Scoring Systems in Critical Care

Sandra D.W. Buttram, Paul R. Bakerman,
and Murray M. Pollack

Abstract

Today's health care environment is focused on providing both high quality and error-free care. Transparency is becoming an expectation, with many outcomes reported publically. Scoring systems are an objective measure which can be used to assess quality of care, assist with the evaluation and modification of complex systems of care, improve patient outcomes and predict morbidity and mortality. Their role has become secure in critical care because physician's judgments are too subjective for quality assessment in large samples.

The development of scoring systems began as external influences in the 1960s favored the assessment of outcome. Concerns about the quality of medical care escalated following the Institute of Medicine Report in 1999 which ultimately led to the Patient Safety and Quality Improvement Act of 2005. A successful scoring system must include a model with carefully defined predictor variables and outcome. The model must be reliable and the scoring system requires both internal and external validation.

Scoring systems were developed to predict mortality in adults (APACHE: Acute Physiology and Chronic Health Evaluations Score) and children (PRISM: Pediatric Risk of Mortality Score). They have also been used to predict morbidity and functional outcome. Scoring systems that apply to specific patient populations such as trauma and congenital heart disease have been developed.

Clinical scoring systems provide a standardized method for intensive care benchmarking and have increased in number and utility over the past 30 years. They are progressively more applicable to clinicians and health services researchers and in the future may be pertinent to individual patients. It is in the intensivist's best interest to understand scoring systems, their applications and implications.

S.D.W. Buttram, MD (✉) • P.R. Bakerman, MD
Critical Care Medicine, Phoenix Children's Hospital,
1919 E. Thomas Road, Phoenix, AZ 85016, USA
e-mail: sbuttram@phoenixchildrens.com;
pbakerm@phoenixchildrens.com

M.M. Pollack, MD (✉)
Department of Child Health,
University of Arizona College of Medicine – Phoenix,
1919 E. Thomas Road, Phoenix, AZ 85016, USA
e-mail: mpollack@phoenixchildrens.com

Introduction

Today's health care environment is focused on providing both high quality and error-free care. Transparency is becoming an expectation, with many outcomes reported publically. Comparative data bases with case-mix adjusted outcomes are available to many children's hospitals. High-profile programs such as pediatric cardiovascular surgery and pediatric critical care have access to national and international benchmarks

for outcomes. Parallel to the movement focusing on quality, there has been a national effort to reduce errors, especially those falling into the category of "never events". This movement was sparked by the Institute of Medicine's (IOM) 1999 report highlighting the need to reduce medical errors with subsequent recommendations in 2001 to improve quality and promote evidence-based practice [1, 2].

A common adage is "you can't improve what you can't measure". Scoring systems are usually an objective measure which can assess quality of care, assist with the evaluation and modification of complex systems of care, improve patient outcomes, and predict morbidity and mortality. Their role has become secure in critical care because physician's judgments are too subjective to be used for quality assessment in large samples, as well as the fact that there is a need for severity of illness assessments in clinical studies. Physician prognostication may be inaccurate for a number of reasons [3–11]. First, there are differences in a physician's ability to predict outcome based on the stage of the practitioner's career. Second, there is a tendency to overly weigh recent experience, particularly when experience with a specific condition is limited. Third, physicians may be unable to continuously account for all relevant clinical components that are important in predicting outcome. Finally, literature supporting evidence-based medicine is ever expanding and may exceed many clinicians' ability to remain current [12–19].

History

External influences have played a significant role in stimulating an environment that favors the assessment of outcome and the development of scoring systems designed to accomplish case-mix adjustment. In the 1960s, the US federal government focused its attention on social issues, including the healthcare safety net. Medicare and Medicaid programs were developed as fee for service plans, with access to healthcare services regardless of the ability to pay. Universal access to care and the concept of patient entitlement changed the perception of healthcare from a privilege to a right. Medical advances in areas such as dialysis and mechanical ventilation led to increasingly "high-tech care" and advanced the emerging specialty of critical care medicine. The result of these changes was increased utilization and cost. As healthcare costs exploded, there was increased focus on appropriate utilization of resources, the quality of these services, and the relationship between cost and quality (e.g. the value equation). These concerns also stimulated the need for objective scoring systems.

Concerns about the quality of medical care have escalated over the last 20 years. In the 1990s, the New England Journal of Medicine published a series of articles highlighting medical errors [20, 21]. Public interest in quality became more

visible following the Institute of Medicine Report in 1999 [1]. This and subsequent reports ultimately led to the Patient Safety and Quality Improvement Act of 2005, which led to the establishment of a system of patient safety organizations and a national patient safety database.

Intensive Care Units (ICUs) were early leaders in developing methods of quality assessment using accurate and reliable adjustments for case-mix differences. Prognostication in the ICU is essential to the debates about quality and cost. It is essential that the technologically advanced care provided in ICUs results in a meaningful outcome to patients and families. Mortality, morbidity and functional outcome prediction are central to these discussions. Thus, prognostic methods are a logical focus for intensive care physicians.

Early scoring systems such as the Glasgow Coma Scale and the APGAR score were developed to assess outcome in select populations. In the 1970s and 1980s, subsequent scoring systems were developed to appraise global ICU care and outcomes which allowed for evaluation of quantity of care, quality of care, and cost. The physiology-based scoring systems enabled case-mix adjustments and comparisons. Importantly, these developments led to the conclusion that there were differences in the practice patterns and quality among ICUs [22].

Physiology-based scoring systems were initially built on the concept that there is a direct relationship between mortality and the number of failing organ systems. Organ system failure results from physiologic derangements that were modeled to produce relatively accurate mortality risk estimates. Mortality prediction scores included the Acute Physiology and Chronic Health Evaluations Score (APACHE) and the Pediatric Risk of Mortality Score (PRISM). These early scoring systems have been modified and adapted, while others have been developed that evaluate morbidity or apply to select patient populations.

Use of Scoring Systems

Physicians and hospitals use internal and external benchmarking to assess quality of care. Benchmarking establishes an external standard reference to which performance levels can be compared and may be used to define "best practice". Internal benchmarking allows an organization to compare performance within itself while external benchmarking compares performance between hospitals or services such as ICUs. Scoring systems may also be incorporated into clinical pathways to remove subjective assessments and incorporate evidence based medicine. These pathways are developed to improve care quality, decrease variability between individual providers and improve the efficiency through which care is delivered. The ultimate goal is to deliver high quality care in the most cost effective manner.

Clinical trials often include scoring systems. Scoring systems can be used to control for case-mix index, control for severity of illness between treatment groups or aid in risk stratification of enrolled subjects. They can also be used to compare expected to observed outcomes. Mortality prediction scores such as APACHE and PRISM have commonly been used in this manner.

While it may seem attractive to apply probabilities to direct patient care, this practice is potentially problematic. Risk assessment is less reliable when applied to an individual patient. In particular, the real range (i.e. the 95 % confidence interval) of the computed estimate is often much larger than the user appreciates. This is especially relevant when the computed mortality risk is very high, but the confidence interval is very wide (i.e. imprecise). Perhaps most importantly, the prognostic performance of physicians is approximately equivalent to the performance of scoring systems for individual patients [7, 8, 23]. At this time, scoring systems are primarily intended to provide objective assessment of quality of care, to assess the effects of interventions provided by a healthcare system [19], and for use in severity of illness assessment for individual trials but should not be applied to individual patients.

Elements of a Scoring System

The important elements of a successful scoring system include outcome, predictor variables, and model. The outcome should be objective, clearly defined and relevant. Historically, mortality has been the primary outcome measure for ICU prognostication, with more recent interest in morbidity and functional outcomes.

Predictor (independent) variables should also be objective, clearly defined, reliably measured, mutually exclusive, applicable across institutions, and as free from lead-time bias as possible. To minimize bias associated with model development, these variables should be defined and collected a priori. Data elements may include diagnoses, physiologic status, physiologic reserve, response to therapy and intensity of interventions [24]. These elements must be logical for the intended use of the scoring system. For example, a recent effort to use Pediatric Index of Mortality (PIM2) for cardiovascular ICU patients led to poor predictor performance [25]. This was not surprising given, the paucity of acute physiological variables and cardiac diagnoses included in the score.

Model development is the next element in scoring system design. Typically, individual predictor variables are tested with a univariate analysis for statistical association to the outcome. The variables that are "loosely associated" with outcome (e.g. p<0.30) from the univariate analysis are then combined in a multivariate analysis. The type of multivariate analysis is outcome specific. Logistic regression is used for dichotomous outcomes such as survival/death. Linear regression is used for continuous variables such as length of stay. Multivariate linear or quadratic discriminate function analysis is most often used for categorical outcomes such as diagnoses [26, 27]. For each independent variable included in the model, a general guideline suggests that there should be at least ten outcome events (e.g. deaths) in the analysis.

Reliability

Reliability of the data elements and the model are vital to a successful scoring system. Clearly defined data elements, precise timing of data collection and standardized training for data collectors all contribute to high quality data acquisition. Reliability of the score can be measured within (*intra-rater*) or between (*inter-rater*) observers [24, 27]. The kappa (κ) statistic can be used to measure the level of agreement with 0 representing chance and 1 representing perfect agreement. The type of data determines the most appropriate reliability measurement. Dichotomous data uses the κ statistic, ordinal data uses the weighted κ statistic, and interval data uses the intraclass correlation coefficient [27].

Validity

Validation of a scoring system is the final test to determine whether the score measures what it was designed to measure. A scoring system is often first validated internally. Internal validation can be accomplished by using subsets of the population from which the score was derived. Three common techniques for internal validation include: data-splitting, cross-validation and bootstrapping [27, 28]. Data-splitting involves randomly dividing the sample into a training set and a validation set, with the training set used for initial model development. Cross-validation generates multiple training and validation sets through repeated data-splitting. Bootstrapping tests the model's performance on a large number of randomly drawn samples from the original population. If a score has good internal validity then it can be externally validated. External validation requires application of the scoring system to a patient population separate from the initial study.

Discrimination and *calibration* are two common statistical methods that are used to assess the model performance [24]. Discrimination is the ability of a model to distinguish between outcome groups, assessed by the area under the receiver operating characteristic curve. As the area under the curve (AUC) approaches one, the discrimination of the scoring system approaches perfection. The AUC of chance performance is 0.5 and for perfect performance is 1.0.

Calibration measures the correlation between the predicted outcomes and actual outcome over the entire range of risk prediction. The most accepted method for assessing calibration is the goodness-of-fit statistic proposed by Lemeshow and Hosmer [29].

Types of Scoring Systems

Scoring systems allow objective quantification of complex clinical states. They can be categorized by the type of predictor variables used to predict outcome. Examples include intervention specific, physiology specific, disease or condition specific and functional outcome scoring systems.

Intervention Specific Scoring Systems

Conceptually, more therapies are provided to sicker patients. The number of interventions that patients receive during their hospitalization can be associated with severity of illness. The Therapeutic Intervention Scoring System (TISS) [30] is an example of this type of scoring system and has been applied to pediatric patients [31]. The initial TISS score included 76 different therapeutic and monitoring interventions scored on a scale of 1–4 based on complexity and invasiveness. Interventions increase with severity of illness, thereby increasing the TISS score which predicts the risk of mortality. However, individual and institutional practice regarding use of interventions may vary which will affect the TISS score independent of the patient's physiology.

Physiology Specific Scoring Systems

Adult Mortality Scores

The most common adult ICU mortality scores are the APACHE, Mortality Probability Model (MPM), and Simplified Acute Physiology Score (SAPS). APACHE IV uses the worst physiologic values from the first 24 h of ICU admission. Weighted variables including age, physiologic data and chronic co-morbid conditions combined with major disease categories are used to predict hospital mortality [32]. The MPM III collects information at time of ICU admission (MPM$_0$). Age, physiologic data and acute and chronic diagnoses are included in the MPM III. The MPM III adds a "zero factor" term for elective surgical patients that have no risk factors identified by the MPM variables to accommodate for the low mortality risk in this patient population [33]. SAPS 3 is similar to MPM III collecting data at time of ICU admission to predict the probability of hospital mortality. SAPS 3 includes variables relating to patient characteristics before ICU admission and the circumstances

of ICU admission in addition to age and physiologic data [34]. Table 6.1 summarizes model characteristics for each of these adult mortality scoring systems.

Pediatric Mortality Scores

The two most common pediatric ICU mortality scores are the PRISM and the PIM. Pediatric scoring systems are similar to their adult counterparts in many respects. The most recent version, PRISM III, was developed from over 11,000 patients in 32 centers in the United States [35]. It consists of 17 physiologic variables collected within the first 12 h (PRISM III-12) or first 24 h (PRISM III-24) of ICU admission (Table 6.1). The most abnormal values at 12 and 24 h are recorded and used to predict the risk of mortality. PRISM has been externally validated [36] and is the first pediatric scoring system to be protected by site licenses.

The PIM2 was developed from over 20,000 patients in 14 centers in Australia, New Zealand and Great Britain (Table 6.1) [37]. Ten variables are collected within the period of time from initial ICU team patient contact, regardless of location, up to 1 h after ICU admission. Conceptually, data is collected early in treatment to more accurately reflect the patient's physiologic state rather than the quality of ICU treatment rendered. While lead time bias theoretically applies to all scoring systems, there is no evidence that data collected during the first 12 or 24 h of ICU admission adversely affects the validity of a model.

Pediatric Morbidity Scores

Morbidity occurs more commonly than mortality and offers an alternative method for assessment of quality of care. The Pediatric Multiple Organ Dysfunction Score (PEMOD) and Pediatric Logistic Organ Dysfunction (PELOD) were modified from previously developed adult scores [38–40] to describe complications and morbidity in pediatric ICU patients [41]. These scoring systems quantify organ system dysfunction based on objective criteria of severity. However, these scores were developed (PEMOD and PELOD) and validated (PELOD only) on relatively small patient populations [42].

Disease or Condition Specific Scoring Systems

Trauma

The most widely used and validated trauma scoring system is the Revised Trauma Score (RTS) [43]. Components of the RTS include Glasgow Coma Scale, systolic blood pressure and respiratory rate. The RTS was initially designed for use by prehospital care personnel to identify adult trauma patients that would benefit from care at a designated trauma center. Subsequently, it has been used to predict mortality from blunt and penetrating injuries and has been validated for use in pediatric trauma patients [44–46].

Table 6.1 Adult and pediatric physiology specific mortality score model design

	Adult			Pediatric	
	APACHE IV	MPM III	SAPS 3	PRISM III	PIM2
Outcome measure	Mortality	Mortality	Mortality	Mortality	Mortality
Developed/revised (year)	1980/2006	1985/2007	1984/2005	1988/1996	1997/2003
Population					
Development set	110,558	124,885	16,784	11,165	20,787
Internal validation set	40 %	40 %	20 %	10 %	7 centers
# ICU/# Hospitals	104/45	135/98	307/NA	NA/32	NA/14
Time of data collection	First 24 h ICU admission	ICU admission	ICU admission	12 h and 24 h from ICU admission	ICU admission
AUC/goodness of fit (validation set)	0.88/p=0.08	0.823/p=0.31 (MPM$_0$)	0.848/p=0.39	0.941/p=0.4168 (PRISM III-12) 0.944/p=0.5504 (PRISM III-24)	0.9/p=0.42
Comments	Worst physiologic value in 1st 24 h			Most abnormal values 1st 12 h and 2nd 12 h of ICU stay PRISM III-12 -1st 12 h of ICU stay PRISM III-24 – 1st 24 h of ICU stay	
References	Zimmerman 2006	Higgins 2007	Moreno 2005	Pollack 1996	Slater 2003

APACHE Acute Physiology and Chronic Health Evaluations, *MPM* Mortality Probability Model, *SAPS* Simplified Acute Physiology Score, *PRISM* Pediatric Risk of Mortality Score, *PIM* Pediatric Index of Mortality, *NA* not available, *ICU* intensive care unit, *AUC* area under the curve

The Abbreviated Injury Scale (AIS) was originally developed to quantify injuries sustained in motor vehicle accidents [47]. The Injury Severity Score (ISS) was adapted from the AIS [48]. The ISS categorizes injuries to six regions of the body and rates the injury severity on a scale of 1–5 (1=minor, 5=critical/survival uncertain). The Trauma Score and Injury Severity Score (TRISS) method combines the RTS and ISS to create a mortality prediction score [49]. TRISS is commonly used for trauma benchmarking and can be applied to the pediatric population [44, 45, 50].

Congenital Heart Disease

Congenital heart disease (CHD) patients are a classic example of a population that requires coordination of multiple services and care systems to achieve a good outcome. Scoring systems for congenital heart disease are based on the diagnosis rather than patient specific physiologic variables. There are three scoring systems that have been developed to assess outcome of these patients. The earliest were the Risk Assessment in Congenital Heart Surgery (RACHS-1) and the Aristotle Basic Complexity (ABC) scores. Both were developed by consensus of experts. In contrast, the most recently developed score, the Society of Thoracic Surgery and the European Association for Cardiothoracic Surgery (STS-EACTS) Congenital Heart Surgery mortality score was empirically derived by retrospective analysis of outcomes from the procedure list defined by the ABC score.

RACHS-1 was the first pediatric congenital heart surgery scoring system developed [51]. This scoring system was developed by an 11 member panel of pediatric cardiologists and cardiovascular surgeons. Data was initially analyzed from over 9,000 patients in the Pediatric Cardiac Care Consortium [52] and hospital discharge data. This score was subsequently refined utilizing multi-institutional databases [53]. Similarly, the ABC score utilized the expert opinions of 50 internationally based congenital heart surgeons to evaluate 145 surgical procedures based on potential mortality, morbidity and technical difficulty [54]. The ABC score was subsequently validated using more than 35,000 cases from the STS-EACTS database [55].

The STS-EACTS score was developed to measure mortality associated with CHD based on analysis of more than 77,000 cases in the STS-EACTS database between 2002 and 2007 [56]. In total, 148 procedures were classified in mortality risk categories from 1 to 5 (1=low mortality risk, 5=high mortality risk). The score was externally validated in 27,700 operations, and showed a higher degree of discrimination for predicting mortality than the RACHS-1 or ABC score [56]. As with previous scores, a major limitation is that the STS-EACTS score does not allow adjustment for patient-specific risk factors.

Other

There are many scoring systems for specific conditions that are potentially relevant to the management of ICU patients. Some examples include scoring systems for croup [57], asthma [58], bronchiolitis [59] and meningococcemia [60, 61]. These scores have been used for triage decision making and severity of illness measurement. However, the ability to validate these models is limited by small patient populations.

Functional Outcome Scores

Functional outcome scores such as Pediatric Cerebral Performance Category (PCPC) and Pediatric Overall Performance Category (POPC) are modified from the Glasgow Outcome Scale [62, 63]. These scores are used to assess short-term changes in cognition (PCPC) and physical disabilities (POPC). Their major drawback is the need for the observer to project a functional status. While the PCPC has been correlated to 1 and 6 month neuropsychological tests such as the Bayley and IQ testing, there is a very large distribution of the neuropsychological tests in PCPC categories. The lack of discrimination would necessitate very large sample sizes if the test were used for long term outcome [64].

The Functional Status Scale (FSS) was recently developed to objectively measure functional outcome across the entire pediatric age range [65]. FSS was developed through a consensus process of pediatric experts and measures six domains of functioning. Domains include mental status, sensory functioning, communication, motor functioning, feeding and respiratory status which are scored from 1 to 5 (1=normal, 5=very severe dysfunction). The Adaptive Behavior Assessment System II (ABAS II) was used to establish construct validity and calibration within each functional domain. The FSS showed very good discrimination with ABAS II categories.

The Future of Scoring Systems

The future of scoring systems is likely to be largely influenced by the changes occurring in the structure and organization of care. Very large databases will become available to health services researchers that will enable more accurate and reliable mortality predictions. More reliable outcome predictions with general models such as PRISM and with disease or condition specific models will be used to assess and benchmark the quality of care in individual pediatric ICUs (PICU). These databases may become large enough to build models with sufficient performance characteristics that they have applicability to the individual pediatric critical care patient. The interest in transparency may dictate public disclosure of this information.

More reliable morbidity assessment methods such as the FSS will shift the paradigm of critical care outcome measurement from mortality prediction to a more global evaluation of functional outcome and morbidity. This will focus quality assessment of PICU therapies on morbidity as well as mortality which will further stimulate advances in quality research, case-mix adjustment methods, and forecasting outcomes. PICU admission data combined with quality of care data will be used to forecast long-term pediatric disability. When the likelihood of new morbidity as well as death becomes part of the outcome prediction models, they will have much greater applicability and utility to individual patients and individual patient decisions.

Conclusion

Scoring systems in intensive care medicine have increased in number and utility over the past 30 years. They are increasingly applicable to clinicians and health services researchers and in the future may be applicable to individual patients. Clinical scoring systems provide a standardized method for ICU benchmarking and are required by governing health care bodies. Additionally, benchmarking information is used by health care payers in creating managed care contracts. Consumer demands for error-free medicine, quality improvement and transparency may lead to public scoring of hospitals and individual physicians. These forces will continue to increase the need for risk adjusted outcomes and institutional benchmarking. It is in the intensivist's best interest to understand scoring systems, their applications and implications.

References

1. Kohn LT, Corrigan JM, Donaldson MS, editors. To err is human: building a safer health system. Washington, DC: National Academy Press; 1999.
2. Institute of Medicine. Crossing the quality chasm: a new health system for the twenty-first century. Washington, DC: National Academies Press; 2001.
3. Perkins HS, Jonsen AR, Epstein WV. Providers as predictors: using outcome predictions in intensive care. Crit Care Med. 1986;14:105–10.
4. Knaus WA, Wagner DP, Lynn J. Short-term mortality predictions for critically ill hospitalized adults: science and ethics. Science. 1991;254:389–94.
5. Poses RM, Bekes C, Copare FJ, et al. The answer to "what are my chances, doctor?" Depends on whom is asked: prognostic disagreement and inaccuracy for critically ill patients. Crit Care Med. 1989;17:827–33.
6. Poses RM, Bekes C, Winkler RL, et al. Are two (inexperienced) heads better than one (experienced) head? Averaging house officers' prognostic judgments for critically ill patients. Arch Intern Med. 1990;150:1874–8.
7. Stevens SM, Richardson DK, Gray JE, et al. Estimating neonatal mortality risk: an analysis of clinicians' judgments. Pediatrics. 1994;93:945–50.
8. McClish DK, Powell SH. How well can physicians estimate mortality in a medical intensive care unit? Med Decis Making. 1989;9:125–32.
9. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. Science. 1989;243:1668–74.
10. Kruse JA, Thill-Baharozian MC, Carlson RW. Comparison of clinical assessment with APACE II for predicting mortality risk in patients admitted to a medical intensive care unit. JAMA. 1988;260:1739–42.
11. Marcin JP, Pollack MM, Patel KM, et al. Prognostication and certainty in the Pediatric Intensive Care Unit. Pediatrics. 1999; 104(4):868–73.

12. Hanson CW, Marshall BE. Artificial intelligence applications in the intensive care unit. Crit Care Med. 2001;29:427–35.
13. Iberti TJ, Fischer EP, Leibowitz AB, et al. A multicenter study of physicians' knowledge of the pulmonary artery catheter. JAMA. 1990;264:2928–32.
14. Levetown M, Pollack MM, Cuerdon TT, et al. Limitations and withdrawals of medical intervention in pediatric critical care. JAMA. 1994;272:1271–5.
15. Morris AH. Developing and implementing computerized protocols for standardization of clinical decisions. Ann Intern Med. 2000;132:373–83.
16. Tversky A, Kahneman D. Availability. A heuristic for judging frequency and probability. In: Kahneman D, Slovic P, Tversky A, editors. Judgment under uncertainty: heuristics and biases. New York: Cambridge University Press; 1982. p. 163–78.
17. Jennings D, Amabile T, Ross L. Informal covariation assessment: data-based versus theory-based judgments. In: Kahneman D, Slovic P, Tversky A, editors. Judgment under uncertainty: heuristics and biases. New York: Cambridge University Press; 1982. p. 211–30.
18. Miller G. The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol Rev. 1956;63:81–97.
19. Pearson GA. Mathematical morbidity in paediatric intensive care. Lancet. 2003;362:180–1.
20. Brennan TA, Leape LL, Laird N, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. N Engl J Med. 1991;324:370–6.
21. Leape LL, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients: results of the Harvard Medical Practice Study II. N Engl J Med. 1991;324:377–84.
22. Pollack MM, Ruttimann UE, Getson PR. Accurate prediction of the outcome of pediatric intensive care. A new quantitative method. N Engl J Med. 1987;316:134–9.
23. Marcin JP, Pollack MM, Patel KM, et al. Combining physician's subjective and physiology-based objective mortality risk predictions. Crit Care Med. 2000;28:2984–90.
24. Pollack MM. Prediction of outcome. In: Fuhrman BP, Zimmerman JJ, editors. Pediatric critical care. 2nd ed. St. Louis: Mosby; 1998.
25. Pollack MM. It is what it was. Pediatr Crit Care Med. 2011;12:228–9.
26. Hand DJ. Statistical methods in diagnosis. Stat Methods Med Res. 1992;1:49–67.
27. Marcin JP, Pollack MM. Review of the methodologies and applications of scoring systems in neonatal and pediatric intensive care. Pediatr Crit Care Med. 2000;1:20–7.
28. Ruttimann UE. Statistical approaches to development and validation of predictive instruments. Crit Care Clin. 1994;10:19–35.
29. Lemeshow S, Hosmer DW. A review of goodness-of-fit statistics for use in the development of logistic regression models. Am J Epidemiol. 1982;115:92–106.
30. Cullen DJ, Civetta JM, Briggs BA, Ferrara LC. Therapeutic intervention scoring system: a method for quantitative comparison of patient care. Crit Care Med. 1974;2:57–60.
31. Pollack MM, Yeh TS, Ruttimann UE, et al. Evaluation of pediatric intensive care. Crit Care Med. 1984;12:376–83.
32. Zimmerman JE, Kramer AA, McNair DS, et al. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006;34:1297–310.
33. Higgins TL, Teres D, Copes WS, et al. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM$_0$-III)*. Crit Care Med. 2007;35:827–35.
34. Moreno RP, Metnitz PG, Almeida E, et al. SAPS 3-from evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med. 2005;31:1345–55.
35. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. Crit Care Med. 1996;24:743–52.
36. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. Crit Care Med. 1988;16:1110–6.
37. Slater A, Shann F, Pearson G, et al. PIM2: a revised version of the Paediatric Index of Mortality. Intensive Care Med. 2003;29:278–85.
38. Marshall JC, Cook DJ, Christou NV, et al. Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. Crit Care Med. 1995;23:1638–52.
39. LeGall JR, Klar J, Lemeshow S, et al. The logistic organ dysfunction system a new way to assess organ dysfunction in the intensive care unit. JAMA. 1996;276:802–10.
40. Vincent JL, Moreno R, Takala J, et al. 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;22:707–10.
41. Leteurtre S, Martinot A, Duhamel A, et al. Development of a pediatric multiple organ dysfunction score: use of two strategies. Med Decis Making. 1999;19:399–410.
42. Leteurtre S, Martinot A, Duhamel A, et al. Validation of the pediatric logistic organ dysfunction score: prospective, observational, multicentre study. Lancet. 2003;362:192–7.
43. Champion HR, Sacco WJ, Copes WS, et al. A revision of the trauma score. J Trauma. 1989;29:623–9.
44. Kaufmann CR, Maier RV, Kaufmann EJ, et al. Validity of applying adult TRISS analysis to injured children. J Trauma. 1991;31:691–8.
45. Orliaguet G, Meyer P, Blanot S, et al. Validity of applying TRISS analysis to paediatric blunt trauma patients managed in a French paediatric level I trauma centre. Intensive Care Med. 2001;27:743–50.
46. Eichelberger MR, Bowman LM, Sacco WJ, et al. Trauma score versus revised trauma score in TRISS to predict outcome in children with blunt trauma. Ann Emerg Med. 1989;18:939–42.
47. Baker SP, O'Neill B, Haddon W, et al. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma. 1974;14:187–96.
48. Copes WS, Lawrick M, Champion HR, et al. A comparison of abbreviated injury scale 1980 and 1985 version. J Trauma. 1988;28:78–86.
49. Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. J Trauma. 1987;27:370–8.
50. Eichelberger MR, Mangubat EA, Sacco WS, et al. Comparative outcomes of children and adults suffering blunt trauma. J Trauma. 1988;28:430–4.
51. Jenkins KJ, Gauvreau K, Newburger JW, et al. Consensus-based method for risk adjustment for surgery for congenital heart disease. J Thorac Cardiovasc Surg. 2002;123:110–8.
52. Moller JH. Perspectives in pediatric cardiology, Surgery of congenital heart disease: pediatric cardiac care consortium 1984–1995, vol. 6. New York: Futura Publishing; 1998. p. 279–329.
53. Jenkins KJ, Gauvreau K. Center specific differences in mortality: preliminary analyses using the Risk Adjustment in Congenital Heart Surgery (RACHS-1) method. J Thorac Cardiovasc Surg. 2002;124:97–104.
54. Lacour-Gayet F, Clarke D, Jacobs J, et al. The Aristotle score for congenital heart surgery. Pediatr Card Surg Annu Sem Thorac Cardiovasc Surg. 2004;7:185–91.
55. O'Brien SM, Jacobs JP, Clarke DR, et al. Accuracy of the Aristotle basic complexity score for classifying the mortality and morbidity potential of congenital heart surgery operations. Ann Thorac Surg. 2007;84:2027–37.
56. O'Brien SM, Clarke DR, Jacobs JP, et al. An empirically based tool for analyzing mortality associated with congenital heart disease. J Thorac Cardiovasc Surg. 2009;138:1139–53.

57. Jacobs S, Shortland G, Warner J, et al. Validation of a croup score and its use in triaging children with croup. Anaesthesia. 1994; 49:903–6.

58. Wood DW, Downes JJ, Lecks HI. A clinical scoring system for the diagnosis of respiratory failure. Preliminary report on childhood status asthmaticus. Am J Dis Child. 1972;123:227–8.

59. Walsh P, Rothenberg SJ, O'Doherty S, et al. A validated clinical model to predict the need for admission and length of stay in children with acute bronchiolitis. Eur J Emerg Med. 2004;11:265–72.

60. Alistair PJT, Sillis JA, Hart CA. Validation of the Glasgow Meningococcal Septicemia Prognostic Score: a 10-year retrospective survey. Crit Care Med. 1991;19:26–30.

61. Hachimi-Idrissi S, Corne L, Ramet J. Evaluation of scoring systems in acute meningococcaemia. Eur J Emerg Med. 1998;5:225–30.

62. Fiser DH. Assessing the outcome of pediatric intensive care. J Pediatr. 1992;121:68–74.

63. Marcin JP, Pollack MP. Review of the acuity scoring systems for the pediatric intensive care unit and their use in quality improvement. J Intensive Care Med. 2007;22:131–40.

64. Fiser DH, Long N, Roberson PK, Hefley G, Zolten K, Brodie-Fowler M. Crit Care Med. 2000;28(7):2616–20.

65. Pollack MM, Holubkov R, Glass P, et al. Functional status scale: new pediatric outcome measure. Pediatrics. 2009;124:e18–28.