



clinical investigations in critical care

Community-Wide Assessment of Intensive Care Outcomes Using a Physiologically Based Prognostic Measure*

Implications for Critical Care Delivery From Cleveland Health Quality Choice

Carl A. Sirio, MD; Laura B. Shepardson, MS; Armando J. Rotondi, PhD;
Greg S. Cooper, MD; Derek C. Angus, MB, ChB, MPH, FCCP;
Dwain L. Harper, DO; and Gary E. Rosenthal, MD

Study objectives: To examine the applicability of a previously developed intensive care prognostic measure to a community-based sample of hospitals, and assess variations in severity-adjusted mortality across a major metropolitan region.

Design: Retrospective cohort study.

Setting: Twenty-eight hospitals with 38 ICUs participating in a community-wide initiative to measure performance supported by the business community, hospitals, and physicians.

Patients: Included in the study were 116,340 consecutive eligible patients admitted to medical, surgical, neurologic, and mixed medical/surgical ICUs between March 1, 1991, and March 31, 1995.

Main outcome measures: The risk of hospital mortality was assessed using a previous risk prediction equation that was developed in a national sample, and a reestimated logistic regression model fit to the current sample. The standardized mortality ratio (SMR) (actual/predicted mortality) was used to describe hospital performance.

Results: Although discrimination of the previous national risk equation in the current sample was high (receiver operating characteristic [ROC] curve area = 0.90), the equation systematically overestimated the risk of death and was not as well calibrated (Hosmer-Lemeshow statistic, 2407.6, 8 *df*, $p < 0.001$). The locally derived equation had similar discrimination (ROC curve area = 0.91), but had improved calibration across all ranges of severity (Hosmer-Lemeshow statistic = 13.5, 8 *df*, $p = 0.10$). Hospital SMRs ranged from 0.85 to 1.21, and four hospitals had SMRs that were higher or lower ($p < 0.01$) than 1.0. Variation in SMRs tended to be greatest during the first year of data collection. SMRs also tended to decline over the 4 years (1.06, 1.02, 0.98, and 0.94 in years 1 to 4, respectively), as did mean hospital length of stay (13.0, 12.4, 11.6, and 11.1 days in years 1 to 4; $p < 0.001$). However, excluding the increasing ($p < 0.001$) number of patients discharged to skilled nursing facilities attenuated much of the decline in standardized mortality over time.

Conclusions: A previously validated physiologically based prognostic measure successfully stratified patients in a large community-based sample by their risk of death. However, such methods may require recalibration when applied to new samples and to reflect changes in practice over time. Moreover, although significant variations in hospital standardized mortality were observed, changing hospital discharge practices suggest that in-hospital mortality may no longer be an adequate measure of ICU performance. Community-wide efforts with broad-based support from business, hospitals, and physicians can be sustained over time to assess outcomes associated with ICU care. Such efforts may provide important information about variations in patient outcomes and changes in practice patterns over time. Future efforts should assess the impact of such community-wide initiatives on health-care purchasing and institutional quality improvement programs. (CHEST 1999; 115:793-801)

Key words: critical care medicine; health services research; hospital mortality; hospital utilization; intensive care units; length of stay; outcome assessment; quality of health care; severity of illness; skilled nursing

Abbreviations: APACHE = acute physiology and chronic health evaluation; APS = acute physiology score; CHQC = Cleveland Health Quality Choice; LOS = length of stay; ROC = receiver operating characteristic; SMR = standardized mortality ratio

Public accountability for outcomes derived from delivered health services is a growing expectation of consumers and purchasers.¹⁻³ Many informed business leaders now advocate the use of outcomes

For editorial comment see page 613

data that are appropriately adjusted for case-mix in their selection of health-care providers.⁴ However, despite the wide geographic and temporal variability that exists in the processes and outcomes from care and the growing recognition of the need for valid data to assess health-care delivery systems, such data are generally not available in most health-care markets.⁵⁻⁹ Moreover, typically no single corporate employer or purchaser has sufficient numbers of beneficiaries to independently influence the health-care delivery system or induce changes in practice patterns.¹⁰ Thus, in recent years, employer-driven coalitions to implement value-based health-care purchasing strategies have emerged in several health-care markets.¹¹⁻¹³

In addition to the trends noted above, changing social priorities, such as the growth of patient autonomy in health-care decision-making, the growing realization that the application of high technology care must be balanced with the ability to improve quality of life, and the pressure to reshape health-care financing and limit total growth of hospital expenditures, have resulted in conflicting challenges for those specifically responsible for delivering critical care services.¹⁴ Addressing these challenges requires an enhanced ability to evaluate the

incremental benefit of intensive care services on patient outcomes, as well as an understanding of the impact of changes in the organization and delivery of intensive care that are made in response to these pressures.^{15,16} Prognostic scoring and risk-adjustment tools that establish predicted rates for selected outcomes have been advocated as a marker for assessing quality performance within hospitals.¹⁷ Comparative ratios between observed and predicted outcomes (eg, hospital death rates) have been utilized to provide insight into areas of relative strength and weakness in clinical performance. In addition, the application of validated and methodologically sound prognostic models has been advocated as one basis for linking societal values with the provision of clinical care.^{14,18}

In 1989, a collaborative and cooperative strategy formulated by hospitals, physicians, and employers in a large metropolitan region led to the implementation of a standardized measurement system to evaluate patient outcomes of intensive care and other services in acute care hospitals serving the region.¹⁰ This effort, Cleveland Health Quality Choice (CHQC), can be traced to the coalescence of concurrent initiatives undertaken by seven local founding organizations in the provider and business communities to examine the cost and quality of care. The underlying philosophy of this ongoing initiative is that availability of timely health-care outcomes data will lead to improved quality of care and more effective decision-making by purchasers. Importantly, this focus emphasizes the importance of quality indicators in appropriately evaluating the value derived from health-care services.

To measure patient outcomes in ICUs, hospitals have collected data to determine the severity of illness of each patient admitted to the ICU. The data are then used to determine a predicted risk of in-hospital death, on the basis of a validated prognostic system—APACHE III (acute physiology and chronic health evaluation). For each hospital, actual and predicted mortality rates are then compared. The current study reports important observations from the first 4 years of data collection and analysis. Specifically, we examine the predictive validity of the APACHE III method, as applied in actual practice in a large community-based cohort of patients, and the utility of applying a previously derived normative risk prediction equation to this cohort. In addition, we examined variations in standardized mortality across individual hospitals and changes in mortality rates that have occurred over the 4-year study period. The findings provide important insights into using physiologically based prognostic tools for ICU patients and interpreting currently accepted clinical end points, such as hospital mortality, in the evaluation of ICU performance.

*From the Department of Anesthesiology and Critical Care Medicine (Drs. Sirio, Rotondi, and Angus), Health Delivery and Systems Evaluation Team, University of Pittsburgh School of Medicine, Pittsburgh, PA; Divisions of General Internal Medicine (Ms. Shepardson) and Gastroenterology (Dr. Cooper), Institute for Health Care Research, Department of Medicine, Cleveland Veterans Affairs Medical Center, University Hospitals of Cleveland, and Case Western Reserve University School of Medicine, Cleveland, OH; Quality Information Management Corporation, Cleveland Health Quality Choice (Dr. Harper), Cleveland, OH; and Division of General Internal Medicine (Dr. Rosenthal), Department of Medicine, University of Iowa College of Medicine, Iowa City, IA.

Dr. Rosenthal was supported by a Career Development Award from the Health Services Research and Development Service, US Department of Veterans Affairs.

Financial disclosure: Drs. Sirio and Harper have provided consulting services to APACHE Medical Systems, Inc (AMS). Dr. Harper is Executive Director of the Cleveland Health Quality Choice Program. AMS holds the commercial copyright on the hospital mortality equations. APACHE and APACHE III are trademarks of AMS. While both the equations and the APACHE database are protected by commercial copyright, they are available to researchers for independent verification and further analysis by contacting the authors or AMS.

Manuscript received February 20, 1998; revision accepted September 4, 1998.

Correspondence to: Carl A. Sirio, MD, University of Pittsburgh Medical Center, 612A Scaife Hall, 200 Lothrop St, Pittsburgh, PA 15213; e-mail: sirio@smtp.anes.upmc.edu

MATERIALS AND METHODS

Hospitals

The study was conducted in 38 ICUs in 28 hospitals participating in CHQC.¹⁰ Nineteen of the study ICUs were mixed medical and surgical units, eight ICUs were medical, eight ICUs were surgical, and three were neurologic and/or neurosurgical. Thirteen additional ICUs in study hospitals that specialized in coronary care ($n = 11$) or cardiovascular surgery ($n = 2$) were excluded from the study, as per CHQC protocols. Five hospitals were members of the Council of Teaching Hospitals of the Association of American Medical Colleges during the period of data collection, and were considered major teaching hospitals for the current study. Other characteristics of participating hospitals and ICUs have been described previously.^{10,19}

Patients

The eligible sample was drawn from 134,402 consecutive ICU admissions from March 1, 1991 to March 31, 1995. Patient data were collected on all ICU admissions with the following exceptions: individuals < 16 years of age; burn patients; patients admitted solely for hemodialysis or peritoneal dialysis; patients who die within 1 h of admission to the ICU or within the first 4 h after admission to the ICU in cardiopulmonary arrest; and patients undergoing coronary artery bypass, cardiac valve, or heart transplant surgery. These exclusions are consistent with those reported in the original APACHE III or were directed by CHQC data collection policies. Of the eligible sample, 8,114 (6.0%) who were readmitted to an ICU during a single episode of hospitalization, 9,167 (6.8%) who were discharged to other acute-care hospitals for further management, and 781 (0.6%) who had missing severity of illness or outcome information were excluded from analysis. The final sample consisted of 116,340 patients.

Data

Data were abstracted from patients' ICU records by trained data reviewers using standard forms and data collection software (provided by APACHE Medical Systems, Inc; McLean, VA). Variables included elements necessary to determine an APACHE III acute physiology score (APS): age; presence of specific severe comorbid conditions (*eg*, AIDS, metastatic cancer, hepatic failure, lymphoma); and the most abnormal value during the first 24 h of ICU admission for 17 specific physiologic variables (*eg*, mean arterial BP, serum sodium and BUN, arterial oxygen tension, arterial pH, abbreviated Glasgow Coma Score). Physiologic variables with missing data were assumed to have normal values, although a minimum of nine physiologic measurements were required for inclusion in the analysis. Other data elements included the following: gender; primary diagnosis prompting admission to the ICU, as classified according to a prior taxonomy¹⁹; admission source, based on eight mutually exclusive categories (*eg*, emergency department, operating room, other hospital ward); dates of ICU and hospital admission and discharge; vital status at ICU and hospital discharge (*ie*, alive/dead); discharge destination (*eg*, home, skilled nursing facility, other acute-care hospital); and ICU and hospital length of stay. As previously reported,¹⁰ several steps were taken to ensure the reliability of study data, including explicit definitions of each variable that were consistent with prior applications of the APACHE method, formal training sessions for data abstractors, manual and electronic edits to identify out-of-range or discrepant data (*eg*, patients with a diagnosis of shock and normal vital signs),

and independent reabstraction of data for randomly selected patients from each hospital to determine interrater reliability.²⁰

Severity of Illness

Measurement of admission severity of illness was based on the APACHE III method. For each patient, an APACHE III APS was determined on the basis of age, comorbidity, and physiologic abnormalities. Scores have a possible range of 0 to 299 and were determined using previously validated weights for each variable.¹⁷ A predicted risk of in-hospital death (0 to 100%) was then determined using a previously developed multivariable equation that considered the APS score, admission source, and ICU diagnosis. Variable weightings used in the equation were derived from an analysis of 16,622 ICU admissions in 1988 and 1989 to 40 US hospitals, in which the rate of in-hospital death was 16.5%.²¹ Thus, the risk predictions provided an expected probability of death, based on the prior national normative sample, and an external benchmark to which the study hospitals could be compared.

In addition to predictions based on the national normative sample, the predicted risk of in-hospital death was reestimated based on the current data set. This entailed the development of a second multivariable equation using logistic regression that included the APS score, admission source, and ICU admission diagnosis. Admission source and diagnosis were expressed as $n-1$ indicator variables, where n is equal to the number of categories for each of these elements. Because the risk of death did not increase in a linear manner with increasing APS, we entered APS into the logistic regression model as both a continuous variable and a series of 15 indicator variables for specific ranges of scores (*eg*, < 20, 20 to 29, 30 to 39). The addition of the indicator variables better fit the curvilinear relationship between APS and in-hospital death, and improved model calibration as described below when compared to a model with just the continuous variable.

To examine the potential impact of interhospital variation in discharge triage practices, a further set of risk predictions was developed based on the local sample, but excluding 14,192 patients who were discharged to skilled nursing facilities, physical rehabilitation centers, and nursing homes. The predictions were also determined from a logistic regression model that included the same independent variables as the prior model.

Analysis

Discrimination of the risk predictions from the national and local samples was by receiver operating characteristic (ROC) curve analysis, which is a measure of the proportion of times the risk of death was higher in patients who died than patients who were discharged alive.^{22,23} Calibration (*ie*, goodness of fit) of the two sets of risk predictions was assessed by the Hosmer-Lemeshow statistic, which compares observed and predicted numbers of deaths in deciles of increasing risk.²⁴

Performance of individual hospitals was summarized using a standardized mortality ratio (SMR). The SMR was the observed hospital mortality rate divided by the mean predicted mortality rate, as determined by aggregating patient-level risk predictions from the locally derived multivariable equations. An SMR > 1.0 indicates an observed death rate that is higher than expected (*ie*, lower performance), whereas an SMR < 1.0 indicates an observed death rate that is lower than expected (*ie*, higher performance). Confidence intervals around hospital SMRs were estimated by calculating exact 99% limits around the observed hospital mortality rate, and dividing this by the mean predicted mortality rate, which was assumed to be a constant. Because of

the large sample sizes, we chose a conservative cutoff of $p < 0.01$ to determine which hospital SMRs statistically differed from 1.0.

RESULTS

The mean age of study patients was 63 years and 52% were men (Table 1). Forty-one percent of patients were admitted through the emergency department; 37% of patients ($n = 42,416$) were postoperative (*ie*, admitted to the ICU after undergoing a surgical procedure), and 63% of patients ($n = 73,924$) were nonoperative. The 10 most common ICU admission diagnoses accounted for nearly 50% of admissions and included the following: angina ($n = 10,046$); congestive heart failure ($n = 8,007$); trauma to the head, spine, chest, abdomen, or

extremities that was managed surgically or nonsurgically ($n = 5,913$); GI hemorrhage ($n = 5,446$); carotid endarterectomy ($n = 4,912$); stroke or CNS hemorrhage ($n = 4,902$); lower extremity bypass graft ($n = 3,950$); acute myocardial infarction ($n = 3,778$); COPD or asthma ($n = 3,588$); and cardiac arrhythmias ($n = 3,454$).

The median APS was 41 and ranged from 29 to 51 in the 28 hospitals. The overall observed in-hospital mortality rate was 11.3%. APSs were strongly related ($p < 0.001$) to in-hospital mortality rates, which ranged from 0.1% in patients with scores of ≤ 10 to 95% in patients with scores > 150 (Fig 1). The ROC curve area of risk predictions based on the previously validated APACHE III national normative risk-adjustment model was 0.901, indicating excellent discrimination. However, the national model consistently overestimated the risk of death across all deciles of predicted risk (Fig 2, *top*). Thus, the overall mean predicted risk of death, based on the APACHE III national model, of 12.5% was roughly 1.2% higher than the observed death rate of the current population, and the Hosmer-Lemeshow statistic, a sensitive measure of differences between predicted and observed outcomes in large data sets, indicated miscalibration ($\chi^2 = 2407.6$, 8 *df*, $p < 0.001$).

As would be expected, risk predictions based on the locally derived model exhibited much better calibration (Fig 2, *bottom*). The mean predicted risk of death of 11.3% was identical to the observed death rate and the Hosmer-Lemeshow statistic was not significant ($\chi^2 = 13.5$, 8 *df*, $p = 0.10$), indicating that the model was well calibrated across all ranges of risk. The ROC curve area of the model was 0.907.

The variation in observed mortality rates across hospitals was wide, ranging from 4.3 to 15.8%. Based on the locally derived model, mean predicted risks of death ranged from 3.6 to 15.8%. At a hospital level, the mean predicted risk of death explained 87% of the variance in observed hospital mortality rates (*ie*, $R^2 = 0.87$). As a result, the variation in SMRs was less than the variation in observed mortality. SMRs ranged from 0.85 to 1.21 (Fig 3). The correlation between SMRs and mean ICU length of stay (LOS) and mean hospital LOS were not significant ($R = 0.07$, $p = 0.72$; and $R = 0.06$, $p = 0.75$; respectively).

Only one hospital had an SMR that was lower ($p < 0.01$) than 1.0, while three hospitals had SMRs that were greater ($p < 0.01$) than 1.0. However, none of the four hospitals that were statistical outliers with respect to SMRs were outliers during each of the 4 study years when years 1 to 4 were examined individually. Although the lack of consistency across each year may be due in part to smaller sample sizes, hospital SMRs across years exhibited some degree of

Table 1—Characteristics of 116,340 Study Patients

Characteristic	Value
Age (y), mean (SD)	62.9 (20.0)
APACHE III APS (SD)	
Mean (SD)	46.9 (27.6)
Median (interquartile range)	41.0 (28.0–59.0)
Predicted risk of death (APACHE III national model)	
Mean (SD)	0.125 (0.201)
Median (interquartile range)	0.040 (0.015–0.126)
Predicted risk of death (local model)	
Mean (SD)	0.113 (0.195)
Median (interquartile range)	0.028 (0.008–0.113)
ICU LOS, d	
Mean (SD)	3.9 (4.8)
Median (interquartile range)	3 (2–4)
Hospital LOS, d	
Mean (SD)	12.0 (12.7)
Median (interquartile range)	8 (5–14)
	% of Patients (No.)
Gender, male	51.9 (60,417)
Admission source	
Operating room	16.5 (19,225)
Recovery room	20.7 (24,073)
Emergency department	40.9 (47,636)
Hospital ward	11.3 (13,152)
Other ICU	3.2 (3,683)
Other acute care hospital	2.7 (3,180)
Direct admission	4.5 (5,289)
Non-ICU holding area	0.1 (100)
ICU mortality rate	6.3 (7,285)
In-hospital mortality rate	11.3 (13,166)
Comorbid conditions	
AIDS	0.2 (193)
Cirrhosis	2.0 (2,309)
End-stage renal disease*	2.9 (3,349)
Hepatic failure	0.6 (641)
Immunosuppressive therapy	3.2 (3,736)
Lymphoma	0.7 (831)
Leukemia/multiple myeloma	0.6 (681)
Solid tumor with metastasis	3.7 (4,278)

*Not used in determining APACHE III APSs.

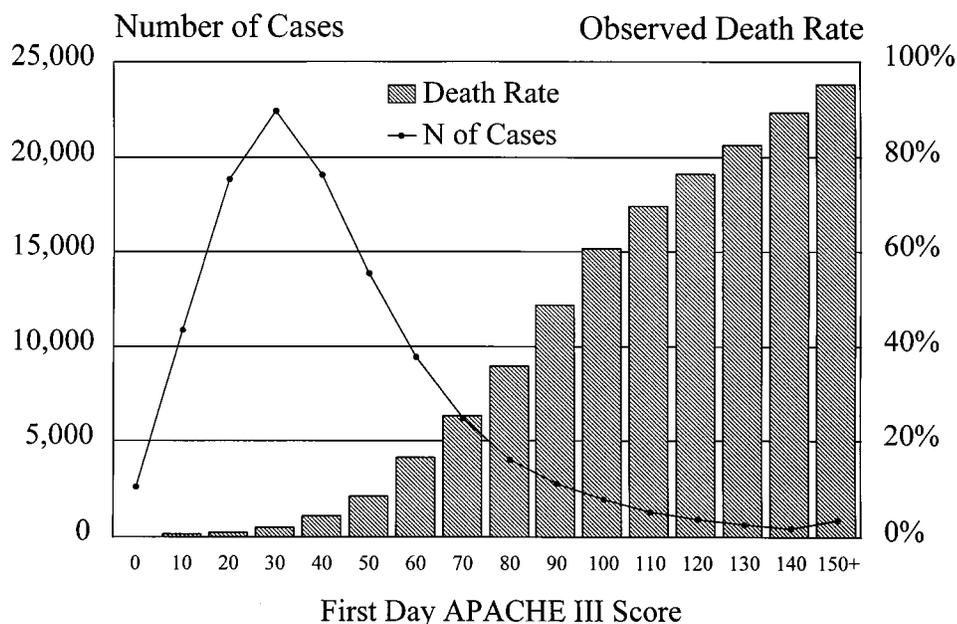


FIGURE 1. In-hospital mortality rates according to APACHE III acute physiology scores. Mortality rates ranged from 0.1% in patients with scores of < 10 to 95.2% in patients with scores of ≥ 150 .

variability, particularly for hospitals with smaller patient volumes. For example, the range of SMRs across years 1 to 4 was > 0.30 for six hospitals. In addition, variation in hospital SMRs was greater in year 1 than in subsequent years. For example, the difference between the highest SMR (1.64) and lowest (0.83) was 0.81 in year 1 and 0.42, 0.44, and 0.48 in years 2, 3, and 4, respectively, while interquartile differences were 0.19, 0.17, 0.15, and 0.11 in years 1, 2, 3, and 4, respectively.

SMRs tended to be lower in the five major teaching hospitals than in other hospitals (mean SMRs 0.95 ± 0.07 vs 1.04 ± 0.10 , respectively; $p = 0.09$).

Observed mortality rates were similar over the 4-year period of data collection (11.4%, 11.4%, 11.4%, and 11.1% in years 1, 2, 3, and 4, respectively; $p = 0.45$). However, mean predicted mortality tended to increase (10.8%, 11.1%, 11.6%, and 11.7% in years 1 to 4, respectively; $p < 0.001$), resulting in a decline in the overall SMR (1.06, 1.02, 0.98, and 0.95 in years 1 to 4, respectively). Coincident with the decline in SMRs, declines were observed in total hospital LOS (13.0, 12.4, 11.6, and 11.1 days in years 1, 2, 3, and 4, respectively; $p < 0.001$) and ICU LOS (4.0, 3.9, 3.8, and 3.8 days in years 1 to 4, respectively; $p < 0.001$), while the proportion of patients discharged to skilled nursing rehabilitation facilities and nursing homes increased (9.5%, 11.2%, 12.3%, and 15.8%, in years 1 to 4, respectively; $p < 0.001$). Further analysis indicated that patients discharged to

these facilities had higher admission severity of illness than patients discharged to home, as measured by mean APACHE III scores (57.3 vs 38.9; $p < 0.001$) and predicted risk of death (15.9% vs 5.7%; $p < 0.001$), suggesting that such patients may have been more likely to die after discharge than patients who were discharged to home.

To examine the potential impact of these changes in discharge triage on mortality performance, a further multivariable model was developed using 102,148 patients that excluded the 14,192 patients discharged to skilled nursing rehabilitation facilities and nursing homes. The ROC curve area of this model was 0.925, and the Hosmer-Lemeshow statistic was not significant ($\chi^2 = 9.2$, 8 *df*, $p = 0.33$). The aggregate in-hospital mortality for this population was 12.9%. Applying risk predictions from this model, overall SMRs remained relatively stable over the 4 years of data collection (1.02, 1.01, 0.98, and 1.00 in years 1, 2, 3, and 4), suggesting that the decline in standardized mortality may have reflected changes in hospital discharge practices. However, variability in hospital SMRs did not decline. For example, hospital SMRs ranged from 0.80 to 1.25; eight hospitals had SMRs that were lower ($p < 0.01$) than 1.0, and three hospitals had SMRs that were greater ($p < 0.01$) than 1.0. Moreover, the difference in mean SMRs between major teaching and other hospitals remained (mean SMRs 0.94 ± 0.06 vs 1.06 ± 0.13 , respectively; $p = .04$).

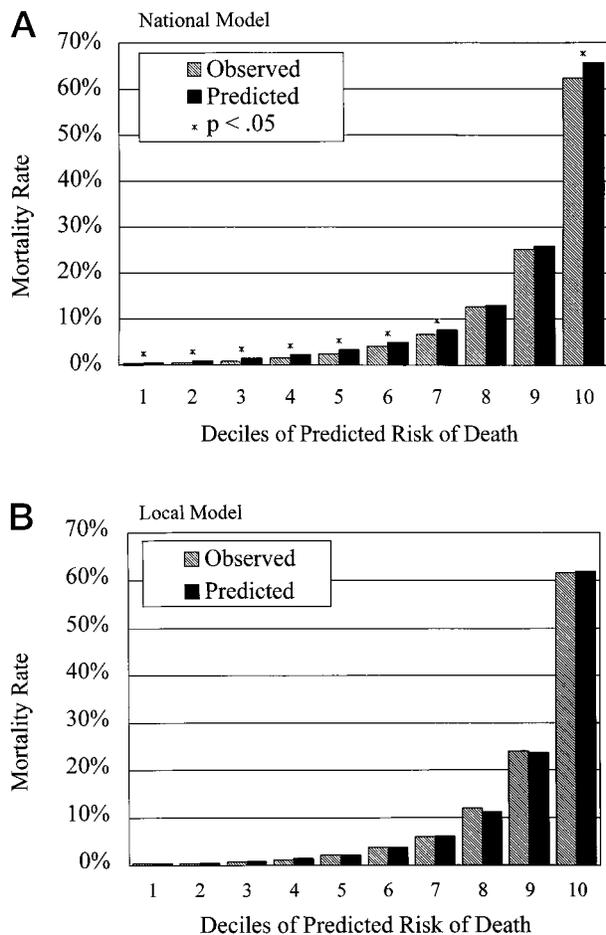


FIGURE 2. Differences in observed and predicted death rates across deciles of increasing risk, based on risk predictions from (top, A) the APACHE III national normative model²¹ and (bottom, B) the locally derived model (see “Materials and Methods” section). Top, A: observed and predicted death rates differed ($p < 0.05$) in all deciles except for deciles 8 and 9. Bottom, B: observed and predicted death rates were similar ($p \geq 0.05$) in all deciles.

DISCUSSION

The current study represents one of the largest evaluations of variations in ICU mortality, and the first study (to our knowledge) to include all hospitals providing critical care services in a single metropolitan region. In analyses of $> 116,000$ patients admitted to ICUs in 28 hospitals over a 4-year period, several important findings emerge. First, an existing ICU risk stratification tool can be successfully implemented in a diverse spectrum of hospitals, as part of an ongoing collaboration between purchasers and providers to evaluate health-care quality and reward high performing institutions. The longevity of this initiative may, in part, be due to the use of a previously validated method. Other important and explicit steps were taken under the initiative to

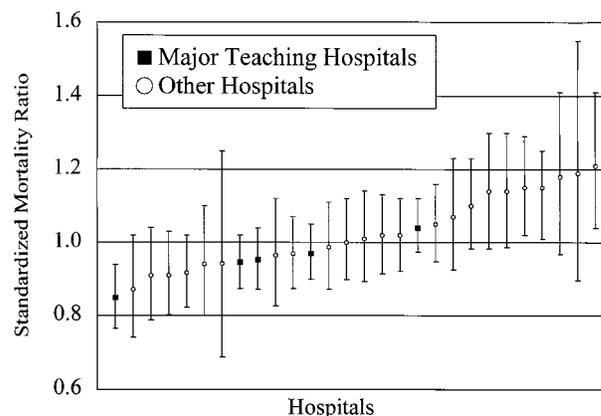


FIGURE 3. Variation in SMRs (*ie*, observed/predicted hospital mortality) across hospitals. Error bars indicate 99% confidence intervals. SMRs > 1.0 indicate higher than expected mortality, while SMRs < 1.0 indicate lower than expected mortality. Major teaching hospitals are denoted by the black squares.

ensure long-term participation, including the following: rigorous independent examination of the validity and reliability of the method in the current population prior to public reporting; a commitment to refine data collection and analysis based on feedback by local ICU physicians; development of training workshops for purchasers and local media to review important methodologic limitations of the data being disseminated; and allowing adequate time to scrutinize results prior to public release.¹⁰

Second, a substantial amount of the variation in observed mortality rates was explained by the APACHE III normative risk predictions that were developed previously. The data confirm the powerful predictive value of weighted abnormal physiology for patients admitted to the ICU. Moreover, the discrimination of the APACHE III method in our community-based cohort was nearly identical to its discrimination in a development cohort assembled using strict research protocols.¹⁷

Calibration analyses indicated that APACHE III systematically overestimated the risk of death in the current population. This may reflect the likelihood that a prediction model will almost always perform better in the data set from which it was derived and not as well when applied to a new population. It is possible that the APACHE III model would be better calibrated than the current model if both were prospectively applied to a new population or there may be no appreciable differences in performance. It is also important to note that, in large data sets, the Hosmer-Lemeshow test is an extremely sensitive method for discerning differences in observed and predicted outcome rates. Nevertheless, overestimation may reflect general improvements in ICU care and patient outcomes since the development of

APACHE III, differences in discharge practices of the current hospitals with triage of ill patients to skilled nursing facilities, differences in ICU utilization with higher numbers of discretionary admissions, or the provision of more effective care in the current hospitals.²⁵ Although the clinical ramifications of these calibration differences is likely small, the overestimation of risk in the current sample highlights the need to recalibrate validated predictive methods for local applications to reflect potential improvements in care or differences in clinical practice.²⁶

Third, severity of illness at the time of ICU admission, as measured by predicted death rates, varied widely across hospitals, suggesting that thresholds for admitting and treating patients in ICUs vary, and likely depend on institutional practice styles and the availability of critical care resources.²⁷⁻³⁹ However, despite the wide variation in admission severity, variation in standardized mortality was relatively small. Indeed, the range in hospital SMRs reflected less than a 1.5-fold difference. In addition, variations in SMRs tended to decline over time. The degree of variation observed in the current study was somewhat lower than the nearly twofold variation in hospital SMRs observed in an earlier national study of 40 US hospitals, and may reflect the larger hospital sample sizes of the current study or its examination of a single geographic area.²¹ Nevertheless, SMRs tended to be somewhat lower in the major teaching hospitals than other hospitals, confirming an earlier observation using a different risk-adjustment method that was based on patients admitted with six medical diagnoses (myocardial infarction, heart failure, pneumonia, stroke, GI bleeding, and obstructive airway disease).⁴⁰

Fourth, the decline in overall standardized mortality over the 4 years of data collection and dissemination may suggest improvements in care over time.^{41,42} However, further analyses indicated that rates of discharge to skilled nursing facilities also increased over time, and that these patients had markedly higher severity of illness than patients discharged home. Moreover, excluding such discharges from analysis attenuated temporal declines in mortality. These findings have important ramifications regarding the utility of hospital discharge status as an accurate or appropriate end point to measure outcomes of ICU care. Traditionally, ICU discharge status has been viewed as an inadequate marker of ICU performance as it may be significantly influenced by discharge timing, especially among terminal patients who are transferred out of the ICU to die in non-ICU settings. The increasing use of skilled nursing facilities, rehabilitation centers, and nursing homes may signal that the adequate evalua-

tion of critical care requires long-term follow-up of patients after hospital discharge.

When interpreting our findings, several methodologic issues should be considered. First, our analysis was limited to a single outcome measure. Although hospital mortality is a widely used indicator, quality of care encompasses multiple dimensions.^{41,43} Thus, the implications of our findings for other aspects of the quality of critical care, such as the appropriateness and process of care, functional outcomes, patient satisfaction, and long-term mortality, are uncertain.

Second, we did not directly examine responses by clinicians and hospital administrators to the reported ICU outcomes data. Thus, the degree to which changes in outcomes over time can be attributed to dissemination of these data is unknown. Indeed, a recent study found similar declines in mortality rates for coronary artery bypass surgery in states with and without initiatives to publicly disseminate severity-adjusted mortality data.⁴⁴

Third, the lack of a unique patient identifier in the database precluded our ability to follow up patients beyond hospital discharge, and to examine hospital readmission and postdischarge mortality. Although such information would allow for a more thorough assessment of the effectiveness of ICU care, routine collection of identifiers, such as the social security number, raises important concerns about patient confidentiality.^{45,46} However, as care is increasingly delivered in settings outside the hospital, the ability to link individual encounters to a single episode of care will become increasingly important to evaluating patient outcomes and health-care quality, particularly in the context of the current analysis in which hospital mortality may no longer be an adequate marker of ICU performance.

Fourth, while our measure of severity of illness exhibited excellent discrimination, it is possible that variations in mortality may be due to prognostic factors, such as functional status, mental health, or social support, that are not assessed by physiologically based methods, such as APACHE III.⁴⁷⁻⁴⁹ In addition, health insurance status may contribute to outcomes. Lastly, no information was collected regarding the goals of ICU treatment, patient and/or family preferences for specific ICU treatments, or resuscitation status. As shown previously, the use of do not resuscitate orders and other treatment limitations have increased over time, and may differ across institutions.^{50,51} Such practices may confound the interpretation of our data with respect to hospital performance.

The findings of the current study have important implications for the delivery of critical care and the assessment of health-care quality, but also raise

further questions. While some studies have suggested that greater ICU specialization may be associated with better outcomes, current uncertainty about the value of regionalizing high-intensity, critical care services stems, in part, from the lack of valid ICU performance data in most regions.^{52,53} The current findings suggest that small, but clinically meaningful differences may exist across institutions. Directing care preferentially to such institutions may be associated with improved outcomes on a community-wide basis. Nevertheless, important questions regarding the feasibility of such practices need to be explicitly explored. For example, which core ICU services should be provided by all hospitals? If some services are, in fact, regionalized, will ICU volumes be adequate in other hospitals to maintain clinicians' skills in providing such services and to financially support these ICUs? Such questions may be best addressed by the establishment of community-based outcome initiatives that can track changes in patient outcomes and the costs of critical care over time.

In addition, differences in calibration of risk adjustment methods based on established national normative models or on locally derived models raise important questions regarding which yardstick is most appropriate to publicly compare institutional performance. For example, consumers may be most interested in local benchmarks that provide relative performance of facilities in a single region, given that for many services (particularly ICU care) it is not feasible to seek care elsewhere. Furthermore, because of changes in practice, national benchmarks that are not frequently updated may lose their clinical relevance. However, large, national corporations purchasing care in many health-care markets may believe a national reference point is not appropriate to compare outcomes across regions in which they operate. The use of national benchmarks may also highlight areas in which similar local practices lead to suboptimal outcomes, and for which local standards may not facilitate appropriate changes in care.

In summary, the current study suggests that physiologically based risk stratification methods can be successfully applied in nonresearch settings, are likely to be highly discriminatory, and can explain a substantial amount of the differences in observed ICU outcomes across institutions. The implementation of community-based initiatives that are based on such methods may provide important information about variations in patient outcomes and changes in practice patterns that occur over time. Such programs may also provide important insight into the impact of changes in the financing and organization of health care. The impact of current programs on

health-care purchasing and on improving quality of care should be studied further.

REFERENCES

- 1 Sirio CA, McGee JL. Public reporting of clinical outcomes—the data needs of health care stakeholders. *Am J Med Qual* 1996; 11:S78–S81
- 2 Chassin MR. Quality improvement nearing the 21st century: prospects and perils. *Am J Med Qual* 1996; 11:S4–S7
- 3 Schroeder J, Lamb S. Data initiatives: HEDIS and the New England Business Coalition. *Am J Med Qual* 1996; 11:S58–S62
- 4 Tilley S, Bomberger D, Ackroyd T, et al. Data initiatives: building a managed care network utilizing clinical performance data. *Am J Med Qual* 1996; 11:S22–S25
- 5 Wennberg J, Gittelsohn A. Small area variations in health care delivery. *Science* 1973; 182:1102–1108
- 6 Chassin MR, Brook RH, Park RE, et al. Variations in the use of medical and surgical services by the Medicare population. *N Engl J Med* 1986; 314:285–290
- 7 Siu AL, Sonnenberg FA, Manning WG, et al. Inappropriate use of hospitals in a randomized trial of health insurance plans. *N Engl J Med* 1986; 315:1259–1266
- 8 Wennberg JE, Freeman JL, Culp WJ. Are hospital services rationed in New Haven or over-utilised in Boston? *Lancet* 1987; 1:1185–1189
- 9 Wennberg JE, Cooper MM. *The Dartmouth atlas of health care*. Chicago, IL: American Hospital Publishers, 1998
- 10 Rosenthal GE, Harper DL. Cleveland health quality choice: a model for collaborative community-based outcomes assessment. *J Comm J Qual Improv* 1994; 20:425–442
- 11 McClure W. Competition and the pursuit of quality: a conversation with Walter McClure [interview by John K. Iglehart]. *Health Affairs* 1988; 7:79–90
- 12 Shaller DV, Woods P. Reforming the market for value in health care: the Cleveland experience. *J Occup Med* 1991; 33:358–362
- 13 Winslow R. Cincinnati firms cutting costs with hospital-ranking system. *Wall Street Journal* 1993; B1
- 14 Knaus WA, Wagner DP, Lynn J. Short-term mortality predictions for critically ill hospitalized adults: science and ethics. *Science* 1991; 254:389–394
- 15 Lemeshow S, Le Gall JR. Modeling the severity of illness of ICU patients: a systems update. *JAMA* 1994; 272:1049–1055
- 16 Shortell SM, Rousseau DM, Gillies RR, et al. Organizational assessment in intensive care units (ICUs): construct development, reliability, and validity of the ICU nurse-physician questionnaire. *Med Care* 1991; 29:709–726
- 17 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
- 18 Blois MS. Medicine and the nature of vertical reasoning. *N Engl J Med* 1988; 318:847–851
- 19 Rosenthal GE, Sirio CA, Shepardson LB, et al. Use of intensive care units for patients with low severity of illness. *Arch Intern Med* 1998; 158:1144–1151
- 20 Damiano AM, Bergner M, Draper EA, et al. Reliability of a measure of severity of illness: acute physiology and chronic health evaluation—II. *J Clin Epidemiol* 1992; 45:93–101
- 21 Knaus WA, Wagner DP, Zimmerman JE, et al. Variations in mortality and length of stay in intensive care units. *Ann Intern Med* 1993; 118:753–761
- 22 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29–36

- 23 Iezzoni LI, Shwartz M, Moskowitz MA, et al. Illness severity and costs of admissions at teaching and nonteaching hospitals. *JAMA* 1990; 264:1426-1431
- 24 Hosmer DW Jr, Lemeshow S. *Applied logistic regression*. New York, NY: John Wiley, 1989
- 25 Miller MG, Miller LS, Fireman B, et al. Variation in practice for discretionary admissions: impact on estimates of quality of hospital care. *JAMA* 1994; 271:1493-1498
- 26 Junker C, Zimmerman JE, Wagner DP, et al. Independent validation of APACHE III outcome prediction among ICU admissions for head trauma. *Crit Care Med* 1996; 24(suppl): A50
- 27 Sirio CA, Tajimi K, Tase C, et al. An initial comparison of intensive care in Japan and the United States. *Crit Care Med* 1992; 20:1207-1215
- 28 Reynolds HN, Haupt MT, Thill-Baharozian MC, et al. Impact of critical care physician staffing on patients with septic shock in a university hospital medical intensive care unit. *JAMA* 1988; 260:3446-3450
- 29 Pollack MM, Katz RW, Ruttimann UE, et al. Improving the outcome and efficiency of intensive care: the impact of an intensivist. *Crit Care Med* 1988; 16:11-17
- 30 Brown JJ, Sullivan G. Effect on ICU mortality of a full-time critical care specialist. *Chest* 1989; 96:127-129
- 31 Li TC, Phillips MC, Shaw L, et al. On-site physician staffing in a community hospital intensive care unit. Impact on test and procedure use and on patient outcome. *JAMA* 1984; 252:2023-2027
- 32 Ron A, Aronne LJ, Kalb PE, et al. The therapeutic efficacy of critical care units: identifying subgroups of patients who benefit. *Arch Intern Med* 1989; 149:338-341
- 33 Petty TL, Lakshminarayan S, Sahn SA, et al. Intensive respiratory care unit: review of 10 years' experience. *JAMA* 1975; 233:34-37
- 34 Hook EW III, Horton CA, Schaberg DR. Failure of intensive care unit support to influence mortality from pneumococcal bacteremia. *JAMA* 1983; 249:1055-1057
- 35 Knaus WA, Le Gall JR, Wagner DP, et al. A comparison of intensive care in the USA and France. *Lancet* 1982; 2:642-646
- 36 Zimmerman JE, Knaus WA, Judson JA, et al. Patient selection for intensive care: a comparison of New Zealand and United States hospitals. *Crit Care Med* 1988; 16:318-326
- 37 Abizanda R. The facilities. In: Miranda DR, Williams A, Loirate P, eds. *Management of the ICU: guidelines for better use of resources*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1990; 55-81
- 38 Knaus WA, Draper EA, Wagner DP, et al. An evaluation of outcome from intensive care in major medical centers. *Ann Intern Med* 1986; 104:410-418
- 39 Sirio C, Coleman M, McGrath B, et al. The impact of intensivists and ICU occupancy on rates of low risk-monitor-only (LRM) admissions to the ICU. *Chest* 1997; 112:16S
- 40 Rosenthal GE, Harper DL, Quinn LM, et al. Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals: results of a regional study. *JAMA* 1997; 278:485-490
- 41 Hannan EL, Kilburn H Jr, Racz M, et al. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA* 1994; 271:761-766
- 42 Rosenthal GE, Quinn L, Harper DL. Declines in hospital mortality associated with a regional initiative to measure hospital performance. *Am J Med Qual* 1997; 12:103-112
- 43 Steen PM, Brewster AC, Bradbury RC, et al. Predicted probabilities of hospital death as a measure of admission severity of illness. *Inquiry* 1993; 30:128-141
- 44 Ghali WA, Ash AS, Hall RE, et al. Statewide quality improvement initiatives and mortality after cardiac surgery. *JAMA* 1997; 277:379-382
- 45 Gostin LO, Lazzarini Z, Neslund VS, et al. The public health information infrastructure: a national review of the law on health information privacy. *JAMA* 1996; 275:1921-1927
- 46 Gostin LO. Health care information and the protection of personal privacy: ethical and legal considerations. *Ann Intern Med* 1997; 127:683-690
- 47 Rosenthal GE, Halloran EJ, Kiley M, et al. Development and validation of the nursing severity index: a new method for measuring severity of illness using nursing diagnoses: Nurses of University Hospitals of Cleveland. *Med Care* 1992; 30: 1127-1141
- 48 Reuben DB, Rubenstein LV, Hirsch SH, et al. Value of functional status as a predictor of mortality: results of a prospective study [published errata appear in *Am J Med* 1993 Feb;94(2):232 and 1993 May;94(5):560]. *Am J Med* 1992; 93:663-669
- 49 Covinsky KE, Justice AC, Rosenthal GE, et al. Measuring prognosis and case mix in hospitalized elders: the importance of functional status. *J Gen Intern Med* 1997; 12:203-208
- 50 Jayes RL, Zimmerman JE, Wagner DE, et al. Do-not-resuscitate orders in intensive care units: current practices and recent changes. *JAMA* 1993; 270:2213-2217
- 51 Shepardson LB, Younger S, Speroff T, et al. A community-based analysis of patient and hospital variation in the use of do-not-resuscitate orders. *Arch Intern Med* 1997; 157:1841-1847
- 52 Shortell SM, Zimmerman JE, Rousseau DM, et al. The performance of intensive care units: does good management make a difference. *Med Care* 1994; 32:508-525
- 53 Hannan EL, O'Donnell JF, Kilburn H Jr, et al. Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. *JAMA* 1989; 262:503-510