

Jean-Roger LE GALL

Service de Réanimation Médicale, Hôpital Saint Louis
Paris, France

Sunday June 1, 2003

Euroanaesthesia 2003 - Glasgow

The third generation of ICU scoring systems consists of the APACHE III,¹ the MPM II,² and the SAPS II³ systems, all of which are based on logistic regression modelling techniques. These systems are now recommended for use by their developers. Scoring systems have been proposed for individual patient prediction, for evaluating the performance of intensive care units, and for therapeutic trials. In general, their use has been proposed for scores or probabilities at the level of both the individual patient and of the group. Thus, one may use a score to make a prediction about a particular patient, or about groups of patients. As serious consequences may arise from actions taken in response to such predictions, a conservative approach to the application of scores to individuals is called for. In spite of all of the careful research that has produced the various severity scoring systems, their uses are not yet universally agreed upon.^{4,5}

PREDICTIONS FOR INDIVIDUAL PATIENTS

Using either an objective risk of death or a clinical assessment, Meyer *et al*¹⁰ showed that, among the patients who were predicted to die by either method, over 40% actually survived. They concluded that no method could reliably predict the mortality of surgical ICU patients. This illustrates the difficulty that exists in predicting whether an individual patient will live or die from an estimated probability of mortality. A good system will provide an accurate estimate of the number of patients expected to die among a group of similar patients ; it cannot provide a prediction of which particular patients will actually die. Using a well-calibrated severity model, we can reasonably expect that approximately 75% of patients with a probability of mortality of 0.75 will die, but we cannot know in advance which patients will form the 25% who survive. That 25% will not have defied the odds — they will have confirmed the validity of the probabilities.

The possibility of augmenting clinical decisions by having an objective (although not always more accurate) assessment of a patient's severity of illness is appealing. Physicians are interested in severity systems for individual patients as an adjunct to their informed but subjective opinion. Using these tools as a part of the decision-making pathway is a reasonable and prudent choice, although using them in a vacuum to dictate individual patient decisions is not appropriate. Decisions will and should remain the responsibility of the individual physician and should be based on a number of criteria - one of which is severity as estimated by a well-calibrated scoring system.

EVALUATION OF ICU PERFORMANCE

Using the APACHE III system, Knaus *et al*¹¹ estimated the probabilities of hospital mortality in a consecutive sample of 16,622 patients from 42 ICUs and noted the actual outcome. They observed that the ratio of observed to expected number of deaths varied from .67 to 1.21 across the ICUs. Thus in some units the observed mortality was lower than expected by the models, and in others it was higher. Similarly, using the SAPS II system, Le Gall *et al*¹² estimated the probabilities of hospital mortality and noted the actual outcome in ICUs in several countries. The ratio varied among the countries from .74 to 1.31, i.e. some countries had fewer deaths than expected and others had more.

One cannot conclude from these findings, however, that clinical performance in certain ICUs or certain countries is necessarily below par when the observed mortality is higher than expected, or that it is necessarily superior when the observed mortality is lower than expected. To interpret these ratios, we need to know to what extent they are being influenced by factors other than clinical performance. They are best interpreted as indicators of the need to investigate the various ICUs more deeply to identify factors associated with the observed mortality. The probabilities themselves do not effectively control for all of the differences that may have an impact on outcome, e.g. differences in patient mix or disparities in availability of technical and therapeutic resources. Neither can they control for administrative differences or staffing levels (nurses per bed, for example). Only after taking such factors into consideration can meaningful evaluations and comparisons be made.

THERAPEUTIC TRIALS

This discussion is specifically oriented toward therapeutic trials for sepsis, but the issues involved can be applied to clinical trials involving any disease or condition and any proposed new therapy. While some authors¹³ have outlined the importance of preexisting comorbidity for prognosis of septicemia in critically ill patients, others¹⁴ have shown by multivariate analysis that, using initial score, etiology (urosepsis or other), and treatment localization prior to ICU admission provided the greatest degree of discrimination (ROC = .82) of patients for risk of hospital death.

A complex model for sepsis derived from a large data base, using physiology, primary disease, previous intensive care, age, history of cirrhosis, along with other variables, has been recently published.¹⁵ It has been proposed for use in clinical trials in which sepsis is the sole pathology. In the data base from which the model was developed, however, the manner in which disease, spectrum of disease, and inclusion criteria were specified, may differ from that specified for a proposed trial of a new therapy for sepsis. In general, it is unlikely that precise inclusion or exclusion criteria for a specific trial would have affected the compilation of the original data base from which the model was developed. Nor is it reasonable to expect that a large, general medical/surgical data base would contain all the information that could address all the requirements of actual and future trials. Although this should not deter use of such models, it should make investigators wary of comparisons between an expected mortality rate, given by a model derived from a large data base, and an observed mortality rate in a more precisely defined treatment group. The probability can be used to stratify patients by level of severity at the onset of the trial, but conclusions about observed and expected outcome should be drawn with care.

In a recent critique of scoring systems, Loirat¹⁶ suggested using a simpler tool that did not assign weights for acute diseases. Such a disease-independent assessment of severity could be used to derive a disease-specific model using half of the patients in a control group. The model would then be applied to the patients in the other half of the control group and the patients in the treatment group, and comparisons of observed and expected outcome between the two groups could then be made.

It must also be noted that the present general models have all been developed for use at very specific time periods, either at admission to the ICU (MPM₀), during the first 24 hours of the ICU stay (SAPS II and APACHE III), or at three 24-hour time points of the ICU stay (MPM₂₄, MPM₄₈, and MPM₇₂). These models are not automatically transferable for use in stratifying patients for randomization in a clinical trial, at times outside the limits for which the models are applicable. Some research is necessary to confirm that severity at the time of randomization is accurately measured by these models, *i.e.*, to confirm that they are well-calibrated at the intended time period.

CONCLUSION

In an editorial, Selker¹⁷ stated that the desirable characteristics of risk-adjusted mortality predictors are that they be:

- time-insensitive predictive instruments
- based on the first minutes of hospital presentation
- not affected by whether a patient is hospitalized
- based on data collected during normal patient care
- precisely calibrated
- integrated into computer systems
- independent of the Diagnostic Related Groups system
- open for inspection and testing.

These criteria are probably utopian, and the ideal scoring system remains to be discovered. The available ICU scoring systems reviewed in this paper are, however, based on rigorous research and have reported excellent calibration and discrimination. The uses to which they can best and most appropriately be put may be subject to some debate, but all are useful in both research and clinically when used in the proper context.

A. COMPUTING THE SCORES FOR APACHE III AND SAPS II

Calculate either score by summing the points associated with the levels of the physiology variables, age, and chronic health comorbidities. The maximum point assignment for each variable in the two systems is shown in the table below.

TABLE 1

VARIABLES	SAPS II	APACHE III
CHRONIC HEALTH STATUS*		
AIDS	17	23
Cirrhosis		4
Lymphoma		13
Hematologic malignancy	10	
Leukemia/multiple myeloma		10
Hepatic failure		16
Metastatic cancer	9	11
Immunosuppression		10
PHYSIOLOGY		
Temperature	3	20
Heart Rate	11	17
Respiratory Rate		18
Blood pressure (mean or systolic)	13	23
Hematocrit		3
White blood cell count	12	19
Albumin		11
Bilirubin	9	16
Glucose		9
Serum Na	5	4
Serum K	3	
Serum HCO ₃	6	
Blood urea or BUN	10	12
Creatinine		10
Urine Output	11	15
PaO ₂ or (A-a)DO ₂		15
PaO ₂ /FiO ₂ (if ventilated or CPAP)	11	
pH & pCO ₂ (acid-base disturbances)		12
Glasgow Coma Scale or modified	26	48
OTHER		
Age	18	24
Type of admission	8	

*APACHE III: if two or more comorbid conditions are present, use the condition with the highest point assignment

B. COMPUTING THE LOGITS FOR THE RISK OF DEATH CALCULATIONS FOR ALL MODELS

APACHE III:

- Calculate the APACHE III score as described in Appendix A. The coefficient for the APACHE III score is 0.0537.
- Select one of 78 major disease categories, each of which has an associated coefficient, b_i , obtained from the developer.
- Determine whether the patient was a surgical or nonoperative admission:
 - >> For nonoperative patients, select a patient origin category to indicate patient location prior to ICU treatment, each location having an associated coefficient, b_j , obtained from the developer.
 - >> For patients admitted to the ICU after surgery, determine whether the surgery was performed on an emergency basis (Yes=1, No=0). The coefficient for emergency surgery is 0.0752.
- Calculate the logit as:

$$\text{logit} = b_i + (b_j \text{ or } 0.0752 * \text{Emergency Surgery}) + 0.0537 * \text{APACHE III score.}$$

SAPS II:

- Calculate the SAPS II score as described in Appendix A. The coefficient for the SAPS II score is 0.0737
- Calculate $\ln(\text{SAPS II score} + 1)$, for which the coefficient is 0.9971.
- Calculate the logit as follows, where -7.7631 is a constant term in the equation:

$$\text{logit} = -7.7631 + 0.0737 * \text{SAPS II Score} + 0.9971 * \ln(\text{SAPS II score} + 1).$$

MPM II:

- For each variable with a threshold value, determine whether the threshold value has been exceeded. For each variable coded as present or absent, determine the appropriate response.
- If a variable exceeds the threshold value or is present, the variable is assigned a value of 1, so when it is multiplied by its associated coefficient it takes on the value of the coefficient. If the threshold value is not exceeded or the variable is absent, it is assigned a value of 0, so when it is multiplied by its associated coefficient it takes on the value of 0, making no contribution to the logit.
- The coefficients for each variable are shown in the table below. Except for the constant term, they are the same for MPM₂₄, MPM₄₈, and MPM₇₂.
- Calculate the logit as the sum of the products of the variables and coefficients plus the correct constant term for the specific model.

VARIABLES	MPM II	MPM II		
	Admission	24, 48, 72 Hours		
CHRONIC HEALTH STATUS				
Cirrhosis	1.137		1.087	
Metastatic cancer	1.200		1.161	
Chronic renal insufficiency	0.919			
PHYSIOLOGY				
Heart Rate	0.456			
Systolic blood pressure	1.061			
Creatinine			0.723	
Urine Output			0.823	
PO ₂			0.467	
Prothrombin time			0.554	
Glasgow Coma Scale or modified	1.486		1.688	
ACUTE DIAGNOSES				
Acute renal failure	1.482			
Cardiac dysrhythmia	0.281			
Cerebrovascular incident	0.213			
Gastrointestinal bleeding	0.397			
Infection			0.497	
Intracranial mass effect	0.865		0.913	
OTHER				
Age	0.031		0.033	
Type of admission	1.191		0.834	
CPR prior to ICU admission	0.570			
Mechanical ventilation	0.791		0.808	
Vasoactive drug therapy			0.716	
CONSTANT	-5.468	-5.646	-5.392	-5.238
	Admission	24 Hrs	48 Hrs	72 Hrs

REFERENCES

1. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991;100:1619-1636.
2. 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.
3. 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.
4. Hadorn D, Keeler E, Rogers W, Brook R: Assessing the Performance of Mortality Prediction Models. Rand/UCLA/Harvard Center for Health Care Financing Policy Research. 1993.
5. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: John Wiley & Sons; 1989.
6. Cleveland WS. LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*. 1981;35:54.
7. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
8. Teres D, Lemeshow S. Why severity models should be used with caution. In: Schuster DP, Kollef MH, eds. *Critical Care Clinics*. Philadelphia: W.B. Saunders Company; 1994:93-110.
9. Watts CM, Knaus WA. Comment on "Why severity models should be used with caution." In: Schuster DP, Kollef MH, eds. *Critical Care Clinics*. Philadelphia: W.B. Saunders Company; 1994:111-115.
10. Meyer AA, Messick WJ, Young R, et al. Prospective comparison of clinical judgment and APACHE II score in predicting the outcome in critically ill surgical patients. *J Trauma*. 1992;32(6):747-753.
11. Knaus WA, Wagner DP, Zimmerman JE, Draper EA. Variation in mortality and length of stay in intensive care units. *Ann Intern Med*. 1993; 118:753-761.
12. Le Gall JR, Artigas A, Lemeshow S, Saulnier F, Avrunin J. Une comparaison internationale des unités de réanimation. *Rean Urg*. 1993; 6:656(Abstract).
13. Pittet D, Thievent B, Wenzel RC, Gurman G, Suter PM. Importance of preexisting comorbidities for prognosis of septicemia in critically ill patients. *Int Care Med*. 1993; 19:265-272.
14. Knaus WA, Sun X, Nystrom PR, Wagner DP. Evaluation of definitions for sepsis. *Chest*. 1992;101:1656-1662.
15. Knaus WA, Harrel FE, Fisher CJ, et al. The clinical evaluation of new drugs for sepsis. A prospective study design based on survival analysis. *JAMA*. 1993;270:1233-1241.
16. Loirat P. Critique of existing scoring systems: Admission scores. *Intensive Care Med*. 1994
17. Selker HP. Systems for comparing actual and predicted mortality rates: Characteristics to promote cooperation in improving hospital care. *Annals of Intern Med*. 1993;118:820-822.