance of selecting appropriate scoring systems in the risk-stratification of patients undergoing NSCLC surgery. While *Eurolung 2* and *Eurolung 2b* have demonstrated their effectiveness in this regard, careful consideration of their limitations and integration with clinical judgment is essential for informed decision-making and optimal patient care.

6 Bibliography

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a cancer journal for clinicians 2021; 71:209-49.

[2] Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology 2017;28:iv1-iv21.

[3] de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. The New England journal of medicine 2020;382:503-13.

[4] Kato H, Oizumi H, Suzuki J, Hamada A, Watarai H, Nakahashi K et al. Thoracoscopic wedge resection and segmentectomy for small-sized pulmonary nodules. Journal of visualized surgery 2017;3:66.

[5] Nagoya A, Kanzaki R, Kanou T, Ose N, Funaki S, Minami M et al. Validation of Eurolung risk models in a Japanese population: a retrospective single-centre analysis of 612 cases. Interactive cardiovascular and thoracic surgery 2019;29:722-28.

[6] Shao W, Zhang Z, Zhang J, Feng H, Liang C, Liu D. Charlson comorbidity index as a predictor of short-term outcomes after pulmonary resection. Journal of thoracic disease 2020;12:6670-79.

[7] Licker MJ, Widikker I, Robert J, Frey JG, Spiliopoulos A, Ellenberger C et al. Operative mortality and respiratory complications after lung resection for cancer: impact of chronic obstructive pulmonary disease and time trends. The Annals of thoracic surgery 2006;81:1830-7.

[8] Wang P, Wang S, Liu Z, Sui X, Wang X, Li X et al. Segmentectomy and Wedge Resection for Elderly Patients with Stage I Non-Small Cell Lung Cancer: A Systematic Review and Meta-Analysis. Journal of clinical medicine 2022;11.

[9] Lin YJ, Chiang XH, Lu TP, Hsieh MS, Lin MW, Hsu HH et al. Thoracoscopic Lobectomy Versus Sublobar Resection for pStage I Geriatric Non-Small Cell Lung Cancer. Frontiers in oncology 2021;11:777590.

[10] Brunelli A, Refai M, Salati M, Xiumé F, Sabbatini A. Predicted versus observed FEV_1 and DLCO after major lung resection: a prospective evaluation at different postoperative periods. The Annals of thoracic surgery 2007;83:1134-9.

[11] Schlachtenberger G, Doerr F, Menghesha H, Hagmeyer L, Leschczyk T, Gaisendrees C et al. A Modified Calculation Improves the Accuracy of Predicted Postoperative Lung Function Values in Lung Cancer Patients. Lung 2021;199:395-402.

[12] Falcoz PE, Conti M, Brouchet L, Chocron S, Puyraveau M, Mercier M et al. The Thoracic Surgery Scoring System (*Thoracoscore*): risk model for in-hospital death in 15,183 patients requiring thoracic surgery. The Journal of thoracic and cardiovascular surgery 2007;133:325-32.

[13] Bernard A, Rivera C, Pages PB, Falcoz PE, Vicaut E, Dahan M. Risk model of in-hospital mortality after pulmonary resection for cancer: a national database of the French Society of Thoracic and Cardiovascular Surgery (*Epithor*). The Journal of thoracic and cardiovascular surgery 2011;141:449-58.

[14] Brunelli A, Salati M, Rocco G, Varela G, Van Raemdonck D, Decaluwe H et al. European risk models for morbidity (EuroLung1) and mortality (EuroLung2) to predict outcome following anatomic lung resections: an analysis from the European Society of Thoracic Surgeons database. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2017;51:490-97.

[15] Brunelli A, Cicconi S, Decaluwe H, Szanto Z, Falcoz PE. 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. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2020;57:455-61.

[16] Doerr F, Badreldin AM, Heldwein MB, Bossert T, Richter M, Lehmann T et al. A comparative study of four intensive care outcome prediction models in cardiac surgery patients. Journal of cardiothoracic surgery 2011;6:21.

[17] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45.

[18] Pompili C, Shargall Y, Decaluwe H, Moons J, Chari M, Brunelli A. Risk-adjusted performance evaluation in three academic thoracic surgery units using the Eurolung risk models. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2018;54:122-26.

[19] Bobbio M, Pollock BH, Cohen I, Diamond GA. Comparative accuracy of clinical tests for diagnosis and prognosis of coronary artery disease. The American journal of cardiology 1988;62:896-900.

[20] Qadri SS, Jarvis M, Ariyaratnam P, Chaudhry MA, Cale AR, Griffin S et al. Could *Thoracoscore* predict postoperative mortality in patients undergoing pneumonectomy? European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2014;45:864-9.

[21] Die Loucou J, Pagès PB, Falcoz PE, Thomas PA, Rivera C, Brouchet L et al. Validation and update of the thoracic surgery scoring system (*Thoracoscore*) risk model. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2020;58:350-56.

[22] Taylor M, Szafron B, Martin GP, Abah U, Smith M, Shackcloth M et al. External validation of six existing multivariable clinical prediction models for short-term mortality in patients undergoing lung resection. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2021;59:1030-36.

[23] Kim AW, Faber LP, Warren WH, Basu S, Wightman SC, Weber JA et al. Pneumonectomy after chemoradiation therapy for non-small cell lung cancer: does "side" really matter? The Annals of thoracic surgery 2009;88:937-43; discussion 44.

[24] Yang CJ, Shah SA, Lin BK, VanDusen KW, Chan DY, Tan WD et al. Right-Sided Versus Left-Sided Pneumonectomy After Induction Therapy for Non-Small Cell Lung Cancer. The Annals of thoracic surgery 2019;107:1074-81.

[25] Brunelli A, Salati M, Pompili C, Gentili P, Sabbatini A. Intraoperative air leak measured after lobectomy is associated with postoperative duration of air leak. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2017;52:963-68.

[26] Xie N, Meng X, Wu C, Lian Y, Wang C, Yu M et al. Both left upper lobectomy and left pneumonectomy are risk factors for postoperative stroke. Scientific reports 2019;9:10432.

[27] Cui Y, Grogan EL, Deppen SA, Wang F, Massion PP, Bailey CE et al. Mortality for Robotic- vs Video-Assisted Lobectomy-Treated Stage I Non-Small Cell Lung Cancer Patients. JNCI cancer spectrum 2020;4:pkaa028.

[28] Shen Y, Wang H, Feng M, Xi Y, Tan L, Wang Q. Single- versus multiple-port thoracoscopic lobectomy for lung cancer: a propensity-matched study†. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2016;49 Suppl 1:i48-53.

7 Appendix

7.1 Table of figures

Figure 1.

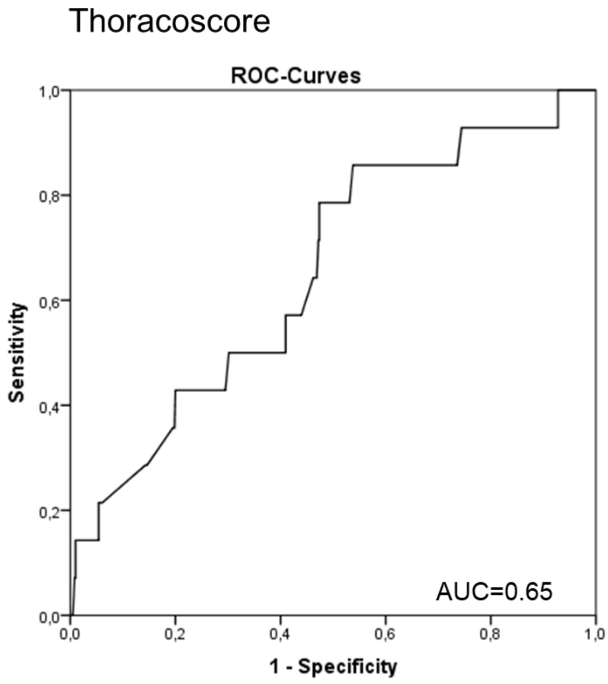


Figure 2.

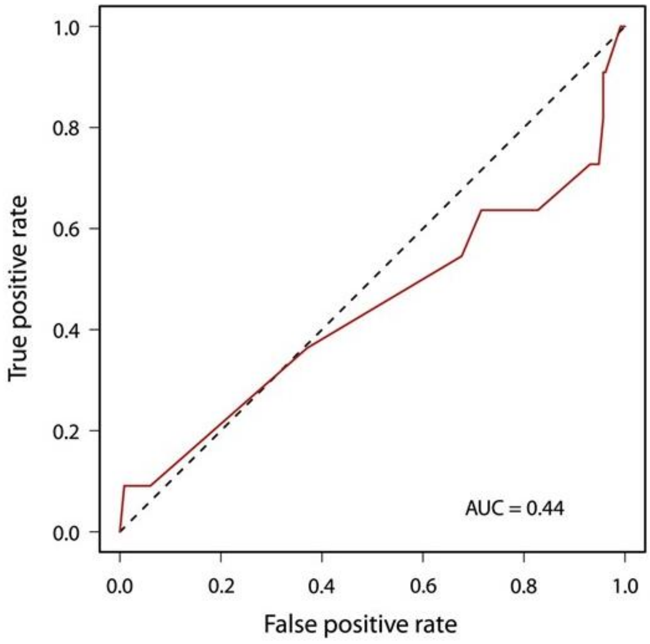


Figure 3.

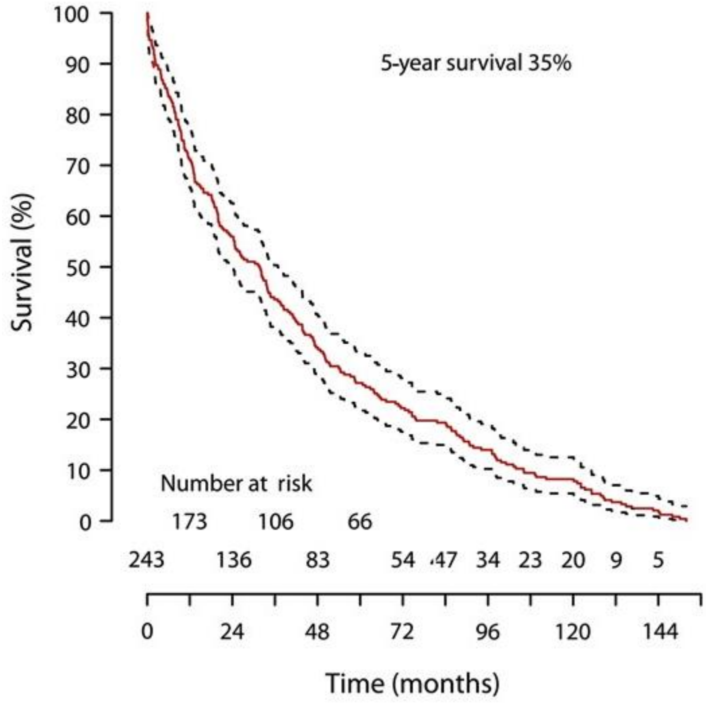


Figure 4.

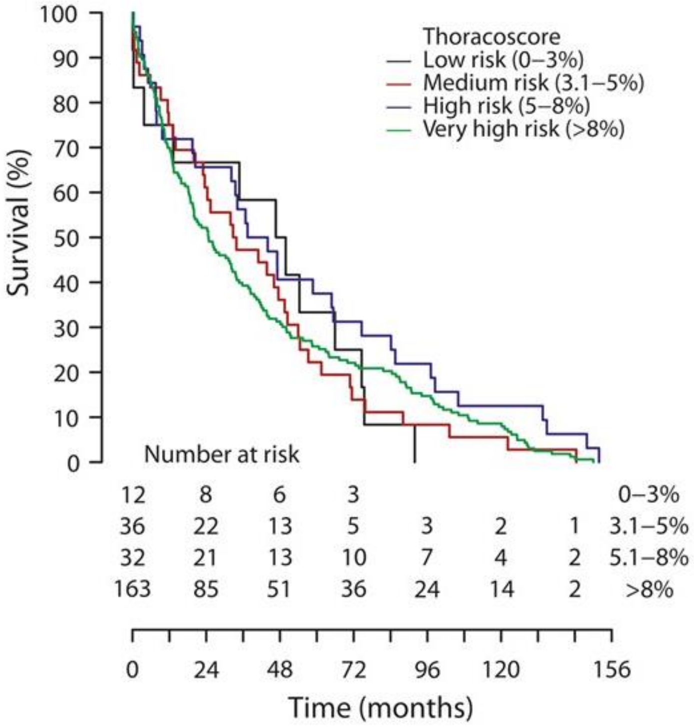


Figure 5.

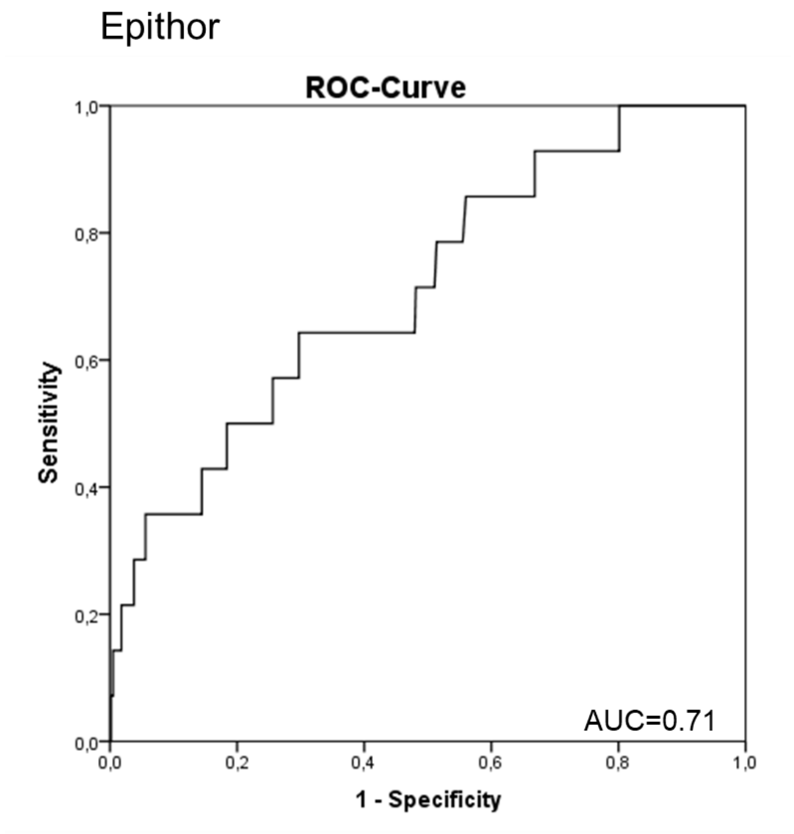


Figure 6.

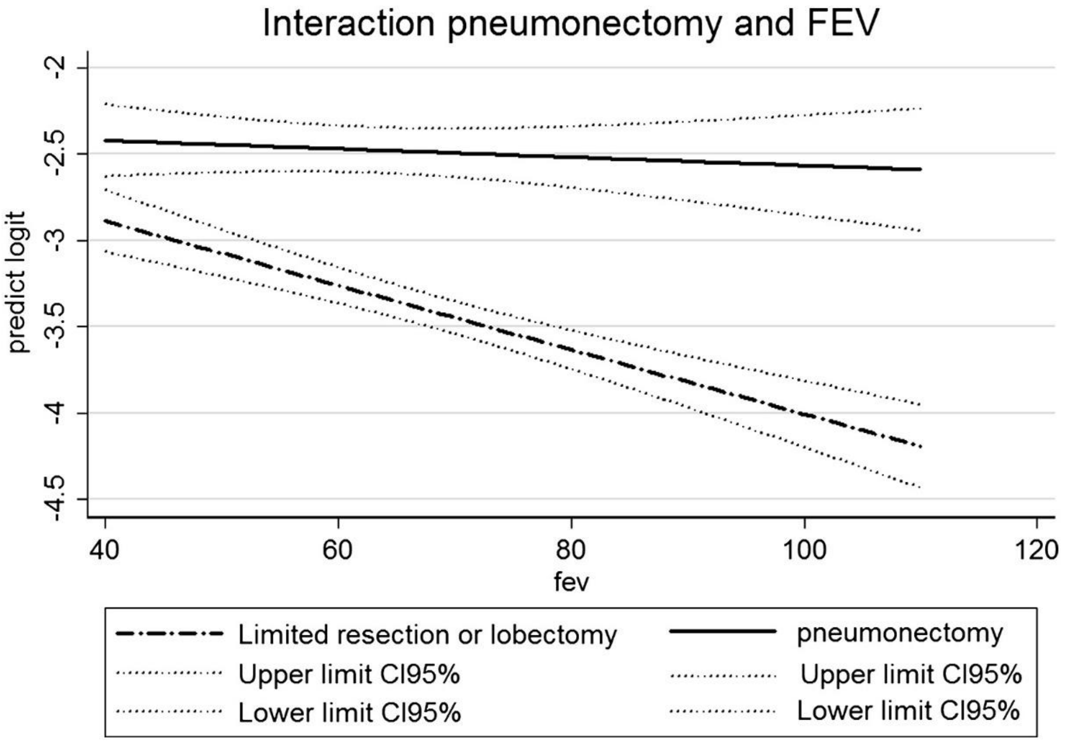


Figure 7.

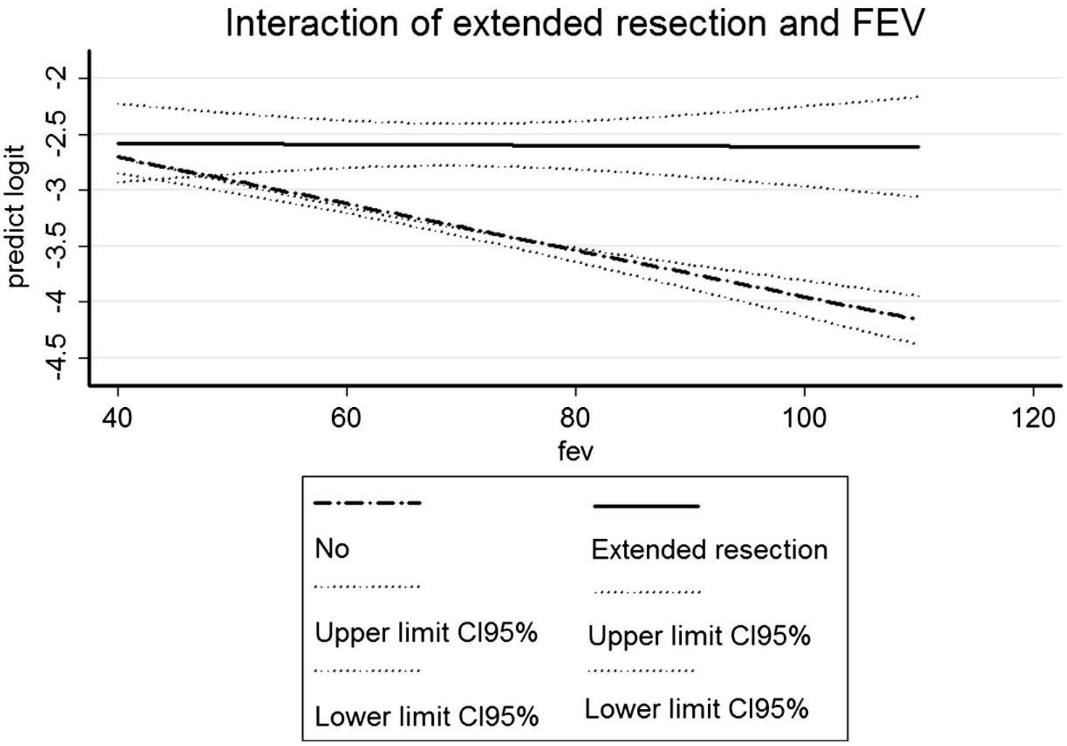


Figure 8.

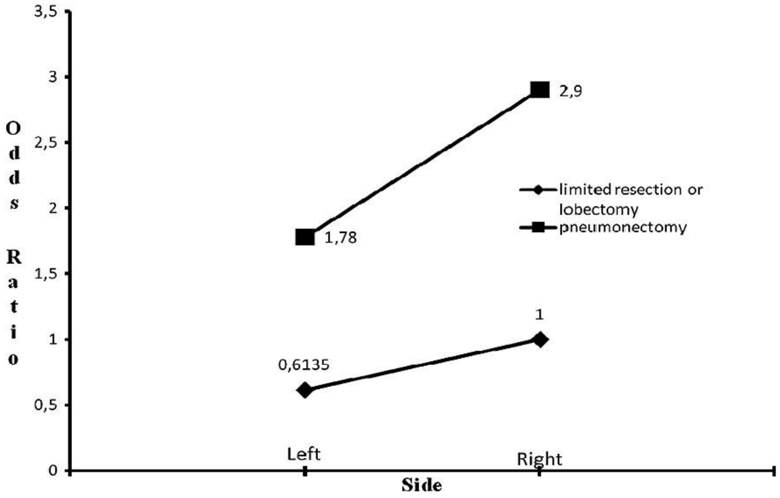


Figure 9.

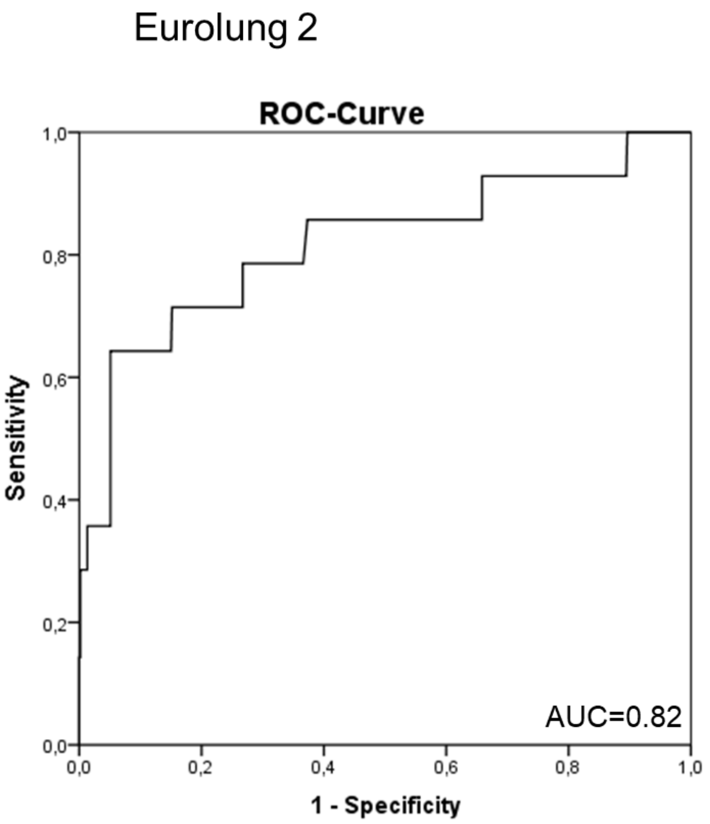


Figure 10.

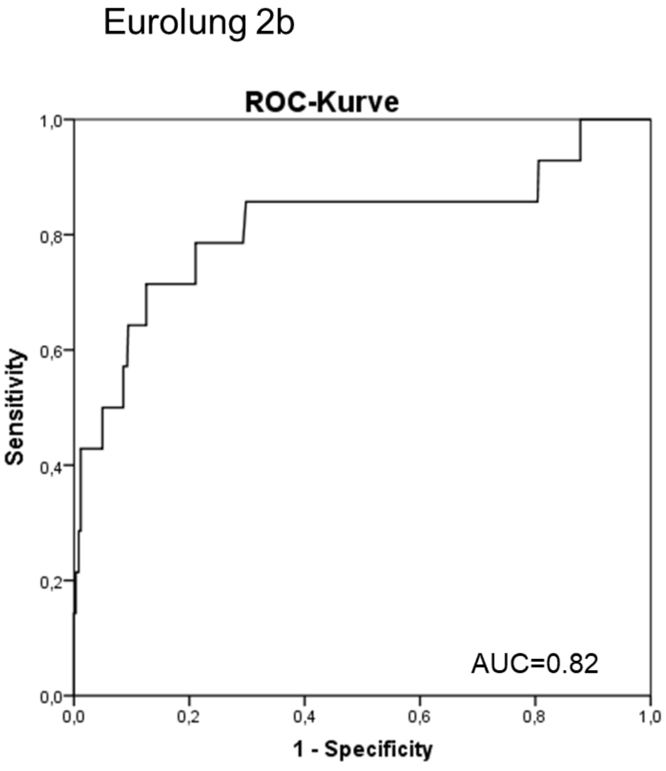


Figure 11 (A) und (B)

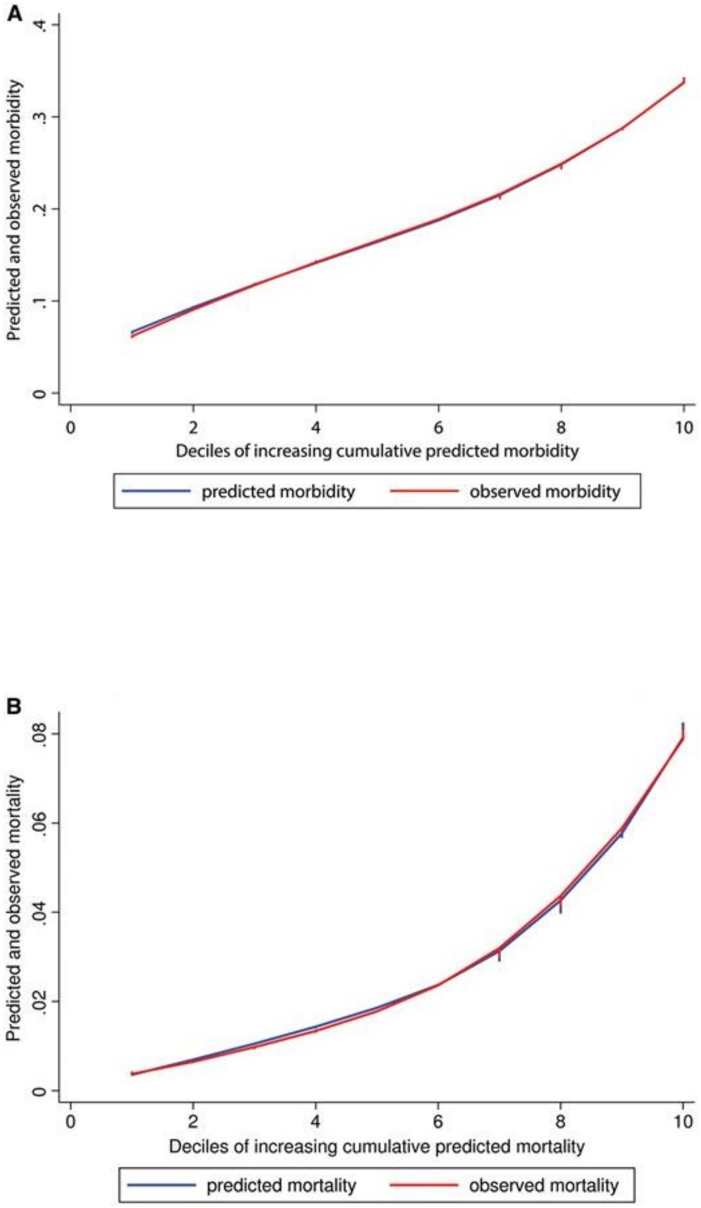


Figure 12.

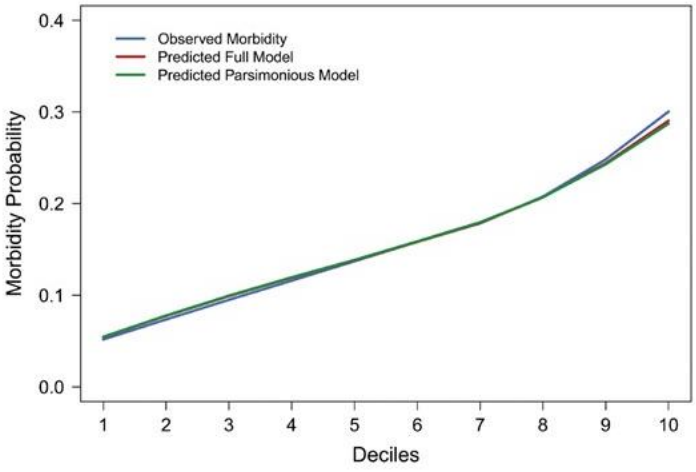
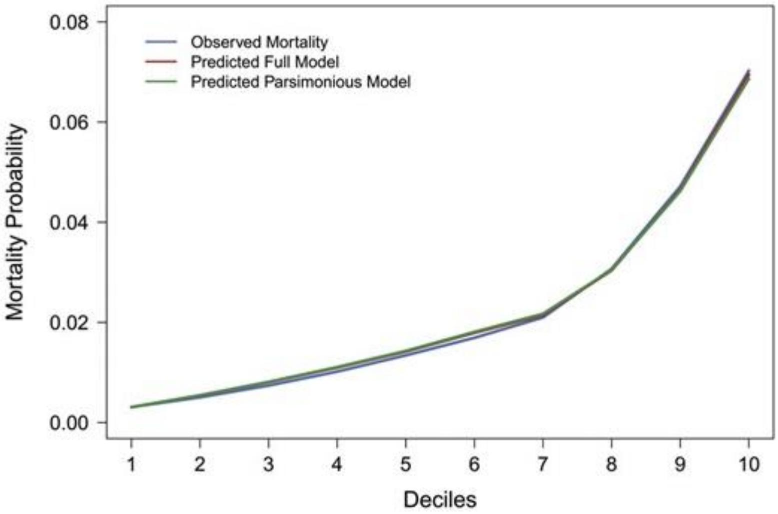


Figure 13.



7.2 Table index

Table 1: Summary of variables included in the different scoring systems

	Thoracoscore	*Epithor*	*Eurolung 2*	*Eurolung 2b*
Preoperative parameters:				
Age	X	X	X	X
Gender	X	X	X	X
BMI		X	X	X
ASA Score	X	X		
Performance status classification	X	X		
Dyspnea score	X			
FEV_1%		X		
ppoFEV_1			X	X
Comorbidities	X	X		
CHD			X	
CVD			X	
Benign/malign	X			
Tumor stage		X		
Surgical approach:				
Priority of surgery (elective/urgent)	X			
Operation side		X		
Pneumonectomy (other)	X	X	X	X
Extended resection (sleeve)		X	X	
Thoracotomy/VATS			X	X

Abbreviations: ASA classification: American Society of Anesthesiology classification; BMI: Body mass index; CHD: coronary heart disease; CVD: cardiovascular disease; FEV_1: forced expiratory volume in the first second; ppoFEV_1: predicted postoperative FEV_1; VATS: Video-assisted thoracic surgery

Table 2: Mortality scores in our cohort

	Total cohort n=624	30-days mortality (n=14)	30-days survival (n=610)	univariate p-value	multivariate p-value
Preoperative parameters:					
Age	64.0 ± 11.0	67.7 ± 15.9	63.9 ± 10.8	0.08	
Male gender (%)	59.5	85.7	58.9	**0.01**	**0.03**
BMI	25.9 ± 4.9	26.1 ± 2.7	25.8 ± 4.9	0.72	
ASA score >3 (%)	8.0	9.2	7.9	0.17	
Performance status classification >3 (%)	9.1	14.3	7.1	0.14	
Dyspnea score >2 (%)	10.1	13.1	9.3	0.12	
FEV_1 %	79.3 ± 18.1	77.3 ± 18.2	83.3 ± 22.6	**0.01**	0.3
$ppoFEV_1$ l	1.79 ± 0.6	1.52 ± 0.52	1.80 ± 0.6	**0.05**	0.4
Comorbidity score >3 (%)	35	37	35	0.17	
CHD (%)	21.8	21.6	21.9	0.63	
CVD (%)	3.0	3.0	7.1	0.36	
Benign/malign (%)	0/100	0/100	0/100	/	
Tumor stage >IIB (%)	25.9	26.1	25.8	0.81	
Surgical approach:					
Priority of surgery (elective/urgent) (%)	100/0	100/0	100/0	/	
Operation side (right/left) (%)	59.7/40.3	57.1/42.9	59.7/40.3	0.76	
Pneumonectomy (other) (%)	6.6	21.4	6.2	**<0.001**	**<0.001**
Extended resection (%)	14.7	21.8	14.6	0.46	
Thoracotomy/VATS (%)	28/72	36/64	28/72	0.23	

Abbreviations: ASA classification: American Society of Anesthesiology classification; BMI: body mass index; CHD: coronary heart disease; CVD: cardiovascular disease; FEV_1: forced expiratory volume in the first second; $ppoFEV_1$: predicted postoperative FEV_1; VATS: video-assisted thoracic surgery

Table 3: Summary of variables included in the different scoring systems

Scoring model	Logistic regression		OCC	HL-test		ROC-Analysis	
	OR	CI 95%	%	x^1	p-value	AUC	CI 95%
Thoracoscore	1.1	1.0-1.2	97.8	14.3	0.67	0.65	0.51-0.79
Epithor	1.06	1.03-1.09	97.8	4.7	0.78	0.71	0.57-0.81
Eurolung 2	1.2	1.1-1.3	97.8	4.4	0.82	0.82	0.67-0.96
Eurolung 2b	1.4	1.2-1.5	97.8	12.5	0.13	0.82	0.6-0.85

Abbreviations: AUC: area under the curve; CI: confidence interval; HL: Hosmer-Lemeshow test; OR: Odds ratio, OCC: overall correct classification; ROC: receiver-operating-characteristic