

Commentary

Predictive mortality models are not like fine wine

Andrew A Kramer

Senior Biostatistician, Cerner Corporation, 1953 Gallows Road, Suite 570, Vienna, VA 22182, USA

Corresponding author: Andrew Kramer, akramer@cerner.com

Published online: 26 October 2005

This article is online at <http://ccforum.com/content/9/6/636>

© 2005 BioMed Central Ltd

Critical Care 2005, 9:636-637 (DOI 10.1186/cc3899)

See related research by Le Gall *et al.* in this issue [<http://ccforum.com/content/9/6/R645>]

Abstract

The authors of a recent paper have described an updated simplified acute physiology score (SAPS) II mortality model developed on patient data from 1998 to 1999. Hospital mortality models have a limited range of applicability. SAPS II, Acute Physiology, Age, and Chronic Health Evaluation (APACHE) III, and mortality probability model (MPM)-II, which were developed in the early 1990s, have shown a decline in predictive accuracy as the models age. The deterioration in accuracy is manifested by a decline in the models' calibration. In particular, mortality tends to get over predicted when older models are applied to more contemporary data, which in turn leads to 'grade inflation' when benchmarking intensive care unit (ICU) performance. Although the authors claim that their updated SAPS II can be used for benchmarking ICU performance, it seems likely that this model might already be out of calibration for patient data collected in 2005 and beyond. Thus, the updated SAPS II model may be interesting for historical purposes, but it is doubtful that it can be an accurate tool for benchmarking data from contemporary populations.

Le Gall *et al.* [1] have described an updated simplified acute physiology score (SAPS) II mortality model that was customized and expanded using 1998 to 1999 patient data from France. The original SAPS II model [2] has been used to predict hospital mortality in Europe and other parts of the world. SAPS II shares many elements in common with other methodologies such as Acute Physiology, Age, and Chronic Health Evaluation (APACHE) III [3] and mortality probability model (MPM)₀-II [4], which have been more commonly used for US populations. Studies employing these models, which were developed in the early 1990s, to predict mortality in more contemporary patient databases from the US [5] and the UK [6] show that the accuracy of these mortality predictions has deteriorated. The deterioration has not been as much in discrimination (the ability to distinguish survivors and non-survivors) as in calibration (the correspondence of observed and predicted mortality). In particular, mortality

tends to get over predicted when older models are applied to more contemporary data, which in turn leads to 'grade inflation' when benchmarking intensive care unit (ICU) performance [7]. It is thus not surprising that Le Gall *et al.* [1] found similar results when applying the original SAPS II model (based on data from 1991 to 1992) to a 'newer' data set (1998 to 1999). A mortality model developed for US Veterans Administration patients [8] and a new generation of mortality models (APACHE IV, MPM₀-III, and SAPS III) have been developed to address this well-documented phenomenon of 'model fade'.

It is thus puzzling why the authors claim that their model is "a tool suitable for benchmarking" [1]. Instead it seems likely that the updated and expanded model presented by Le Gall *et al.* might already be out of calibration for patient data collected in 2005 and beyond. The authors concede as much when they apologize for the age of their data and state that, "Nevertheless, for *historical comparisons* (emphasis mine), the expanded SAPS II can be easily obtained from existing databases". Further, the authors also acknowledge that a different SAPS model, SAPS III "the more recent and sophisticated model", is currently under evaluation. Although the patient sample used to develop SAPS III is not large [9], it is based on more contemporary data.

There are some serious concerns about the patient mix in this study. First, Le Gall *et al.* state that some ICUs were in fact "intermediate units with only monitored patients" [1]. Mortality at these units is likely to be different from that at ICUs, resulting in models with coefficients optimized for this diluted population [10]. This would compound the effects caused by the age of the data and make benchmarking to contemporary ICUs even more problematic. Second, there is the potential for bias from inadequate collection of cohort data; "Among the 106 ICUs, 22 (21%) failed to provide the SAPS II score

APACHE = Acute Physiology, Age, and Chronic Health Evaluation score; ICU = intensive care unit; MPM = mortality probability model; SAPS = simplified acute physiology score.

for over 20% of admissions" [1]. What are the characteristics of these ICUs and how do they compare with the 84 ICUs that provided more complete data? Were certain patient groups more likely to have a missing SAPS II score and, if so, then would this bias the results? These questions were not addressed in the paper. Third, the frequency of drug overdose patients is very high (11%) and mortality was greatly overestimated in this group. Because of these findings the authors make an exception to their rule of not including diagnostic variables and add a binary variable for the drug overdose patients. In effect, they are acknowledging that diagnostic information is useful in mortality models. They are correct in this assumption as demonstrated by the accuracy among diagnostic subgroups shown in the APACHE models, and they should seriously consider adding more of such variables to their model. The authors go on to state, however, that the inclusion of diagnostic group variables will result in poor calibration across patient groups. This contradicts their including a variable for drug overdose patients.

In summary, unlike fine wine, models for predicting ICU mortality do not age well. The article by Le Gall *et al.* provides an interesting footnote in the history of critical care mortality models. Beyond that it is equivocal whether their 'updated' model provides any tangible benefit.

Competing interests

Dr Kramer is an employee of and shareholder in Cerner Corporation, which owns the rights to the APACHE and MPM predictive models.

References

1. Le Gall JR, Neumann A, Hemery F, Bleriot JP, Fulgencio JP, Garrigues B, Gouzes C, LePage E, Moine P, Villers D: **Mortality prediction using SAPS II: an update for French intensive care units.** *Critical Care* 9:R645-R652.
2. Le Gall JR, Lemeshow S, Saulnier F: **A new Simplified Acute Physiology Score (SAPS II) based on an European/North American multicenter study.** *J Am Med Assoc* 1993, **270**:2957-2963.
3. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, Harell FE: **The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults.** *Chest* 1991, **100**: 1619-1636.
4. 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 patients.** *J Am Med Assoc* 1994, **270**:2478-2486.
5. Glance LG, Osler TM, Dick A: **Rating the quality of intensive care units: Is it a function of the intensive care unit scoring system?** *Crit Care Med* 2002, **30**:1976-1982.
6. Livingston BM, MacKirdy FN, Howie JC, Jones R, Norrie JD: **Assessment of the performance of five intensive care scoring models within a large Scottish database.** *Crit Care Med* 2000, **28**:1820-1827.
7. Popovich MJ: **If most intensive care units are graduating with honors, is it genuine quality or grade inflation?** *Crit Care Med* 2002, **30**:2145-2146.
8. Render ML, Kim M, Deddens J, Sivaganesin S, Welsh DE, Bickel K, Freyberg R, Timmons S, Johnston J, Connors AF, *et al.*: **Variation in outcomes in Veterans Affairs intensive care units with a computerized severity measure.** *Crit Care Med* 2005, **33**: 930-939.
9. Metnitz PGH, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzi M, Le Gall JR: **SAPS3 – From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description.** *Intensive Care Med* 2005, **31**:1336-1344.
10. Junker C, Zimmerman JE, Alzola C, Draper EA, Wagner DP: **A multicenter description of intermediate-care patients: comparison with ICU low-risk monitor patients.** *Chest* 2002, **121**: 1253-1261.