

The Thoracic Surgery Scoring System (Thoracoscore): Risk model for in-hospital death in 15,183 patients requiring thoracic surgery

Pierre Emmanuel Falcoz, MD, PhD,^a Massimo Conti, MD,^b Laurent Bouchet, MD,^c Sidney Chocron, MD, PhD,^a Marc Puyraveau, BSc,^d Mariette Mercier, MD, PhD,^e Joseph Philippe Etievent, MD,^a and Marcel Dahan, MD^c



Dr
Falcoz

Supplemental material is available online.

From the Department of Thoracic and Cardiovascular Surgery, Jean-Minjoz Hospital, Besançon,^a the Department of Thoracic Surgery, Calmette Hospital, Lille,^b the Department of Thoracic Surgery, Larrey Hospital, Toulouse,^c the Clinical and Biological Research Center, Saint-Jacques Hospital, Besançon,^d and the Department of Biostatistics and Epidemiology, Medical School, Besançon,^e France.

Read at the Eighty-sixth Annual Meeting of The American Association for Thoracic Surgery, Philadelphia, Pa, April 29–May 3, 2006.

Received for publication May 17, 2006; revisions received Aug 27, 2006; accepted for publication Sept 29, 2006.

Address for reprints: Pierre Emmanuel Falcoz, MD, Department of Thoracic and Cardiovascular Surgery, Hôpital Jean-Minjoz, Boulevard Fleming, 25000 Besançon, France (E-mail: pierre-emmanuel.falcoz@wanadoo.fr).

J Thorac Cardiovasc Surg 2007;133:325-32
0022-5223/\$32.00

Copyright © 2007 by The American Association for Thoracic Surgery

doi:10.1016/j.jtcvs.2006.09.020

Objective: This study was undertaken to determine factors associated with in-hospital mortality among patients after general thoracic surgery and to construct a risk model.

Methods: Data from a nationally representative thoracic surgery database were collected prospectively between June 2002 and July 2005. Logistic regression analysis was used to predict the risk of in-hospital death. A risk model was developed with a training set of data (two thirds of patients) and validated on an independent test set (one third of patients). Model fit was assessed by the Hosmer–Lemeshow test; predictive accuracy was assessed by the c-index.

Results: Of the 15,183 original patients, 338 (2.2%) died during the same hospital admission. Within the data used to develop the model, these factors were found to be significantly associated with the occurrence of in-hospital death in a multivariate analysis: age, sex, dyspnea score, American Society of Anesthesiologists score, performance status classification, priority of surgery, diagnosis group, procedure class, and comorbid disease. The model was reliable (Hosmer–Lemeshow test 3.22; $P = .92$) and accurate, with a c-index of 0.85 (95% confidence interval 0.83-0.87) for the training set and 0.86 (95% confidence interval 0.83-0.89) for the test set of data. The correlation between the expected and observed number of deaths was 0.99.

Conclusions: The validated multivariate model Thoracoscore, described in this report for risk of in-hospital death among adult patients after general thoracic surgery was developed with national data, uses only 9 variables, and has good performance characteristics. It appears to be a valid clinical tool for predicting the risk of death.

The decision to perform an operative procedure requires careful assessment of the potential risks and benefits involved. Risk is usually assessed by applying knowledge on the one hand from results of both surgical series published in the literature and registry data and on the other from the personal experience (clinical acuity) of the physician or physicians who are about to perform the procedure. A number of biases, however, may contribute to the difficulty of predicting the likelihood of an event.¹

Quantitative methods that discriminate factors associated with in-hospital mortality and the integration of this information by clinical prediction rules² may benefit both clinician and patient. Several scoring systems have been adapted for use among patients undergoing lung resection^{3,4} or have been developed either to stratify patients according to risk of complications⁵⁻⁹ or risk of in-hospital death after lung resection.¹⁰ To the best of our knowledge, however, there is currently no accepted general risk model for thoracic surgery that can be used to assess the risk of in-hospital death for thoracic surgical patients (esophageal operations excluded).

Abbreviations and Acronyms

| | |
|-----|---|
| ASA | = American Society of Anesthesiologists |
| CI | = confidence interval |
| OR | = odds ratio |
| PS | = performance status |

The manifest need for risk-adjusted outcome evaluation in general thoracic surgery led us to develop and validate a risk model derived from and tested on a nationally representative thoracic surgery database. Thus the aim of this study was twofold, to identify factors associated with in-hospital mortality among adult patients after general thoracic surgery and to construct a risk model that could be used prospectively to inform decisions (patient counseling concerning operative risk) and retrospectively to enable fair comparisons of outcomes (planning of postoperative advance care management, comparison of performances stratified by risk groups).

Materials and Methods**Data Collection and Processing**

Data were collected with a computer database (Epithor) developed for the French Society of Thoracic and Cardio-Vascular Surgery and coordinated by one of the authors (M.D.). The Epithor database is financially independent (not related to any surgical firm or tobacco manufacturer). Thoracic surgery units that applied to participate in this database were visited and validated by the coordinator and then were sent a confidential code. This enabled them to download the database application and instructions from a website (<http://www.epithor.net/>). The database application was password protected, allowing each unit to have multiple users with their own passwords.

Patient data were recorded exclusively by means of a hierarchic pull-down menu. Through the deliberate choice of the database programmer, most of the entries (such as in-hospital death) were mandatory. To ensure that records were analyzed independently of one another, only those representing each patient's main surgical procedure were used in the risk analysis. Analysis was restricted to adult patients (older than 16 years). Data were exported from within each unit's database with encryption, automatically attached to an e-mail, and sent to the central coordinating center to be put into a common database. Units could submit data whenever they wished and were sent a return e-mail certifying reception by the central database. Moreover, each participating unit had access to the national database on condition that they had sent in data within the previous 2 months.

Data were collected on the following variables: age, sex, body mass index, American Society of Anesthesiologists (ASA) score, dyspnea score (Medical Research Council), performance status (PS) classification (World Health Organization), comorbid diseases, reoperative surgery, preoperative forced expiratory volume in 1 second, side (left, right, or bilateral), surgical approach (open thoracotomy, video-assisted thoracoscopy, cervicotomy, or other), localization (lung, pleura, or mediastinum), diagnosis group (ma-

lignant or otherwise), procedure class (wedge resection, lobectomy, pneumonectomy, mediastinoscopy, or other diagnostic procedure), and priority of surgery. For patients with cancer, additional information was reported concerning pathologic staging, type of lymphadenectomy, type of histologic resection, and any adjuvant chemotherapy or radiotherapy received. The data set included information on postoperative complications and in-hospital mortality (deaths within 30 days and deaths within the same hospital admission). Status at discharge was used as the outcome measure of interest for the risk analysis; the dependent variable under study was in-hospital mortality.

Statistical Analyses

To determine independent predictive factors for in-hospital death with the dependent variable being binary, we first performed a univariate analysis with the Fisher exact test and then a multivariate analysis by logistic regression.¹¹ Variables with a level of significance less than or equal to .20 in the univariate analysis were included in the multivariate model, which was analyzed with a stepwise logistic regression. Interaction effects were sought for all variables included in the model. One feature of logistic regression analysis is that only cases with data for all of the variables considered are included in the analysis. For the purpose of the regression analysis, the variable age was divided into three groups (<55 years, 55-64 years, ≥ 65 years) and the variable comorbidity into three groups (0, 1-2, ≥ 3). The other variables were binary: sex (male vs female), ASA score (1-2 vs ≥ 3), dyspnea score (0-2 vs ≥ 3), PS class (1-2 vs ≥ 3), localization (lung or pleura vs mediastinum), diagnosis group (malignant vs otherwise), procedure class (pneumonectomy vs other), and priority of surgery (elective vs urgent or emergency).

Two thirds of the records were randomly selected to contribute to model development (training set). The remaining third (test set) were reserved for model testing on data other than those from which it was developed.

Model discrimination was assessed by the c-index, which is identical to the area under the receiver operating characteristic curve.¹² Calibration was assessed by the Hosmer-Lemeshow goodness-of-fit statistic.¹³ In addition, on the basis of the final model, we determined risk groups, whose choice of threshold with regard to in-hospital mortality was made according to the most clinically relevant cut point values. Observed and expected numbers of deaths by group of predicted risk were calculated and compared.

Discrete variables are expressed as counts with percentages and continuous variables as mean and range, unless otherwise stated. All statistical analyses were performed with SAS software, version 8.02 (SAS Institute, Inc, Cary, NC).

Results

The data reported in this study were collected prospectively from 59 French hospitals between June 1, 2002, and July 1, 2005. The data represent 22,127 patients (corresponding to 25,313 surgical procedures). Because data entry was optional, a 12.2% of missing values existed in the database. To enhance the robustness of results, we decided to include only patients with more than 95% of completed data in the final analysis. After discarding questions with too many

inconsistent or missing values and patients with unknown information on variables otherwise suitable for study, a group of 15,183 patients who were older than 16 years and undergoing a first thoracic surgical procedure was selected for further analysis. From these 15,183 records, 10,122 were randomly selected to be the training set for model development and 5061 were used to test the model. Random selection produced no significant differences between these two groups (data not shown).

There were 338 deaths (2.2%, 95% confidence interval [CI] 2.0%-2.5%), 218 in the training set and 120 in the test set. The mean age of the patients was 54.7 ± 17.4 years (\pm SD). The baseline patient characteristics in the training set and the in-hospital mortalities for the different variables are given in Table 1.

Development of the Risk Model

Predictors of in-hospital death were assessed by calculating odds ratios (ORs) and χ^2 tests. Multivariate logistic regression analysis results are summarized in Table 2. Age was significantly and positively associated with the risk of in-hospital death. The risk for an individual at least 65 years old was almost 3 times greater (OR 2.738, $P < .0001$) than for those younger than 55 years. Male sex was a significant predictor of in-hospital death (OR 1.569, $P < .0141$). High ASA score, high PS classification, and high dyspnea score were also associated with higher in-hospital mortalities: OR 1.833 ($P < .0023$) for ASA score of at least 3, OR 1.992 ($P < .0008$) for PS classification of at least 3, and OR 2.478 ($P < .0001$) for dyspnea score of at least 3. Priority of surgery influenced the risk of in-hospital death. Urgent or emergency surgery was associated with a greater risk than was elective surgery ($P < .0001$). This risk was almost 2.5 times (OR 2.326) that associated with elective surgery. Procedure class was associated with in-hospital death ($P < .0001$). Patients who had undergone a pneumonectomy had a mortality approximately 3.5 times as high those who had undergone other thoracic surgery procedures (OR 3.379). Concerning diagnosis group, patients with a malignant pathology had approximately 3.5 times the risk (OR 3.464, $P < .0001$) of in-hospital death of patients with a benign pathology. Comorbidity was associated with increased mortality. Patients with a score of at least 3 had nearly 2.5 times the risk (OR 2.476, $P < .0003$) of in-hospital death of patients with a score of 0. In these data, 94.6% of the comorbidity was related to 10 major diagnoses: smoking addiction (27.1%), history of cancer (11.8%), chronic obstructive pulmonary disease (11.5%), arterial hypertension (10.5%), heart disease (10.4%), diabetes mellitus (7.6%), peripheral vascular disease (6.1%), obesity (4.9%), and alcoholism (4.7%). No other diagnoses, with the exception of hyperlipidemia (2.9%), were mentioned for more than 1.0% of patients.

Parameterization of the variables and the calculation of predicted risk are detailed in Table 3. Coefficients from the logistic regression equation were used to calculate predicted probability of in-hospital death for an individual patient. The dependent variable was in-hospital mortality. The regression model significantly predicted the occurrence of in-hospital death in the training data set ($\chi^2_{11} = 73.1$, $P < .0001$).

Validation of the Risk Model

Receiver operating characteristic curves for the multivariate risk prediction were calculated. The performance of the prediction model for the training set (the data from which it was derived) and its ability to predict in-hospital mortality for the test set were compared. The c-indices were 0.85 (95% CI 0.83-0.87) for the training set and 0.86 (95% CI 0.83-0.89) for the test set. In addition, to study the performance of the prediction model on subgroups of patients, the data were stratified by age, sex, ASA score, PS classification, dyspnea score, priority of surgery, procedure class, diagnosis group, and comorbidity; the respective c-indices were then calculated (Table 4). For the 6% of patients undergoing pneumonectomy, the c-index was 0.65 with a large CI; otherwise, c-index was consistent, ranging from 0.74 to 0.92, which indicates that the prediction model performs well among most subgroups of patients.

The Hosmer–Lemeshow goodness-of-fit statistic, a comparison of observed and expected deaths by decile of risk, was not statistically significant (Hosmer–Lemeshow 3.22; $P = .92$), indicating a nearly perfect fit. Table 5 displays observed versus expected mortalities by incremental risk group, along with the exact binomial CIs for the observed mortalities. The observed death rates in the test set were compared with the predicted death rates by category of risk: the correlation of observed to expected mortality was 0.99.

Table 6 is a tabulation of the characteristics of the patients in the four risk groups. It reveals that the percentage (or range) under consideration increased steadily and consistently across the risk groups irrespective of the variable considered, save for two exceptions. First, the percentage of urgent or emergency cases in the risk group less than 1% is superior to those of risk groups 1% to 3% and 3% to 7%. Second, the percentage of patients with lung cancer was superior in the risk group 1% to 3% to that of the risk group 3% to 7%.

Discussion

In this national, prospective study, a multivariate clinical model of risk was developed and validated. This mathematical model is a highly significant predictor of the dependent variable, in-hospital mortality. This statistical method calculates the independent OR associated with each of the variables and allows combination of this information in a

TABLE 1. Baseline patient characteristics in the training set (n = 10,122)

| Variable | All patients (n = 10,122) | Alive (n = 9904) | In-hospital deaths (n = 218) | P value* |
|---|------------------------------|---------------------|---------------------------------|----------|
| Age (y) | | | | <.0001 |
| <55 | 4541 (44.9%) | 4508 (99.3%) | 33 (0.7%) | |
| 55-65 | 2342 (23.1%) | 2285 (97.6%) | 57 (2.4%) | |
| ≥65 | 3239 (32.0%) | 3111 (96.0%) | 128 (4.0%) | |
| Sex | | | | .003 |
| Male | 6932 (68.5%) | 6763 (97.6%) | 169 (2.4%) | |
| Female | 3190 (31.5%) | 3141 (98.5%) | 49 (1.5%) | |
| Reoperative surgery† | | | | .007 |
| Yes | 215 (2.1%) | 204 (94.9%) | 11 (5.1%) | |
| No | 9903 (97.9%) | 9696 (97.9%) | 207 (2.1%) | |
| American Society of Anesthesiologists score† | | | | <.0001 |
| ≤2 | 6879 (71.5%) | 6813 (99.0%) | 66 (1.0%) | |
| ≥3 | 2738 (28.5%) | 2595 (94.8%) | 143 (5.2%) | |
| Performance status† | | | | <.0001 |
| ≤2 | 7815 (81.9%) | 7723 (98.8%) | 92 (1.2%) | |
| ≥3 | 1722 (18.1%) | 1605 (93.2%) | 117 (6.8%) | |
| Dyspnea score | | | | <.0001 |
| ≤2 | 9054 (89.5%) | 8923 (98.5%) | 131 (1.5%) | |
| ≥3 | 1068 (10.5%) | 981 (91.8%) | 87 (8.2%) | |
| Body mass index (kg/m ²) | | | | .3267 |
| <30 | 9162 (90.5%) | 8961 (97.8%) | 201 (2.2%) | |
| ≥30 | 960 (9.5%) | 943 (98.2%) | 17 (1.8%) | |
| Preoperative forced expiratory volume in 1 s† | | | | <.0001 |
| <50% | 1279 (13.4%) | 1216 (95.1%) | 63 (4.9%) | |
| ≥50% | 8282 (86.6%) | 8102 (97.8%) | 180 (2.2%) | |
| Priority of surgery | | | | .0002 |
| Urgent or emergency | 1582 (15.6%) | 1527 (96.5%) | 55 (3.5%) | |
| Elective | 8540 (84.4%) | 8377 (98.1%) | 163 (1.9%) | |
| Localization | | | | <.0001 |
| Lung | 4797 (47.4%) | 4682 (97.6%) | 115 (2.4%) | |
| Pleura | 3188 (31.5%) | 3106 (97.4%) | 82 (2.6%) | |
| Mediastinum | 2137 (21.1%) | 2116 (99.0%) | 21 (1.0%) | |
| Procedure class | | | | <.0001 |
| Mediastinoscopy or other mediastinal surgery | 2642 (26.1%) | 2605 (98.6%) | 37 (1.4%) | |
| Wedge resection | 4389 (43.4%) | 4307 (98.1%) | 82 (1.9%) | |
| Lobectomy or bilobectomy | 2484 (24.5%) | 2429 (97.8%) | 55 (2.2%) | |
| Pneumonectomy | 607 (6.0%) | 563 (92.7%) | 44 (7.3%) | |
| Diagnosis group | | | | <.0001 |
| Benign | 4339 (42.9%) | 4302 (99.1%) | 37 (0.9%) | |
| Malignant | 5783 (57.1%) | 5602 (96.9%) | 181 (3.1%) | |
| Surgical approach | | | | .0107 |
| Thoracotomy | 6357 (62.8%) | 6211 (97.7%) | 146 (2.3%) | |
| Video-assisted thoracoscopy | 1792 (17.7%) | 1761 (98.3%) | 31 (1.7%) | |
| Cervicotomy or other | 1973 (19.5%) | 1902 (96.4%) | 71 (3.6%) | |
| Comorbidities | | | | <.0001 |
| None | 3333 (32.9%) | 3307 (99.2%) | 26 (0.8%) | |
| 1-2 | 4852 (47.9%) | 4747 (97.8%) | 105 (2.2%) | |
| ≥3 | 1937 (19.2%) | 1850 (95.5%) | 87 (4.5%) | |
| Preoperative treatment | | | | .0137 |
| Yes | 1272 (12.6%) | 1232 (96.9%) | 40 (3.14%) | |
| No | 8850 (87.4%) | 8672 (98.0%) | 178 (2.0%) | |

Data are expressed as numbers and percentages of patients. Because of rounding, percentages may not sum to 100. *Comparison of patients dying in the hospital versus patients staying alive. †Indicates variables with missing value. Because of missing data, the regression analysis included 9035 patients.

TABLE 2. Determinants of in-hospital mortality for adult patients after thoracic surgery

| Variable | Odds ratio | 95% Confidence interval | P value |
|---|------------|-------------------------|---------|
| Age (y) | | | |
| <55 | 1 | | |
| 55-65 | 2.155 | 1.352-3.435 | .0012 |
| ≥65 | 2.738 | 1.782-4.206 | <.0001 |
| Sex | | | |
| Female | 1 | | |
| Male | 1.569 | 1.095-2.248 | .0141 |
| American Society of Anesthesiologists score | | | |
| ≤2 | 1 | | |
| ≥3 | 1.833 | 1.241-2.707 | .0023 |
| Performance status classification | | | |
| ≤2 | 1 | | |
| ≥3 | 1.992 | 1.333-2.976 | .0008 |
| Dyspnea score | | | |
| ≤2 | 1 | | |
| ≥3 | 2.478 | 1.680-3.655 | <.0001 |
| Priority of surgery | | | |
| Elective | 1 | | |
| Urgent or emergency | 2.326 | 1.533-3.530 | <.0001 |
| Procedure class | | | |
| Other | 1 | | |
| Pneumonectomy | 3.379 | 2.277-5.016 | <.0001 |
| Diagnosis group | | | |
| Benign | 1 | | |
| Malignant | 3.464 | 2.270-5.285 | <.0001 |
| Comorbidity score | | | |
| 0 | 1 | | |
| ≤2 | 2.106 | 1.319-3.361 | .0018 |
| ≥3 | 2.476 | 1.513-4.050 | .0003 |

No interaction effects were found to be significant for the variables included in the model.

quantitative manner, the predicted probability of in-hospital death. The model uses patient and clinical data that are available before surgery. The model of risk had good performance characteristics (c-index 0.85), and performed equally well when used to predict outcomes on a test set of data and within clinically relevant subgroups of patients.

Two limitations do, however, need to be mentioned. First, this model deals only with in-hospital mortality, which is an imperfect surrogate for the risk of death attributable to surgery and only one factor in decision making. Second, unmeasured variables are likely to contribute to imprecision in this prediction model, which describes risk with only 9 variables. The magnitude of this effect is difficult to determine. An indication that the effect is small in these data is the relatively small remaining area under the receiver operating characteristic curve.

TABLE 3. Prediction of risk of in-hospital mortality

| Variable | Value | Code | β coefficient |
|---|---------------------|------|---------------|
| Age (y) | <55 | 0 | |
| | 55-65 | 1 | 0.7679 |
| | ≥65 | 2 | 1.0073 |
| Sex | Female | 0 | |
| | Male | 1 | 0.4505 |
| American Society of Anesthesiologists score | ≤2 | 0 | |
| | ≥3 | 1 | 0.6057 |
| Performance status classification | ≤2 | 0 | |
| | ≥3 | 1 | 0.689 |
| Dyspnea score | ≤2 | 0 | |
| | ≥3 | 1 | 0.9075 |
| Priority of surgery | Elective | 0 | |
| | Urgent or emergency | 1 | 0.8443 |
| Procedure class | Other | 0 | |
| | Pneumonectomy | 1 | 1.2176 |
| Diagnosis group | Benign | 0 | |
| | Malignant | 1 | 1.2423 |
| Comorbidity score | 0 | 0 | |
| | ≤2 | 1 | 0.7447 |
| | ≥3 | 2 | 0.9065 |
| Constant | — | | -7.3737 |

Methods for using the logistic regression model to predict the risk of in-hospital death: (1) The odds are calculated with the patient values, and the coefficients are determined from the regression equation: Odds = $\exp[-7.3737 + (0.7679 \text{ if code of age is 1 or } 1.0073 \text{ if code of age is 2}) + (0.4505 \times \text{sex score}) + (0.6057 \times \text{American Society of Anesthesiologists score}) + (0.6890 \times \text{performance status classification}) + (0.9075 \times \text{dyspnea score}) + (0.8443 \times \text{code for priority of surgery}) + (1.2176 \times \text{procedure class}) + (1.2423 \times \text{diagnosis group}) + (0.7447 \text{ if code of comorbidity is 1 or } 0.9065 \text{ if code of comorbidity is 2})$. (2) The odds for the predicted probability of in-hospital death are calculated: Probability = odds/(1 + odds).

Despite these limitations, our analysis found 9 variables significantly associated with in-hospital mortality. This finding is in accordance with the publications of Harpole and colleagues¹⁴ and Berrisford and associates,¹⁰ who modeled the risk of in-hospital death after lung resection in the frameworks of the multi-institutional National Veterans Affairs Surgical Quality Improvement Program and the European Thoracic Surgery Database Project, respectively. It is important to note that Harpole and colleagues' study¹⁴ contained laboratory values that were not included in the Epithor database. However, 3 predictors were systematically identified in our study and in those of Harpole and colleagues¹⁴ and Berrisford and associates¹⁰: age, procedure group, and dyspnea. Moreover, 1 additional predictor, ASA score, was identical between the study of Berrisford and associates¹⁰ and our own study, and 2 other predictors were significant in our study but not in that of Berrisford and associates¹⁰: sex and diagnosis group (benign vs malignant).

TABLE 4. c-Index by patient and disease characteristics in the training set (n = 10,122)

| Subgroup | % | c-Index | 95% Confidence interval |
|---|-------|---------|-------------------------|
| Total | 100% | 0.85 | 0.83-0.87 |
| Age (y) | | | |
| <55 | 44.9% | 0.90 | 0.85-0.95 |
| 55-65 | 23.1% | 0.78 | 0.73-0.84 |
| ≥65 | 32.0% | 0.79 | 0.74-0.83 |
| Sex | | | |
| Female | 28.5% | 0.89 | 0.84-0.94 |
| Male | 71.5% | 0.83 | 0.80-0.85 |
| American Society of Anesthesiologists score | | | |
| ≤2 | 71.5% | 0.82 | 0.78-0.86 |
| ≥3 | 28.5% | 0.76 | 0.72-0.80 |
| Performance status classification | | | |
| ≤2 | 81.9% | 0.82 | 0.79-0.86 |
| ≥3 | 18.1% | 0.74 | 0.70-0.79 |
| Dyspnea score | | | |
| ≤2 | 89.5% | 0.82 | 0.80-0.86 |
| ≥3 | 10.5% | 0.74 | 0.69-0.79 |
| Priority of surgery | | | |
| Elective | 84.4% | 0.83 | 0.80-0.85 |
| Urgent or emergency | 15.6% | 0.89 | 0.86-0.93 |
| Procedure class | | | |
| Other | 93.8% | 0.86 | 0.83-0.88 |
| Pneumonectomy | 6.2% | 0.65 | 0.55-0.75 |
| Diagnosis group | | | |
| Benign | 42.9% | 0.92 | 0.88-0.97 |
| Malignant | 57.1% | 0.78 | 0.75-0.81 |
| Comorbidity score | | | |
| 0 | 32.9% | 0.92 | 0.89-0.96 |
| ≤2 | 47.9% | 0.83 | 0.80-0.86 |
| ≥3 | 19.2% | 0.75 | 0.70-0.80 |

It is especially noteworthy that neither the localization of the procedure (mediastinum, lung, pleura) nor the preoperative forced expiratory volume in 1 second were taken into account in the final model. One finding that might be somewhat disturbing to readers is that 17.9% of urgent or emergency operations belong to risk group 1, as shown in Table 6. This counterintuitive finding is due simply to the fact that patients with spontaneous pneumothorax are considered to need urgent operation, yet most have a very low operative risk. Moreover, with respect to patients with a malignant disease, further detailed studies with very large sample sizes are needed to elucidate reasons for the nonlinear increase in percentage of in-hospital mortality.

The reliability of any data collection remains the key issue when the information is used for further outcome analysis and comparison. Data must therefore be complete

TABLE 5. Observed versus expected in-hospital mortality by predicted risk category among those in the test set

| Risk group | Expected mortality (%) | Observed mortality (%) | 95% Confidence interval |
|------------|------------------------|------------------------|-------------------------|
| <1% | 0.2 | 0.29 | 0.15-0.43 |
| 1%-3% | 2.47 | 2.12 | 1.56-2.68 |
| 3%-7% | 4.84 | 4.73 | 3.46-5.99 |
| ≥7% | 13.2 | 15.60 | 12.76-18.43 |

and accurate. Administrative registries, such as those used by Medicare for billing, provide more exhaustive data than do clinical registries. However, Mack and associates¹⁵ recently demonstrated that models developed from administrative database could have lessened medical relevance because of the absence of many clinical variables. In France, all hospitalized patients are assigned to a specific diagnosis related group on the basis of diagnosis and the procedures performed. The Epithor database currently represents 66% of all diagnosis related groups concerning thoracic surgery performed in France each year.¹⁶ Consequently, this database could be considered representative of French thoracic surgical activity. Within the 48 hospitals participating in the Epithor network, 33 are public (20 of which are university hospitals) and 15 are private. In all, 38 hospitals have entered more than 100 patients into the database and 25 have entered more than 500 patients. Participating centers were free to choose the patients and the procedures they entered. Minor procedures were generally entered less than major ones, thus allowing more physician time to enter data for the major procedures.

In the Epithor database, the completeness and accuracy of each center's data were neither audited nor validated at this time. Nevertheless, recent studies with registries coordinated by national and international medical societies have demonstrated that there are no major differences in what can be assessed from controlled data and uncontrolled data. Herbert and colleagues,¹⁷ in a detailed audit of a clinical outcomes registry database, showed that the major fields within this specific database, including operative mortality, major complications, and the significant factors in a risk algorithm, were highly accurate. Maruszewski and associates,¹⁸ in an attempt at data verification in the EACTS Congenital Database, showed no statistically significant differences between verified and unverified data for 30-day mortality. In the Epithor database, 6944 of the 22,127 initial patient records collected were excluded for incomplete data; hence 15,183 patients were included in the final analysis. The percentages of in-hospital mortality, however, were nearly identical (2.2%, 95% CI 2.0%-2.5% for the 15,183 patients in the final analysis and 2.0%, 95% CI 1.8%-2.2% for the 22,127 initial patients), and there was no statistical difference in the baseline patients characteristics between

TABLE 6. Characteristics of the risk groups in the training set of patients (n = 10,122)

| | Risk group | | | | All groups |
|--|------------|-----------|-----------|-----------|------------|
| | <1% | 1%-3% | 3%-7% | ≥7% | |
| No. of patients | 5883 | 2530 | 1080 | 629 | 10,122 |
| Age (y) | | | | | |
| Mean | 47 | 63.8 | 66.3 | 68.8 | 54.6 |
| Range | 16.2-94.8 | 18.0-95.8 | 18.5-95.8 | 19.9-96.1 | 16.2-96.1 |
| Male (%) | 61.9 | 74.6 | 82.1 | 81.9 | 68.5 |
| American Society of Anesthesiologists ≥3 (%) | 10.2 | 31.3 | 76.2 | 91.3 | 28.5 |
| Performance status ≥3 (%) | 4.3 | 16.6 | 50.0 | 85.2 | 18.1 |
| Dyspnea ≥3 (%) | 3.1 | 7.5 | 26.5 | 65.2 | 10.6 |
| Urgent or emergency (%) | 17.9 | 6.6 | 14.4 | 33.2 | 15.7 |
| Pneumonectomy (%) | 1.3 | 5.7 | 21.6 | 23.9 | 6.0 |
| Cancer (%) | 36.7 | 86.2 | 82.6 | 88.2 | 57.1 |
| Comorbidity | | | | | |
| Mean | 0.84 | 1.6 | 1.74 | 2.2 | 0-3 |
| Range | 0-3 | 0-3 | 0-3 | 1.2 | 0-3 |

these two populations (data not shown). Thus there was no selection bias. In a subsequent version of Epithor, a quality audit currently under experimentation in pilot centers corresponding to an automatic feedback system for incomplete data entries and based on Fine and colleagues' work¹⁹ on validation and feedback in the UK Cardiac Surgery Experience, will be implemented to enhance the quality (completeness and accuracy) of each center's data.

In conclusion, the validated multivariate model for risk of in-hospital death among adult patients after general thoracic surgery (Thoracoscore) described in this report was developed with national data, uses only 9 variables, and has good performance characteristics. It appears to be a valid clinical tool for predicting the risk of death. It would therefore be useful both for calculating the death risk of an individual patient and for contrasting expected and observed mortalities for an institution or independent clinician. As ever, caution is required in interpreting the prediction of a risk model in the case of an individual patient.

We thank Prof Pascal Thomas for his invaluable comments and Nancy Richardson-Peuteuil for her editorial assistance. We are grateful to all the French thoracic surgeons who participated in this study for their essential collaboration: Dr J. F. Andro (Quimper), Dr M. Aubert (Saint Martin d'Hères), Prof J. Azorin (Bobigny), Dr J. Bellamy (Saint Cloud), Dr F. Bellenet (Cergy Pontoise), Prof A. Bernard (Dijon), Dr P. Boitet (Le Havre), Prof J. Borrelly (Maxéville), Prof P. Y. Brichon (Grenoble), Dr G. Cardot (Boulogne sur Mer), Dr T. Charpentier (Marseille), Dr F. Clement (Besançon), Prof P. Corbi (Poitiers), Prof M. Dahan (Toulouse), Dr C. Delaisement (Agen), Dr E. De Laroche (Lyon), Dr A. Dujon (Bois Guillaume), Prof P. Dumont (Tours), Prof G. Escande (Clermont Ferrand), Prof J. P. Gamondes (Lyon), Prof R. Giudicelli (Marseille), Dr D. Gossot (Paris), Prof G. Grosdidier (Nancy), Dr B. Guibert (Pierre Bénite), Dr O. Hagry (Chalon sur Saône), Dr S.

Jaillard (Lille), Dr J. M. Jarry (Aix en Profovenne), Dr Y. Laborde (Pau), Dr D. Lefant (La Rochelle), Dr L. Lombart (Béziers), Dr F. Mazeres (Bayonne), Dr E. Mensier (Marcq en Baroeul), Dr D. Metois (Orléans), Dr M. Monteau (Reims), Prof J. Mouroux (Nice), Dr P. Mulsant (Lyon), Prof C. Peillon (Rouen), Prof F. Pons (Clamart), Dr E. Pouliquen (Ploemeur), Prof M. Riquet (Paris), Dr N. Santelmo (Avignon), Dr G. Seban (Annecy), Dr B. Sevray (Vannes), Dr O. Tiffet (Saint Etienne), Prof J. F. Velly (Pessac), Dr B. Wack (Metz), Prof G. Massard (Strasbourg), and Prof A. Wurtz (Lille).

References

1. Taversky A, Kahnemann D. Judgement under uncertainty: heuristics and biases. *Science*. 1974;185:1124-31.
2. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical predictions rules: application and methodological standards. *N Engl J Med*. 1985;313:793-9.
3. Brunelli A, Fianchini A, Gesuita R, Carle F. POSSUM scoring system as an instrument of audit in lung resection surgery. *Ann Thorac Surg*. 1999;67:329-31.
4. Birim O, Maat AP, Kappentein AP, Van Meerbeeck JP, Damhuis RA, Bogers AJ. Validation of the Charlson comorbidity index in patients with operated non-small cell lung cancer. *Eur J Cardiothorac Surg*. 2003;23:30-4.
5. Epstein SK, Faling SJ, Daly BD, Celli BR. Predicting complications after pulmonary resection. Preoperative exercise testing vs a multifactorial cardiopulmonary risk index. *Chest*. 1993;104:694-700.
6. Izbicki JR, Knoefel WT, Passlick B, Haberkost M, Karg O, Thetter O. Risk analysis and long-term survival in patients undergoing extended resection of locally advanced lung cancer. *J Thorac Cardiovasc Surg*. 1995;110:386-95.
7. 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;150:947-55.
8. Ferguson MK, Durkin AE. A comparison of three scoring systems for predicting complications after major lung resection. *Eur J Cardiothorac Surg*. 2003;23:35-42.

9. Melendez JA, Barrera R. Predictive respiratory complication quotient predicts pulmonary complications in thoracic surgical patients. *Ann Thorac Surg.* 1998;66:220-4.
10. Berrisford R, Brunelli A, Rocco G, Treasure T, Utlei M. The European Thoracic Surgery Database project: modeling the risk of in-hospital death following lung resection. *Eur J Cardiothorac Surg.* 2005;28:306-11.
11. Hosmer DW Jr, Lemeshow S. Applied logistic regression. 2nd ed. New York: Wiley-Interscience; 2000. p. 1-373.
12. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143:29-36.
13. Hosmer DW Jr, Lemeshow S. A goodness-of-fit test for the multiple regression model. *Commun Statistics.* 1980;A10:1043-69.
14. Harpole DH Jr, DeCamp MM Jr, Daley J, Hur K, Oprian CA, Henderson WG, et al. Prognostic models of thirty-day mortality and morbidity after major pulmonary resection. *J Thorac Cardiovasc Surg.* 1999;117:969-79.
15. Mack MJ, Herbert M, Prince S, Dewey TM, Magee MJ, Edgerton JR. Does reporting of coronary artery bypass grafting from administrative databases accurately reflect actual clinical outcomes? *J Thorac Cardiovasc Surg.* 2005;129:1309-17.
16. Agence Technique de l'Information sur l'Hospitalisation. French official website of the Healthcare Program Information Systems [cited 2006 Mar 27]. Available from: URL:<http://www.stats.atih.sante.fr/mco/statone.php>.
17. Herbert MA, Prince SL, Williams JL, Magee MJ, Mack MJ. Are unaudited records from an outcomes registry database accurate? *Ann Thorac Surg.* 2004;77:1960-5.
18. Maruszewski B, Lacour-Gayet F, Monro JL, Keogh BE, Tobota Z, Kansy A. An attempt at data verification in the EACTS Congenital Database. *Eur J Cardiothorac Surg.* 2005;28:400-6.
19. Fine LG, Keogh BE, Cretin S, Orlando M, Gould MM. How to evaluate and improve the quality and credibility of an outcomes database: validation and feedback study on the UK Cardiac Surgery Experience. *BMJ.* 2003;326:25-8.

Discussion

Dr David H. Harpole (Durham, NC). I have no conflicts to disclose. I congratulate Falcoz and coworkers on an excellent presentation and article. These data were collected with a computer database developed by the French Society of Thoracic and Cardiovascular Surgery and coordinated by Dr Dahan in Toulouse, France. It is noted in the article that the Epithor database is financially independent and not related to any surgical or tobacco firm. It includes 40-plus institutions of all types in France and is a good cross-section of thoracic surgery in France during the study interval, in fact including two thirds of all thoracic cases performed in France during that time. The statistical methods were sound,

with a 10,000-patient training set and a 5000-patient test set, both of which were random and allowed excellent model fit coefficients. The data included more than 22,000 patients from 2002 to 2005, a recent interval. My first question concerns the amount and quality of the data.

First, more than 12% of the patients had missing data. So ultimately the analysis was only on 15,000 patients, although this is still a very large data set. What are your plans to improve this? Are you going to include site audits, better data manager training, or another method for data entry in your computerized model?

Second, although good, your hospital data were only on mortality at 30 days. Do you plan, because you are collecting morbidity data, to create morbidity models? Are there plans for long-term analyses in your cancer population, which are our ultimate goals for risk adjustment?

Finally, what are your future plans with this excellent database? Do you plan to analyze these data with other databases in Europe, and do you plan to increase the scope of the data that you are collecting, including other types of surgical procedures?

Dr Falcoz. Thank you, Dr Harpole, for the important points you raised. First, concerning the missing values, in Epithor, patient data entry was recorded exclusively by means of a pull-down menu. By deliberate choice, an initial deliberate choice of the database programmer, most of the entries are mandatory, such as in-hospital mortality. Unfortunately, however, not all data entries are mandatory, and this explains this large numbers of missing values. We thought that it was better to have a large database with lots of patient files, even if incomplete, than fewer patient files.

The second question was about the morbidity. I think dealing with mortality, even just in-hospital mortality, was difficult, and to add morbidity may be a much more difficult task. So for now we only want to make a predictive score for in-hospital mortality.

This brings me to your question about our future plans. We have validated the Thoracoscore in France by the methodology we used, but the next plan will be to validate the Thoracoscore in other countries, including North America and the rest of Europe.

Concerning your point about the patient with cancer, it is in our plan to make a specific score for primary lung surgery to individually assess these patients. Most of the important factors in the multivariate analysis are those dealing with cancer, such as the procedure class (pneumonectomy or other) and the diagnosis group (malignant or other).

Appendix E1: Glossary of Statistical Terms

c-Index.¹² The c-index is identical to the area under the receiver operating characteristic curve for binary outcomes. The area measures discrimination, that is, the ability of the test to correctly classify those with and without the event. Its value represents the probability, from 0 to 1, that a randomly selected dead patient will have a higher calculated risk of dying than a randomly selected surviving patient.

Hosmer–Lemeshow statistic.¹³ The Hosmer–Lemeshow statistic is used to test the calibration of a model. It groups the ordered predictions from the logistic regression model into deciles and produces a Pearson-like statistic that has a χ^2 distribution with 8 *df*. The larger the *P* value, the better the reliability.

Logistic regression.¹¹ Logistic regression is a standard method of data analysis concerned with describing the relationship between a binomial response variable and one or more explanatory variables. The model can predict the probability of occurrence of a binary outcome, in our case, the probability of in-hospital death.

Odds ratio.¹¹ The odds (*O*) of an event (eg, death) is related to the probability (*P*) of the event by the following equation: $O = P/(1 - P)$. When *P* is small, they are approximately equal. Logistic regression produces an odds ratio as a measure of the strength of a risk factor. The odds ratio is the odds of the event with the risk factor present divided by the odds of the event with the factor absent. The risk factor is considered predictive when the odds ratio confidence interval inferior limit is more than 1.

Receiver operating characteristic curves.¹² These graphical techniques are used to assess the accuracy of diagnosis systems. They consist in plotting true-positive rates (sensitivity) versus false-positive rates ($1 - \text{specificity}$) at different cutoff points. The receiver operating characteristic curves have three desirable characteristics as a performance metric for multivariate prediction. (1) The area under the receiver operating characteristic curve is independent of the relative frequencies of the events (eg, death). (2) It also is unaffected by the diagnostic system's decision biases or decision threshold. (3) It allows the comparison of diagnostic systems by putting them on a common scale.