Research

Performance of six severity-of-illness scores in cancer patients requiring admission to the intensive care unit: a prospective observational study

Márcio Soares[1], Flávia Fontes[1], Joana Dantas[1], Daniela Gadelha[1], Paloma Cariello[1], Flávia Nardes[2], César Amorim[2], Luisa Toscano[3] and José R Rocco[4]

[1]Attending physician, Intensive Care Unit, Instituto Nacional de Câncer, and Programa de Pós-Graduação em Clínica Médica, Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
[2]Medical student, Hospital Universitário Clementino Fraga Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
[3]Director, Intensive Care Unit, Instituto Nacional de Câncer, Rio de Janeiro, Brazil
[4]Professor, Hospital Universitário Clementino Fraga Filho, Universidade Federal do Rio de Janeiro, and Programa de Pós-Graduação em Clínica Médica, Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil Correspondence: Márcio Soares

Corresponding author: Márcio Soares, marciosoaresms@globo.com

Abstract

Introduction The aim of this study was to evaluate the performance of five general severity-of-illness scores (Acute Physiology and Chronic Health Evaluation II and III-J, the Simplified Acute Physiology Score II, and the Mortality Probability Models at admission and at 24 hours of intensive care unit [ICU] stay), and to validate a specific score – the ICU Cancer Mortality Model (CMM) – in cancer patients requiring admission to the ICU.

Methods A prospective observational cohort study was performed in an oncological medical/surgical ICU in a Brazilian cancer centre. Data were collected over the first 24 hours of ICU stay. Discrimination was assessed by area under the receiver operating characteristic curves and calibration was done using Hosmer–Lemeshow goodness-of-fit H-tests.

Results A total of 1257 consecutive patients were included over a 39-month period, and 715 (56.9%) were scheduled surgical patients. The observed hospital mortality was 28.6%. Two performance analyses were carried out: in the first analysis all patients were studied; and in the second, scheduled surgical patients were excluded in order to better compare CMM and general prognostic scores. The results of the two analyses were similar. Discrimination was good for all of the six studied models and best for Simplified Acute Physiology Score II and Acute Physiology and Chronic Health Evaluation III-J. However, calibration was uniformly insufficient ($P < 0.001$). General scores significantly underestimated mortality (in comparison with the observed mortality); this was in contrast to the CMM, which tended to overestimate mortality.

Conclusion None of the model scores accurately predicted outcome in the present group of critically ill cancer patients. In addition, there was no advantage of CMM over the other general models.

Keywords: cancer, mortality, outcome, severity-of-illness scores

Introduction

Advances in oncological and supportive care have improved survival rates in cancer patients to the point that many of them can now be cured or have their disease controlled. However,

---

APACHE = Acute Physiology and Chronic Health Evaluation; AUROC = area under the receiver operating characteristic curve; BMT = bone marrow transplant; CMM = Cancer Mortality Model; ICU = intensive care unit; MPM = Mortality Probability Model; MV = mechanical ventilation; SAPS = Simplified Acute Physiology Score; SMR = standardized mortality ratio.

such advances have often been achieved through aggressive therapies and support, at high expense [1]. Some of these patients require admission to the intensive care unit (ICU) for acute concurrent illness, postoperative care, or complications of their cancer or its therapy [2]. In general hospitals, intensivists frequently consider these patients as having a poor prognosis and tend to oppose their admission to the ICU [3]. Recent studies [4,5] have indicated that this reluctance to offer ICU care to cancer patients with severe illness is unjustified, and is usually based on inequitable parameters in comparison with other severe and chronic diseases that share a similarly poor prognosis [6,7]. Hence, efforts have been made to identify parameters that are associated with worse prognosis and to improve allocation of ICU resources [4,5,8-11].

Prognostic scores have been used to predict outcome in patients admitted to ICUs. Although none of these models should be used to predict individual outcomes, they can assist physicians in discussions of prognosis and in clinical decision making to improve allocation of resources in intensive care [12]. Stratification of patients for clinical research and assessment of quality of intensive care are other potential applications [12-14].

However, the performance of a prognostic score must be validated before it may be used in an ICU, where there is a specific mix of patients such as cancer patients. Few studies have addressed adequately the performance (calibration and discrimination properties) of prognostic scores in cancer patients [4,9,15-17]. Some years ago a specific model with which to predict outcome among critically ill cancer patients – the Cancer Mortality Model (CMM) – was developed by Groeger and coworkers [9]. To the best of our knowledge, only one external validation for the CMM has been conducted recently [17], and few studies have compared different scores in cancer patients [4,16,17]. The present study evaluated the performance of five general prognostic models and validated the CMM in predicting outcome in a large prospective cohort of cancer patients requiring intensive care.

Methods
Patients and setting
The study was conducted from May 2000 to July 2003 at the Instituto Nacional de Câncer, a public tertiary hospital for referral of cancer patients in Rio de Janeiro, Brazil. Its ICU is a 10-bed medical/surgical unit exclusively for oncological patients, with full-time medical and nurse directors, and medical, physiotherapy and nursing staff who are qualified in intensive care; facilities such as haemodynamic monitoring, microprocessor-controlled mechanical ventilation and dialysis are available and can be offered for each bed. At least two senior intensivists and one junior intensivist are on duty 24 hours a day. In each shift (two per day), at least two postgraduate and four undergraduate nurses work regularly in the ICU. Most of the postgraduate nurses have a special diploma in oncology and/or

intensive care, and take part in regular training of oncology and intensive care nurses. The nurse/patient ratio ranges from 1.2 to 1.7. Routine clinical rounds, including medical, nurse and physiotherapy staff, and meetings with oncologists, are done each day in the ICU. Approximately 600 patients are admitted each year to the ICU. The ICU uses a patient data management system, which allows automatic capture and registration of physiological data.

The decision regarding whether a patient should be admitted to the ICU is made jointly by the senior intensivist and the oncologist who is responsible for the patient's care. To be admitted to the ICU, patients should be considered to have a chance of being cured or having their cancer controlled. The ICU medical staff makes decisions regarding discharge during daily clinical rounds. Patients are always discharged to wards. End-of-life care is offered in the ICU when a patient does not recover from their acute illness despite ICU care. Occasionally, patients (with a diagnosis of cancer) may be admitted because of life-threatening illness identified during assessment of the extent of their cancer and/or consideration of the therapeutic options. This assessment is conducted as soon as is possible, and end-of-life care is started if specific treatments aimed at cancer cure or control can no longer be justified.

All consecutive patients with a definite diagnosis of cancer (pathologically proven) admitted to the ICU because of severe illness were included in the present study. In those patients with multiple hospital admissions, the most recent was considered. For patients readmitted to the ICU during the same hospital stay, only the first ICU admission was considered. Patients younger than 18 years ($n = 284$), those with burn injuries ($n = 0$), those with an ICU stay of less than 6 hours ($n = 29$) and those with definite diagnosis of acute coronary syndrome or in whom such a disorder could not be ruled out were excluded ($n = 15$). Patients who had been considered cured of their cancer for more than 5 years ($n = 20$) and those with noncancer disease ($n = 36$) were also excluded. Bone marrow transplant (BMT) patients are treated at a separate unit, even in case of life-threatening complications; therefore, BMT patients were not studied. The study was approved by the institutional review board, which waived the need for informed consent because the study did not interfere with clinical decisions regarding patient care.

Measurements
At admission and over the first 24 hours of ICU stay, various demographic, clinical and laboratory variables were assessed. In the calculation of scores, the most disturbed values were assigned for vital signs and laboratory data. For sedated patients, Glasgow Coma Scale score before sedation was used [18]. Zero points or normal values were inserted where data were missing [19]. There were no missing variables for physiological data. Among laboratory variables, normal values were inserted for albumin in 623 (49.6%), prothrombin

time in 274 (21.8%) and bilirubin in 676 (53.8%) patients. No patient with jaundice on physical examination lacked serum bilirubin measurements. Severe chronic comorbidities were considered as defined in the assessment of each scoring system. Patients were classified, based on reason for ICU admission, into medical, scheduled surgical and emergency surgical groups. We also recorded the underlying malignancy (solid tumour versus haematological malignancy), disease status (newly diagnosed/controlled versus recurrence/progression; locoregional versus metastatic), treatments over the 6 months before ICU admission (chemotherapy, radiation therapy and surgery, excluding biopsies and catheter insertions) and Eastern Cooperative Oncology Group performance status [20] during the week before hospital admission. Neutropenia was defined as a neutrophil count of below 1000/mm³. The Sequential Organ Failure Assessment score was used to assess acute organ dysfunctions/failures [21].

The following general prognostic scores were measured: the Simplified Acute Physiology Score (SAPS) II [22], the Mortality Probability Models at admission (MPM II_0) and at 24 hours (MPM II_{24}) [23], and Acute Physiology and Chronic Health Evaluation (APACHE®; a registered trademark of Cerner Corporation, Kansas, MO, USA) versions II and III-J [19,24]. Each model was applied as described in their original reports. The mortality equations of the APACHE III-J have recently become available for use worldwide. The CMM [9] is a cancer specific, multivariable, logistic regression model that was specifically designed to predict the probability of hospital death in patients at admission to the ICU. Briefly, it comprises 16 easily evaluable clinical variables: cardiac arrest before admission, endotracheal intubation, intracranial mass effect, allogeneic BMT, cancer recurrence/progression, performance status, respiratory rate, systolic blood pressure, arterial oxygen tension/fractional inspired oxygen ratio, Glasgow Coma Scale score, platelet count, prothrombin time, serum albumin, bilirubin, blood urea nitrogen, and number of hospital days before ICU admission (lead time). Hospital mortality was the main endpoint of interest.

Data management and statistical analysis

Data were entered into a computer database by a single author (MS). In order to ensure data consistency, another single author (JRR) cross-checked every variable entered, and a final recheck procedure was conducted for a 10% random sample of patients. All documented data were also evaluated for implausible and outlying values. Statistical analyses were carried out using SPSS software for Windows, version 10.0 (SPSS Inc., Chicago, IL, USA). Continuous variables are presented as mean ± standard deviation or median (25–75% interquartile range) and compared, respectively, using Student's t-test or Mann–Whitney U-test. Categorical variables were reported as absolute numbers (frequency percentages) and analyzed using χ^2 test (with Yates correction where applicable).

Validation of the prognostic scores was performed using standard tests to measure discrimination and calibration for each of the predictive models. The area under the receiver operating characteristic curve (AUROC) was used to evaluate the ability of each model to discriminate between patients who lived from those who died (discrimination) [25]. Hosmer–Lemeshow goodness-of-fit H statistic was used to evaluate the agreement between the observed and expected number of patients who did or did not die in the hospital across all of the strata of probabilities of death (calibration) [26]. A high P value (>0.05) would indicate a good fit for the model. Calibration curves were constructed by plotting predicted mortality rates stratified by 10% intervals of mortality risk (x axis) against observed mortality rates (y axis). Standardized mortality ratios (SMRs) with 95% confidence intervals were calculated for each model by dividing observed by predicted mortality rates. A two-tailed P value < 0.05 was considered statistically significant.

Results

During the period of study 1357 adult patients were admitted to the ICU, and of those 1257 (92.6%) met the eligibility criteria. Sources of admission were distributed as follows: wards ($n = 234$ [18.6%]), emergency room ($n = 156$ [12.4%]), operating room ($n = 853$ [67.9%]) and other hospital ($n = 14$ [1.1%]). Patients were referred from another hospital because of a severe medical complication and had to have a prior diagnosis of malignancy. Based on the reason for ICU admission, there were 404 (32.1%) medical, 715 (56.9%) scheduled surgical and 138 (11.0%) emergency surgical patients. At admission and during the first day of ICU stay, 468 (37.2%) patients required mechanical ventilation (MV), 302 (24.0%) received therapy with vasopressors, and 64 (5.1%) received haemodialysis. Within 2 hours before and the first 24 hours after ICU admission, 39 (3.1%) patients presented with cardiac arrest and 362 (28.8%) patients had a definite or probable diagnosis of infection. Median (25–75% interquartile range) lead time was 2 (1–5) days, hospital stay was 12 (8–25) days and ICU stay was 2 (1–6) days. The patients' demographical and clinical characteristics are shown in Table 1 and their cancer related data are summarized in Table 2. Global hospital mortality was 28.6% (360/1257) and the global ICU mortality rate was 20.8% (261/1257). As expected, hospital mortality was significantly higher for medical (69.4%) and emergency surgical patients (49.3%) than for scheduled surgical ones (5.7%; $P < 0.001$).

The performance of each individual mortality prediction system among all patients is presented in Table 3. All models exhibited excellent discriminatory power but calibration was poor. General prognostic scores underestimated the observed mortality (SMR > 1). By contrast, the CMM tended to overestimate (SMR = 0.51, 95% confidence interval 0.46–0.57).

Table 1

Patients' demographical and clinical characteristics (n = 1257)

Parameter	Value
Age (years)	56.0 ± 16.7 (18–93)
Male sex	660 (52.5%)
SOFA (points)	3 (2–7, 1–19)
APACHE II (points)	12 (8–18, 0–48)
APACHE III-J® (points)	44 (27–71, 3–199)
SAPS II (points)	29 (18–47, 0–121)
Primary reason for ICU admission	
Scheduled surgical patients	715 (56.9%)
Craniotomy	281 (22.3%)
Gastrointestinal surgery	146 (11.6%)
Head and neck surgery	116 (9.2%)
Lung resection	101 (8.0%)
Genitourinary surgery	41 (3.3%)
Other	30 (2.4%)
Emergency surgical patients	138 (11.0%)
Complications of previous GI surgery	72 (5.7%)
GI perforation/rupture	12 (1.0%)
GI obstruction	12 (1.0%)
Intracranial haemorrhage	8 (0.6%)
Cholangitis/cholecystectomy	7 (0.6%)
Other	27 (2.1%)
Medical patients	404 (32.1%)
Sepsis	221 (17.6%)
Pulmonary	183
Urinary	19
Other/unknown	19
Respiratory failure (excluding sepsis)	77 (6.1%)
Cancer	37
Pulmonary embolism	28
Other	12
Cardiopulmonary arrest	19 (1.5%)
Metabolic disturbances	15 (1.2%)
Cardiac arrhythmias (excluding coronary disease)	12 (1.0%)
Stroke	12 (1.0%)
Status epilepticus	12 (1.0%)
GI bleeding	10 (0.8%)
Other	26 (2.1%)

A total of 1257 patients were included. Values are expressed as mean ± standard deviation (range); median (interquartile range, range); or *n* (%). APACHE, Acute Physiology and Chronic Health Evaluation; GI, gastrointestinal; ICU, intensive care unit; SAPS, Simplified Acute Physiology Score; SOFA, Sequential Organ Failure Assessment.

Table 2

Cancer related characteristics

Characteristic	*n* (%)
Type of cancer	
Solid tumour	1116 (88.8%)
Brain tumour	286 (22.8%)
Gastrointestinal cancer	282 (22.4%)
Head and neck cancer	189 (15.0%)
Lung cancer	129 (10.3%)
Urogenital cancer	83 (6.6%)
Breast cancer	58 (4.6%)
Other	89 (7.1%)
Haematological	141 (11.2%)
Non-Hodgkin's lymphoma	75 (6.0%)
Hodgkin's disease	13 (1.0%)
Leukaemia	28 (2.2%)
Multiple myeloma	22 (1.8%)
Other	5 (0.4%)
Extent (solid tumours only)	
Locoregional	893 (80.0%)
Metastatic	223 (20.0%)
Cancer status	
Newly diagnosed/remission	888 (70.6%)
Recurrence/progression	369 (29.4%)
Performance status	
0–1	900 (71.6%)
2–3	112 (8.9%)
4	245 (19.5%)
Neutropenia	49 (3.9%)
Treatments prior to ICU admission (past 6 months)	
Chemotherapy	273 (21.7%)
Radiation therapy	314 (25.0%)
Surgery	814 (64.8%)

A total of 1257 patients were included. ICU, intensive care unit.

To better compare the performances of the CMM and of the general prognostic scores, all scheduled surgical patients were excluded, and therefore 542 (43.1%) patients were included in this analysis. A total of 411 (75.8%) patients had solid tumours and 131 (24.2%) had haematological malignancies. Their mean age was 58.7 ± 16.7 years and their mean Sequential Organ Failure Assessment score was 7.6 ± 4.2 points. Out of these patients, 380 (70.1%) had acute respiratory failure. Hospital and ICU mortality rates were 58.7% (318/

Table 3

Performance of each mortality prediction system for all patients

Prognostic score	ROC curve		Goodness-of-fit H-test		Predicted mortality (mean ± SD)	SMR (CI 95%)
	AUROC ± SE	95% CI	χ^2	P		
SAPS II	0.916 ± 0.009	0.899–0.933	29.400	<0.001	24.4 ± 29.2	1.17 (1.03–1.34)
APACHE III	0.915 ± 0.009	0.898–0.933	117.206	<0.001	20.3 ± 28.2	1.41 (1.23–1.62)
MPM II$_{24}$	0.909 ± 0.009	0.891–0.926	114.713	<0.001	19.1 ± 23.4	1.50 (1.30–1.73)
CMM	0.892 ± 0.011	0.871–0.913	517.662	<0.001	55.9 ± 27.5	0.51 (0.46–0.57)
APACHE II	0.888 ± 0.010	0.868–0.907	78.181	<0.001	20.4 ± 23.0	1.41 (1.22–1.62)
MPM II$_0$	0.854 ± 0.012	0.830–0.878	373.317	<0.001	13.5 ± 18.7	2.12 (1.80–2.50)

Shown are area under receiver operating curves (AUROCs), Hosmer–Lemeshow goodness-of-fit H statistics, and standardized mortality ratios (SMRs) for individual mortality prediction models (degrees of freedom = 8). A total of 1257 patients were included. The observed hospital mortality was 28.6%. APACHE, Acute Physiology and Chronic Health Evaluation; AUROC, area under receiver operating characteristic curve; CI, confidence interval; CMM, Cancer Mortality Model; MPM, Mortality Probability Model; SAPS, Simplified Acute Physiology Score; SD, standard deviation; SE, standard error; SMR, standardized mortality rate.

Table 4

Performance of each mortality prediction system for medical and emergency surgical patients (excluding scheduled surgical patients)

Prognostic score	ROC curve		Goodness-of-fit H-test		Predicted mortality (mean ± SD)	SMR (95% CI)
	AUROC ± SE	95% CI 95%	χ^2	P		
SAPS II	0.815 ± 0.018	0.780–0.851	49.315	<0.001	47.9 ± 29.9	1.23 (1.09–1.37)
APACHE III	0.812 ± 0.018	0.776–0.847	113.113	<0.001	42.6 ± 30.2	1.38 (1.22–1.55)
CMM	0.795 ± 0.019	0.758–0.833	150.411	<0.001	78.7 ± 20.7	0.75 (0.69–0.81)
MPM II$_{24}$	0.792 ± 0.019	0.754–0.830	124.237	<0.001	37.7 ± 24.9	1.56 (1.37–1.77)
APACHE II	0.754 ± 0.021	0.713–0.794	129.729	<0.001	38.2 ± 24.1	1.54 (1.35–1.75)
MPM II$_0$	0.729 ± 0.022	0.686–0.771	645.464	<0.001	25.0 ± 23.0	2.35 (2.00–2.77)

Shown are areas under receiver operating curve (AUROCs), Hosmer–Lemeshow goodness-of-fit H statistics, and standardized mortality ratios (SMRs) for individual mortality prediction models (degrees of freedom = 8). A total of 542 patients were included. The observed hospital mortality was 58.7%. APACHE, Acute Physiology and Chronic Health Evaluation; CI, confidence interval; CMM, Cancer Mortality Model; MPM, Mortality Probability Model; SAPS, Simplified Acute Physiology Score; SD, standard deviation; SE, standard error; SMR, standardized mortality rate.

542) and 43.9% (238/542), respectively. The ICU (47.8% versus 32.6%; P = 0.003) and hospital (61.9% versus 49.3%; P = 0.013) mortality rates for medical patients were significantly greater than for emergency surgical patients. Patients with haematological malignancies had higher mortality than did those with solid tumours (67.2% versus 56.0%; P = 0.030). Their median scores were 18 (25–75% interquartile range 13–25, range 4–48) for APACHE II, 74 (55–99, range 7–199) for APACHE III-J and 50 (37–64, range 6–121) for SAPS II. Results for the performance of the six prognostic scores are shown in Table 4. As was observed for all patients combined, among medical and emergency surgical patients SAPS II exhibited the best discriminative ability (AUROC = 0.815) and MPM II$_0$ the poorest (AUROC = 0.729), and all of the scores were poorly calibrated. Statistically significant differences between observed and predicted mortality rates,

using goodness-of-fit H statistics, were obeserved for all scores. Significant underestimation of actual mortality by general scores and overestimation by the CMM were again observed. The impacts of the differences between actual and predicted mortality rates are demonstrated in the calibration curves (Figs 1 and 2).

Discussion
Many severity-of-illness scores have been developed and used to predict outcome in critically ill patients. During the past few years a series of studies dealing with the application of outcome prediction models in general critically ill patients demonstrated a similar pattern – good discrimination with poor calibration. This pattern has been observed in different settings and with different instruments [27]. Information regarding the usefulness of these general scores in cancer patients
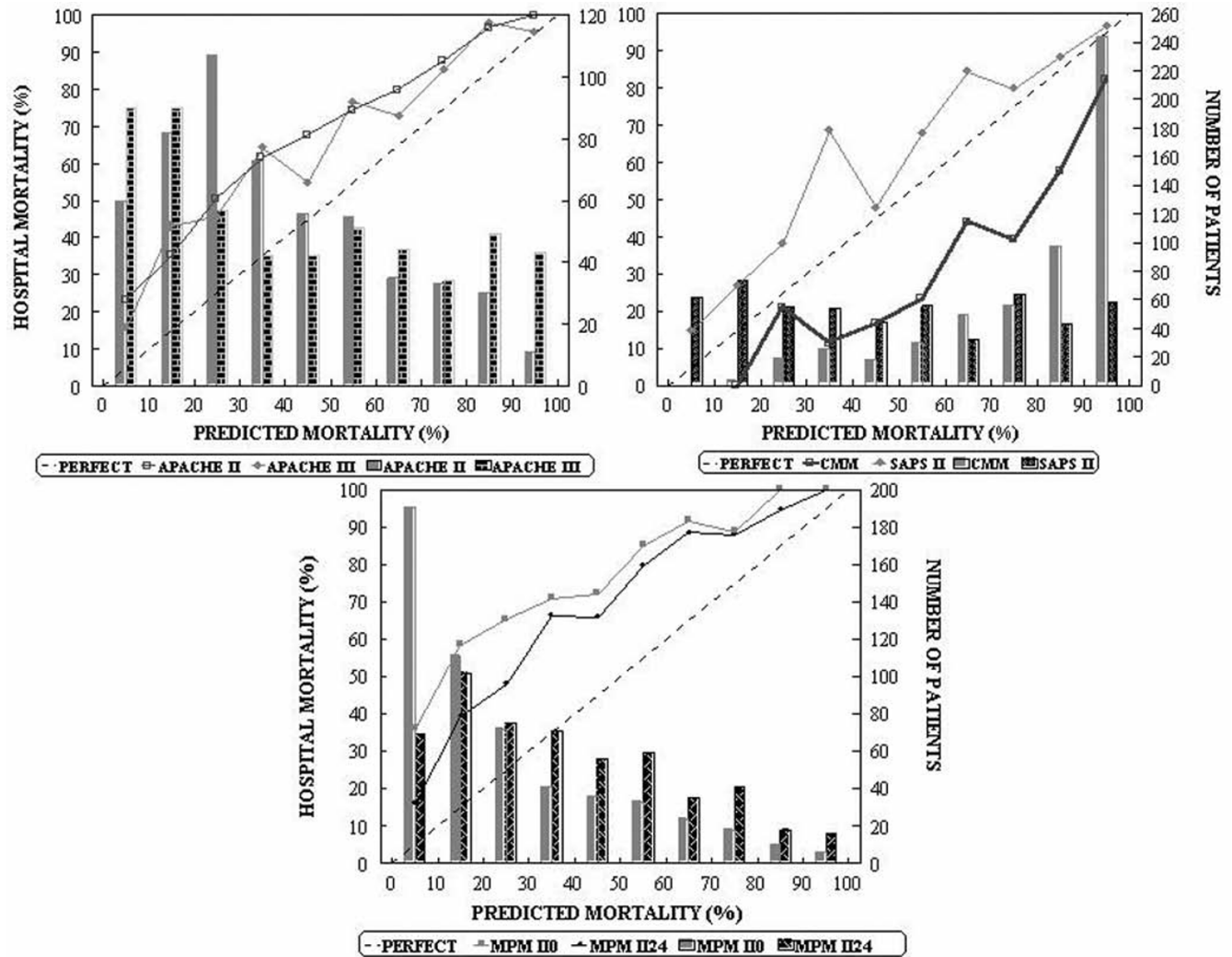
Figure 1
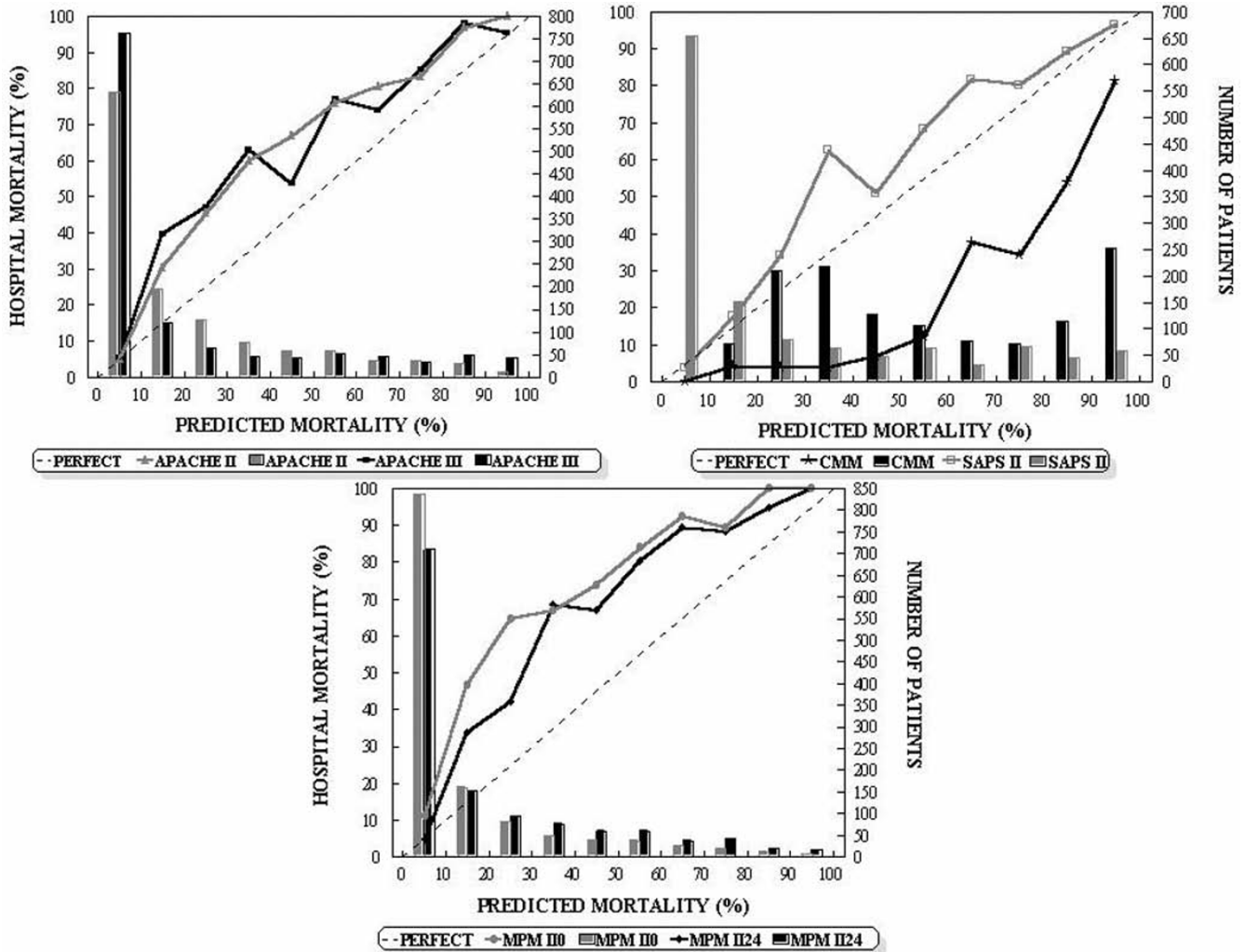
Calibration curves for the six severity-of-illness scores (solid lines) for all 1257 patients. The diagonal dotted line represents the line of ideal prediction. Columns represent the number of patients in each stratum (10% of probability). APACHE, Acute Physiology and Chronic Health Evaluation; CMM, Cancer Mortality Model; MPM, Mortality Probability Model; SAPS, Simplified Acute Physiology Score.

requiring ICU care is still restricted and most reports are limited by relatively small sample sizes and/or the statistical analyses used in the assessment of models' performance [28-32]. In order to better address these issues, we conducted the present study to evaluate simultaneously the performance of five general prognostic scores and to validate the CMM in a large prospective cohort of cancer patients requiring ICU admission. The hospital mortality (28.6%) for the group of ICU cancer patients evaluated here seems to be low at a first glance. However, two thirds of our patients were admitted for routine postoperative care following elective surgery. When these patients were excluded, the hospital mortality (58.7%) was similar to that in previous studies dealing with large cohorts of critically ill cancer patients (33–58.7%) [8-11,16,17]. Staudinger and coworkers [5] reported that ICU mortality was 47% and 1-year mortality was 77%. Mortality

may vary with respect to the mix of patients (e.g. type of tumour, number of BMT patients, disease status and extent, and level of ICU support). In particular, the prognosis for cancer patients receiving MV is very poor. In a large prospective study conducted in 782 patients requiring MV, 76% died in the hospital [33]. In the present cohort about 37% of patients received MV.

Whether studying the entire population or the subgroup of nonscheduled surgical patients, all of the general models tested in the present study had comparatively similar levels of performance. As expected, they significantly underestimated the mortality rate. In general, discrimination was satisfactory (especially for the SAPS II and the APACHE III-J scores), but calibration was inadequate. Studying all patients, AUROC values were remarkably high (>0.850). The higher proportion of

Figure 2

Calibration curves for the six severity-of-illness scores (solid lines) for the sample (excluding scheduled surgical patients; $n = 542$). The diagonal dotted line represents the line of ideal prediction. Columns represent the number of patients in each stratum (10% of probability). APACHE, Acute Physiology and Chronic Health Evaluation; CMM, Cancer Mortality Model; MPM, Mortality Probability Model; SAPS, Simplified Acute Physiology Score.

scheduled surgical patients (very low mortality), in contrast to patients with a severe illness (whether medical or emergency surgical), could be responsible for this finding. When those patients were excluded, AUROC values were similar to those reported in the literature [4,9,15-17]. To our knowledge, there is no conventional method for comparing goodness-of-fit χ^2 tests, but it seems that this statistic was considerably lower for the SAPS II score than for the other models. This can be better appreciated in the calibration curves, which indicate significant underestimates in practically all of the strata of predicted mortality. Nevertheless, the line of observed mortality for the SAPS II score was closer to the line of equality when compared with other general scores. Assessments of both calibration and discriminatory abilities of general prognostic scores in cancer patients were reported in recent years, and yielded

conflicting results [4,9,15-17]. These scores usually tend to underestimate the observed mortality [9,15,16,34]. Groeger and coworkers [9] tested the MPM II_0 model in the first 805 patients included in the sample from which the CMM was developed. The MPM II_0 model exhibited both poor calibration and poor discrimination, and underestimated the mortality. Sculier and coworkers [16] reported similar findings from their evaluation of the APACHE II and the SAPS II scores in a cohort of 261 patients. Guiguet and coworkers [15], studying 98 neutropenic cancer patients, found a reasonable discrimination (AUROC = 0.78) and good calibration for SAPS II. In a retrospective study conducted in 124 patients with haematological cancer, Benoit and colleagues [4] recently reported similar results for the SAPS II (AUROC = 0.765) and the APACHE II (AUROC = 0.712) scores. However, the results of

calibration analyses in the latter two studies should be interpreted with caution because of the relatively small numbers of patients included, so that differences between predicted and observed mortalities may not reach statistical significance. In an elegant study, Zhu and coworkers [14] analyzed the impact of sample size on the accuracy of MPM II models by performing computer simulations. They showed that the smaller the sample size, the better the model calibration, as demonstrated by lower values of the goodness-of-fit χ^2 statistics. In contrast, discrimination was not affected by sample size.

The limitations of general prognostic models in predicting outcome in cancer patients motivated investigators to develop a specific model. Reported in 1998, the CMM was developed in a multicentre study from a cohort of 1483 critically ill cancer patients to predict hospital mortality at admission to the ICU, and it was further validated in another 230 patients [9]. By containing variables specific to oncology (disease progression/recurrence, performance status and allogeneic BMT group), this model was expected to be a more accurate scoring system in cancer patients [5,16]. SAPS II, APACHE II, APACHE III-J and MPM II$_{24}$ models also take into account the presence of some cancer diagnostic categories, but they were not derived exclusively from cancer patients. The performance of the CMM was studied in medical and emergency surgical patients separately (i.e. excluding elective surgical patients) in order to minimize selection bias. There was no mention in the intial CMM report that elective surgical patients had been included in its development. At our ICU, CMM exhibited good discrimination and the AUROC value (0.795) is similar to values observed in both generation (0.812) and validation (0.802) groups of patients. However, CMM was poorly calibrated and, in contrast to general scores, exhibited a tendency to overestimate the observed mortality. Recently, Schellongowski and coworkers [17] compared the levels of performance of CMM, SAPS II and APACHE II in 242 ICU cancer patients [17]. In that study, the ability of SAPS II to discriminate between survivors and nonsurvivors (AUROC = 0.825) was superior to those of APACHE II (AUROC = 0.776) and CMM (AUROC = 0.698). All scores had acceptable calibration, although the statistical significance for the Hosmer–Lemeshow goodness-of-fit tests was borderline. The authors emphasized the limitations imposed by relatively small sample size on the results of calibration analyses.

The present study also has potential limitations. Ideally, a prognostic score should be employed in populations with similar characteristics to the sample of patients in which it was developed. Because we did not study BMT patients, it can be argued that our patients were less severely ill than those studied by Groeger and coworkers [9]. In that study, 11.3% and 5.8% of the sample were allogeneic and autologous BMT patients, respectively. These patients are considered to have the worst prognosis among cancer patients requiring intensive care, and prognosis is particularly poor when such patients

need MV [16,35,36]. Our patients (excluding elective surgical patients) actually had a higher hospital mortality rate (58.7% versus 42%), but it was not feasible to make reliable comparisons of acute physiological disturbances (e.g. organ failures) between groups. In addition, whenever case mix adjustments are attempted, possible selection bias – resulting from different approaches to care (e.g. do-not-resuscitate orders) and from ICU admission/discharge policies – cannot be ruled out, especially in a single centre. Decisions to forgo life-sustaining therapy were demonstrated to independently predict hospital death in ICU patients [37]. Our ICU policies, including decisions to offer end-of-life care, appear similar to those reported in the literature [16,17].

Another issue that deserves mention is the impact of missing data on the performance of models; in the present study prothrombin time, and serum albumin and bilirubin were not obtained in all patients. The differences between the predicted mortality with each score and the observed mortality were considerable, but there is a possible impact of missing data in the unsatisfactory performance of the models. As stated above, the study did not interfere with clinical decisions, including request for laboratory tests. In particular, the poor performance of the CMM cannot be attributed to missing data because it significantly overestimated the mortality rate.

Finally, we should be cautious when using SMR findings to evaluate the quality of intensive care. The prognostic scores that are already available do not take into consideration multidimensional parameters (ICU organizational and economic aspects in addition to clinical variables) in evaluating ICU performance [38].

In conclusion, none of the severity-of-illness scores evaluated in the present study were accurate in predicting outcome for critically ill cancer patients. Moreover, similar to a recent report [17], we found no advantage of CMM over the general prognostic models. It must be re-emphasized that any prognostic model should not be the only parameter taken into account when predicting outcome, and neither should they be used for triage and cost containment in individual patients. After all, prognostic scores were constructed based on patients who have been effectively admitted to the ICU. Otherwise, an accurate score may be helpful in enroling patients in clinical trials and enriching discussions about prognosis in intensive care.

Key messages

None of the severity-of-illness scores evaluated in the present study were accurate in predicting outcome for critically ill cancer patients. There was no advantage of CMM over the general prognostic models. Prognostic scores should not be the only parameters taken into account when predicting outcome, and neither should they be used for triage and cost containment in individual patients.

Competing interests

Author contributions

Acknowledgements

References

1. Schapira DV, Studinicki J, Bradham DD, Wolff P, Jarrett A: **Intensive care, survival, and expense of treating critically ill cancer patients.** *JAMA* 1993, **269**:783-786.
2. Sculier J-P: **Intensive care and oncology.** *Support Care Cancer* 1995, **3**:93-105.
3. Azoulay E, Pochard F, Chevret S, Vinsonneau C, Garrouste M, Cohen Y, Thuong M, Paugam C, Apperre C, De Cagny B, Brun F, Bornstain C, Parrot A, Thamion F, Lacherade JC, Bouffard Y, Le Gall J-R, Herve C, Grassin M, Zittoun R, Schlemmer B, Dhainaut JF, for the PROTOCETIC Group: **Compliance with triage to intensive care recommendations.** *Crit Care Med* 2001, **29**:2132-2136.
4. Benoit DD, Wandewoude KH, Decruyenaere JM, Hoste EA, Colardyn FA: **Outcome and early prognostic indicators in patients with a hematologic malignancy admitted to the intensive care unit for a life-threatening complication.** *Crit Care Med* 2003, **31**:104-112.
5. Staudinger T, Stoiser B, Mullner M, Locker GJ, Laczika K, Knapp S, Burgmann H, Wilfing A, Kofler J, Thalhammer F, Frass M: **Outcome and prognostic factors in critically ill cancer patients admitted to the intensive care unit.** *Crit Care Med* 2000, **28**:1322-1328.
6. Watcher RM, Luce JM, Hearst N, Lo B: **Decisions about resuscitation: inequities among patients with different diseases but similar prognoses.** *Ann Intern Med* 1989, **111**:525-532.
7. Tanvetyanon T, Leighton JC: **Life-sustaining treatments in patients who died of chronic congestive heart failure compared with metastatic cancer.** *Crit Care Med* 2003, **31**:60-64.
8. Kress JP, Christenson J, Pohlman AS, Linkin DR, Hall JB: **Outcomes of critically ill cancer patients in a university hospital setting.** *Am J Respir Crit Care Med* 1999, **160**:1957-1961.
9. Groeger JS, Lemeshow S, Price K, Nierman DM, White P, Klar J, Granovsky S, Horak D, Kish SK: **Multicenter outcome study of cancer patients admitted to the intensive care unit: a probability of mortality model.** *J Clin Oncol* 1998, **16**:761-770.
10. Azoulay E, Moreau D, Aberti C, Leleu G, Adrie C, Barboteu M, Cottu P, Levy V, Le Gall J-R, Schlemmer B: **Predictors of short-term mortality in critically ill patients with solid malignancies.** *Intensive Care Med* 2000, **26**:1817-1823.
11. Maschmeyer G, Bertschat FL, Moesta KT, Häusler E, Held TK, Nolte M, Osterziel K-J, Papstein V, Peters M, Reich G, Schmutzler M, Sezer O, Stula M, Wauer H, Wörtz T, Wischnewsky M, Hohenberger P: **Outcome analysis of 189 consecutive cancer patients referred to the intensive care unit as emergencies during a 2-year period.** *Eur J Cancer* 2003, **39**:783-792.
12. Lemeshow S, Klar J, Teres D: **Outcome prediction for individual intensive care patients: useful, misused, or abused.** *Intensive Care Med* 1995, **21**:770-776.
13. Knaus WA, Wagner DP, Zimmerman JE, Draper EA: **Variations of mortality and length of stay in intensive care units.** *Ann Intern Med* 1993, **118**:753-761.
14. Zhu B-P, Lemeshow S, Hosmer DW, Klar J, Avrunin J, Teres D: **Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization. A simulation study.** *Crit Care Med* 1996, **24**:57-63.
15. Guiguet M, Blot F, Escudier B, Antoun S, Leclercq B, Nitenberg G: **Severity-of-illness scores for neutropenic cancer patients in an intensive care unit. Which is the best predictor? Do multiple assessment times improve the predictive value?** *Crit Care Med* 1998, **26**:488-493.
16. Sculier J-P, Paesmans M, Markiewicz E, Berghmans T: **Scoring systems in cancer patients admitted for an acute complication in a medical intensive care unit.** *Crit Care Med* 2000, **28**:2786-2792.
17. Schellongowski P, Benesch M, Lang T, Traunmüller F, Zauner C, Laczika K, Locker GJ, Frass M, Staudinger T: **Comparison of three severity scores for critically ill cancer patients.** *Intensive Care Med* 2004, **30**:430-436.
18. Bastos PG, Sun X, Wagner DP, Wu AW, Knaus WA: **Glasgow Coma Scale in the evaluation of outcome in the intensive care unit: findings from the Acute Physiology and Chronic Health Evaluation III study.** *Crit Care Med* 1993, **21**:1459-1465.
19. Knaus WA, Draper EA, Wagner DP, Zimmerman JE: **APACHE II: a severity of disease classification system.** *Crit Care Med* 1985, **13**:818-829.
20. Zubrod CG, Schneiderman M, Frei III E, Brindley C, Gold GL, Shnider B, Oviedo R, Gorman J, Jones R Jr, Jonsson U, Colsky J, Chalmers T, Ferguson B, Dederick M, Holland J, Selawry O, Regelson W, Lasagna L, Owens AH Jr: **Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and triethylene thiophosphoramide.** *J Chron Dis* 1960, **11**:7-33.
21. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG: **The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine.** *Intensive Care Med* 1996, **22**:707-710.
22. Le Gall J-R, Lemeshow S, Saulnier F: **A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study.** *JAMA* 1993, **270**:2957-2963.
23. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J: **Mortality probability models (MPM II) based on an international cohort of intensive care unit patients.** *JAMA* 1993, **270**:2478-2486.
24. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, Harrel FE: **The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults.** *Chest* 1991, **100**:1619-1636.
25. Hanley JA, McNeil BJ: **The meaning and use of the area under receiver operating characteristic (ROC) curve.** *Radiology* 1982, **143**:29-36.
26. Lemeshow S, Hosmer DW: **A review of goodness of fit statistics for use in the development of logistic regression model.** *Am J Epidemiol* 1982, **115**:92-106.
27. Moreno R, Matos R: **The 'new' scores: what problems have been fixed, and what remain?** *Curr Opin Crit Care* 2000, **6**:158-165.
28. Lloyd-Thomas AR, Wright I, Lister TA, Hinds CJ: **Prognosis of patients receiving intensive care for lifethreatening medical complications of haematological malignancy.** *BMJ* 1988, **296**:1025-1029.
29. Giangiuliani G, Gui D, Bonatti P, Tozzi P, Caracciolo F: **APACHE II in surgical lung carcinoma patients.** *Chest* 1990, **98**:627-630.
30. Abbott RR, Setter M, Chan S, Choi K: **APACHE II: prediction of outcome of 451 ICU oncology admissions in a community hospital.** *Ann Oncol* 1991, **2**:571-574.
31. Headley J, Theriault R, Smith TL: **Independent validation of APACHE II severity of illness score for predicting mortality in**

patients with breast cancer admitted to the intensive care unit. *Cancer* 1992, **70:**497-503.

32. Van le L, Fakhry S, Walton LA, Moore DH, Fowler WC, Rutledge R: **Use of the APACHE II scoring system to determine mortality of gynecologic oncology patients in the intensive care unit.** *Obstet Gynecol* 1995, **85:**53-56.

33. Groeger JS, White P Jr, Nierman DM, Glassman J, Shi W, Horak D, Price K: **Outcome for cancer patients requiring mechanical ventilation.** *J Clin Oncol* 1999, **17:**991-997.

34. Massion PB, Dive AM, Doyen C, Bulpa P, Jamart J, Bosly A, Installé E: **Prognosis of hematologic malignancies does not predict intensive care unit mortality.** *Crit Care Med* 2002, **30:**2260-2270.

35. Price KJ, Thall PF, Kish SK, Shannon VR, Andersson BS: **Prognostic indicators for blood and marrow transplant patients admitted to an intensive care unit.** *Am J Respir Crit Care Med* 1998, **158:**876-884.

36. Bach PB, Schrag D, Nierman DM, Horak D, White P Jr, Young JW, Groeger JS: **Identification of poor prognostic features among patients requiring mechanical ventilation after hematopoietic stem cell transplantation.** *Blood* 2001, **98:**3234-3240.

37. Azoulay E, Pochard F, Garrouste-Orgeas M, Moreau D, Montesino L, Adrie C, de Lassence A, Cohen Y, Timsit J-F: **Decisions to forgo life-sustaining therapy in ICU patients independently predict hospital death.** *Intensive Care Med* 2003, **29:**1895-1901.

38. Moreno R, Matos R: **Outcome prediction in intensive care. Solving the paradox.** *Intensive Care Med* 2001, **27:**962-964.