Institutional report - Pulmonary

Risk-adjusted morbidity, mortality and failure-to-rescue models for internal provider profiling after major lung resection

Alessandro Brunelli*, Francesco Xiume’, Majed Al Refai, Michele Salati, Rita Marasco, Armando Sabbatini

Unit of Thoracic Surgery, ‘Umberto I’ Regional Hospital, Via S. Margherita 23, Ancona, 60129 Italy

Received 19 August 2005; received in revised form 1 November 2005; accepted 5 December 2005

Abstract

This work was aimed at developing risk-adjusted outcome models for profiling the internal quality of care after major lung resection. One thousand and sixty-two patients submitted to lobectomy (845) or pneumonectomy (217) from 1994 through 2004 at our unit were analyzed. Risk-adjusted models of 30-day or in-hospital morbidity, mortality and failure-to-rescue (death/complication ratio) were developed by stepwise logistic regression analyses and validated by bootstrap procedures. The regression equations were then used to estimate the outcome risks in three successive periods of activity (early: 1994–1997; intermediate: 1998–June 2001; late: July 2001–2004). Observed and predicted morbidity, mortality and failure-to-rescue rates were compared within each period by the z-test. The following regression models were developed: Predicted morbidity: \[ \ln R = -2.1 + 0.035 \times \text{age} - 0.02 \times \text{FVC} + 0.6 \times \text{extended resection} + 0.7 \times \text{cardiac co-morbidity} \] (c-index = 0.68). Predicted mortality: \[ \ln R = -7.6 + 0.08 \times \text{age} - 0.04 \times \text{pooFEV1} + 1.6 \times \text{extended resection} + 1.2 \times \text{cardiac co-morbidity} + 1.1 \times \text{cerebrovascular co-morbidity} \] (c-index = 0.83). Predicted failure-to-rescue: \[ \ln R = -6.7 + 0.06 \times \text{age} + 1.5 \times \text{extended resection} + 1.2 \times \text{cerebrovascular co-morbidity} \] (c-index = 0.71). No differences were noted between observed and predicted outcome rates within each period, despite apparent unadjusted differences between periods. The use of risk-adjusted outcome models prevented misleading information derived from the unadjusted analysis of performance. We are currently using these models for internal quality-of-care audit purposes.

© 2006 Published by European Association for Cardio-Thoracic Surgery. All rights reserved.

Keywords: Lung cancer surgery; Outcomes; Morbidity; Mortality; Risk modelling; Quality of care

1. Introduction

Data about outcomes that include risk adjustment using detailed clinical variables from specialized databases provide the best measurement of quality of care.

For the purpose of internal single provider evaluation, newly-developed models derived from local subjects have been shown to perform better than ‘ready-made’ models derived from external databases [1–4].

Thus, the objective of the present analysis was to develop multiple risk-adjusted outcome models (morbidity, mortality and failure-to-rescue) to assess the performance of our general thoracic surgery unit during three successive periods of activity in patients submitted to major lung resections for non-small cell lung cancer (NSCLC) and to serve as a temporal benchmarking for future quality improvement activities.

2. Patients and methods

We analyzed all patients (1062) submitted to major lung resections for NSCLC (845 lobectomies or bilobectomies or 217 pneumonectomies, 879 males and 183 females) at our unit from January 1994 through December 2004. This is a retrospective analysis performed on a prospective, periodically audited, electronic database. A designated Clinical Audit Lead was responsible for the accuracy and consistency of data collection and recording. The study was approved by the institutional review board after the informed consent was obtained from each patient.

The same surgical team (compounded by 4 certified staff general thoracic surgeons) performed the following operations through a muscle sparing thoracotomy whenever possible: 235 right upper lobectomies; 232 left upper lobectomies; 138 left lower lobectomies; 104 right lower lobectomies; 41 middle lobectomies; 52 right lower bilobectomies; 45 right upper bilobectomies; 115 left pneumonectomies; 100 right pneumonectomies. Completion pneumonectomies (11 cases) were counted as pneumonectomies.

Perioperative treatment was standardized and focused on the control of post-thoracotomy chest pain, chest physio-
therapy and early as possible mobilization, antibiotic and anti-thrombotic prophylaxis.

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. From January 2000, symptom-limited stair climbing test was systematically used in our patients for risk stratification before operation. Operation was contraindi-
cated in those patients with a predicted postoperative forced expiratory volume in one second (ppoFEV₁) less than
30% of predicted [5], and with hemodynamic instability. Starting from 2000, even those patients with a ppoFEV₁,
lower than 30% but with a satisfactory exercise tolerance (height climbed at the stair climbing test above 12 m) were considered for operation [6].

Postoperative morbidity, mortality and failure-to-rescue (proportion of death among the complicated patients) were considered as those occurring within 30 days postoperative-
ly or during a longer period if the patient was still in the hospital.

A number of preoperative and operative variables were tested for possible association with outcome variables (see Appendix for explanation of variables). We started to systematically perform carbon monoxide diffusion lung capacity (DLCO), by using the single breath method, since January 2000. As only 474 patients (44.6%) in this series had this test performed and all of them were in the last 5 years of the study, DLCO was not utilized as a variable in the present analysis.

2.1. Statistical analysis

The entire database was initially used to develop the predictive logistic models. For each measure of outcome
(morbidity, mortality or failure-to-rescue rates), variables were initially screened by univariate analyses. The univariate comparisons of outcomes 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-square test or the Fisher’s exact test, whenever appropriate.

Variables with a \( P < 0.10 \) at univariate analysis were then used as independent variables in the stepwise logistic regression analyses. The presence or absence of one or more complications, of mortality or of death among com-
plicated patients (failure-to-rescue) were used as dependent
variables in each respective model. All data were at least 95% complete. Sporadic missing data were imputed by averaging the non-missing values (numerical variables) or taking the most frequent response category (categorical variables). To avoid multicollinearity, only one variable in a set of variables with a correlation coefficient greater than 0.5 was selected (by bootstrap procedure) and used in the regression model.

A \( P < 0.05 \) was selected for retention of variables in the final model. The c-index and the Hosmer-Lemeshow goodness-of-fit were used to study the discrimination and the calibration of the models, respectively. Furthermore, the multivariate procedures were validated by bootstrap bag-
ging with 1000 samples [7,8]. In the bootstrap procedure, repeated samples of 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 \( P < 0.1 \) at univariate analysis. The stability of the final stepwise model can be assessed by viewing the variables that enter most frequently in the repeated bootstrap mod-
els 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 models were then used to predict morbidity, mortality and failure-to-rescue rates in the patients operated on during 3 successive periods with approximately the same number of patients (‘early’: 1994–1997, 359 patients; ‘intermediate’: 1998–June 2001, 347 patients; ‘late’: July 2001–2004, 356 patients). Predicted and observed outcome rates in each period were then compared by means of the z-test for comparison of proportions.

All the statistical tests were two-tailed and a significance level of 0.05 was accepted. The analysis was performed by using the STATA 8.2 (Stata Corp., College Station, TX) statistical software.

3. Results

Total observed morbidity, mortality and failure-to-rescue rates were 21.8% (232 cases), 3.6% (37 cases) and 15.9%, respectively. Pulmonary complications occurred in 119 patients (11.2%) and cardiovascular complications occurred in 131 patients (12.3%).
Table 2
Results of the logistic regression analysis (dependent variable: mortality). Parsimonious model (1062 patients)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>S.E.</th>
<th>OR</th>
<th>95% C.L.</th>
<th>P-value</th>
<th>Bagging (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−7.6</td>
<td>1.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.08</td>
<td>0.03</td>
<td>1.08</td>
<td>1.02−1.1</td>
<td>0.004</td>
<td>68.2</td>
</tr>
<tr>
<td>ppoFEV, (%)</td>
<td>−0.04</td>
<td>0.01</td>
<td>0.96</td>
<td>0.94−0.99</td>
<td>0.003</td>
<td>74.5</td>
</tr>
<tr>
<td>Extended resection</td>
<td>1.6</td>
<td>0.35</td>
<td>4.8</td>
<td>2.4−9.7</td>
<td>&lt;0.0001</td>
<td>96.9</td>
</tr>
<tr>
<td>Cardiac co-morbidity</td>
<td>1.2</td>
<td>0.4</td>
<td>2.9</td>
<td>1.6−6.3</td>
<td>0.008</td>
<td>79.9</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>1.1</td>
<td>0.5</td>
<td>3</td>
<td>1.1−8.7</td>
<td>0.04</td>
<td>58.9</td>
</tr>
</tbody>
</table>

Coefficients for age and ppoFEV, are calculated for 1 unit change (1 year change or 1% change, respectively). Bagging %: frequency of significance (P<0.05) in 1000 bootstrap samples.

Tables 1, 2 and 3 show the results of the stepwise logistic regression analyses for each outcome variables (morbidity, mortality and failure-to-rescue), respectively.

The developed regression models, validated by bootstrap analyses, were the following:

Predicted morbidity: ln R/1−R = −2.1 + 0.035 × ppoFEV + 0.6 × extended resection + 0.7 × cardiac co-morbidity (R² 0.07, Hosmer-Lemeshow statistic 2.9 (P = 0.9), c-index 0.68).

Predicted mortality: ln R/1−R = −7.6 + 0.08 × age − 0.04 × ppoFEV + 1.6 × extended resection + 1.2 × cardiac co-morbidity + 1.1 × cerebrovascular co-morbidity (R² 0.19, Hosmer-Lemeshow statistic 6.3 (P = 0.6), c-index 0.83).

Predicted failure-to-rescue: ln R/1−R = −6.7 + 0.06 × age + 1.5 × extended resection + 1.2 × cerebrovascular co-morbidity (R² 0.10, Hosmer-Lemeshow statistic 6.6 (P = 0.6), c-index 0.71).

Extended resection, cardiac co-morbidity and cerebrovascular co-morbidity were coded as 0 for absence and 1 for presence of the variables, respectively.

Table 4 shows the comparison of baseline and operative characteristics between patients operated on in the 3 periods taken into consideration for the analysis.

Table 5 shows the results of the comparisons between the observed and the predicted outcome rates within each period. Despite apparent unadjusted differences between periods, the observed outcome rates were in line with the predicted ones within each period of activity.

4. Discussion

In managed care systems the analysis of the performance is of utmost importance for monitoring and improving the quality of care. However, raw outcome rates are misleading since they do not take into consideration the baseline and operative characteristics of the patients. Therefore, risk-adjustment, preferably using detailed clinical variables from specialized databases, should be used for clinical audit processes.

Recently, the European Association for Cardiothoracic Surgery and the European Society of Thoracic Surgeons have agreed the establishment of the European Cardiovascular and Thoracic Surgery Institute of Accreditation (www.ectsia.org) aiming to monitor and recognize good practice in cardiovascular and thoracic surgery. A satisfactory performance against the targets of risk-adjusted clinical performance indicators is one of the criteria for accreditation.

Moreover, for single provider profiling, outcome models derived from an internal database have been shown to perform better than models derived from external databases [1–4].

Therefore, the objective of this study was to develop different risk-adjusted outcome models (morbidity, mortality and failure-to-rescue) to evaluate the performance of our unit during 3 successive periods of time and to serve as temporal benchmarking for future quality improvement activity.

Most of the studies on predictive models in lung resection surgery have taken into consideration morbidity as outcome measure [9–13]. When used as an outcome measure, however, complications may have inherent problems: their definition may be complex and subjective and their recording may have variations.

This study was a single-institution 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, and no variation in the recording of complications has occurred.

Given its low rate after lung resection, mortality has been less frequently taken into consideration as an outcome

Table 3
Results of the logistic regression analysis (dependent variable: failure-to-rescue). Parsimonious model (235 patients)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>S.E.</th>
<th>OR</th>
<th>95% C.L.</th>
<th>P-value</th>
<th>Bagging (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−6.7</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.06</td>
<td>0.03</td>
<td>1.07</td>
<td>1.01−1.1</td>
<td>0.02</td>
<td>64.5</td>
</tr>
<tr>
<td>Extended resection</td>
<td>1.5</td>
<td>0.4</td>
<td>4.4</td>
<td>2.1−9.6</td>
<td>&lt;0.0001</td>
<td>93.2</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>1.2</td>
<td>0.6</td>
<td>3.4</td>
<td>1.1−11.8</td>
<td>0.049</td>
<td>51.1</td>
</tr>
</tbody>
</table>

Coefficients for age are calculated for 1 year change. Bagging %: frequency of significance (P<0.05) in 1000 bootstrap samples.
variable in risk-stratification modeling [12–14]. Recently, the EACTS/ESTS European Thoracic Surgery Database project generated a model of risk-adjusted in-hospital mortality for lung tumors (approximately 7% of the patients analyzed in this study were included in the European database study as well). This model was developed from 1694 patients (33 deaths) and included, as significant factors, age and ppoFEV [14]. Even though single-center and multi-center databases are difficult to compare, these two variables resulted in reliable predictors of mortality even in our mortality model.

Finally, failure-to-rescue has been studied in other surgical specialties [15], but there are no models for risk-adjusting this variable in lung resection patients. The use of more than one risk-adjusted outcome measures has been recommended in order to obtain a more complete and reliable analysis of the quality of care [15]. In fact, different outcome variables may measure different aspects of quality. The outcome variables selected for this study investigated only partial aspects of the quality of care. In the future, risk-modeling may be applied to other important endpoints such as technical complications (i.e. bronchopleural fistula, prolonged air leak, chylothorax, hemothorax, empyema, recurrent nerve injury, etc.), residual functional status, quality of life and long-term survival to allow a more accurate profiling of the centers or surgeons.

In this study, differences of unadjusted outcome rates suggested a poorer performance in the late period. However, observed outcome rates were in agreement with the predicted ones within each period of activity. Thus, the use of the risk-adjusted models revealed that the increased morbidity and mortality rates in the late period were not a consequence of a poorer performance, but a reflection of a more compromised physiological state of patients at the time of surgery and of a greater frequency of extended resection (Table 4).

Some important clinical variables, such as DLCO or ergonomic parameters, were not included in our models since they were systematically available only in the most recent years (<50% of patients in this series). In the future, the addition of these variables may contribute to improve the discrimination ability of our models. In fact, when the analysis was limited to the 474 patients with DLCO performed, ppoDLCO resulted significantly associated to morbidity and increased the discrimination of the model (c-index 0.73, R² 0.11), compared to a model without ppoDLCO developed from the same group of patients (c-index 0.68, R² 0.07).

The models in this study were developed for internal audit only and were not intended for selecting individual patients for operation, a process which should be based mostly on clinical judgment and should take into account other important factors such as DLCO and exercise tolerance. This work may serve as a methodological example for internal model building. However, the validity of our models in other settings or their use for multi-institutional quality assessment should be verified by other independent studies. For the purpose of setting international or nation-wide standards of care and for multi-institutional comparison of outcomes, the European Thoracic Database project [14] would represent the ideal instrument.

In conclusion, the use of multiple risk-adjusted outcome models prevented misleading information derived from the unadjusted analysis of performance. We are currently using these models for implementing a continuous quality improving program.
Appendix

Preoperative and operative variables: The following variables were tested for possible association with outcome: age, arterial oxygen tension (PaO₂), arterial carbon dioxide level (PaCO₂), smoking history (pack-years), preoperative hemoglobin concentration, serum albumin level, type of operation (lobectomy vs. 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, poFEV₁; residual volume to total lung capacity, RV/TLC ratio), cerebrovascular (stroke or transient ischemic attack) or symptomatic peripheral vascular co-morbidity, presence of concomitant chronic renal insufficiency (serum creatinine 2 mg/dl).

Pulmonary function tests were performed according to the American Thoracic Society criteria. Results of spirometry were collected after bronchodilator administration. FEV₁, FVC, poFEV₁ were expressed as percentages of predicted for age, sex and height. PoFEV₁ was calculated by the following formula: (preoperative FEV₁/number of preoperative functioning segments) X 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 poFEV₁ less than 50% of predicted and in all pneumonectomy candidates a quantitative perfusion lung scan was used, according to Marks et al. [5]. The simple calculation of poFEV₁ was previously shown to be as accurate as perfusion lung scanning [5].

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 (pleuralectomy resection), chest wall, mediastinal structures or diaphragm. In this series, 40% of the extended resections were extrapleural ones, 30% were associated with chest wall resection, 30% were associated with resection of the diaphragm, pericardium or other mediastinal structures.

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.

Outcome variables: For the purpose of this study, according to previous studies [11,12] and to the EACTS/ESTS thoracic surgery database [14], the following major complications were included: respiratory failure requiring mechanical ventilation for more than 48 h; pneumonia; atelectasis requiring bronchoscopy; adult respiratory distress syndrome (ARDS); pulmonary edema; pulmonary embolism; myocardial infarction; hemodynamically unstable arrhythmia requiring medical treatment; cardiac failure; acute renal failure; stroke. For the purpose of this study, which was an internal evaluation of the quality of care, only those complications that increased the complexity of the postoperative management were selected as outcome measure. The inclusion of minor morbidities and technical complications would have, in our view, flawed the model.

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 [15].

References