The development of intensive patient care has increased our ability to monitor, diagnose, and treat critically ill and injured patients; consequently, critical care practice has grown substantially in the past 4 decades. However, there is a strong perception by some that intensive care unit (ICU) efforts introduce an inappropriate financial burden. Intensive care unit costs are high and account for a disproportionately large fraction of health care resources. Furthermore, the true impact of the delivery of such services has yet to be defined. The detractors perceive that ICU care is a way of prolonging life without improving long-term survival and thus is a waste of resources. In addition, there is increasing recognition of the wide variation in health care practices and, more importantly, of the potential effect of this variance on health care delivery and outcomes. Consequently, critical care is under increasing pressure to improve ICU performance and quality of care in ways that will reduce costs. As a result, clinical decisions must often be based on a simultaneous evaluation of clinical outcomes and resource consumption.

The ability to identify critically ill patients who will not survive to hospital discharge may yield substantial cost savings. A variety of instruments are now available for severity evaluation and outcome prediction in critical care. Outcome prediction models are usually classified as disease specific (ie, applied to patients with a particular condition) or generic (ie, designed to be applied to all critically ill patients). These instruments typically assign points according to severity of illness and purport to predict an outcome, providing the user with a numeric estimate of the probabilities of an outcome (eg, in-hospital mortality) for that patient or group of patients. The most common outcome prediction models used in adult intensive care include the Mortality Probability Model (MPM) II, the Simplified Acute Physiology Score (SAPS) II, and the Acute Physiology and Chronic Health Evaluation (APACHE) II and APACHE III. In general, these systems are based on logistic regression models that use a set of clinical and physiological variables—evaluated and registered on admission or on the first day of intensive care—to predict hospital mortality. These models differ considerably in the number and types of variables used, as well as the time frame for data collection. For instance, the APACHE III system consists of 2 parts—an APACHE III score, which is based on 17 physiological variables, age, and chronic health status, and the predictive equation APACHE III, a series of predicted equations linked to ICU admission diagnosis, patient selection criteria, and the APACHE III database.

These predictive systems are being used more commonly. At the patient level, severity scores describe severity of illness and are being used in some centers to support clinical decision making and to help guide discussions with patients and families concerning withdrawal of life support. At the ICU level, these outcome prediction models compare actual and expected outcome for groups of patients (ie, the standardized mortality ratio); they are used to compare ICU performance, to ensure that patient groups in research studies have similar severity of illness, to determine optimal allocation of critical care resources, and to evaluate the effect of new therapies, procedures, or ICU organization.

Appropriate use of scoring systems depends on several factors, including consistent and accurate data collection and interpretation. Periodic system calibration helps define how closely mortality prognosis fits the observed mortality. These systems have been validated by 2 techniques.
The first, cross-validation, evaluates the goodness of fit of the model by examining a model’s calibration and discrimination on a subset of the data set different from the subset used to develop the model. The second, external validation, evaluates the goodness of fit of the model on an entirely new data set.

In the current issue of Mayo Clinic Proceedings, Berge et al report the results of a retrospective study that evaluated the process of care, resource utilization, and outcomes in a cohort of 248 gravely ill patients as defined by the APACHE III scoring system model. Their study highlights the difficulties of applying scoring systems for clinical management of individual patients.

Most investigators acknowledge the limitations of APACHE III and other scoring systems for predicting individual prognosis based on physiological data collected during the first 24 hours of the patient’s ICU stay. Other investigators have proposed prognostic models that incorporate daily APACHE III values from subsequent days for predicting mortality. Conceptually, an approach that includes changes in physiological variables over time should more accurately predict the chance of survival. In a novel approach, Berge et al used a 95% or higher probability of hospital death APACHE III prediction score obtained on 2 consecutive days (median time after ICU admission, 4.5 days; range, 1-76 days) rather than the APACHE III score obtained on the first day of ICU care or a daily prognostic model. Surprisingly, there was a striking difference between the observed and the expected survival rates (23% vs ≤5%). Several factors may explain such difference. First, when data are not obtained at approximately the same time in the course of an acute illness or within a similar period, prognostic estimates will be inaccurate because physiological measures reflect different phases of critical illness.9,10 Second, for an individual patient, mortality predictions calculated with use of scoring systems are often inaccurate in populations other than those in which the scales were developed. Moreover, the application of a different model to the same patient often results in different outcome predictions. Mackenzie et al summarized the factors that may be responsible for significant variations in standardized mortality ratios: (1) the characteristics of the health care system, (2) the characteristics of the individual population, (3) the patterns of care in an individual ICU, (4) the intrinsic deficiencies of the model, (5) inconsistent applications of the model, (6) the size of the study population, and (7) variations in the quality of care in the ICU and/or the hospital.

Several studies have evaluated the ability of critical care providers to predict both futility of care and outcomes. Theoretically, prognostic scoring systems should offer more accurate predictions than an individual physician’s judgment because they are based in large databases, as opposed to being limited to a single clinician’s experience. As an aid to the process of clinical decision making, evidence suggests that outcome prediction models perform better than clinicians in prognostic predictions.12 Esserman et al developed a model based on the product of APACHE III risk estimates on ICU days 1 and 5 to identify patients receiving potentially ineffective care, whereas Afessa et al identified potentially ineffective care in the sickest critically ill patients with a predicted hospital mortality rate of 80% or higher by showing an increase in APACHE III score on the third ICU day compared with the first ICU day. Berge et al evaluated whether an APACHE III score that would predict an in-hospital mortality of 95% or higher on 2 consecutive days could also be used to identify patients for whom care would be potentially ineffective or futile and would potentially use more resources. They discovered that their use of APACHE III scores woefully underestimated observed survival at the time of hospital discharge.

Often, physicians are confronted with decisions about what type of care to offer, how aggressively to pursue diagnostic and therapeutic interventions, and the potential futility of these actions. Clinicians often include input from patients, families, surrogates, and members of the health care team in the decision-making process.3 The preconceived notion is that, in the sickest critically ill patient, the process of care and resource utilization are often driven by the family’s unrealistic expectations. In the study by Berge et al, unrealistic expectations of a good outcome correlated with increased resource utilization without obvious survival benefit. These investigators confirmed previous observations that the sickest critically ill patients consume a high proportion of resources. However, they were unable to ascertain whether the higher resource utilization was physician or family driven. This study also confirmed previous studies that showed that clinicians have differing opinions about futility of care, often interpreting the same information in diverging ways.

The study by Berge et al provides further evidence that quantitative risk estimates should not be deterministic in clinical decision making. Current prognostic systems will never be able to predict outcome with a 100% specificity, and therefore they will never be indicative of absolute irreversibility of disease or impossibility of survival.13 However, generic ICU prognostic systems can be useful in comparing ICU performance with respect to a wide variety of end points, including ICU mortality. Unless APACHE III or other scoring systems can be shown to retain the same level of predictive power across different ICUs with different mixes and survival rates or can be calibrated to different ICU populations, it is unlikely that
current prognostic systems will be used as the single basis to withdraw care or allocate resources. As stated by Becker and Zimmerman,15 “Future research will improve the accuracy of individual patient predictions but, even with the highest degree of precision, such predictions are only useful in support of, and not as a substitute for, good clinical judgment.”

Pedro A. Mendez-Tellez, MD
Todd Dorman, MD
Johns Hopkins University
Baltimore, Md