ELSEVIER

European Journal of Cardio-thoracic Surgery 23 (2003) 35–42

EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY

www.elsevier.com/locate/ejcts

A comparison of three scoring systems for predicting complications after major lung resection☆

Mark K. Ferguson*, Amy E. Durkin

Section of Cardiac and Thoracic Surgery, Department of Surgery, The University of Chicago, Chicago, IL, USA

Received 28 August 2002; accepted 10 September 2002

Abstract

Objectives: Although complications occur frequently after major lung resection, current predictive models are not entirely satisfactory. We devised a new predictive scoring system and compared it to two existing systems. **Methods**: We performed an initial retrospective review of 400 patients who underwent major resection for lung cancer from 1980 to 1995. Predictive covariates (age, spirometry, diffusing capacity) associated with three or more complication groups were used to develop a scoring system. This system (EVÁD) was then evaluated against the Physiological and Operative Severity Score for Enumeration of Mortality and Morbidity (POSSUM) and Cardiopulmonary Risk Index (CPRI) systems for patients operated between 1996 and 2001. **Results**: Major resection for lung cancer included lobectomy (188) and pneumonectomy (30). Complication categories were: pulmonary (23; 10.5%); cardiovascular (24; 11.0%); infectious (8; 3.6%); other (29; 13.2%); nonfatal (45; 20.6%); and any (53; 24.2%). Death occurred in ten patients (4.6%). Mean EVÁD scores were significantly different between groups with and without complications in all categories except infectious complications and death, whereas mean CPRI scores differed only for pulmonary complications, nonfatal complications, and death, and mean POSSUM scores did not appropriately differ for any complications. EVÁD predicted incremental risk in all complication categories except cardiovascular, infectious, and death, whereas CPRI predicted incremental risk only for nonfatal and possibly any complications, and POSSUM did not predict incremental risk for any complication category. Receiver operating characteristic analysis demonstrated the EVÁD system to be equivalent to or better than CPRI and POSSUM for all complication categories. **Conclusions**: A simple scoring system (EVÁD) that utilizes pulmonary function test data and patient age predicts the likelihood of complications after major lung resection. It is easier to use and at least as accurate as other scoring systems currently in use.

Keywords: Lung cancer; Lung resection; Postoperative complications; Diffusing capacity; Spirometry; Scoring system

1. Introduction

Lung cancer is the leading cause of cancer death in the United States, and almost 170,000 new cases of lung cancer were predicted for 2002 [1]. Based on tumor stage and other clinical factors, up to 40% of patients with lung cancer are candidates for potentially curative resection. Postoperative cardiopulmonary complications occur in 20–30% of patients after resection for lung cancer [2–6]. The development of these and other complications after major lung resection is associated with an increase in the duration of hospital stay, an increased incidence of operative mortality, and an increase in the cost of hospitalization.

Algorithms for identifying patients who are at increased risk for developing major postoperative complications after lung resection include the Cardiopulmonary Risk Index (CPRI) and the Physiological and Operative Severity Score for Enumeration of Mortality and Morbidity (POSSUM) [6–9]. These have been shown to have variable ability to predict complications and mortality after major lung resection. Such scoring systems are sometimes cumbersome to use because large amounts of information are required for each patient and score calculations are sometimes complex. Simplified scoring systems intended to be more specific for lung resection patients have been proposed, including the Predictive Respiratory Quotient (PRQ) and the Predicted Postoperative Product (PPP) [10,11]. Although these systems appear to have predictive ability for pulmonary complications and operative mortality, respectively, in small numbers of patients, their utility in

☆ Presented in part at the 16th Annual Meeting of the European Association for Cardio-thoracic Surgery, Monte Carlo, Monaco, September 22–25, 2002.

* Corresponding author. Department of Surgery, The University of Chicago Medical Center, 5841 S. Maryland Avenue MC5035, Chicago, IL 60637, USA. Tel.: +1-773-702-3551; fax: +1-773-702-2642.

E-mail address: mferguso@surgery.bsd.uchicago.edu (M.K. Ferguson).

1010-7940/02/$ - see front matter
PII: S1010-7940(02)00675-9

Table 1
Complication definitions and categories

Pulmonary
Initial postoperative ventilatory support >24 h
Reintubation for primary respiratory failure
Pneumonia (pulmonary infiltrate and fever requiring antibiotic therapy)
Lobar collapse on plain chest radiograph
Cardiovascular
Use of inotropic agent other than renal dose dopamine for maintenance of systolic blood pressure or urine output
Arrhythmia requiring treatment
Myocardial infarction
Pulmonary embolism
Cardiopulmonary
Any pulmonary or cardiovascular complication
Infectious
Wound infection
Empyema
Other
Recurrent nerve injury
Bronchopleural fistula
Bleeding requiring reoperation or postoperative transfusion
Miscellaneous
Overall nonfatal complications
Death (death during hospitalization for lung resection or within 30 days of operation)
Any complication

predicting other types of complications is limited and they have not been adopted for general use.

We and others have previously shown that operative risks are related to diffusing capacity for carbon monoxide expressed as a percent of predicted (DLCO%), maximum oxygen consumption during exercise, age, and forced expiratory volume in 1 s expressed as a percent of predicted (FEV_1%) [2,12,13]. Additional work from our medical center has demonstrated that long-term survival is related to age, FEV_1%, cancer stage, and possibly gender, but is not related to diffusing capacity or the type of resection performed [14–17]. Building on this experience, we elected to use this expertise to develop a new and simpler scoring system that will help identify patients who are at increased risk for operative complications after major resection for lung cancer and to compare this system to others that are already in use.

2. Materials and methods

Patients who underwent thoracotomy for major lung resection (lobectomy, bilobectomy, or pneumonectomy) between January 1980 and December 2001 were entered into this retrospective study, which was approved by our institutional review board (IRB). Inclusion criteria were: proven non-small cell lung cancer, no prior major lung resection or distant thoracic irradiation, age ≥21 years, and elective operation. Exclusion criteria were: age <21 years, prior major lung resection, distant prior thoracic irradiation, and emergency or urgent operation. Clinical information and laboratory data for these patients were abstracted from the office and hospital clinical records. Complications for the period of hospitalization were recorded and categorized (Table 1).

Using data from patients operated on during 1980–1995, univariate analyses were performed and potential covariates were identified for each complication category. These were entered into stepwise backward-elimination multivariable analyses to identify important covariates for each complication category. The most frequently identified covariates across the spectrum of complications were selected to create a new scoring system. A weighted score was assigned to each of the value ranges of the variables and the scores were added to obtain an overall risk score.

In order to validate the new scoring system, data from patients operated on during 1996–2001 were used. The new scoring system was analyzed to determine whether scores were distributed normally, whether the mean scores were different between patients with and without complications (unpaired *t*-test), and whether the scores were able to predict incremental risk of complications after collapsing the scores into four categories (chi-squared analysis or Kruskal–Wallis one-way ANOVA; Systat 8.0, Systat Software, Inc., Richmond, CA).

Scores also were calculated according to established techniques for POSSUM and CPRI using 1996–2001 patient data [7,18]. Missing data were imputed. The physical status category of POSSUM was used for statistical comparisons. The three scoring systems also were analyzed to determine whether scores were distributed normally, whether the mean scores were different between groups of patients with and without complications, and whether the scores were able to predict incremental risks of complications. For assessing incremental risk, the physical status component of POSSUM was collapsed into four categories. The utilities of the three scoring systems were compared using receiver operating characteristic (ROC) analysis with calculation of the area under the curve (AUC) using the trapezoidal rule (nonparametric method), adjusting for the correlation between the curves (STATA 7.0, Stata Corporation, College Station, TX) [19,20]. A chi-squared test on one degree of freedom was used to test the pairwise equality of the AUCs for the scoring systems. Data are expressed as mean ± SEM.

3. Results

During the period 1980–1995, 400 patients who met the inclusion criteria for this study underwent major lung resection. There were 235 men and 165 women with a mean age of almost 62 years (range 22–87 years; Table 2). Operations performed included lobectomy or bilobectomy in almost 75% of patients (Table 3). Nearly 70% of patients had pathological stage I disease. The overall incidence of

Table 2
Demographics and laboratory values for the initial (1980–1995) and validation (1996–2001) study groups[a]

	1980–1995 patients[b]	1980–1995 values or affected patients	1996–2001 patients[b]	1996–2001 values or affected patients	P value
Men/women	400	235/165	219	115/104	0.13
Age (years)	400	61.6 ± 0.5	219	64.2 ± 0.8	0.005
Smoking	398	246	219	73	< 0.001
Diabetes	399	50	219	38	0.10
Hypertension	398	125	218	78	0.27
Prior MI	394	48	219	16	0.058
Preoperative RT	297	29	219	16	0.33
Preoperative CT	288	8	201	18	0.006
PS	387	0.43 ± 0.04	219	0.75 ± 0.05	< 0.001
Serum creatinine (mg/dl)	221	1.14 ± 0.07	217	0.94 ± 0.03	0.006
Serum hemoglobin (g/dl)	228	13.1 ± 0.1	216	13.2 ± 0.1	0.5
Serum albumin (g/dl)	327	4.0 ± 0.4	163	4.0 ± 0.05	0.4
FVC%	363	86.2 ± 0.9	219	87.1 ± 1.2	0.5
FEV_1%	362	82.9 ± 1.1	219	81.9 ± 1.5	0.6
DLCO%	319	87.1 ± 1.2	208	83.8 ± 1.6	0.11

[a] Smoking, cigarette smoker within 6 weeks of operation; MI, myocardial infarction; RT, radiotherapy; CT, chemotherapy; PS, ECOG (Eastern Cooperative Oncology Group) performance status; FVC%, forced vital capacity expressed as a percent of predicted; FEV_1%, forced expiratory volume in the first second expressed as a percent of predicted; DLCO%, diffusing capacity for carbon monoxide expressed as a percent of predicted.
[b] Patients available for analysis.

complications was over 35%, mostly representing pulmonary or cardiovascular problems.

The univariate analyses of factors potentially associated with the development of complications are shown in Table 4. Those with P values <0.2 were entered into multivariable analyses to identify covariates associated with the development of complications. Variables with values missing for more than 25% of the patients were not entered into multivariable analyses. The most frequent important covariates identified were DLCO%, FEV_1%, age, and serum albumin (Table 5). Three preliminary scoring systems were devised that included: (a) DLCO%, FEV_1%, and age; (b) DLCO%, FEV_1%, and serum albumin; and (c) DLCO%, FEV_1%, age, and serum albumin. Age was assigned a point for each decade of increasing age beginning at age 50. DLCO% and FEV_1% were assigned a point for each 10 percentage point decrement beginning at 90%. Albumin was assigned a point for each 0.5 decrement beginning at 5.0. The maximum possible score for any variable was 4. The scores within each preliminary system were added to achieve a total score, and scores subsequently were collapsed into either four (three-variable system) or five (four-variable system) categories.

Within each system, mean scores were significantly different for each complication category comparing patients with and without complications except for the category of 'other complications'. Using categorized scores, chi-squared analysis demonstrated that each scoring system

Table 3
Types of operations and postoperative complications

	1980–1995 patients (%)	1996–2001 patients (%)	P value
Procedure			0.001
Lobectomy	259 (64.8)	172 (78.5)	
Bilobectomy	38 (9.5)	17 (7.8)	
Pneumonectomy	103 (25.7)	30 (13.7)	
Stage			0.001
I	183 (45.9)	135 (61.6)	
II	92 (23.0)	39 (17.8)	
III, IV	124 (31.1)	45 (20.6)	
Complication			
Pulmonary	63 (15.9)	23 (10.5)	0.066
Cardiovascular	79 (19.9)	24 (11.0)	0.004
Infectious	18 (4.5)	8 (3.6)	0.6
Other	56 (14.1)	29 (13.2)	0.8
Overall nonfatal	120 (30.1)	45 (20.5)	0.01
Death	32 (8.0)	10 (4.6)	0.10
Any complication	145 (36.6)	53 (24.2)	0.002

Table 4
Univariate analyses of factors associated with complications[a]

Factor	Pulmonary	Cardiovascular	Cardiopulmonary	Infectious	Other	Nonfatal	Death	Any complication
Age	0.071*	< 0.001*	< 0.001*	0.87	0.35	0.20*	0.26	0.048*
Sex	0.97	0.035*	0.16*	0.098*	0.76	0.64	0.42	0.36
PS	0.046*	0.032*	0.013*	0.24	0.25	0.28	0.074*	0.079*
Smoker	0.92	0.78	0.96	0.05*	0.065*	0.26	0.66	0.37
Hemoglobin	0.17	0.24	0.23	0.56	0.01	0.43	0.049	0.10
Albumin	0.012*	0.099*	0.023*	0.82	0.28	0.53	0.005*	0.032*
Creatinine	0.32	0.50	0.21	0.014	0.21	0.65	0.32	0.20
Diabetes	0.092*	0.15*	0.64	0.21	0.67	0.56	0.27	0.54
Hypertension	0.83	0.54	0.41	0.73	0.61	0.83	0.12*	0.85
Prior MI	0.34	0.16*	0.15*	0.38	0.61	0.24	0.53	0.23
Preop RT	0.45	0.58	0.53	0.97	0.24	0.78	0.087	0.51
Preop chemo	0.97	0.014	0.07	0.63	0.22	0.60	0.44	0.31
pO_2	0.66	0.44	0.41	0.15*	0.20*	0.42	0.46	0.20*
pCO_2	0.17*	0.69	0.64	0.46	0.19*	0.67	0.12*	0.30
FEV_1%	0.11*	0.0043*	0.01*	0.39	0.23	0.21*	0.099*	0.025*
DLCO%	< 0.001*	0.13*	0.002*	0.002*	0.02*	0.014*	0.019*	0.001*

[a] PS, ECOG (Eastern Cooperative Oncology Group) performance status; smoker, cigarette smoker within 6 weeks of operation; hemoglobin, serum hemoglobin (g/dl); albumin, serum albumin (g/dl); creatinine, serum creatinine (mg/dl); prior MI, prior myocardial infarction; preop RT, preoperative radiotherapy; preop chemo, preoperative chemotherapy; FEV_1%, forced expiratory volume in the first second expressed as a percent of predicted; DLCO%, diffusing capacity for carbon monoxide expressed as a percent of predicted. *These values were entered into subsequent multivariable analyses.

predicted incremental risk of most complications (pulmonary, cardiovascular, cardiopulmonary, nonfatal, death, and any complication) but there was no reliable prediction of incremental risk for the categories of infectious and other complications. Using ROC analysis the AUCs for complication categories of pulmonary, cardiovascular, cardiopulmonary, nonfatal, death, and any were similar and were generally best for the preliminary scoring system that used F*EV*$_1$%, *a*ge, and *D*LCO%, which was termed the EVÁD scoring system.

The validity of the EVÁD scoring system was assessed using the population of patients who underwent major lung resection from 1996 to 2001. These 219 patients were somewhat older and had a worse mean performance status than the first group of patients, but in most other respects their preoperative data and demographics were similar (Table 2). Compared to the initial patient group, the validation group of patients underwent more lobectomies and fewer pneumonectomies, and had a higher incidence of stage I cancers. The validation group also experienced a lower incidence of many complication categories (Table 3). Mean EVÁD scores for the initial and validation groups were 7.0 ± 0.1 and 7.7 ± 0.2, respectively ($P < 0.001$).

CPRI scores were calculated for the validation patient group. A total of 3066 data points were necessary and 73 (2.4%) were missing. These were all values for pCO_2, which were imputed using the formula $pCO_2 = 40.8 - (0.036 \times FEV_1\%)$, which was derived empirically from our data. The POSSUM scores also were calculated for the validation patient group. For the physical status component of the POSSUM scoring system a total of 3942 data points were necessary and 39 (1.0%) were missing. The values for systemic blood pressure, heart rate, serum hemoglobin, white blood cell count, and serum chemistries were imputed using the mean values for the remaining patients. For the EVÁD system a total of 668 values were required, of which

Table 5
Significant covariates for complications[a]

Complication category	Age	Albumin	Diabetes	DLCO%	FEV_1%	PS
Pulmonary	–	2.34 (1.04–5.28)	–	1.46 (1.20–1.78)	–	–
Cardiovascular	1.91 (1.36–2.69)	–	6.81 (1.52–30.37)	1.14 (0.99–1.30)	1.23 (1.05–1.44)	–
Cardiopulmonary	1.68 (1.26–2.23)	–	–	1.24 (1.10–1.41)	1.16 (1.02–1.33)	–
Infectious	–	–	–	1.56 (1.18–2.06)	–	–
Other	–	–	–	1.22 (1.04–1.43)	–	–
Nonfatal	–	–	–	1.17 (1.04–1.32)	–	–
Death	–	3.11 (1.13–8.57)	–	1.32 (1.05–1.66)	–	1.71 (1.10–2.67)
Any complication	–	–	–	1.28 (1.13–1.45)	1.12 (0.99–1.26)	–

[a] Values represent odds ratios; numbers in parentheses represent 95% confidence intervals. PS, ECOG (Eastern Cooperative Oncology Group) performance status; albumin, serum albumin (g/dl); DLCO%, diffusing capacity for carbon monoxide expressed as a percent of predicted; FEV_1%, forced expiratory volume in the first second expressed as a percent of predicted.

Table 6
Mean scores for patients without and with complications for three different scoring systems[a]

Complication	EVÁD	CPRI	POSSUM[b]
Pulmonary	7.5 ± 0.2 vs. 8.7 ± 0.4 (0.009)	2.7 ± 0.1 vs. 3.0 ± 0.2 (0.05)	21.0 ± 0.3 vs. 22.1 ± 1.0 (0.30)
Cardiovascular	7.6 ± 0.2 vs. 8.5 ± 0.4 (0.04)	2.7 ± 0.1 vs. 2.9 ± 0.2 (0.19)	20.9 ± 0.3 vs. 22.7 ± 0.8 (0.05)
Cardiopulmonary	7.5 ± 0.2 vs. 8.5 ± 0.4 (0.01)	2.6 ± 0.1 vs. 3.0 ± 0.2 (0.04)	21.0 ± 0.3 vs. 21.7 ± 0.7 (0.39)
Infectious	7.7 ± 0.2 vs. 7.5 ± 0.5 (0.79)	2.7 ± 0.1 vs. 2.9 ± 0.3 (0.56)	21.2 ± 0.3 vs. 18.9 ± 0.6 (0.009)
Other	7.5 ± 0.2 vs. 8.7 ± 0.4 (0.004)	2.7 ± 0.1 vs. 2.8 ± 0.2 (0.37)	21.0 ± 0.3 vs. 22.2 ± 1.0 (0.24)
Nonfatal	7.4 ± 0.2 vs. 8.7 ± 0.3 (<0.001)	2.6 ± 0.1 vs. 3.1 ± 0.1 (0.005)	21.0 ± 0.3 vs. 21.5 ± 0.7 (0.51)
Death	7.6 ± 0.2 vs. 8.7 ± 0.6 (0.10)	2.7 ± 0.1 vs. 2.7 ± 0.2 (0.99)	21.0 ± 0.3 vs. 23.0 ± 1.5 (0.24)
Any complication	7.4 ± 0.2 vs. 8.6 ± 0.3 (<0.001)	2.6 ± 0.1 vs. 3.0 ± 0.1 (0.005)	20.9 ± 0.4 vs. 21.8 ± 0.7 (0.26)

[a] Unpaired *t*-test *P* values are in parentheses.
[b] The physical status component was used.

11 (1.7%) were missing. These were all values for DLCO%, which were imputed using the formula $\text{DLCO\%} = 60.1 + (0.289 \times \text{FEV}_1\%)$, which was derived empirically from our data.

Mean EVÁD scores for patients without and with complications differed significantly for all categories except infection and operative mortality (Table 6). The differences were greater for the EVÁD system than for the CPRI or POSSUM systems. None of the systems were able to predict infectious complications with any accuracy, and only the EVÁD system demonstrated a tendency towards predicting operative mortality. After collapsing scores into four categories, the EVÁD system was shown to predict incremental risk of complications for most categories of complications (Fig. 1). Based on chi-squared analyses using collapsed scores, the ability to predict incremental risk was equivalent to or better for the EVÁD system than for the CPRI or POSSUM systems for all complication categories (data not shown). ROC analysis produced AUCs that were greater for the EVÁD system than for the CPRI or POSSUM systems for all complication categories except infection, for which the CPRI system had a higher AUC (Table 7). The EVÁD system approached or reached statistical difference compared to the CPRI system for predicting other complications and compared to the POSSUM system for predicting nonfatal and any complications. In general, all AUCs were less than ideal.

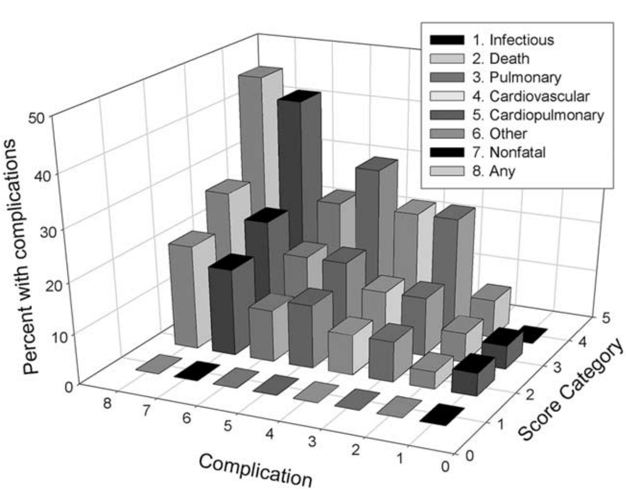

Fig. 1. Incidence of complications according to score and complication categories. Score categories are: 1 = score 0–4 (20 patients); 2 = score 5–7 (87 patients); 3 = score 8–10 (85 patients); 4 = score 11–16 (26 patients).

4. Discussion

Major pulmonary resection often is associated with the development of important complications. Such complications dramatically increase the length of hospital stay and the cost of hospitalization [2]. The ability to predict which patients are at increased risk for the development of complications may help select patients who are most appropriate for surgery, institute preoperative interventions to decrease operative risks, and provide increased therapeutic resources postoperatively for higher risk patients in an effort to decrease the incidence of complications. A number of individual factors have been associated with an increased risk of complications and mortality after major lung resection including age, performance status, nutritional status, exercise capacity, cardiovascular factors, spirometric values, and measures of gas exchange or consumption such as diffusing capacity and oxygen consumption during exercise.

No single measure has been found to have adequate predictive value in the identification of the risk of complications after major lung resection. In an effort to improve predictive ability, various risk factors have been combined into scoring systems. Some systems have been developed that are specific for lung resection, including the CPRI, the PRQ, and the PPP [7,10,11]. Among these, the CPRI system has been evaluated most often in the assessment of complications after major lung resection, with mixed results. Two publications suggested that postoperative complications were predicted by a high CPRI score [6,7], whereas one found that CPRI was possibly useful in predicting complications after pneumonectomy but not after lobectomy [8]. In the latter study the AUCs for pulmonary, cardiac, and any complications ranged from 0.51 to 0.53, which were somewhat lower than in the present study. In contrast, other

Table 7
Area under the curve for ROC analysis comparing three different scoring systems for their ability to predict complications

Complication category	EVÁD	CPRI	POSSUM[a]	EVÁD vs. CPRI[b]	EVÁD vs. POSSUM
Pulmonary	0.652	0.608	0.602	0.47 (−0.08, 0.17)	0.53 (−0.11, 0.21)
Cardiovascular	0.613	0.573	0.583	0.56 (−0.10, 0.18)	0.68 (−0.12, 0.18)
Cardiopulmonary	0.625	0.597	0.549	0.62 (−0.08, 0.14)	0.23 (−0.05, 0.20)
Infectious	0.475	0.551	0.383	0.45 (−0.13, 0.28)	0.44 (−0.15, 0.33)
Other	0.648	0.547	0.577	0.06 (−0.01, 0.21)	0.34 (−0.08, 0.22)
Nonfatal	0.656	0.631	0.534	0.61 (−0.07, 0.12)	0.02 (0.01, 0.23)
Death	0.641	0.506	0.647	0.17 (−0.06, 0.33)	0.96 (−0.24, 0.25)
Any complication	0.654	0.621	0.552	0.48 (−0.06, 0.13)	0.06 (0.00, 0.21)

[a] The physical status component was used.
[b] *P* value for pairwise equality (95% confidence interval for difference).

systems have been developed that are applicable to general surgical procedures and subsequently have been used to assess issues related to pulmonary resection procedures, such as POSSUM and APACHE II [9,21]. Of these, POSSUM has been evaluated most often in the assessment of complications after major lung resection [9,22,23]. One group in particular has reported that POSSUM is useful for predicting postoperative complications with an AUC of 0.66, which is substantially higher than the AUC for the any complication category in the present study [22].

Despite their potential utility, the existing scoring systems are cumbersome to use, requiring the collection of large amounts of data and time-consuming calculations in order to generate a risk score for an individual patient. We attempted to improve upon these scoring techniques by devising a system based on just a few variables that are importantly associated with postoperative complications after major lung surgery. Prior work from our institution has documented the importance of age and diffusing capacity in the development of postoperative complications [2,12,13,15]. These, along with spirometry, which is a universally recognized predictor of risk of major lung resection, were selected for inclusion in the scoring system. The resultant system (EVÁD) appeared to have good predictive ability for postoperative complications.

The EVÁD system was validated using a new patient population and was then compared to CPRI and POSSUM to determine their relative utility in predicted complications. We chose to use a group of patients operated on more recently for validation and comparison purposes rather than selecting a contemporaneous group. The amount of data needed to calculate scores for CPRI and POSSUM was prohibitively large and was not available for the group of patients operated on prior to 1996. Validation of the EVÁD system using a separate set of patients confirmed its predictive ability. Interestingly, the mean EVÁD score was significantly higher for the validation group than for the initial group despite the fact that the initial group had a higher incidence of complications including operative mortality. These findings suggest that improvements in patient selection, surgical techniques, and postoperative care have resulted in a decreased incidence of complications in spite of the fact that patients in the validation group were at higher risk.

The comparisons among the systems demonstrated that the EVÁD system had a better ability to quantitate relative risk than CPRI or POSSUM. In addition, the AUCs for the EVÁD system were better than for CPRI and POSSUM for almost all complication categories. The AUC for the EVÁD system was significantly better than POSSUM for the category of other complications and approached significance for the category of any complications. The EVÁD system requires the use of data that are normally available to the surgeon at the time of the patient's preoperative visit, making the assessment of risk straightforward. Thus, the ease of use of the EVÁD system, combined with its equal or superior ability to predict risk, make it a preferable scoring system for predicting the risk of complications after major lung resection.

The EVÁD system is far from ideal, however. There is no single value that can serve as a threshold for differentiating among patients who are at normal risk compared to those who are at increased risk. In addition, the AUCs for the new system do not approach the ideal 0.8–0.9 or greater values that would make its predictive ability invaluable. What the system currently permits is quantitation of a surgeon's impression of an individual patient's relative risk for major lung resection based on the factors included in the scoring system. This information, along with the surgeon's general impression of the patient, may enhance a surgeon's ability to select patients for major lung resection and thus reduce risk to the overall population of patients who are candidates for surgery.

There are manifold reasons for trying to develop a useful scoring system. An accurate scoring system would help stratify patients into average, increased, and prohibitive risk categories as an aid in selecting patients for lung resection. Patients who are in increased or prohibitive risk categories may benefit from preoperative cardiopulmonary rehabilitation. Knowledge of an individual patient's risk will assist in the discussion of risk with that patient, enhancing the informed consent process. Patients at increased risk may be afforded more resources for postoperative care, possibly reducing the incidence of complications. Finally, risk stratification may be of use in comparing outcomes among surgeons or institutions for quality improvement purposes.

Improvements in current scoring systems are necessary before their routine use can be recommended. It may be that variations in patient populations among institutions or among countries are so great that no single scoring system will be found globally useful. In addition, vagaries of postoperative complications are such that a certain percentage of them cannot be predicted based on preoperative factors. Nevertheless, we feel it is important to refine current scoring parameters and further assess their utility in predicting which patients are at risk for postoperative complications. In this light, we are currently performing a prospective evaluation of these scoring systems in our lung resection candidates.

Acknowledgements

We wish to thank Ted Karrison and Kristen Kasza of the Department of Health Studies, The University of Chicago, as well as the Cancer Research Center of The University of Chicago for statistical support.

References

[1] Jemal A, Thomas A, Murray T, Thun M. Cancer statistics, 2002. CA Cancer J Clin 2002;52:23–47.
[2] Wang J, Olak J, Ultmann RE, Ferguson MK. Assessment of pulmonary complications after lung resection. Ann Thorac Surg 1999;67:1444–1447.
[3] Hazelrigg SR, Landreneau RJ, Boley TM, Priesmeyer M, Schmaltz RA, Nawarawong W, Johnson JA, Walls JT, Curtis JJ. The effect of muscle-sparing versus standard posterolateral thoracotomy on pulmonary function, muscle strength, and postoperative pain. J Thorac Cardiovasc Surg 1991;101:394–400.
[4] Busch E, Verazin G, Antkowiak JG, Driscoll D, Takita H. Pulmonary complications in patients undergoing thoracotomy for lung resection. Chest 1994;106:1930–1931.
[5] 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;151:1472–1480.
[6] Epstein SK, Faling LJ, Daly BD, Celli BR. Inability to perform bicycle ergometry predicts increased morbidity and mortality after lung resection. Chest 1995;107:311–316.
[7] Epstein SK, Faling LJ, Daly BD, Celli BR. Predicting complications after pulmonary resection. Preoperative exercise testing vs a multifactorial cardiopulmonary risk index. Chest 1993;104:694–700.
[8] Melendez JA, Carlon VA. Cardiopulmonary risk index does not predict complications after thoracic surgery. Chest 1998;114:69–75.
[9] Brunelli A, Fianchini A, Gesuita R, Carle F. POSSUM scoring system as an instrument of audit in lung resection surgery. Ann Thorac Surg 1999;67:329–331.
[10] Pierce RJ, Copland JM, Sharpe K, Barter CE. Preoperative risk evaluation for lung cancer resection: predicted postoperative product as a predictor of surgical mortality. Am J Respir Crit Care Med 1994;150:947–955.
[11] Melendez JA, Barrera R. Predictive respiratory complication quotient predicts pulmonary complications in thoracic surgical patients. Ann Thorac Surg 1998;66:220–224.
[12] Ferguson MK, Little L, Rizzo L, Popovich KJ, Glonek GF, Leff A, Manjoney D, Little AG. Diffusing capacity predicts morbidity and mortality after pulmonary resection. J Thorac Cardiovasc Surg 1988;96:894–900.
[13] Ferguson MK, Reeder LB, Mick R. Optimizing selection of patients for major lung resection. J Thorac Cardiovasc Surg 1995;109:275–281.
[14] Ferguson MK, Skosey C, Hoffman PC, Golomb HM. Sex-associated differences in presentation and survival in patients with lung cancer. J Clin Oncol 1990;8:1402–1407.
[15] Wang J, Olak J, Ferguson MK. Diffusing capacity predicts operative mortality but not long-term survival after resection for lung cancer. J Thorac Cardiovasc Surg 1999;117:581–587.
[16] Ferguson MK, Karrison T. Does pneumonectomy for lung cancer adversely influence long-term survival? J Thorac Cardiovasc Surg 2000;119:440–448.
[17] Ferguson MK, Wang J, Hoffman PC, Haraf DJ, Olak J, Masters GA, Vokes EE. Sex-associated difference in survival of patients undergoing resection for lung cancer. Ann Thorac Surg 2000;69:245–250.
[18] Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. Br J Surg 1991;78:356–360.
[19] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–845.
[20] Song HH. Analysis of correlated ROC areas in diagnostic testing. Biometrics 1997;53:370–382.
[21] Giangiuliani G, Gui D, Bonatti P, Tozzi P, Caracciolo F. APACHE II in surgical lung carcinoma patients. Chest 1990;98:627–630.
[22] Brunelli A, Fianchini A, Xiume F, Gesuita R, Mattei A, Carle F. Evaluation of the POSSUM scoring system in lung resection surgery. Physiological and operative severity score for the enumeration of mortality and morbidity. Thorac Cardiovasc Surg 1998;46:141–146.
[23] Brunelli A, Fianchini A, Al Refai M, Gesuita R, Carle F. Internal comparative audit in a thoracic surgery unit using the physiological and operative severity score for the enumeration of mortality and morbidity (POSSUM). Eur J Cardiothorac Surg 2001;19:924–928.

Appendix A. Conference discussion

Dr W.S. Walker (*Edinburgh, UK*): Could I just ask you for clarification. You devised the score by determining the most significant variables from analyzing your own data. Is it then appropriate to compare that with the other scoring systems in the same data set when you've already picked out the most significant data elements for your own system?

Dr Ferguson: There is concern about the appropriateness of using a data set and then going back retrospectively and looking at it again, which is why we developed the scoring system using the initial group from 1980 through 1995 and then applied that scoring system as well as the other two prospectively to the patients from 1996 through 2001 to hopefully avoid that problem.

Dr O. Kshivets (*Siauliai, Lithuania*): My question is, do you use only the statistical method? Did you use neural network computing and Monte Carlo simulation? Your precision of your scoring system is only 65%. Its not enough for individual prognosis. If you use a combined approach, I think ROC may be 0.99, 0.98 maybe, and the precision may be 85% and higher. Its enough for individual prognosis.

Dr Ferguson: I'm sorry, I didn't understand what you meant by a combined approach.

Dr Kshivets: Statistical artificial intelligence and system analysis in complex, mixed. Your scoring system is based only on the statistical method. This method is not enough for individual prognosis. This prognosis is only for a group of patients. For individual prognosis its not enough, because your precision is only 65%.

Dr Ferguson: Yes, I agree completely that this system isn't useful for predicting complications for individual patients. We have not applied a neural network approach to this problem.

Dr D. Wood (*Seattle, WA, USA*): I disagree with one of your conclusions, which is that your risk stratification is very useful for patient selection, because we already make those judgments – not in as scientific a manner as you have, but in our surgical judgments. The difficulty when we don't have an extremely accurate way of predicting complications, which this still isn't, is that we still have the difficulty of directing patients who have otherwise fatal disease to palliative rather than curative treatment. I think you are absolutely right that for defining resources or anticipating complications it is useful, but I'm wondering, couldn't we especially apply it in developing risk stratification outcomes analysis that we do not do very well now. That is where I'm excited about this. Our cardiac colleagues have provided risk stratification that we do not have in general thoracic surgery, and I think this is elegant, simple, and that might be the most useful application, developing it in our systems for risk stratification.

Dr Ferguson: For risk stratification and surgical audit, yes, I think that that's quite useful.

Let me just comment on the patient selection issue. I agree that its difficult to turn down patients who are at increased risk for complications or mortality after major lung resection. In fact, polls have identified operative mortality as very low on the list of concerns of patients who are at increased risk. Elderly patients, for example, are more concerned about quality of life postoperatively. But it does assist us in discussing with patients what the risks may be, and we and the patients are comfortable that informed consent has taken place.