Intensive Care Med (2000) 26: 263–264

EDITORIAL

D. Teres
P. Pekow

Assessing data elements in a severity scoring system

Accepted: 3 January 2000

D. Teres (✉)
Tufts University School of Medicine,
Boston, Mass, 02111, USA
e-mail: danielteres@aol.com
Tel.: + 1-781-444-7006,
Fax: + 1-781-444-7006

P. Pekow
University of Massachusetts School of Public Health
and Health Sciences, Amherst, Mass, USA
e-mail: penny.pekow@bhs.org

The study by Rue et al., "Interobserver variability of the measurement of the Mortality Probability Models (MPM 11) in the assessment of severity-of-illness," is a careful analysis of the quality of data in the Mortality Probability Mortality system [1]. There are implications from this study for all severity systems.

In Rue's study design, the assessment of the physician attending at the bedside was compared to a research physician's retrospective chart review after the patient had been discharged from the hospital. MPM had been utilized in these hospitals before for previous studies; in addition, the bedside intensivist went through a formal training program before the start of data collection in order to standardize the study process.

There was generally high interobserver agreement for the majority of parameters. One notable exception was the term "chronic renal insufficiency," while the more objective measures of kidney dysfunction (elevated creatinine, low urine output and acute renal failure) had a very good rate of agreement. There may be multiple reasons for the poor field testing of the term "chronic renal insufficiency", including low frequency in the population, poor operational definition, inconsistent location of charting in the medical record or over-interpretation by the bedside physician. The clinician recognizes that there are differences in prognosis ascribed to pre-renal failure, obstructive uropathy, chronic dialysis, acute on chronic renal failure, acute tubular necrosis as well as chronic renal insufficiency. However, from the modeling perspective only three factors: elevated creatinine, low urine output and acute renal failure (whatever the cause) were "recognized". Rue's study suggests that in the next upgrade of the risk-based severity system, chronic renal insufficiency should probably be dropped, if only because there is no consistent location for charting this variable.

The second "worst" variable was a physiological one, low partial pressure of oxygen, while the other acute and chronic diagnoses had acceptable levels of interobserver agreement. This study is important because it shows that it may be just as difficult to collect physiological data in a consistent fashion as it is to collect complex risk factors and conditions such as coma/deep stupor and confirmed infection. The ritual of taking hourly vital signs and medical charting may not be as standardized as one would presume [2].

Hypoxia is an important factor in all of the severity models and here there must be better ways of standardizing the process of data collection. It should be understood that in physiology-based severity models there is a presumption that screening by experienced ICU nurses is routinely carried out to minimize the "noise" associated with the continuous monitoring of especially extreme, transient, non-clinically relevant values. For a patient on a ventilator or mask oxygen, there are fluctuations in the partial pressure of oxygen during suctioning, movement, chest physiotherapy, tube disconnection, spells of agitation or simply if he/she pulls off the mask or cannulae. Which value should be recorded on the hourly vital sign sheet? Which value should be recorded for the severity model? It would not be surprising if the bedside physician recorded a clinically rele-

vant value for the severity model while the nurse recorded a different value on the hourly vital sign sheet. In addition, the partial pressure of oxygen can be estimated in multiple ways (transcutaneous, arterial blood).

Rue's study does not address what is the best approach to data collection. Other studies have compared bedside nurses and house staff [3]. In this era of cost containment, it would be inappropriate to ask an overly burdened ICU nurse or house officer also to record data to maintain an ICU database, even though consistency in charting of the medical record is an essential element of quality care. Project IMPACT [4] was validated by having a trained research nurse retrospectively evaluate key elements of the system compared to ICU data collectors.

One "best practice" approach would follow phase lll clinical drug trials, where an independent research nurse collects data at regular intervals while a research monitor provides a constant overview of missing or inconsistent data. Perhaps an approach worth investigating in the future would be to have the clinical team summarize clinical data on a computer with an Internet-based data management system [5]. With a simple user interface, a common data set can be created for the clinical record as well as for a registry or centralized data system. In this approach the data collection would be the responsibility of the clinical care team while the overview and monitoring would come from the research nurse or current data collector to guarantee that the data entries conform to the definitions.

One concern for manual systems is the amount of data needed. Daily organ failure systems, trauma registries, APACHE lll compared to APACHE ll, PRISM lll to PRISM ll all have large data requirements. There may be many reasons for the shift to more detailed or complex models. Clinicians may be more comfortable with more detailed models and, of course, more clinical research can be produced with a larger database.

Risk- and condition-based models such as MPM and preoperative cardiac surgery models do not have such extensive data element requirements. To its credit, the Simplified Acute Physiology Score (SAPS ll) has remained simple. Both MPM II and SAPS II form the required severity elements in the mandatory data components in Project IMPACT. Although clinicians prefer physiology-based models, the risk-based approach should be viewed as complementary and can also be used for assessing case-mix [6]. The original Burn score was simple and remarkably successful even though a major factor, inhalation injury, was not included [7].

As strongly suggested by Rue, constant and systematic monitoring of the reliability of severity measures is mandatory to maintain the integrity, consistency and acceptance of widely used severity systems. Careful studies like this one can also improve severity models.

References

1. Rue M, Valero C, Quintana S, et al. (2000) Interobserver variability of the measurement of the mortality probability models (MPM II) in the assessment of severity-of-illness. Intensive Care Med 26: 286–291
2. East TD, Young WH, Gardner RM (1992) Digital electronic communication between ICU ventilators and computers and printers. Respir Care Sep 37 (9): 1113–1123
3. Holt AW, Kimbury L, Bersten AD, et al. (1992) Prospective evaluation of residents and nurses as severity score data collectors. Crit Care Med 20: 1688–1691
4. Cook SF, Visscher WA, Hobbs CL, Williams RL and the Project IMPACT Clinical Implementation Committee (1999) Project IMPACT: preliminary results from a pilot validity study of a new observational database [abstract]. Crit Care Med 27 (suppl): A36
5. Gliklich, RE, Taghizadeh F (1999) Enabling technologies and the internet. In: Isenberg SF, Gliklich RE (eds) Profiting from Quality: Outcomes Strategies for Medical Practice. Jossey-Bass, San Francisco, pp 95–103
6. Murphy-Filkins RL, Teres D, Lemeshow S, Hosmer DW (1996) Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model: how to distinguish a general from a specialty intensive care unit. Crit Care Med 24: 1968–1973
7. Feller I, Tholen D, Cornell RG (1980) Improvements in burn care 1965–1979. JAMA 244: 2074–2078