A model for the internal evaluation of the quality of care after lung resection in the elderly
Alessandro Brunelli*, Aroldo Fianchini, Majed Al Refai, Michele Salati

Unit of Thoracic Surgery, Department of Respiratory Diseases, “Umberto I” Regional Hospital, Ancona, Italy

Received 14 November 2003; received in revised form 28 January 2004; accepted 29 January 2004

Abstract

Objectives: The objective of this study was to develop a logistic model for internal audit in a population of elderly patients submitted to lung resection. Methods: Three hundred twenty-six patients older than 70 years of age and submitted to lung resection for lung carcinoma were retrospectively analyzed. Univariate and logistic regression analyses yielded a model for the prediction of postoperative complications that was validated by bootstrap resampling analysis. The model was then used to assess the performance of our unit during two successive periods of activity (‘early’, 1993–1999; ‘late’, 2000–2003). Results: Significant independent predictors of postoperative complications were a low ppoFEV1 ($P < 0.0001$), the presence of concomitant cardiac disease ($P = 0.01$) and extended resection ($P = 0.03$). The observed morbidity rate in the late period was higher than that in the early period (48.3 vs. 33.8%; $P = 0.008$). The predicted morbidity rate was also higher in the late period, compared to that in the early period (44 vs. 39%; $P = 0.003$). Moreover, no differences were noted between predicted and observed morbidity rates in each of the two periods (early, $P = 0.4$; late, $P = 0.5$). Conclusions: We showed that applying a model of risk-adjustment in elderly patients submitted to lung resection was useful for the internal evaluation of the quality of care and prevented misleading information derived by the comparison of the crude rates of the observed morbidity.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Postoperative complications; Lung resection; Lung cancer; Audit; Quality of care

1. Introduction

Risk-adjusted models of outcomes are increasingly used to evaluate the quality of care of surgical units. In subsets of patients with an increased frequency of underlying co-morbidities, such as the elderly, the use of such predictive models seems a logical choice for quality assessment and quality improvement processes.

The objective of this study was to develop a logistic model to be used as an instrument to assess the performance of our thoracic surgery unit over two successive periods of activity in a population of elderly patients submitted to pulmonary resection for non-small cell lung cancer (NSCLC).

2. Patients and methods

Three hundred forty-six patients older than 70 years of age underwent lung resection for NSCLC at our institution from January 1993 to August 2003, representing 32.4% of all the patients submitted to lung resection for lung cancer during the same period. Twenty patients were excluded from the analysis for incomplete data. Thus, 326 patients (275 males, 51 females) formed the dataset of the present study. Twenty-nine wedge resections/segmentectomies, 263 lobectomies, and 34 pneumonectomies were performed by the same surgical team. The same medical staff administered the perioperative treatment.

This is a retrospective analysis performed on a prospective database.

Resectability was assessed by means of computed tomography, bronchoscopy, and, when indicated, cervical mediastinoscopy. Operability was evaluated as follows: pulmonary function tests, blood gas analysis, electrocardiogram, echocardiography, and more invasive cardiologic tests when needed. In the last 3 years, a symptom-limited stair climbing test was systematically used in our patients for risk stratification before operation. Surgery was contraindicated in those patients with a predicted postoperative forced expiratory volume in one second (ppoFEV1) less than 30% of predicted, according to Markos et al. [1],
and with hemodynamic instability (symptomatic coronary artery disease, cardiac failure, symptomatic arrhythmia). In the last 3 years, even those patients with a ppoFEV1 lower than 30% but with a total height climbed at the stair climbing test above 12 m were considered for operation.

Postoperative complications were considered as those occurring within 30 days postoperatively or for a longer period if the patient was still in the hospital.

According to previous studies [2–7] and the EACTS/ESTS thoracic surgery database [8], the following complications were included: prolonged air leak (more than 7 days); bronchopleural fistula; respiratory failure requiring mechanical ventilation for more than 48 h; pneumonia; atelectasis requiring bronchoscopy; adult respiratory distress syndrome (ARDS); empyema; hemotherax requiring reoperation; chylothorax; pulmonary edema; pulmonary embolism; myocardial infarction; hemodynamically unstable arrhythmia requiring medical treatment; cardiac failure; fever higher than 38 °C for more than 3 days without any apparent cause; acute renal failure; stroke. Mortality was not separately analyzed due to small numbers, however, since all patients who died also had complications, they were included in the analysis as complicated patients.

We included only those complications that added complexity to the management of the patients and prolonged the hospital stay. Furthermore, we considered the occurrence of any of the aforementioned complications as a single outcome measure, in accordance with most of the work done with complications at the present time that does not weigh complications [9].

The following variables were tested for the association with complications: age, arterial oxygen tension (PaO2), arterial carbon dioxide level (PaCO2), smoking history (pack-years), preoperative hemoglobin concentration, serum albumin level, type of operation (wedge/lobectomy/pneumonectomy), type of resection (extended vs. non-extended), pathologic T status (pT1 vs. >pT1), pathologic N status (pN + vs. pN −), presence of concomitant cardiac disease, diabetes, neoadjuvant chemotherapy, preoperative pulmonary function tests (forced expiratory volume in one second, FEV1; forced vital capacity, FVC; FEV1/FVC ratio; predicted postoperative FEV1, pppoFEV1; residual volume to total lung capacity, RV/TLC ratio). We started to systematically perform carbon monoxide diffusion lung capacity (DLCO), using the single breath method, from January 2000. Therefore, only 158 patients (48.5%) in this series had this test performed. Due to the low prevalence of the test we did not use DLCO in the logistic regression models, however, this variable was evaluated in the univariate analysis to assess its effect on the occurrence of complications.

We computed the number of pack-years of smoking as the total number of years smoked multiplied by the average number of cigarettes smoked per day, divided by 20.

For the purpose of the present study a resection was considered extended when associated with resection of parietal pleura (extrapleural resection), chest wall, mediastinal structures or diaphragm. In this series, 40% of the extended resections were extrapleural, 30% were associated with chest wall resection, 30% were associated with resection of the diaphragm, pericardium or other mediastinal structures.

Pathologic T and N descriptors were classified according to the 1997 International System for Staging Lung Cancer [10].

A concomitant cardiac disease was defined as follows: previous cardiac surgery, previous myocardial infarction, history of coronary artery disease, current treatment for arrhythmia, cardiac failure or hypertension.

Pulmonary function tests were performed according to the American Thoracic Society criteria. Results of spirometry were collected after bronchodilator administration.

FEV1, FVC, ppoFEV1 and DLCO were expressed as percentages of predicted age, sex and height. PpoFEV1 was calculated by the following formula: (preoperative FEV1/number of preoperative functioning segments) × number of postoperative functioning segments. 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. [1]. The simple calculation of ppoFEV1 was previously shown to be as accurate as perfusion lung scanning [1].

2.1. Statistical analysis

The entire database was initially used to develop the predictive logistic model. The comparisons between patients with and without complications were performed by means of the unpaired Student’s t-test for numerical variables with normal distribution and by means of the Mann–Whitney U-test for numerical variables without a normal distribution. Categorical variables were compared by means of the $\chi^2$-test or Fisher’s exact test. Significant variables at univariate analysis were then used as independent variables in a stepwise logistic regression analysis. The presence or absence of one or more complications was used as dependent variable. All data were more than 99% complete. To avoid multicollinearity, only one variable in a set of variables with a correlation coefficient greater than 0.5 was used in the regression model. The area under the receiver operating characteristics (ROC) curve was used to study the discrimination ability of the model. Model calibration was examined by comparing observed and predicted morbidity rates within four subgroups arranged in increasing order of patient’s risk of complications ($< 0.2$, $0.2–0.4$, $0.4–0.6$, $> 0.6$). Hosmer–Lemeshow goodness-of-fit was used to assess the calibration of the model. Furthermore, the multivariate procedure was validated by bootstrap bagging with 1000 samples. In the bootstrap procedure, repeated samples of 326 observations (the same
number of observations as the original database) were selected with replacement from the original set observations. For each sample, stepwise logistic regression was performed, entering the variables with \( P < 0.05 \) at univariate analysis. The stability of the final stepwise 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 stepwise model. If the final stepwise model variables occur in a majority (>50%) of the bootstrap models, the original final stepwise regression model can be judged to be stable.

The logistic model was then used to predict morbidity in the patients operated on during two successive periods, which were arbitrarily selected (early: 1993–1999, 152 patients; late: 2000–2003, 174 patients). Predicted and observed morbidity rates in each period were then compared by means of the \( \chi^2 \)-test.

All the statistical tests were two-tailed and a significance level of 0.05 was accepted. The analysis was performed by using the statistical softwares Statview 5.0 (SAS Institute, Cary, NC) and STATA 8.0 (Stata Corp., College Station, TX).

3. Results

The morbidity rate in the entire database was 41.4% (135 patients) and 13.3% of complicated patients died (18 cases; overall mortality rate 5.5%).

There were 181 complications in 135 patients. Forty-one (30.3%) patients had more than one complication. The complications in order of frequency were: prolonged air leak (52 cases), arrhythmia (27 cases), pneumonia (22 cases), respiratory failure (19 cases), empyema (10 cases), fever (8 cases), pulmonary edema (7 cases), atelectasis (7 cases), bronchopleural fistula (6 cases), hemotherax (5 cases), myocardial infarction (5 cases), cardiac failure (5 cases), acute renal failure (3 cases), chylothorax (2 cases), stroke (2 cases), ARDS (1 case).

Table 1 shows the results of the univariate comparison between patients with and without complications. In particular, compared with patients without complications, those with complications had a lower FEV1 (\( P = 0.0002 \)), FVC (\( P = 0.009 \)), FEV1/FVC ratio (\( P = 0.002 \)), ppoFEV1 (\( P = 0.0002 \)) and serum albumin level (\( P = 0.004 \)). Moreover, they had a greater frequency of cardiac co-morbidities (\( P = 0.02 \)) and extended resections (\( P = 0.007 \)). DLC0 was lower in patients with complications with respect to those without complications (69.3% vs. 76.1%, respectively; \( P = 0.02 \)).

Logistic regression analysis showed that significant independent predictors of complications were ppoFEV1 (\( P < 0.0001 \)), the presence of a concomitant cardiac disease (\( P = 0.01 \)) and the extended resection (\( P = 0.03 \)) (Table 2). The area under the ROC curve was 0.67 (95% CI, 0.61–0.72). No differences were noted between predicted and observed morbidity rates among patient subgroups of risk

<table>
<thead>
<tr>
<th>Variables</th>
<th>Complicated</th>
<th>Non-complicated</th>
<th>( P )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>74.6 (2.9)</td>
<td>74.6 (3.0)</td>
<td>0.7*</td>
</tr>
<tr>
<td>PaO2 (mmHg)</td>
<td>79.2 (10.4)</td>
<td>80.0 (10.9)</td>
<td>0.6</td>
</tr>
<tr>
<td>PaCO2 (mmHg)</td>
<td>37.4 (4.2)</td>
<td>37.9 (5.5)</td>
<td>0.9*</td>
</tr>
<tr>
<td>Pack-years</td>
<td>43.9 (26.6)</td>
<td>40.6 (29.9)</td>
<td>0.3</td>
</tr>
<tr>
<td>Preop Hb (g/dl)</td>
<td>13.4 (1.9)</td>
<td>13.6 (1.5)</td>
<td>0.2</td>
</tr>
<tr>
<td>Albumin (g/dl)</td>
<td>4.0 (0.7)</td>
<td>4.1 (0.4)</td>
<td>0.004*</td>
</tr>
<tr>
<td>Operation (%; W/L/P)</td>
<td>74/79/4/13.2</td>
<td>10/0/81.6/8.4</td>
<td>0.3*</td>
</tr>
<tr>
<td>Extended resection (%)</td>
<td>20.6</td>
<td>10.0</td>
<td>0.007*</td>
</tr>
<tr>
<td>pT stage (%; &gt;pT1)</td>
<td>77.9</td>
<td>74.7</td>
<td>0.6*</td>
</tr>
<tr>
<td>pN stage (%; N +)</td>
<td>37.1</td>
<td>43.2</td>
<td>0.3*</td>
</tr>
<tr>
<td>Cardiac disease (%)</td>
<td>61.8</td>
<td>48.4</td>
<td>0.02*</td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td>9.6</td>
<td>13.2</td>
<td>0.3*</td>
</tr>
<tr>
<td>Neoadjuvant chemotherapy (%)</td>
<td>8.8</td>
<td>7.9</td>
<td>0.8*</td>
</tr>
<tr>
<td>FEV1%</td>
<td>79.7 (18.2)</td>
<td>89.8 (22.2)</td>
<td>0.0002*</td>
</tr>
<tr>
<td>FVC%</td>
<td>90.2 (16.3)</td>
<td>97.3 (20.0)</td>
<td>0.009*</td>
</tr>
<tr>
<td>FEV1/FVC ratio</td>
<td>0.66 (0.12)</td>
<td>0.70 (0.11)</td>
<td>0.002*</td>
</tr>
<tr>
<td>RV/TLC ratio</td>
<td>0.47 (0.08)</td>
<td>0.48 (0.08)</td>
<td>0.5*</td>
</tr>
<tr>
<td>ppoFEV1%</td>
<td>63.1 (15.6)</td>
<td>71.7 (19.4)</td>
<td>0.0002*</td>
</tr>
</tbody>
</table>

Results are expressed as means ± SD unless otherwise specified. Operation (W/L/P): W, wedge resection/segmentectomy; L, lobectomy; P, pneumonectomy. Cardiac disease: concomitant cardiac disease. * Mann–Whitney U-test.  b \( \chi^2 \)-test.

(Fig. 1). This showed that the model was relatively accurate across the ranges of patient risk subgroups. The Hosmer–Lemeshow goodness-of-fit was 7.0 (\( P = 0.5 \)), showing good calibration of the model. The frequency of occurrence of the independent variables in 1000 bootstrap resampling models was: ppoFEV1 (97.8%), concomitant cardiac disease (71.2%), extended resection (57.3%). All the predictors in the final logistic model occurred in more than 50% of the bootstrap samples, indicating their reliability.

The logistic equation for the prediction of the risk of complication was: exp(0.97 – 0.026 × ppoFEV1 + 0.599 × concomitant cardiac disease + 0.704 × extended resection)/(1 + exp(0.97 – 0.026 × ppoFEV1 + 0.599 × concomitant cardiac disease + 0.704 × extended resection)). Concomitant cardiac disease and extended resection

Table 2

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Coefficient</th>
<th>Odds ratio (95% CI)</th>
<th>( P )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.973</td>
<td>0.97 (0.96–0.99)</td>
<td>0.04</td>
</tr>
<tr>
<td>ppoFEV1%</td>
<td>–0.026</td>
<td>0.97 (0.96–0.99)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Cardiac disease</td>
<td>0.599</td>
<td>1.82 (1.13–2.92)</td>
<td>0.18</td>
</tr>
<tr>
<td>Extended resection</td>
<td>0.704</td>
<td>2.02 (1.05–3.88)</td>
<td>0.06</td>
</tr>
</tbody>
</table>

AF, attributable fraction; attributable fraction was estimated from within a logistic regression framework, thus enabling the confounders to be taken into account.
were coded as 0 for absence and 1 for presence of the variables, respectively. This equation was used to predict the risk of complications in two successive periods of activity (early, 1993–1999; late, 2000–2003).

Table 3 shows the comparison between the patients operated on in the early and late periods, respectively. In the late period, there was a significantly higher proportion of patients with a predicted morbidity rate greater than 60% compared to the early period (Table 4).

When the patients were grouped according to their period of operation, those in the early period had a lower incidence of observed complications compared to those in the late period (33.8 vs. 48.3%, respectively; \( P = 0.008 \)). However, the predicted risk of complications was also lower in the early period compared to the late period (39 vs. 44%, respectively; \( P = 0.003 \)). No differences were noted between predicted and observed morbidity rates in each of the two periods (early, \( P = 0.4 \); late, \( P = 0.5 \)).

### 4. Discussion

The objective of the present study was to develop and validate a logistic model to be used as an instrument of internal audit for quality control in a subset of elderly patients submitted to lung resection for NSCLC.

We chose to analyze a cohort of elderly patients since in this population the physiologic changes in the cardiovascular and respiratory systems and the underlying co-morbidities increase the risk of postoperative morbidity. In these high-risk surgical candidates, the use of a risk-adjustment model seems appropriate for audit purposes, since the crude comparison of morbidity rates can be misleading and may encourage surgeons to practice a risk-averse behavior.

Although mortality rate would be an important outcome measure for assessing the performance of a surgical unit, limitations exist for using this clinical endpoint in internal audit analyses for procedures where deaths are rare.

When used as an outcome measure, however, complications may have three inherent problems: first, their definition may be complex and subjective; second, their recording may have variations; finally, morbidity is influenced by the severity of the operation and the patients’ co-morbidities.

This study was a retrospective analysis performed on a prospectively compiled electronic database, in which the complications have been strictly defined before starting the database project. Therefore, the criteria of inclusion of the patients into each of these complications remained constant during the period taken into consideration for analysis.

Being a study performed at a single institution over a relatively short period of time, no variation in the recording of complications should have occurred.

Finally, in this study, a model for risk-adjustment was developed and validated to obviate the third problem.
and an extended resection [4,17] increased the risk of postoperative complications corroborate the conclusions of other authors who found disease and extended resection. Predicted postoperative morbidity was a low ppoFEV1, the presence of a concomitant cardiac disease and extended resection performed (Tables 3 and 4).

A logistic equation was derived for the prediction of the risk of morbidity. For instance, a hypothetical patient with a ppoFEV1 of 40% of predicted, with underlying cardiac co-morbidities, undergoing an extended resection, would have a predicted morbidity rate of 77.5%. In a patient with a ppoFEV1 of 80% of predicted, without concomitant cardiac disease and submitted to a non-extended resection, the risk of postoperative complications would be only 24.8%.

After the logistic model was validated by bootstrap resampling analysis, we used it to compare the observed and predicted morbidity rates in two successive periods of activity of our unit.

Despite the higher rate of observed morbidity in the late period (2000–2003) compared to the early period (1993–1999), which suggest a worsened performance, no differences were noted between observed and risk-adjusted morbidity rates in each of the two periods.

Therefore, in each of the two successive periods, the performance of our unit remained in agreement with the predicted outcomes. The use of the model revealed that the increased morbidity rate found in the late period was not a consequence of a poorer performance but a reflection of a more compromised physiological state of the patients at the time of operation and of a greater frequency of extended resections performed (Tables 3 and 4).

In spite of the fact that in the late period the observed and the predicted morbidity rates were higher than those in the early period, the mortality rates were not statistically different in the two periods (6.9 vs. 4.0%, respectively; \( P = 0.3 \), Fisher’s exact test). However, the analysis of mortality should be interpreted with caution, since small numbers precluded us analyzing risk-adjusted mortality rates.

Although the logistic model was well calibrated and reliable, the discrimination ability was moderate as shown by an area under the ROC curve of 0.67, which was, however, similar to that reported in other studies dealing with complications [6,18]. Presumably other factors may be added in the model to improve this parameter, such as ergometric variables [19] and multifactorial scoring systems [6,18,20].

The use of predictive models, such as the one developed in the present study, allows the discussion of single ‘surgical success’ cases as indicators of the quality of surgical care. Patients in a high-risk group, in whom no complications were observed, may reveal possible successful factors in their perioperative treatment, which may be generalized to future patients.

This model was developed for internal audit only and was not intended to be used for selecting individual patients for operation, a process which should be based mostly on clinical judgment [21]. Furthermore, its validity in other settings or its use for multiinstitutional quality assessment should be verified by other independent studies.

In conclusion, we showed that applying a model of risk-adjustment in elderly patients submitted to lung resection was useful for the internal evaluation of the quality of care, since it prevented misleading information derived by the comparison of the crude rates of morbidity.

**Acknowledgements**

This paper is dedicated to the memory of the late Prof. Nino Masera, who spent most of his life in the study and care of the surgical elderly patients.

**References**


