EUROPEAN JOURNAL OF
CARDIO-THORACIC
SURGERY

Inability to perform maximal stair climbing test before lung resection: a propensity score analysis on early outcome

Alessandro Brunelli, Armando Sabbatini, Francesco Xiume', Alessandro Borri, Michele Salati, Rita Daniela Marasco and Aroldo Fianchini

Eur J Cardiothorac Surg 2005;27:367-372
DOI: 10.1016/j.ejcts.2004.11.012

European Journal of Cardio-thoracic Surgery 27 (2005) 367–372

EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY

www.elsevier.com/locate/ejcts

Inability to perform maximal stair climbing test before lung resection: a propensity score analysis on early outcome

Alessandro Brunelli*, Armando Sabbatini, Francesco Xiume', Alessandro Borri, Michele Salati, Rita Daniela Marasco, Aroldo Fianchini

Unit of Thoracic Surgery, Department of Respiratory Diseases, 'Umberto I°' Regional Hospital, Ancona, Italy

Received 30 July 2004; received in revised form 9 November 2004; accepted 17 November 2004; Available online 13 January 2005

Abstract

Objective: The objective of the present study was to assess whether patients unable to perform a preoperative maximal stair climbing test had an increased incidence of morbidity and mortality after major lung resection compared to patients who were able to exercise. **Methods**: Three hundred and ninety one patients submitted to pulmonary lobectomy or pneumonectomy for lung cancer were analyzed. Forty-five of these patients were unable to perform a preoperative maximal stair climbing test for underlying comorbidities. Unadjusted and propensity score case matched comparisons were performed between patients who could and who could not perform a preoperative stair climbing test. Multivariable analyses were then performed to identify predictors of morbidity and mortality, and were validated by bootstrap bagging. **Results**: Patients who could not perform the stair climbing test had similar morbidity rates (31.1 vs. 35.6%, respectively, $P=0.7$), but higher mortality rates (15.6 vs. 4.4%, respectively, $P=0.08$) and deaths among complicated patients (50 vs. 12.5%, respectively, $P=0.025$), compared to propensity score matched patients who could perform the stair climbing test. Logistic regression analyses showed that the inability to perform the stair climbing test was an independent and reliable predictor of mortality ($P=0.005$) but not of morbidity ($P=0.2$). **Conclusions**: Patients unable to perform a preoperative maximal exercise test had an increased risk of mortality after major lung resection. Half of these patients did not survive postoperative complications, due to their decreased aerobic reserve caused by physical inactivity which made them unable to cope with the increased oxygen demand.

Keywords: Complications of surgery; Mortality; Lung cancer surgery; Risk analysis; Exercise test

1. Introduction

Cardiopulmonary exercise tests are increasingly used in the preoperative evaluation of the lung resection candidates. In our unit, symptom-limited stair climbing test has become part of the preoperative routine risk stratification process [1]. However, some patients are unable to perform a proper maximal exercise test for a number of clinical conditions. The lack of reliable data from the preoperative exercise testing make their risk assessment less accurate. Moreover, the independent role of the inability to perform the exercise in increasing the risk of postoperative complications is poorly defined.

Therefore, the objective of the present analysis was to quantify the risk of morbidity and mortality after major lung resection in patients unable to perform a maximal stair climbing test and compare this risk with that of patients who performed the preoperative exercise test, after adjusting this comparison by using a propensity score matching.

* Corresponding author. Tel.: +39 071 596439; fax: +39 071 5964433. *E-mail address:* alexit_2000@yahoo.com (A. Brunelli).

1010-7940/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejcts.2004.11.012

2. Patients and methods

Three hundred and ninety one patients underwent pulmonary lobectomy (323) or pneumonectomy (68) for non-small cell lung cancer (NSCLC) from January 2000 through March 2004 and were prospectively enrolled in this analysis after giving informed consent.

Resectability was assessed by means of computed tomographic scan, bronchoscopy and, when indicated, cervical mediastinoscopy.

Operability was evaluated by means of maximal stair climbing test, pulmonary function tests, blood gas analysis, electrocardiogram, echocardiography and more invasive cardiologic procedures if needed.

Criteria for inoperability were a predicted postoperative forced expiratory volume (ppoFEV1) less than 30% of predicted in association with a total height climbed at the preoperative stair climbing test lower than 12 m, and/or cardiac instability. Using these criteria only 2 patients were excluded from operation during the period of the study.

Symptom-limited stair climbing test was administered preoperatively (usually the day before the operation, when the patients are admitted to our unit) to all patients, who were able to perform it, according to the criteria described elsewhere [1]. Briefly, the patients were asked to climb the maximum number of stairs and stop only for exhaustion, limiting dyspnea, chest pain or leg discomfort. Contraindications to perform the stair climbing test were the following: unstable cardiac failure, a recent (less than 30 days) myocardial infarction, unstable coronary artery disease, severe arrhythmia, a recent pulmonary embolism, symptomatic aortic stenosis, severe symptomatic peripheral vascular disease, presence of aortic aneurysm, recent cerebrovascular disease, symptomatic muscoloskeletal, neurological and psychiatric disease. We did not consider chronic obstructive respiratory disease as a contraindication to exercise testing [2].

In this series, 346 patients were able to perform a preoperative stair climbing test (SC), whereas 45 patients were unable to do it (NoSC). Reasons for not taking the test were the following: severe symptomatic musculoskeletal disease (1 muscular dystrophy, 15 osteoarthrosis, 2 recent hip fractures, 2 recent femoral fractures); symptomatic neurological disease (3 hemiplegia, 3 recent strokes, 2 poliomyelitis, 1 intracranial aneurysm, 1 spastic tetra paresis, 1 ataxia); severe peripheral vascular disease (5 cases); cardiac ischemic disease (2 cases); blindness (2 cases); psychiatric disease (1 case); morbid obesity (1 case); cachexia (3 cases). Therefore, 9 of these 45 patients had clinical contraindications to perform the test (recent stroke, intracranial aneurysm, cardiac ischemic disease, blindness and psychiatric disease), whereas the other 36 patients were impeded to perform a proper exercise test by severe comorbidities.

An additional 13 patients did not perform the stair climbing test for reasons other than medical ones (refusal to exercise, logistic problems, etc.) and they were excluded from the present study.

Patients who did not perform the stair climbing test were deemed operable by traditional criteria of operability (ppoFEV1 $>$ 30% and hemodynamically stable state).

2.1. Statistical analysis

2.1.1. Unadjusted comparison

The comparison of patients who did and did not perform preoperative stair climbing test in terms of preoperative and postoperative characteristics (Appendix 1) was made by means of the Chi-square test for categorical variables, and by means of the unpaired Student's *t*-test or the Mann–Whitney *U*-test for continuous ones. All statistical tests were 2-tailed, and a significance level of $P<0.05$ was selected.

2.1.2. Propensity score case matched comparison

Selection bias was addressed by constructing propensity scores [3,4]. The aim of the analysis was to match patients who could and those who could not perform the preoperative stair climbing test according to baseline characteristics and compare surgical outcomes between the matched groups. Before matching patients, a parsimonious explanatory model was developed by bootstrap bagging for variables selection [5]. The probability of performing preoperative stair climbing test (propensity score) was estimated by logistic regression analysis incorporating the variables identified in the parsimonious model (whose stability was assessed by bootstrap analysis) plus additional baseline variables. Therefore, the variables used in the model were the following: age, gender, type of operation (lobectomy vs. pneumonectomy), forced expiratory volume in 1 s (FEV1), forced vital capacity (FVC), FEV1/FVC ratio, carbon monoxide diffusion lung capacity (DLCO), arterial carbon dioxide and oxygen tensions, preoperative hemoglobin level, presence of a concomitant cardiac disease, Charlson Comorbidity Index (CCI) [6], POSSUM Physiological Score [7,8], smoking history (pack-years), neoadjuvant chemotherapy, percentage of functioning lung parenchyma removed during operation (Func loss%) (see Appendix 1 for explanation of variables). The method of generating a parsimonious model and then augmenting it with other factors to develop the propensity model was described elsewhere [4]. As stated by its developers [9], the propensity model is not parsimonious. In fact, the goal is to balance patient characteristics by incorporating 'everything' recorded that may relate to either systematic bias or simply bad luck that has otherwise unbalanced the comparison groups of interest, ignoring usual concerns about model overdetermination [4,10]. The c-index, or area under the receiver operator curve (ROC), for this model was 0.75. All variables were at least 95% complete and sporadic missing values were imputed by taking the most frequent response category or averaging non-missing values for continuous variables. Greedy matching techniques were then used to select stair-climbing counterparts to the non-stair-climbing patients by choosing the patient with the nearest propensity score [4].

The procedure yielded 45 well-matched pairs. The two groups of propensity score matched patients were compared by using Chi-square test for categorical variables and the unpaired Student's *t*-test or the Mann–Whitney *U* test for continuous ones.

2.1.3. Risk factors analysis of outcome

Multi-variable analyses were used to identify predictors of postoperative morbidity and mortality on the entire database of 391 patients. Due to the small number of events, only the propensity score and the ability to perform stair climbing test were used as independent variables in the mortality model [10].

In order to identify predictors of morbidity, univariate analysis was initially performed to screen significant variables to be used as independent variables in a stepwise logistic regression analysis, in which the propensity score and the ability to perform stair climbing test were forced in the model irrespective of their *P*-value. Variables initially screened in univariate analysis for a possible association with postoperative morbidity are those listed in Appendix 1. To avoid multi-colinearity only one variable in a set of variables with a correlation coefficient greater than 0.5 was included in the model.

Bootstrap resampling procedures with 1000 samples were used to confirm reliability of the variables included in the final models. In the bootstrap procedure, repeated samples of 391 observations (the same number of observations as the original database) were selected with replacement from

Table 1
Unadjusted comparison of baseline numerical variables between patients who could and those who could not perform preoperative stair climbing test

Variables	No stair (45 patients)	Stair (346 patients)	*P*-value
Age	69.1 (8.3)	67.0 (9.0)	0.15§
FEV1%	83.0 (23.2)	85.7 (18.3)	0.4*
FVC%	87.4 (20.3)	94.7 (17.3)	0.009*
FEV1/FVC ratio%	90.9 (12.7)	88.3 (12.9)	0.3§
DLCO%	76.3 (21.5)	88.5 (13.1)	0.2§
ppoFEV1%	65.5 (17.7)	67.0 (16.2)	0.7§
ppoDLCO%	60.0 (18.7)	59.7 (16.9)	0.6§
PaO2 (mmHg)	78.9 (9.6)	80.2 (10.7)	0.4§
PaCO2 (mmHg)	38.2 (4.1)	37.3 (3.9)	0.04§
Preop Hb (g/dl)	13.3 (2.0)	13.7 (1.7)	0.14*
Pack-years	43.4 (24.0)	39.9 (25.8)	0.4§
Func loss%	17.4 (9.1)	18.7 (9.1)	0.4§
CCI	2.7 (1.7)	2.0 (1.7)	0.003§
POSSUM PS	19.9 (5.1)	17.4 (3.3)	0.001§
Operation length (min)	178.9 (52.6)	171.8 (49.2)	0.5§

Results are expressed as means ± standard deviations. (*) Unpaired Student's *t*-test; (§) Mann-Whitney *U*-test. Func loss%: percentage of functional parenchyma removed during operation; CCI, Charlson comorbidity index; POSSUM PS, POSSUM Physiological Score.

the original set observations. For each sample, stepwise logistic regression was performed entering the preoperative and intraoperative variables with $P<0.05$ at univariate analysis. The stability of the final model can be assessed by identifying the variables that enter most frequently in the repeated bootstrap models and comparing those variables with the variables in the final regression model. If the final model variables occur in a majority (>50%) of the bootstrap models, the original final regression model can be judged to be stable.

The analyses were performed by using the Statview 5.0 (SAS Institute, Cary, NC) and the Stata 8.2 (Stata Corp., College Station, TX) statistical softwares.

3. Results

The group of patients that could not perform preoperative stair climbing test (11.5% of the present series) had a lower FVC ($P=0.009$), and a higher PaCO2 ($P=0.04$), CCI ($P=0.003$) and POSSUM PS ($P=0.001$), compared to the SC patients (Tables 1 and 2). The unadjusted comparison between the 2 groups showed that the NoSC patients had a longer postoperative hospital stay ($P=0.002$), a higher incidence of cardiopulmonary morbidity ($P=0.09$), a higher mortality rate ($P<0.0001$) and a higher proportion of deaths among the complicated patients ($P=0.03$), compared to the SC patients (Table 5).

To minimize the selection bias, we constructed a propensity score case matched analysis yielding 45 well-matched pairs. The two groups were similar in terms of baseline characteristics with the exception of a lower CCI in the NoSC patients (Tables 3 and 4). The propensity score-adjusted analysis of early surgical outcome showed that the NoSC patients had a higher mortality rate ($P=0.08$) and a 4-fold higher incidence of the deaths among complicated patients ($P=0.025$), with respect to the SC patients (Table 5).

Causes of death among the NoSC patients were the following: 2 strokes, 2 cardiac failures, 3 respiratory failures. Causes of death among the SC patients were the following: 5 respiratory failures, 1 bronchopleural fistula that underwent reoperation and subsequently complicated with ARDS and respiratory failure, 1 cardiac failure and 1 cardiac arrest. The 2 propensity score matched SC patients who died had a cardiac arrest and a respiratory failure, respectively.

Stepwise logistic regression analysis (Table 6) showed that significant predictors of cardiopulmonary

Table 2
Unadjusted comparison of baseline categorical variables between patients who could and those who could not perform preoperative stair climbing test

Variables	No stair (45 patients)	Stair (346 patients)	*P*-value*
Gender (males, *n*, %)	36 (80)	273 (79)	0.9
Diabetes (*n*, %)	7 (16)	40 (12)	0.4
Cardiac disease (*n*, %)	23 (51)	171 (49)	0.8
Neoadjuvant chemotherapy (*n*, %)	6 (13)	56 (16)	0.6
Side of operation (right, *n*, %)	25 (56)	163 (47)	0.3
Pneumonectomies (*n*, %)	5 (11)	63 (18)	0.2

(*) Chi-square test.

Table 3
Comparison of baseline numerical variables between propensity score matched patients who could and those who could not perform preoperative stair climbing test

Variables	No stair (45 patients)	Stair (45 patients)	*P*-value
Age	69.1 (8.3)	71.3 (7.4)	0.2§
FEV1%	83.0 (23.2)	82.1 (19.7)	0.9*
FVC%	87.4 (20.3)	83.5 (16.5)	0.3*
FEV1/FVC ratio%	90.9 (12.7)	92.1 (11.9)	0.7§
DLCO%	76.3 (21.5)	78.2 (19.8)	0.7§
ppoFEV1%	65.5 (17.7)	66.1 (16.2)	0.9§
ppoDLCO%	60.0 (18.7)	56.3 (27.3)	0.3§
PaO2 (mmHg)	78.9 (9.6)	78.3 (9.1)	0.8§
PaCO2 (mmHg)	38.2 (4.1)	39.0 (3.5)	0.3§
Preop Hb (g/dl)	13.3 (2.0)	13.0 (2.1)	0.4*
Pack-years	43.4 (24.0)	44.9 (31.0)	0.8§
Func loss%	17.4 (9.1)	16.0 (8.2)	0.7§
CCI	2.7 (1.7)	3.7 (2.2)	0.02§
POSSUM PS	19.9 (5.1)	21.2 (4.0)	0.12§
Operation length (min)	178.9 (52.6)	160.8 (42.0)	0.07§

Results are expressed as means ± standard deviations. (*) Unpaired Student's *t*-test; (§) Mann-Whitney *U*-test. Func loss%, percentage of functional parenchyma removed during operation; CCI, Charlson comorbidity index; POSSUM PS, POSSUM Physiological Score.

Table 4
Comparison of baseline categorical variables between propensity score matched patients who could and those who could not perform preoperative stair climbing test

Variables	No stair (45 patients)	Stair (45 patients)	*P*-value*
Gender (males, *n*, %)	36 (80)	32 (71)	0.3
Diabetes (*n*, %)	7 (16)	10 (22)	0.4
Cardiac disease (n, %)	23 (51)	28 (62)	0.3
Neoadjuvant chemotherapy (*n*, %)	6 (13)	8 (18)	0.6
Side of operation (right, *n*, %)	25 (56)	22 (49)	0.5
Pneumonectomies (*n*, %)	5 (11)	2 (4.4)	0.2

(*) Chi-square test.

Table 5
Comparison of early outcome in unadjusted and propensity-matched groups

Outcome	Unadjusted			Propensity-matched		
	No stair (45 cases)	Stair (346 cases)	*P*-value	No stair (45 cases)	Stair (45 cases)	*P*-value
Postop hospital stay (days)	12.5 (10.4)	9.5 (5.4)	0.002*	12.5 (10.4)	11.3 (6.5)	0.7*
Morbidity (*n*, %)	14 (31)	69 (20)	0.09§	14 (31)	16 (36)	0.7§
Mortality (*n*, %)	7 (16)	8 (2.3)	<0.0001§	7 (16)	2 (4.4)	0.08§
Failure to rescue (%)	7 (50)	8 (12)	0.003§	7 (50)	2 (12.5)	0.025§

Results are expressed as means ± standard deviations unless otherwise specified. (*) Unpaired Student's *t* test; (§) Chi-square test. Failure-to-rescue, proportion of deaths among complicated patients.

complications were a low DLCO ($P=0.01$), a high POSSUM PS ($P=0.04$) and the presence of a concomitant cardiac disease ($P=0.001$). The inability to perform stair climbing test was not associated with an increased risk of postoperative morbidity after its effect was corrected for the other 7 variables in the model. These results were validated by bootstrap bagging.

Table 7 shows that the inability to perform stair climbing test was an independent predictor of mortality after controlling the effect of the propensity score in a logistic regression model. This finding was validated by bootstrap bagging.

4. Discussion

In the present study we wanted to quantify the risk of morbidity and mortality after major lung resection in patients unable to perform a preoperative maximal exercise test such as the stair climbing test and to compare it with that of those patients able to exercise.

Even though the patients that were not able to exercise had severe comorbidities that could predispose by themselves to postoperative morbidity and mortality, we hypothesized that the reduced physical activity or total inactivity caused by these comorbidities may have played a role in reducing the aerobic reserve, increasing the risk of postoperative mortality. In order to test this hypothesis we tried to reduce the selection bias by a propensity score case matched analysis [3,4], which is considered the most rigorous method available for the so-called 'apples-to-apples' investigation of causal effects on outcome in a non-randomized setting [11]. Since associations found through traditional multi-variable regression may be misleading because of under-representation of 1 covariate of interest within levels of another, propensity scores is recommended to adjust for confounding background characteristics [11].

The comparison between propensity score matched groups revealed similar morbidity rates between the NoSC and the SC patients. However, the mortality rate and the proportion of deaths among the complicated patients was higher in the NoSC group.

When the effect of the propensity score and of other variables was controlled in logistic regression analyses, the incapacity to perform the stair climbing test was a significant predictor of mortality but not of morbidity. We were therefore unable to find an effect of the inability to exercise in itself on the occurrence of complications, which were presumably more associated to the presence of comorbidities. On the other hand, the risk of mortality was associated not only with the presence of comorbidities but also with the reduction of the aerobic reserve caused by the reduced physical activity.

Furthermore, a significantly higher proportion of failure-to-rescue cases were found in the NoSC patients compared to the propensity score matched SC counterparts. Fifty percent of the NoSC complicated patients died as a result of their complications, a 4-fold higher proportion compared to the propensity score matched SC patients.

This latter finding indirectly demonstrates that the inability to perform the preoperative exercise test is an important marker of a reduced aerobic reserve, which in turn may be either the cause of the incapacity to perform a physical effort and also the consequence of a deconditioning effect resulting from a limited physical activity.

Under these circumstances, the onset of a complication may increase the oxygen demand beyond that which the patients can provide, inducing an oxygen debt. If this debt remains critically high over specified periods of time, then a multi-organ failure may ensue [12,13].

Our results are similar to those reported by Epstein et al. [14] in a smaller series of patients unable to perform a maximal exercise test on cycle ergometer. However, they

Table 6
Risk factors for cardiopulmonary complications (parsimonious model)

Variables	Estimate±SE	Odds ratio	95% CL	*P*-value	% Bagging
POSSUM PS	0.10 (0.05)	1.11	1.01-1.22	0.04	66
DLCO%	−0.02 (0.008)	0.98	0.96-0.99	0.01	67
Cardiac disease	1.0 (0.3)	2.72	1.5-4.8	0.001	69

Hosmer-Lemeshow goodness-of-fit, 10.9 ($P=0.2$); c-index, 0.73. Independent variables used: FEV1, preoperative Hb, CCI, POSSUM PS, DLCO, concomitant cardiac disease, inability to perform stair climbing test, propensity score. CCI, Charlson comorbidity index; POSSUM PS, POSSUM Physiological Score. Inability to perform stair climbing test and propensity score were not significant predictors of morbidity. % Bagging, frequency of significancy ($P<0.05$) of independent variables in 1000 bootstrap resampling models.

Table 7
Risk factors for mortality

Variables	Estimate±SE	Odds ratio	95% CL	*P*-value	% Bagging
Stair	−1.61 (0.58)	0.20	0.06-0.62	0.005	85
Propensity score	−0.05 (0.02)	0.95	0.91-0.99	0.01	58.1

Hosmer-Lemeshow goodness-of-fit, 9.4 ($P=0.3$); c-index, 0.72. Propensity score is expressed as percent probability of performing stair climbing test. % Bagging, frequency of significancy ($P<0.05$) of independent variables in 1000 bootstrap resampling models.

did not use any balancing score to match the patients that could and could not exercise, and this may explain why they found a higher morbidity rate in the no exercise group, such as in our unadjusted comparison. Similar to us, they found a higher proportion of deaths among the complicated no-exercising patients compared to the exercising ones (27.3 vs. 4.8%, respectively).

In another study, Girish et al. [15] found that 8 out of 9 patients unable to climb any flights of stairs experienced cardiopulmonary complications after thoracic and upper abdominal surgery.

This is not a randomized trial, and even though the propensity score analysis constitutes the most rigorous method for investigating the causal effects in this setting, it cannot account for unknown variables affecting outcome that are not correlated strongly with measured variables. Therefore the results generated by this analysis cannot be interpreted as definitive. Unfortunately a randomized clinical trail on the issue addressed in this paper appears impossible to realize.

In conclusion, we were able to show that the patients that were not able to perform a preoperative maximal exercise test had an increased risk of mortality after major lung resection. In this particular group of patients every effort should be done to prevent the occurrence of a complication that may be fatal in 50% of cases.

We think that the findings of the present study warrant the institution of a specific preoperative physical training (with exercise programs appropriate to the physical limitation that precluded the exercise test) in patients unable to perform an exercise test, in order to increase their aerobic reserve. However, the efficacy of this training program in these selected high risk patients should be verified by other prospective studies.

Furthermore, we currently use the results of this analysis for counseling the patients before the operation and for planning a postoperative intensive care management.

References

[1] Brunelli A, Al Refai M, Monteverde M, Borri A, Salati M, Fianchini A. Stair climbing test predicts cardiopulmonary complications after lung resection. Chest 2002;121:1106–10.

[2] Brunelli A, Fianchini A. Stair climbing test in lung resection candidates with low predicted postoperative FEV1. Chest 2003;124:1179.

[3] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.

[4] Blackstone EH. Comparing apples and oranges. J Thorac Cardiovasc Surg 2002;123:8–15.

[5] Breiman L. Bagging predictors. Mach Learning 1996;24:123–40.

[6] Birim O, Maat APWM, Kappetein AP, van Meerbeeck JP, Damhuis RAM, Bogers AJJC. Validation of the Charlson comorbidity index in patients with operated primary non-small cell lung cancer. Eur J Cardiothorac Surg 2003;23:30–4.

[7] Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. Br J Surg 1991;78:356–60.

[8] Brunelli A, Fianchini A, Gesuita R, Carle F. POSSUM scoring system as an instrument of audit in lung resection surgery. Physiological and operative severity score for the enumeration of mortality and morbidity. Ann Thorac Surg 1999;67:329–31.

[9] Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984;79: 516–24.

[10] Rosenbaum PR. Optimal matching for observational studies. J Am Stat Assoc 1989;84:1024–32.

[11] Koch CG, Khandwala F, Nussmeier N, Blackstone EH. Gender and outcomes after coronary artery bypass grafting: a propensity-matched comparison. J Thorac Cardiovasc Surg 2003;126:2032–43.

[12] Shoemaker WC, Appel PL, Kram HB. Role of oxygen debt in the development of organ failure sepsis and death in high-risk surgical patients. Chest 1992;102:208–15.

[13] Wasserman K. Preoperative evaluation of cardiovascular reserve in the elderly. Chest 1993;104:663–4.

[14] Epstein SK, Faling LJ, Daly BDT, Celli BR. Inability to perform bicycle ergometry predicts increased morbidty and mortality after lung resection. Chest 1995;107:311–6.

[15] Girish M, Trayner E, Damman O, Pinto-Plata V, Celli B. Symptom-limited stair climbing as a predictor of postoperative cardiopulmonary complications after high-risk surgery. Chest 2001;120:1147–51.

[16] Markos J, Mullan BP, Hillman DR, Musk AW, Antico VF, Lovegrove FT, Carter MJ, Finucane KE. Preoperative assessment as a predictor of mortality and morbidity after lung resection. Am Rev Respir Dis 1989; 139:902–10.

[17] Olsen GN, Bolton JWR, Weiman DS, Hornung CA. Stair climbing as an exercise test to predict the postoperative complications of lung resection: two years experience. Chest 1991;99:587–90.

[18] Smith TP, Kinasewitz GT, Tucker WY, Spillers WP, George RB. Exercise capacity as a predictor of post-thoracotomy morbidity. Am Rev Respir Dis 1984;129:730–4.

[19] Bechard D, Wetstein L. Assessment of exercise oxygen consumption as preoperative criterion for lung resection. Ann Thorac Surg 1987;44: 344–9.

[20] Bolliger CT, Jordan P, Soler M, Stulz P, Gradel E, Skarvan K, Elsasser S, Gonon M, Wyser C, Tamm M. Exercise capacity as a predictor of postoperative complications in lung resection candidates. Am J Respir Crit Care Med 1995;151:1472–80.

[21] Silber JH, Rosenbaum PR, Schwartz JS, Ross RN, Williams SV. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. J Am Med Assoc 1995;274:317–23.

Appendix A

A.1. Preoperative and operative variables

For the purpose of the present study the following spirometric variables were considered: forced expiratory volume in 1 s (FEV1); forced vital capacity (FVC); FEV1/FVC ratio; carbon monoxide diffusion lung capacity (DLCO); predicted postoperative FEV1 (ppoFEV1) calculated by the formula, (preoperative FEV1/number of preoperative functioning segments) × number of postoperative functioning segments; predicted postoperative DLCO (ppoDLCO) calculated by the formula (preoperative DLCO/number of preoperative functioning segments) × number of postoperative functioning segments.

Pulmonary function tests were performed according to the American Thoracic Society criteria. DLCO was measured by the single-breath method. Results of spirometry were collected after bronchodilator administration and were expressed as percentage of predicted for age, sex and height.

The number of functioning segments was estimated by means of CT scan and bronchoscopy findings. In patients with a calculated ppoFEV1 less than 50% of predicted and in all pneumonectomy candidates a quantitative perfusion lung scan was used, according to Markos et al. [16].

For the purpose of the present study, a concomitant cardiac disease (cardiac comorbidity) was defined as follows: previous cardiac surgery, previous myocardial infarction, history of coronary artery disease, current treatment for hypertension, arrhythmia, or cardiac failure. All the patients with a concomitant cardiac disease underwent an extensive cardiac evaluation before performing the stair climbing test and they were allowed to perform the test only when deemed in a hemodynamically stable state.

The following additional variables were used for comparison of patients who performed and who did not perform stair climbing test: age, gender, type of operation (lobectomy vs. pneumonectomy), arterial carbon dioxide level (PaCO2), arterial oxygen tension (PaO2), preoperative hemoglobin level, diabetes, concomitant cardiac disease, neoadjuvant chemotherapy, operation time, percentage of functional parenchyma removed during operation (Func loss%), POSSUM Physiological Score [7], smoking history (pack-years), side of resection (right or left), Charlson comorbidity index (CCI). POSSUM Physiological score is a 12-factors four-graded score which was previously validated in thoracic surgery [8]. Likewise CCI is a comorbidity index which was shown to predict postoperative complications after lung resection [6]. We computed the number of pack-years of smoking as the total number of years smoked, times the average number of cigarettes smoked per day, divided by 20.

A.2. Outcome variables

Postoperative cardiopulmonary complications and mortality were considered as those occurring within 30 days from operation or during a longer period if the patient was still in the hospital. Most of the authors [1,17–20] agree that the inclusion of technical complications (i.e. empyema, wound infection, bronchopleural fistula, hemothorax, chylothorax, left recurrent nerve injury, etc.) as outcome variables in the analysis of the efficacy of the exercise tests to predict postoperative morbidity would invalidate the results. In fact, it appears unlikely that a reduced aerobic reserve may be the basis of a technical complication. For this reason and for the sake of comparison with other authors [1,14,17–20], only the following cardiopulmonary complications were included: respiratory failure requiring mechanical ventilation for more than 48 h; pneumonia; atelectasis requiring bronchoscopy; pulmonary edema; pulmonary embolism; myocardial infarction; hemodynamically unstable arrhythmia requiring medical treatment; cardiac failure.

The proportion of deaths among the complicated patients was termed ‘failure to rescue’ [21].