

## A new and feasible model for predicting operative risk

A. Donati<sup>1\*</sup>, M. Ruzzi<sup>1</sup>, E. Adrario<sup>1</sup>, P. Pelaia<sup>1</sup>, F. Coluzzi<sup>2</sup>, V. Gabbanelli<sup>1</sup> and P. Pietropaoli<sup>2</sup>

<sup>1</sup>Department of Neuroscience, Anaesthesia and Intensive Care Unit, Marche Polytechnic University, Ancona, Italy. <sup>2</sup>Department of Anaesthesiology, University 'La Sapienza', Rome, Italy

\*Corresponding author: Anestesia e Rianimazione Clinica, Ospedale Regionale Torrette, Via Conca 1, 60020 Torrette di Ancona, Italy. E-mail: donati@indi.it

**Background.** Although the POSSUM (Physiological and Operative Severity Score for the enumeration of Mortality and Morbidity) score can be used to calculate operative risk, its complexity makes its use unfeasible in the immediate clinical setting. The aim of this study was to create a new model, based on ASA status, to predict mortality.

**Methods.** Data were collected in two hospitals. All types of surgery were included except for cardiac surgery and Caesarean delivery. Age, sex and preoperative information, including the presence of cardiocirculatory and/or lung disease, renal failure, diabetes mellitus, hepatic disease, cancer, Glasgow Coma Score, ASA grade, surgical diagnosis, severity of the procedure and type of surgery (elective, urgent or emergency), were recorded for each patient. The model was developed using a data set incorporating data from 1936 surgical patients, and validated using data from a further 1849 patients. Forward stepwise logistic regression was used to build the model. Goodness of fit was examined using the Hosmer–Lemeshow test and receiver operating characteristic (ROC) curve analyses were performed on both data sets to test calibration and discrimination. In the validation data set, the new model was compared with POSSUM and P-POSSUM for both calibration and discrimination, and with ASA alone to compare discrimination.

**Results.** The following variables were included in the new model: ASA status, age, type of surgery (elective, urgent, emergency) and degree of surgery (minor, moderate or major). Calibration and discrimination of the new model were good in both development and validation data sets. This new model was better calibrated in the validation data set (Hosmer–Lemeshow goodness-of-fit test:  $\chi^2=6.8017$ ,  $P=0.7440$ ) than either P-POSSUM ( $\chi^2=14.4643$ ,  $P=0.1528$ ) or POSSUM, which was not calibrated ( $\chi^2=31.8147$ ,  $P=0.0004$ ). POSSUM and P-POSSUM had better discrimination than the new model, although this was not statistically significant. Comparing the two ROC curves, the new model had better discrimination than ASA alone (difference between areas, 0.077, SE 0.034, 95% confidence interval 0.012–0.143,  $P=0.021$ ).

**Conclusions.** This new, ASA status-based model is simple to use and can be performed routinely in the operating room to predict operative risk for both elective and emergency surgery.

*Br J Anaesth* 2004; **93**: 393–9

**Keywords:** assessment, perioperative; risk, operative; surgery, outcome

Accepted for publication: April 30, 2004

A number of studies investigating indices of operative and/or anaesthetic risk have been published.<sup>1–3</sup> The Physiological and Operative Severity Score for the enumeration of Mortality and Morbidity (POSSUM), developed by Copeland and colleagues,<sup>4</sup> was described as a method for quantifying patient data to enable direct comparison of outcomes. The POSSUM score is derived from a combination of physiological variables measured on admission and operative variables. The operative variables are the type and number of surgical procedures, blood loss, peritoneal soiling, presence of malignancy and mode of surgery, while the

12 physiological variables are age, cardiac status, pulse rate, systolic blood pressure, respiratory status, Glasgow coma score, serum concentrations of urea, potassium and sodium, haemoglobin concentration, white cell count and electrocardiographic findings. These two scores (physiological and operative) are then inserted into formulae allowing the risk of morbidity and mortality to be predicted.

The aims of this study were to develop a new model for assessing operative risk that is easy to both calculate and use, and to validate this new model against both POSSUM and P-POSSUM, an updated system that takes into account some

of the shortcomings of the original POSSUM scoring system. These models are very widely used for the prediction of postoperative mortality, and probably represent the standard that any new and improved modelling process would hope to supersede. Finally, the discriminative ability of the new model was compared with the ASA score alone.

## Patients and methods

The development data set of the model was collected between October 1998 and April 1999. Data were collected from all patients, with no age limits imposed, who underwent any type of elective or emergency surgical procedure in two different hospitals.<sup>5</sup> Patients having cardiac surgery or Caesarean delivery were excluded. Before surgery, the following data were recorded for each patient: age, sex, the presence of cardiocirculatory and/or lung disease, renal failure, diabetes mellitus, hepatic disease, cancer, Glasgow coma score, ASA grade,<sup>6</sup> surgical diagnosis, severity of the proposed procedure, and type of surgery (elective, urgent or emergent). For the determination of surgical severity, the Johns Hopkins

criteria<sup>7</sup> were modified to simplify the new model from the five original levels to three: levels 1 and 2 were combined to form grade 1 (representing minor surgery); level 3 became grade 2 (moderate surgery); and levels 4 and 5 became grade 3 (major surgery) (Table 1). Preoperative and intraoperative data were also collected to calculate the POSSUM and P-POSSUM scores. The ASA grading was performed separately by two anaesthetists, at least one of whom was a consultant; any disagreement was resolved by a senior anaesthetist. Finally, the duration of surgery and the occurrence of postoperative complications were recorded. The patient's postoperative course was followed until discharge from hospital. Death or survival at hospital discharge was the outcome variable defined for the model.

Univariate analyses of the preoperative comorbidities were conducted using the  $\chi^2$  test or Fisher's exact test as appropriate. The odds ratios for the association between each variable and perioperative death were calculated.

To develop the new model, only preoperative variables were used. The model was produced using forward stepwise logistic regression (1990 BMDP Statistical Software Inc., Cork, Ireland, running under DOS and

**Table 1** The modified Johns Hopkins surgical criteria

	General	Includes	Excludes
Grade I	Minimal to mild risk independent to anaesthesia Minimal to moderately invasive procedure Potential blood loss less than 500 ml	Breast biopsy Removal of minor skin or subcutaneous lesions Myringotomy tubes Hysteroscopy Cystoscopy Vasectomy Circumcision Fibre-optic bronchoscopy Diagnostic laparoscopy Dilatation and curettage Fallopian tube ligation Arthroscopy Inguinal hernia repair Laparoscopic lysis of adhesion Tonsillectomy/rhinoplasty	Open exposure of internal body organs Repair of vascular or neurological structures Placement of prosthetic devices Postoperative monitored care setting Open exposure of abdomen, thorax, neck, cranium Resection of major body organs
Grade II	Moderately to significantly invasive procedures Potential blood loss 500–1500 ml Moderate risk to patient independent of anaesthesia	Thyroidectomy Hysterectomy Myomectomy Cystectomy Cholecystectomy Laminectomy  Hip/knee replacement Nephrectomy Major laparoscopic procedures Resection/reconstructive surgery of the digestive tract	Open thoracic or intracranial procedure Major vascular repair (e.g. aortofemoral bypass) Planned postoperative monitored care setting (ICU, ACU)
Grade III	Highly invasive procedure Potential blood loss greater than 1500 ml Major to critical risk to patient independent of anaesthesia Usual postoperative ICU stay with invasive monitoring	Major orthopaedic-spinal reconstruction Major reconstruction of the gastrointestinal tract Major genitourinary surgery (e.g. radical retropubic prostatectomy) Major vascular repair without postoperative ICU stay  Cardiothoracic procedure Intracranial procedure Major procedure on the oropharynx Major vascular, skeletal, neurological repair	

PACU=post anaesthesia care unit; ICU=intensive care unit.

Windows™ platforms). The logistic regression model is explained in the Appendix.

The validation data set was recorded from January to April 2002 of 1849 consecutive patients in the same two hospitals. The operative risk was calculated both for the new model and for POSSUM and P-POSSUM scores. The Hosmer–Lemeshow goodness-of-fit test was used for calibration, comparing the expected and observed numbers of deaths by risk group, and the area under the receiver operating characteristic (ROC) curve was measured for discrimination. Pairwise comparisons of ROC curves from the new model, POSSUM and P-POSSUM were performed (MedCalc 7.1; Medcalc Software, Mariakerke, Belgium). In the validation data set, results from the new model were compared with ASA status alone for discrimination by pairwise comparison of ROC curves.

### Results

Table 2 shows characteristics of the two data sets and Table 3 the type of surgery performed in the development data set.

From univariate analysis, the variables which were significantly correlated with death were anaemia, heart failure [New York Heart Association (NYHA) III–IV] and previous myocardial infarction (Table 4).

Using multivariate logistic regression analysis, the variables that were significantly correlated with death, and therefore included in the model, were age, ASA grade, mode of surgery and severity of surgery (Table 5). Tests for linearity performed for ASA grade and age suggested they could be considered continuous variables (see Appendix), and increments were of 1 yr for age and 1 class for ASA. The coefficients for the categorical variables severity and mode are calculated from the design variables (1) and (2) (see Appendix). No case was lost as a result of missing values.

**Table 2** Patient characteristics of the two data sets

	Development data set		Validation data set	
	Total	Deaths	Total	Deaths
No. patients	1936	33	1849	36
Men	976	12	952	30
Women	960	21	897	6
Age (yr): mean	54		56	
(range)	(1–92)		(5–99)	
Severity				
Grade 1	756	2	850	8
Grade 2	778	18	577	13
Grade 3	402	13	422	15
Mode				
Elective	1833	28	1772	12
Urgency	62	2	62	18
Emergency	41	3	15	6
ASA I	859	2	781	3
ASA II	738	9	711	10
ASA III	318	19	290	9
ASA IV	21	3	67	14

No other variable was independently significantly correlated with death, using logistic regression analysis. The resulting model was well calibrated using the Hosmer–Lemeshow goodness-of-fit test ( $\chi^2=9.1219$ ,  $P=0.5257$ ). The area under the ROC curve was 0.881 (SE 0.025, CI 0.833–0.930), indicating that this new model has good discriminative ability.

The validation data set comprised 1849 patients (Table 2). The new model applied to this data set was also well calibrated using the Hosmer–Lemeshow goodness-of-fit

**Table 3** Type of surgery performed in the development data set

Type of surgery	No.	%
Abdominal surgery	439	22.7
Orthopaedic surgery	312	16.1
Vascular surgery	242	12.5
Urological surgery	165	8.5
Thyroid or parathyroid surgery	143	7.4
Otorhinolaryngological surgery	132	6.8
Neurosurgery	128	6.6
Gynaecological surgery	114	5.9
Ocular surgery	90	4.6
Thoracic surgery	84	4.3
Other types of surgery	87	4.5

**Table 4** Univariate analysis of preoperative morbidities in relation to postoperative death. \*Fisher’s exact test

	Total	Deaths	Odds ratio	95% confidence interval	P*
Anaemia	39	4	7.362	2.456–22.068	0.0038
Heart failure (NYHA 3–4)	58	4	4.723	1.604–13.909	0.0155
Previous myocardial infarction	69	4	3.9	1.332–11.422	0.0277
Arterial hypertension	559	15	2.082	1.042–4.161	0.0503
Diabetes mellitus	105	4	2.461	0.8486–7.136	0.099
Renal failure	55	2	2.252	0.525–9.659	0.24
Hepatic failure	37	1	1.621	0.215–12.193	0.47
Previous stroke	51	1	1.158	0.155–8.648	0.59
Severe bronchopulmonary disease	158	3	1.128	0.340–3.738	0.75

**Table 5** Coefficients of the model with odds ratio and confidence interval of the odds ratio. The P-value in the last column indicates the significance of each term in improving the prediction of the model (likelihood ratio test)

Term	Coefficient	Odds ratio	95% confidence interval		P
			Lower bound	Upper bound	
ASA	1.09	2.97	1.73	5.11	<0.001
Age	0.03346	1.03	1.003	1.07	0.023
Severity					0.026
Severity (1)	0.5317	1.70	0.936	3.09	
Severity (2)	0.6739	1.96	1.04	3.7	
Mode					0.011
Mode (1)	0.00477	1.005	0.9	1.12	
Mode (2)	0.8163	2.26	0.799	6.41	
Constant	–8.087	0.000308	0.000033	0.00287	<0.001

test ( $\chi^2=6.8017$ ,  $P=0.7440$ ). In this data set the POSSUM score showed poor calibration ( $\chi^2=31.8147$ ,  $P=0.0004$ ). Better calibration was seen for the P-POSSUM score, although this was still inferior to our new model ( $\chi^2=14.4643$ ,  $P=0.1528$ ). The discriminatory ability of the POSSUM score, the P-POSSUM score and the new model were assessed using ROC curves (Fig. 1). The area under the ROC curve for the new model was 0.888, SE 0.025, 95% confidence interval (CI) 0.838–0.937. The ROC curve for the POSSUM score was 0.915 (SE 0.016, CI 0.884–0.947) and for the P-POSSUM score it was 0.912 (SE 0.033, CI 0.898–0.924). Pairwise comparison of ROC curves between the new model and the POSSUM score showed a difference between areas of 0.028 (SE 0.035, CI  $-0.040$ – $0.095$ ,  $P=0.423$ ), and between the new model and P-POSSUM score of 0.024 (SE 0.035, CI  $-0.044$ – $0.092$ ,  $P=0.491$ ). Between the POSSUM and P-POSSUM scores, the difference between areas was 0.004 (SE 0.005, CI  $-0.008$ – $0.016$ ,  $P=0.549$ ).

As the ASA grade was such a highly significant predictor variable within the new model, we generated a further logistic regression model using ASA grade as the only predictor variable. An ROC curve derived from ASA grade model was constructed and compared with the ROC curve from the new model. The area under the ROC curve for ASA status gave an area of 0.810 (SE 0.044, CI 0.792–0.828). The difference between areas (new model *vs* ASA) was 0.077 (SE 0.034, CI 0.012–0.143,  $P=0.021$ ).

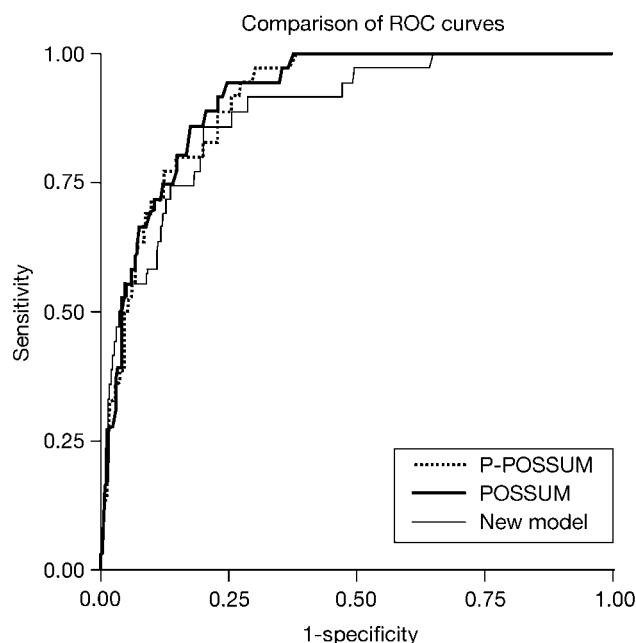
The operative risk was calculated for different age groups on the basis of ASA class and the type of surgery (elective or emergency) for both major (Table 6A) and moderate to minor

surgery (Table 6B). The risk was calculated on the median and range (minimum – maximum) of values for each age group. These tables are provided in order to overcome what would otherwise be the considerable challenge of performing a calculation based on a logistic regression equation at the patient's bedside.

## Discussion

We have developed and validated a new model to predict the operative risk of death. This model is more feasible to apply at the bedside than the POSSUM score. It displays good calibration when examined using the Hosmer–Lemeshow test, but this did not translate into improved discrimination when the ROC curve for the new model was compared with ROC curves generated using the POSSUM and P-POSSUM scores. However, the advantage of this new model is that it can be applied preoperatively and does not require the use of intraoperative data. In any case, for widespread use, a new validation data set from different hospitals would be required. In our hospital a new clinical information system (DEIO; Datex, Helsinki, Finland) has been acquired, into which the equation of the new model has been inserted. This allows operative risk to be calculated automatically during the preoperative anaesthetic assessment. This new model will be useful as an internal quality assessment, allowing annual comparisons of observed *vs* predicted mortality of our surgical patients.

Many studies have been published recommending a variety of scores.<sup>1–3</sup> The perfect index would be one that is easy and quick to use, adoptable by all hospitals, and able to predict the operative risk in all surgical patients, whether elective or urgent/emergent. The Goldman Cardiac Risk Index introduced by Goldman and colleagues in 1977<sup>8</sup> agrees in part with these requirements. It is applicable to all types of surgical operation but it only calculates the risk of onset of cardiovascular complications. Other authors have also analysed cardiovascular risk for both elective and emergency surgery.<sup>9–10</sup> Chung and colleagues proposed a predictive model on the basis of pre-existing medical conditions, but this model is only applicable in day-case surgery.<sup>11</sup> POSSUM (and the derived P-POSSUM) is a good model as it can be used in all types of surgery, both elective and emergency. P-POSSUM (and POSSUM) has been used for many purposes: to compare mortality rates after surgery between patients in the USA and UK,<sup>12</sup> to assess outcome after laparoscopic colectomy<sup>13</sup> or after surgery for colorectal cancer<sup>14</sup> and to predict mortality in infrarenal abdominal aortic aneurysm repair.<sup>15</sup> However, as its use requires intra- and post-operative data it is neither simple nor rapid. Moreover, for its complete calculation blood samples and physiological measurements are necessary. However, its greatest limitation as a prognostic score is its applicability only after the surgical procedure. It cannot be used preoperatively, when the patient (and surgeon) should ideally be aware of the operative risk.



**Fig 1** ROC curves calculated using the validation data set for POSSUM, P-POSSUM and the new model.

**Table 6A** Operative risk calculated for every age group on the basis of the ASA score and the mode (elective or urgent/emergency surgery) of major (grade III) surgical operations. The risk was calculated on the median (the risk calculated for the extreme of every age group)

Elective Age				vs	Urgent/emergency Age			
Class	<50	50–69	≥70	Class	<50	50–69	≥70	
Median (age min–max)	41 (1–49)	58 (50–69)	70 (70–94)	Median (age min–max)	41 (1–49)	58 (50–69)	70 (70–94)	
ASA I	0.3% (0.1–0.4)	0.5% (0.4–0.8)	0.0% (0.0–0.0)	ASA I	1.6% (0.4–2.0)	2% (2.0–4.0)	0.0% (0.0–0.0)	
ASA II	0.9% (0.2–1.1)	1.7% (1.1–2.4)	2.8% (2.4–5.2)	ASA II	4.5% (1.2–5.8)	8.2% (5.8–11.1)	12.9% (11.1–21.9)	
ASA III	2.7% (0.7–3.5)	4.9% (3.5–5.8)	7.9% (5.8–13.9)	ASA III	12.4% (3.6–15.6)	21% (15.6–27.2)	30.6% (27.2–45.5)	
ASA IV	7.6% (2.1–9.7)	13.4% (9.7–17.8)	20.3% (17.8–32.5)	ASA IV	29.6% (9.9–35.5)	44.3% (35.5–52.6)	56.8% (52.6–71.2)	

**Table 6B** Operative risk calculated for every age group on the basis of the ASA score and the mode (elective or urgency/emergency surgery) for minor/moderate (grade I/II) surgical operations. The risk has been calculated on the median (the risk calculated for the extreme of every age group)

Elective Age				vs	Urgent/emergency Age			
Class	<50	50–69	≥70	Class	<50	50–69	≥70	
Median (age min–max)	41 (1–49)	58 (50–69)	70 (70–94)	Median (age min–max)	41 (1–49)	58 (50–69)	70 (70–94)	
ASA I	<0.1% (<0.1)	0.1% (<0.1–0.1)	0.0% (0.0–0.0)	ASA I	0.2% (0.1–0.3)	0.4% (0.3–0.6)	0.0% (0.0–0.0)	
ASA II	0.1% (<0.1–0.2)	0.3% (0.2–0.4)	0.4% (0.4–0.8)	ASA II	0.7% (0.2–0.9)	1.3% (0.9–1.8)	2.2% (1.8–4.1)	
ASA III	0.4% (0.1–0.5)	0.8% (0.5–1.1)	1.3% (1.1–2.4)	ASA III	2.1% (0.6–2.7)	3.9% (2.7–5.4)	6.3% (5.4–11.3)	
ASA IV	1.2% (0.3–1.6)	2.3% (1.6–3.2)	3.7% (3.5–6.9)	ASA IV	6% (1.6–7.7)	10.8% (7.7–14.5)	16.7% (14.5–27.5)	

In agreement with the literature, we found in our study an overestimation of the operative risk for POSSUM that is more important in the lower deciles. Whiteley and colleagues reviewed the POSSUM model, changing the coefficients, and made similar criticisms.<sup>16</sup> The lowest physiological and operative scores are 12 and 6 respectively; when applied to the POSSUM mortality predictor equation this gives a minimum risk of death of 1.1%. This is far too high, given that it represents the fittest individual undergoing the least intricate surgery. Previously published series of fit people undergoing uncomplicated hernia repair suggest that mortality rates are less than 0.001%.<sup>17</sup>

With regard to variables and their strength of prediction of risk, there are ample candidates to be included in a prognostic model. Some relate to the patient and some to the surgical procedure. Age is a significant patient factor and thus enters the model. Indeed, it significantly increases the accuracy of prediction ( $P=0.0228$ ). This agrees with most of the published literature, which considers age to be an important factor for increased mortality risk.<sup>18,19</sup> However, it is important to note that it is not age *per se* but the deterioration of organ function that occurs with age.<sup>18</sup> The odds ratio of 1.03 found in our study is consistent with the study by Wolters and colleagues,<sup>20</sup> who reported an odds ratio of 1.0105 per year of life increment.

The variable best correlated with an increase in operative risk was the physical condition of the subject as represented by the ASA grade, with an odds ratio of 2.97. It is important to note that it not only significantly increases the accuracy of

prediction ( $P<0.0001$ ), but after it entered the model all other preoperative risk factors, such as heart and lung disease and renal failure, were not included in the model as they were independently not significant. Since 1941,<sup>21</sup> and with some subsequent modifications,<sup>6</sup> the ASA grading has been the most important instrument for assessing the patient’s baseline health status. It has also been applied with other variables to predict postoperative complications.<sup>20</sup> Wolters and colleagues examined the strength of association between ASA grade and perioperative risk factors and postoperative outcome, with both univariate analysis and logistic regression.<sup>20</sup> They found that intraoperative blood loss, duration of postoperative ventilation, duration of intensive care stay, rates of pulmonary and cardiac complications, and in-hospital mortality showed significant increases as ASA status advanced from I to IV. In contrast to our present study, their study did not intend to build a mathematical model to predict mortality and/or postoperative complications. However, their results demonstrated not only the association between ASA status and postoperative outcome, but also the great value of this type of statistical analysis in the improvement of patient therapy.

The importance of the type of surgery has been emphasized previously.<sup>22,23</sup> Elective surgery and minor severity surgery reduce operative risk as the greater effect on poor outcome is attributable to emergency and/or high-severity surgery. A patient in poor physical condition who needs emergency surgery may perhaps benefit from a reduction in severity of the surgery, or deferring major surgery until their state

of health has been optimized.<sup>24</sup> Thus, our new model, which includes the mode and severity of surgery, improved on the discriminatory ability of the ASA grade alone.

In conclusion, this new model can be helpful for both surgeons and anaesthetists in daily practice, providing them with a true idea of the operative risk of death of the surgical patient. It will also be useful as an internal quality assessment. The next step will be to include postoperative complications in this model in order to have a more complete score for evaluating surgical patient outcome.

## Acknowledgement

The authors would like to thank Professor Mervyn Singer for his help with the English language in this paper.

## Appendix

The goal of logistic regression analysis is to describe the relationship between a prospectively observed dichotomous outcome (the occurrence of death or not) and a set of predictor variables (preoperative variables in our model). The predictor variables can be binary, categorical (more than two categories) or continuous [25].

If we consider  $\pi_i$  as the population probability of an event (also called the 'expected' value) we can write  $E(y_i)=\pi_i$  ( $E$  is for expected value). If an event has probability  $\pi_i$ , the odds ratio for this is  $\pi_i/(1-\pi_i)$  to 1.

The model is

$$\text{Log}_e\{\pi_i/(1-\pi_i)\}=\text{logit}(\pi_i)=\beta_0+\beta_1x_1+\dots+\beta_nx_n$$

where the predictor variables are  $x_1, \dots, x_n$ .

The term on the left-hand side of the equation is the log of the odds of success, and is called the logistic or logit transform.

The coefficient  $\beta$  is related to the odds ratio in 2x2 tables. If the predictor variable ( $x$ ) is binary, the odds ratio associated with  $x$  is given by  $\exp(\beta)$ . If  $x$  is continuous,  $\exp(\beta)$  is the odds ratio associated with a unit increase in  $x$ . The parameters in the model are estimated with the maximum likelihood function.

Table 5 shows variables entering the model with their respective weights (i.e. the coefficient  $\beta$  for every factor). The threshold for inclusion of the variables in the model was a significance to predict death lower than 0.05, while the removal limit was  $P>0.10$ . Tests for linearity were performed for ASA status and age, which were considered continuous variables. The variables were tested in two ways [25]. First, a quadratic term ( $x^2$ ) was included in addition to the linear term ( $x$ ) in the model. A significant coefficient for  $x^2$  indicates a lack of linearity, but in this case there was not a significant coefficient for  $x^2$ , either for age ( $P=0.11$ ) or for ASA ( $P=0.18$ ). Secondly, ASA grade and age were considered as categorical variables, with four categories for ASA and five for age (dividing age into five quintile groups), and the coefficients were examined. For a linear relationship, the coefficients themselves will increase linearly, and this happened both for ASA and age.

**Table A1** Coding of the design variables for the categorical variables severity and mode

Variable name	Group index	Frequency	Design variables	
			(1)	(2)
Severity	1 (grade 1)	756	-1	-1
	2 (grade 2)	778	1	0
	3 (grade 3)	402	0	1
Mode	1 (elective)	1833	-1	-1
	2 (urgency)	62	1	0
	3 (emergency)	41	0	1

The program generates design variables for each categorical variable. These are used in the model instead of the value or category numbers recorded for the variable. The design variables that are generated either contrast the first category with later categories or are orthogonal polynomial components. Assuming three categories, the program generates by default two design variables, (1) and (2), of the following type (Table A1): design (1) (category one, -1; category two, 1; category three, 0); design (2) (category one, -1; category two, 0; category three, 1). The coefficient entering the model for category one is: design (1) - design (2); for category two is equal to design (1) and for category three is equal to design (2).

For model calibration (how closely the predicted versus the observed outcomes match throughout the range of risk) the Hosmer-Lemeshow goodness-of-fit test was used [26]. Hosmer and Lemeshow proposed a statistic that they show, through simulation, has a  $\chi^2$  distribution when there is no replication in any of the subpopulations. This test is available only for binary response models. The Hosmer-Lemeshow goodness-of-fit statistic is obtained by calculating the Pearson  $\chi^2$  statistic from the 2xg table of observed and expected frequencies, where  $g$  is the number of groups. The statistic is written:

$$\chi_{HL}^2 = \sum_{i=1}^g \frac{(O_i - N_i \bar{\pi}_i)^2}{N_i \bar{\pi}_i (1 - \bar{\pi}_i)}$$

where  $N_i$  is the total frequency of subjects in the  $i$ th group,  $O_i$  is the total frequency of event outcomes in the  $i$ th group, and  $\bar{\pi}_i$  is the average estimated probability of an event outcome for the  $i$ th group. The Hosmer-Lemeshow statistic is then compared with a  $\chi^2$  distribution with  $(g-n)$  degrees of freedom. Large values of  $\chi_{HL}^2$  (and small  $P$  values) indicate a lack of fit of the model.

## References

- 1 Jones HJS, Cossart L. Risk scoring in surgical patients. *Br J Surg* 1996; **86**: 149-57
- 2 Klotz Hp, Candinas D, Platz A, et al. Preoperative risk assessment in elective general surgery. *Br J Surg* 1996; **83**: 1788-91
- 3 Kroenke K, Lawrence VA, Theroux J, Tuley MR, Hilsenbeck S. Postoperative complication after thoracic and major abdominal surgery in patients with and without obstructive lung disease. *Chest* 1993; **104**: 1445-51
- 4 Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991; **52**: 355-60

- 5 Pietropaoli P, Donati A, Arzano S, Messina T. Valutazione comparativa del 'Rischio Operatorio' tra due studi, eseguiti a distanza di 10 anni, presso due identiche strutture. *Minerva Anesthesiol* 1999; **65** (9S): 729–36
- 6 Vacanti CJ, Van Houten RJ, Hil RC. A statistical analysis of the relationship of physical status to postoperative mortality in 68,388 cases. *Anesth Analg* 1970; **49**: 564–6
- 7 Pasternak LR. Preanesthesia evaluation of the surgical patient. *ASA Refresher Courses in Anesthesiology* 1996; **24**: 205–19
- 8 Goldman L, Caldera DL, Nussbaum SR, et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. *N Engl J Med* 1977; **297**: 845–50
- 9 Howell SJ, Sear YM, Yates D, Goldacre M, Foex P. Risk factors for cardiovascular death within 30 days after anaesthesia and urgent or emergency surgery: a nested case–control study. *Br J Anaesth* 1999; **82**: 679–84
- 10 Howell SJ, Sear YM, Yates D, Goldacre M, Sear JW, Foex P. Risk factors for cardiovascular death after elective surgery under general anaesthesia. *Br J Anaesth* 1998; **80**: 700
- 11 Chung F, Mezei G, Tong D. Pre-existing conditions as predictors of adverse events in day-case surgery. *Br J Anaesth* 1999; **83**: 262–70
- 12 Bennett-Guerrero E, Hyam JA, Shaefi S, et al. Comparison of P-POSSUM risk-adjusted mortality rates after surgery between patients in the USA and the UK. *Br J Surg* 2003; **90**: 1593–8
- 13 Senagore AJ, Delaney CP, Duepre HJ, Brady KM, Fazio VW. Evaluation of POSSUM and P-POSSUM scoring systems in assessing outcome after laparoscopic colectomy. *Br J Surg* 2003; **90**: 1280–4
- 14 Menon KV, Farouk R. An analysis of the accuracy of P-POSSUM scoring for mortality risk assessment after surgery for colorectal cancer. *Colorectal Dis* 2002; **4**: 197–200
- 15 Shuhaiber JH, Hankins M, Robless P, Whitehead SM. Comparison of POSSUM with P-POSSUM for prediction of mortality in infrarenal abdominal aortic aneurysm repair. *Ann Vasc Surg* 2002; **16**: 736–41
- 16 Whiteley MS, Prytherch DR, Higgins B, Weaver PC, Prout WG. An evaluation of the POSSUM surgical scoring system. *Br J Surg* 1996; **83**: 812–5
- 17 Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. *Br J Surg* 1998; **85**: 1217–20
- 18 Priebe HJ. The aged cardiovascular risk patient. *Br J Anaesth* 2000; **85**: 763–78
- 19 Jin F, Chung F. Minimizing perioperative adverse events in the elderly. *Br J Anaesth* 2001; **87**: 533–6
- 20 Wolters U, Wolf T, Sttzer H, Schröder T. ASA classification and perioperative variables as predictors of postoperative outcome. *Br J Anaesth* 1996; **77**: 217–22
- 21 Sakland M. Grading of patients for surgical procedures. *Anesthesiology* 1941; **2**: 281–4
- 22 Tiret L, Hatton F, Desmonts JM, Vourc'h G. Prediction of outcome of anaesthesia in patients over 40 years: a multifactorial risk index. *Stat Med* 1988; **7**: 947–54
- 23 Arvidsson S, Ouchterlong J, Sjostedt L, Swardsudd K. Predicting postoperative adverse events. Clinical efficiency of four general classification systems. *Acta Anaesthesiol Scand* 1996; **40**: 783–91
- 24 Curra JE, Grounds RM. Ward versus intensive care management of high-risk patients. *Br J Surg* 1998; **85**: 956–61
- 25 Campbell MJ. *Statistics at Square Two*. London: BMJ Books, 2001; 26–27, 37–58
- 26 Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1980; **115**: 92–106