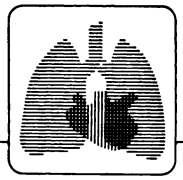


The APACHE III Prognostic System*

Risk Prediction of Hospital Mortality for Critically Ill Hospitalized Adults

William A. Knaus, M.D.; Douglas P. Wagner, Ph.D.; Elizabeth A. Draper, M.S.; Jack E. Zimmerman, M.D.; Marilyn Bergner, Ph.D.; Paulo G. Bastos, M.D.; Carl A. Sirio, M.D.; Donald J. Murphy, M.D.; Ted Lotring, M.S.; Anne Damiano, M.S.; and Frank E. Harrell Jr., Ph.D.

The objective of this study was to refine the APACHE (Acute Physiology, Age, Chronic Health Evaluation) methodology in order to more accurately predict hospital mortality risk for critically ill hospitalized adults. We prospectively collected data on 17,440 unselected adult medical/surgical intensive care unit (ICU) admissions at 40 US hospitals (14 volunteer tertiary-care institutions and 26 hospitals randomly chosen to represent intensive care services nationwide). We analyzed the relationship between the patient's likelihood of surviving to hospital discharge and the following predictive variables: major medical and surgical disease categories, acute physiologic abnormalities, age, preexisting functional limitations, major comorbidities, and treatment location immediately prior to ICU admission. The APACHE III prognostic system consists of two options: (1) an APACHE III score, which can provide initial risk stratification for severely ill hospitalized patients within independently defined patient groups; and (2) an APACHE III predictive equation, which uses APACHE III score and reference data on major disease categories and treatment location immediately prior to ICU admission to provide risk estimates for hospital mortality for individual ICU patients. A five-point increase in APACHE III score (range, 0 to 299) is independently associated with a statistically significant increase in the relative risk of hospital death (odds ratio, 1.10 to 1.78) within each of 78 major medical and surgical disease categories. The overall predictive accuracy of the first-day APACHE III equation was such that, within 24 h of ICU admission, 95 percent of ICU admissions could be given a risk estimate for hospital death that was within 3 percent of that actually observed ($r^2 = 0.41$; receiver operating characteristic = 0.90). Recording changes in the APACHE III score on each subsequent day of ICU therapy provided daily updates in these risk estimates. When applied across the individual ICUs, the first-day APACHE III equation accounted for the majority of variation in observed death rates ($r^2 = 0.90$, $p < 0.0001$).

(Chest 1991; 100:1619-36)

ROC = receiver operating characteristic

The ability to objectively estimate patient risk for mortality or other important outcomes is a new undertaking for clinical research.[1-4] Empirically based risk assessments for important clinical events have been extremely useful in evaluating new therapies, in monitoring resource utilization, and in improving quality assessment.[5,6] Attempts at prediction, however, have been much less successful in forecasting individual patient risk or in reducing the uncertainty of daily clinical decision making.[7]

Objective risk estimates are particularly important in the high-cost, emotional, and technologically demanding environments of intensive care units (ICUs). Because of the high costs of ICUs, precise quality assurance and utilization management strategies are essential.[6] Knowledge of the risk faced by a patient on the day of ICU admission could provide an empiric basis for quality assurance and utilization activities. Estimates during the course of therapy could be useful in investigating the optimal time for discharge or in deciding how long to continue therapy.[8] The demand for intensive treatment is growing, but resources are increasingly constrained, and many ICUs are already overcrowded.[9] One half of all the deaths that now

*From the ICU Research Unit, Department of Anesthesiology, George Washington University Medical Center, Washington, DC (Drs Knaus, Wagner, Zimmerman, Bastos, and Murphy; Mr Lotring); APACHE Medical Systems, Inc (Ms Draper); Health Services Research and Development Center, School of Hygiene and Public Health, Johns Hopkins University, Baltimore (Dr Bergner, Ms Damiano); Department of Critical Care Medicine, National Institutes of Health, Bethesda, Md (Dr Sirio); and Division of Biometry, Duke University Medical Center, Durham, NC (Dr Harrell).

Supported by the Agency for Health Care Policy and Research (grant No. HS05787); The John A. Hartford Foundation (grant No. 87267); the Department of Anesthesiology, George Washington University Medical Center; and APACHE Medical Systems, Inc. Dr Bastos was supported by a grant from the National Council of Scientific and Technology Development (CNPq), Brazil.

Manuscript received May 15; revision accepted August 13.

Reprint requests: Dr. Knaus, 2300 K Street, NW, Washington, DC 20037

occur in ICUs take place only after a decision has been made that further therapy would be futile.[10] Physicians, patients, and society want to ensure that decision making is accurate and compatible with current therapeutic capabilities.[11] In this article, we present the background, development, details of measurement, and validation for the APACHE III system, which is aimed at addressing these various issues and challenges.

The development of APACHE III was based on the association between acute changes in a patient's physiologic balance and short-term risk of death.[12-14] Details of the two previous versions have been reviewed recently.[15]

Briefly, in developing APACHE III, we sought to improve upon the risk prediction available with APACHE II by reevaluating the selection and weighting of physiologic variables while examining how differences in patient selection for and timing of admission to ICUs related to outcome variations across hospitals. We also sought to (1) clarify the distinction between using the APACHE scoring system to stratify by risk of mortality within independently defined patient groups and using it to estimate individual risks of mortality; (2) expand the size and representativeness of our reference data base; and (3) examine issues regarding the selection of patients and the timing of scoring.[16-18] These and additional considerations which led to this investigation are detailed in a comprehensive study design.[15]

Patients and Methods

The two major analytic steps in developing APACHE III were (1) the collection of an appropriate data base and (2) analysis to establish a final system design. First, we assembled a list of candidate variables and questions for each of the five major predictive constructs (major disease categories, acute physiology, age, comorbidities, and origin and timing of patient selection).

Variable Selection

To determine the primary reason for ICU admission, we developed a comprehensive list of 212 disease categories. Each category classified the patient according to medical or surgical status, major organ system involved, and, when possible, specific etiology. Within 24 h of ICU admission, a patient was assigned to one of these major disease categories by using data available in the medical record that indicated the disease process most directly responsible for the patient's ICU admission. On the basis of past experience and clinical judgment, we selected 20 physiologic variables to measure severity of disease. Data collectors obtained physiologic data from ICU flow sheets and recorded the exact values at various times during the ICU study.

Data on the patient's prior health status consisted of chronologic age and the presence of one or more preexisting functional limitations or comorbidities. Chart abstractors collected these data during the initial 24 h of ICU admission and also noted the location (emergency, recovery, hospital, or operating room; ICU readmission; or transfer from another ICU or hospital) and, when applicable, the time between the patient's arrival in the emergency room and ICU admission.

Hospital and ICU Selection

We randomly selected 26 hospitals to be representative of all 1,691 hospitals in the continental United States with 200 or more acute-care beds. Approximately 50 percent of all hospitals and 85 percent of all adult ICU beds in the United States were eligible for participation in the study. Hospitals listed on the 1985 American Hospital Association data tape were grouped into 16 strata based on geographic location (Northeast, Southeast, Midwest, and West), size (200 to 350 beds and >350 beds), and teaching status, as defined by the presence of resident house staff or the existence of an accredited graduate medical training program. A computer-based random number table was used to assign a number to each hospital. We invited the first hospital listed within each of the 16 strata to participate and the second hospital in ten strata. When a hospital declined to participate, a randomly selected alternate was chosen. Among the initial 26 randomly chosen hospitals, 23 agreed to participate. The three reasons for nonparticipation were the sale of a hospital; a severe nursing shortage, making data collection assistance unlikely; and a poor fiscal condition, making bankruptcy and closure imminent. In all three cases randomly selected replacement hospitals were chosen from the same computerized listing.

The 26 randomly selected hospitals were combined with 14 volunteer institutions to complete the 40-hospital data base. The 14 volunteer hospitals were primarily tertiary-care referral institutions with an interest in this project. In hospitals with more than one ICU, data collection took place in the unit with the highest annual admission rate. In two of the volunteer institutions, data collection took place in two separate ICUs.

Patient Selection

The inception cohort at each ICU consisted of approximately 400 ICU patients consecutively admitted following initiation of data collection. When patient volume precluded data collection on consecutive admissions, we used an alternating data collection scheme (*eg*, every second or third patient). A patient had to remain in the ICU for a minimum of 4 h to be included in the study. We did not include patients with burn injuries, patients aged less than 16 years, or individuals with chest pain who were admitted to rule out myocardial infarction. Other cardiac diagnoses, however, were included. Data for coronary artery bypass graft patients were collected in an independent data file that will be reported separately. No patients were excluded because of missing values. All breaks in data collection were reviewed and approved by the data coordinating center, which continuously monitored data quality.

Data collection began in May 1988 and was completed by November 1989. All patients were followed up for survival at hospital discharge. Data on survival after hospitalization were obtained for all Medicare patients and for a 15 percent random sample of all other patients.

Data Collection

After agreeing to participate, each hospital selected data collectors and sent them to Washington, DC, for a three-day training course that included detailed instruction in recording all aspects of disease, physiologic, and chronic health data. As an initial quality assurance test, we reviewed data collected on the first 20 patients. If this review demonstrated that the data collectors were able to complete data forms with accuracy and completeness, data collection began; otherwise, additional training occurred. All patient data were entered into on-site microcomputers by using specially designed software with extensive internal logic and edit checks to increase the accuracy of data collection.

Reliability of Data Collection

Eleven hospitals were selected for a formal reliability study because they reported anecdotal data collection problems.[19] Because

this reliability study was completed in 1989, before analysis of the APACHE III data base was completed, APACHE II, rather than APACHE III, scores were compared. Assignment of the Acute Physiology Score component of APACHE II was highly reproducible (interclass correlation coefficient = 0.90). Percent agreement values for patient age, sex, race, ICU admission date, origin of admission, time from emergency room to ICU, and primary system failure were 91.8, 99.5, 95.4, 97.4, 98.0, 85.7, and 93.9, respectively. In 77.4 percent of cases with one or more chronic health items, both abstractors agreed that the patient had one of the 34 potential chronic health conditions.[19] A reanalysis using the APACHE III weighting approach has demonstrated similar results.

Statistical Analysis

Our specific analytic objectives were to maximize the explanatory power of APACHE III, to improve discrimination at both low and high risk of death, and to maintain maximum measurement reliability. This involved examining weighting of the components of the APACHE III scale as well as investigating their relationship with an APACHE III equation that predicted patient outcome by using the APACHE III score, disease classification, and patient selection practices.

Upon completion of data collection, we first examined differences in mortality between the 14 volunteer hospitals and the 26 randomly selected hospitals, using the APACHE II equation.[20] No significant differences in risk-adjusted mortality rates between these two groups were found ($p = 0.20$). Therefore, we combined data from all 40 hospitals into a common file for all subsequent analyses. Ninety percent of this data file was then randomly divided into estimation and validation halves, with approximately 50 percent of each hospital's patients included in each half. Analyses used in developing weights for the physiologic components of APACHE III were completed on the estimation half of the data file alone. The 40 disease categories from the APACHE II equation[20] served as disease-specific controls. The chronic health, age, and previously described patient location variables were also controlled statistically in this analysis.

Component Variables and Weights

To estimate weights for the physiologic variables, we used multivariable logistic regression analyses to determine the relationship between death rate and each of the 20 candidate physiologic variables, controlling for 19 other physiologic variables, age, chronic health conditions, operative status, and major disease categories by using both categoric and continuous weighting approaches[21] (Appendix A).

We also explored interactions between physiologic variables by evaluating individual and combined variable weighting. For the variables reflecting acid-base disturbances (serum pH, Pco_2, and bicarbonate), we found discrepancies that were not compatible with established physiologic principles. The computer-derived weights for a serum Pco_2 above 50 mm Hg were consistently estimated as having little or no significant relationship to risk of death. We hypothesized that this was because the appropriate weighting for Pco_2 is also dependent on the associated serum pH (*ie*, whether there is a primary or secondary respiratory disorder). Therefore, we developed a combined variable, which included serum pH and Pco_2, to establish weights for common acid-base disorders. The subsequent estimation of weights for this combined variable proceeded as did all others. In addition, we found important interactions between urine output and serum creatinine, and also between respiratory rate, PaO_2, and ventilator use. In each case, we developed combined variables and compared them with individual variables for the clinical validity of their respective weights. All physiologic weights estimated on half of the data base were subsequently validated in the independent half. Details regarding choice of timing, assigning weights, disease-specific weighting,[22,23] and assigning zero weight to missing physiologic values are in Appendix A.

We estimated weights for the comorbid chronic health variables that met our magnitude and statistical criteria and for age divided into five-year ranges using the entire data file (both estimation and validation halves). Weights assigned to the physiologic variables, however, remained those established by the estimation sample. We then converted the weights for physiology, age, and chronic health conditions into points for construction of the final APACHE III score.

The APACHE III Equation

To produce an equation for predicting hospital mortality after the first day of ICU treatment, we combined the disease and patient location coefficients with the relative weights assigned to the first-day values of the three components of the APACHE III score (physiology, age, chronic health). Because of the large number of disease categories in the APACHE III data base, coefficients for each disease classification with a sufficient number of patients and/or deaths were obtained using the entire data file. The coefficients for specific diagnostic categories were examined to assess their stability with regard to the weights derived and used with APACHE II and the clinical experience of the investigators. The clinical homogeneity, the cell size, and the impact of the disease on short-term outcome were the criteria used for assessing stability.

Because of previous evidence[24,25] suggesting that the patient's treatment location immediately prior to ICU admission held important prognostic significance, we next assessed the effect on overall explanatory power of a variable describing selection for intensive care. We did this by assessing the effect of the patient's

Table 1—*Demographic Characteristics of Patients*

Total patients	
26 random hospitals	10,941
14 volunteer hospitals	6,499
	17,440
Nonoperative admissions	
Emergency room	6,199
Floor	2,860
Transfer from other hospital	423
Transfer from other ICU	581
	10,063 (58)*
Postoperative admissions	
Elective surgery	5,811
Emergency surgery	1,566
	7,377 (42)*
Average number of patients in each unit (range)	425 (299-449)
Age, yr (mean, 59 yr)	
<45	23.2†
45-54	10.8
55-64	18.0
65-74	25.5
75-84	17.2
≥85	5.3
Sex†	
Male	44.8
Female	55.2
Race†	
White	80.3
Black	14.1
Hispanic	4.1
Asian	1.2
American Indian	0.3

*Values in parentheses are percentages of total number of patients.
†Values are percentages of total number of patients.

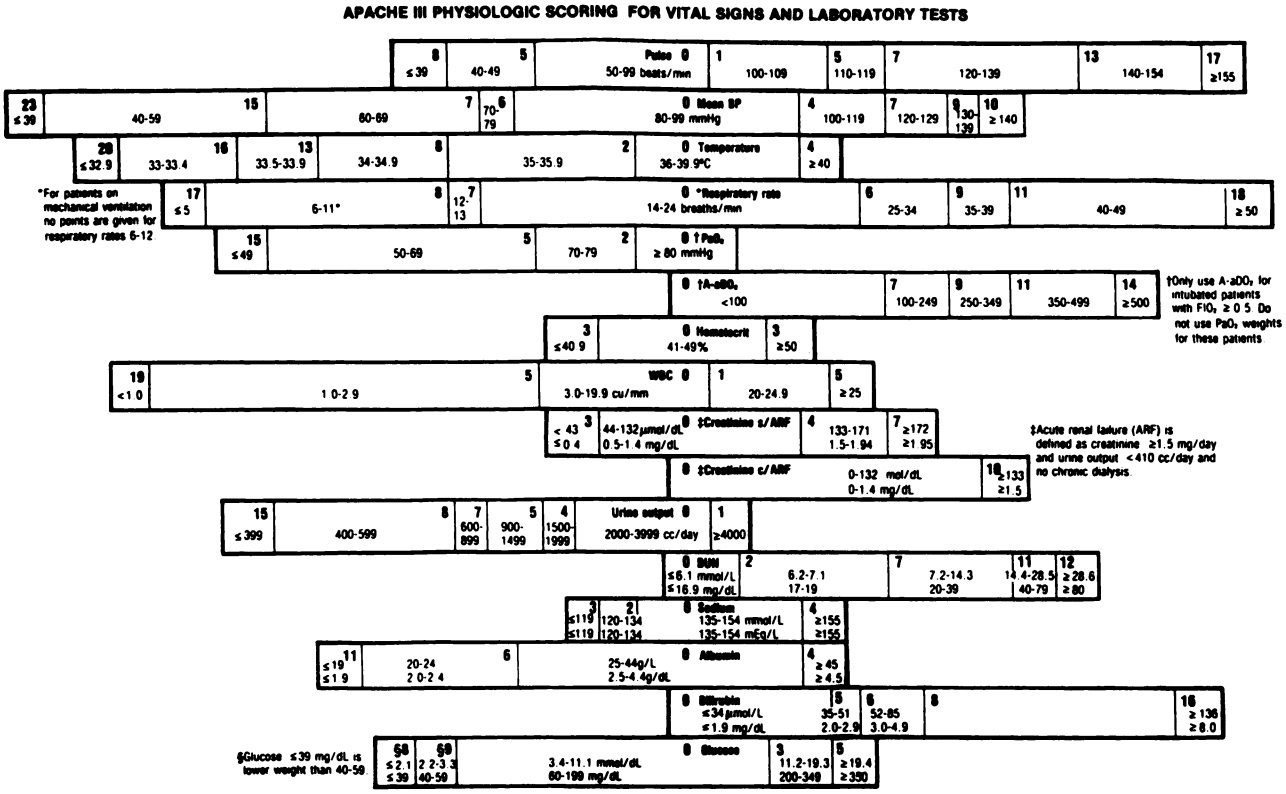


FIGURE 1. APACHE III scoring for vital signs and laboratory abnormalities.

location immediately prior to ICU admission (emergency room, floor, other ICU, other hospital) and the time spent in the emergency room on the entire model. We also investigated the influence of whether surgery was performed on an emergency basis on total explanatory power.

In order to provide an initial basis for predictions over time that require incorporating changes in the patient's physiologic status, we investigated the use of daily APACHE III scores as an additional independent variable within the above equation.

RESULTS

The majority (89 percent) of the 40 hospitals were nonprofit, and 54 percent were affiliated with a medical school. The average number of hospital beds was 359, the average number of ICU beds was 21, and the average number of ICU beds was 13. These characteristics reflect national statistics on the 1,691 US hospitals with 200 or more beds. Of the 42 ICUs studied, 71 percent were mixed medical-surgical, 16 percent were surgical, 10 percent were medical, and 3 percent represented other specialties.

Patient demographics, including location prior to ICU admission, appear in Table 1. A total of 9,195 (53 percent) of the patients had positive answers to one or more of the 34 chronic health and functional status questions. This proportion decreased to 33 percent for patients admitted following elective surgery.

Physiologic Variables and Their Weights

Compared to the weighting system used in APACHE II, increased explanatory power for patient outcome resulted when additional weight was assigned to the extremes of physiologic measurements and when narrower ranges of physiologic measures were assigned a zero or normal weight. For some variables, such as blood pressure, the risk associated with extremely high recordings is different from the risk associated with equally extreme but low recordings. Overall explanatory power for patient outcome also increased when we accounted for the following interactions: serum pH with Pco_2, serum creatinine with urine output, and respiratory rate with ventilator use. As a result of this analysis, the weighting of each remaining physiologic variable used in APACHE II was refined for inclusion in APACHE III, and five new variables (blood urea nitrogen, urine output, serum albumin, bilirubin, and glucose) were included (Fig 1 and 2). Examination of the relationship between death rate and each of the 20 candidate physiologic variables revealed that serum potassium and serum bicarbonate did not meet our minimal statistical inclusion criteria.

Reliability results suggested that reformatting the Glasgow Coma variables to eliminate similar scores with different clinical presentations would be useful. We accomplished this by eliminating the distinctions between incomprehensible words and inappropriate sounds, flexion withdrawal and decorticate rigidity, and decerebrate rigidity and no responses, and by simplifying the evaluation of eye opening (Fig 3).

APACHE III PHYSIOLOGIC SCORING FOR ACID BASE ABNORMALITIES

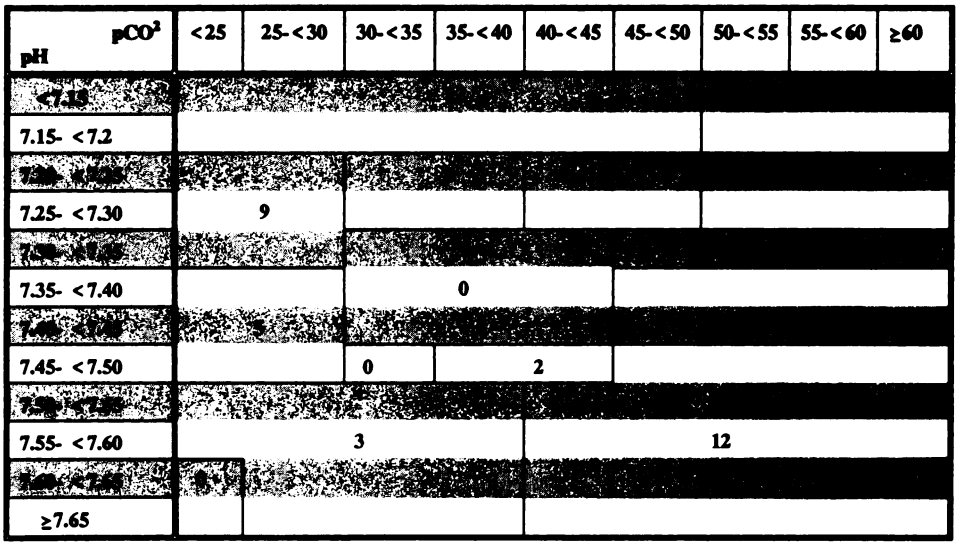


FIGURE 2. APACHE III scoring for acid-base disturbances.

On the basis of comparative results, we retained the worst value over the initial 24 h of ICU care as the most appropriate scoring approach for the physiologic component of APACHE III. The final results for weightings of all physiologic variables are in Figures 1, 2, and 3. The range for total physiologic weight is from 0 to 252; for an individual variable, the range is 0 to 48. A missing physiologic value is assigned a zero weight (Appendix A). In Figure 4, we have compared the explanatory power and discrimination of the final weights of these 17 physiologic variables in the estimation and validation halves of the data base, using

APACHE III PHYSIOLOGIC SCORING FOR NEUROLOGIC ABNORMALITIES

Eyes open spontaneously or to painful/verbal stimulation

verbal / motor	oriented converses	confused conversation	inappropriate words and incomprehensible sounds	no response
obeys verbal command	0	3	10	15
localizes pain	3	8	13	15
flexion withdrawal/ decorticate rigidity	3	13	24	24
decerebrate rigidity/ no response	3	13	29	29

Eyes do not open spontaneously or to painful/verbal stimulation

verbal / motor	oriented converses	confused conversation	inappropriate words and incomprehensible sounds	no response
obeys verbal command				16
localizes pain				16
flexion withdrawal/ decorticate rigidity			24	33
decerebrate rigidity/ no response			29	48

The shaded areas without scores represent unusual and unlikely clinical combinations. There were few or no cases in these cells. For the shaded areas with scores we had data that permit us to extrapolate values. Placing a patient in any of these cells should be done after careful confirmation of clinical findings.

FIGURE 3. APACHE III scoring for neurologic abnormalities according to presence or absence of eye opening.

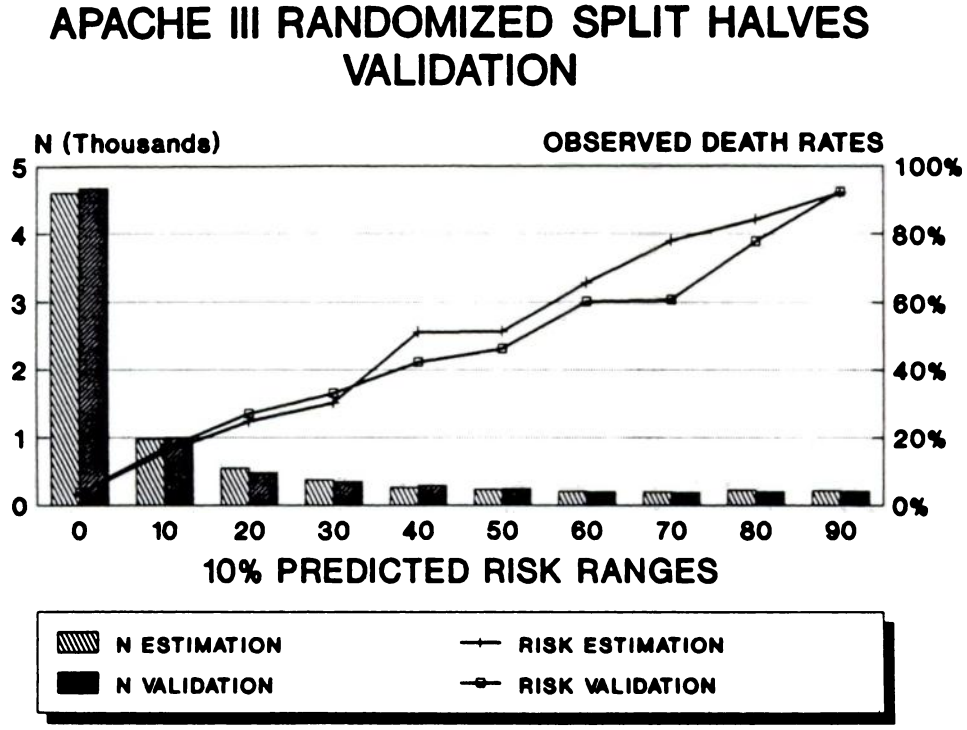


FIGURE 4. APACHE III physiology randomized split halves validation. Patient distribution and observed death rate for each 10 percent increment of predicted death risk are shown for 7,840 patients in the estimation half and 7,840 patients in the validation half, using APACHE III physiologic weights in both. Disease and chronic health weights were from APACHE II.

the 40 APACHE II diagnostic categories as controls for disease. The APACHE III physiologic weights forecast well to the validation sample (receiver operating characteristic [ROC] = 0.88 estimation; 0.87 validation).

Comorbidity and Age Weights

Estimation of the 34 candidate chronic health items yielded seven variables (acquired immunodeficiency syndrome [AIDS], hepatic failure, lymphoma, solid tumor with metastasis, leukemia/multiple myeloma, immunocompromise, and cirrhosis) that met previously established statistical criteria for inclusion (Table 2). Because these variables did not occur frequently with elective postoperative ICU admissions and because they did not improve overall explanatory power within these elective surgical categories, the comorbid conditions are not required for elective postoperative admissions. They are included as part of APACHE III when the patient is an emergency surgery case, defined as surgery to treat an immediate life-threatening condition.

Using the physiologic weights derived from the estimation half of the data base, we then reestimated the remainder of the equation on the entire data base, this time using 78 diagnostic categories adapted from the original 212 based on frequency and death rate (Table 3). Using this equation, we also derived the final weights for the chronic health evaluation (range, 0 to 23) and age (range, 0 to 24) variables (Table 2).

APACHE III Score

The score that results from the addition of the three groups of variables (physiology, age, and chronic health) is a cardinal number with a range of 0 to 299 (physiology, 0 to 252; chronic health evaluation, 0 to 23; age, 0 to 24). It is referred to as the APACHE III score. The mean APACHE III score in this population was 50.

Table 2—*APACHE III Points for Age and Chronic Health Evaluation*

	Points
Age, yr	
≤44	0
45-59	5
60-64	11
65-69	13
70-74	16
75-84	17
≥85	24
Comorbid condition*	
AIDS	23
Hepatic failure	16
Lymphoma	13
Metastatic cancer	11
Leukemia/multiple myeloma	10
Immunosuppression	10
Cirrhosis	4

*Excluded for elective surgery patients.

Table 3—*Major Disease Categories in APACHE III Prognostic System**

Disease Category	Total Patients	Unadjusted Hospital Death Rate, %	Odds Ratio for Increase in Hospital Death Rate Risk for 5-Point Increase in APACHE III Score	95% Confidence Interval
Nonoperative				
Cardiovascular/vascular				
Cardiogenic shock	41	65.9	1.20	1.07-1.34
Cardiac arrest	414	59.9	1.24	1.19-1.29
Aortic aneurysm	53	26.4	1.11	1.00-1.23
Congestive heart failure	891	21.0	1.30	1.24-1.35
Peripheral vascular disease	95	13.7	1.56	1.26-1.93
Rhythm disturbance	340	10.6	1.33	1.22-1.44
Acute myocardial infarction	603	10.1	1.38	1.28-1.48
Hypertension	124	8.1	1.31	1.13-1.52
Other cardiovascular diseases	272	21.0	1.30	1.22-1.40
Respiratory				
Parasitic pneumonia	59	64.4	1.10	1.03-1.18
Aspiration pneumonia	102	50.0	1.18	1.09-1.28
Respiratory neoplasm (including larynx, trachea)	54	51.9	1.12	1.04-1.21
Respiratory arrest	131	38.2	1.17	1.09-1.25
Pulmonary edema (non-cardiogenic)	107	37.4	1.21	1.11-1.33
Bacterial/viral pneumonia	454	32.8	1.21	1.16-1.26
Chronic obstructive pulmonary disease	362	23.8	1.28	1.20-1.36
Pulmonary embolism	165	19.4	1.24	1.15-1.34
Mechanical airway obstruction	68	17.6	1.30	1.13-1.49
Asthma	140	7.9	1.40	1.18-1.66
Other respiratory diseases	319	32.3	1.22	1.16-1.28
Gastrointestinal (GI)				
Hepatic failure	30	50.0	1.12	1.02-1.23
GI perforation/obstruction	87	26.4	1.34	1.16-1.53
GI bleeding due to varices	210	23.3	1.21	1.13-1.28
GI inflammatory disease (ulcerative colitis/crohn's/pancreatitis	168	22.6	1.25	1.15-1.36
GI bleeding due to ulcer/laceration	644	14.4	1.28	1.22-1.34
GI bleeding due to diverticulosis	103	6.8	1.44	1.14-1.81
Other GI diseases	92	28.3	1.27	1.14-1.41
Neurologic				
Intracerebral hemorrhage	194	54.1	1.37	1.27-1.49
Subarachnoid hemorrhage	161	36.6	1.39	1.26-1.52
Stroke	326	30.4	1.25	1.19-1.32
Neurologic infection	51	27.5	1.14	1.03-1.25
Neurologic neoplasm	40	25.0	1.30	1.08-1.55
Neuromuscular disease	45	15.6	1.32	1.07-1.63
Seizure	309	15.2	1.32	1.22-1.42
Other neurologic diseases	115	11.3	1.32	1.15-1.51
Sepsis				
Sepsis (other than urinary tract)	415	52.0	1.18	1.14-1.23
Sepsis of urinary tract origin	104	29.8	1.15	1.07-1.23
Trauma				
Head trauma (with/without multiple trauma)	477	13.4	1.30	1.23-1.37
Multiple trauma (excluding head trauma)	399	2.0	1.44	1.23-1.67
Metabolic				
Metabolic coma	68	35.3	1.31	1.15-1.49
Diabetic ketoacidosis	274	4.4	1.23	1.13-1.34
Drug overdose	646	0.9	1.42	1.22-1.65
Other metabolic diseases	143	20.3	1.34	1.20-1.50
Hematologic				
Coagulopathy/neutropenia/thrombocytopenia	37	45.9	1.37	1.13-1.65
Other hematologic diseases	64	4.7	1.19	1.01-1.40
Renal diseases	92	23.9	1.18	1.06-1.31
Other medical diseases	244	10.2	1.46	1.29-1.64
Postoperative				
Vascular/cardiovascular				
Dissecting/ruptured aorta	104	41.3	1.20	1.12-1.30
Peripheral vascular disease (no bypass graft)	215	14.4	1.28	1.17-1.40
Valvular heart surgery	211	8.1	1.31	1.15-1.49
Elective abdominal aneurysm repair	525	6.5	1.27	1.18-1.37
Peripheral artery bypass graft	535	4.7	1.51	1.32-1.73

(Continued)

Table 3 continued

Carotid endarterectomy	429	2.1	1.78	1.41-2.25
Other cardiovascular diseases	225	9.8	1.29	1.17-1.44
Respiratory				
Respiratory infection	57	8.8	1.64	1.13-2.36
Lung neoplasm	411	5.8	1.40	1.25-1.57
Respiratory/neoplasm (mouth, sinus, larynx, trachea)	119	3.4	1.32	1.08-1.61
Other respiratory diseases	218	6.4	1.47	1.27-1.70
Gastrointestinal				
GI perforation/rupture	260	26.2	1.31	1.22-1.40
GI inflammatory disease	244	20.9	1.28	1.19-1.37
GI obstruction	308	17.2	1.26	1.17-1.36
GI bleeding	109	14.7	1.32	1.16-1.49
Liver transplant	40	12.5	1.32	1.03-1.68
GI neoplasm	500	10.4	1.30	1.20-1.40
GI cholecystitis/cholangitis	170	7.6	1.23	1.09-1.40
Other GI diseases	180	11.1	1.64	1.36-1.97
Neurologic				
Intracerebral hemorrhage	51	43.1	1.17	1.07-1.29
Subdural/epidural hematoma	88	26.1	1.35	1.19-1.52
Subarachnoid hemorrhage	93	9.7	1.34	1.13-1.57
Laminectomy/other spinal cord surgery	214	5.6	1.56	1.29-1.93
Craniotomy for neoplasm	437	5.5	1.36	1.23-1.52
Other neurologic diseases	126	7.1	1.52	1.23-1.88
Trauma				
Head trauma (with/without multiple trauma)	210	22.9	1.26	1.18-1.34
Multiple trauma (excluding head trauma)	381	3.7	1.39	1.24-1.56
Renal†				
Renal neoplasm	225	4.9	1.34	1.12-1.59
Other renal diseases	173	4.6	1.45	1.18-1.79
Gynecologic				
Hysterectomy	65	7.7	1.28	1.06-1.54
Orthopedic				
Hip or extremity fracture	139	12.2	1.19	1.05-1.36

*Disease categories represent the single most specific reason for ICU admission.
†Odds ratio for 46 renal transplants not calculated because of small number of deaths (2.2%).

The direct relationship of the APACHE III score obtained during the first day of ICU treatment within two homogeneous diagnostic groups, congestive heart failure and trauma, is illustrated in Figure 5. Within specific disease categories, the relationship of increases in the APACHE III score to the risk of death is also reported in Table 3, where the odds ratio of increased risk of death relative to APACHE III score is provided for each of 78 major disease categories included in the APACHE III data base. Each of these odds ratios is calculated from a separate logistic regression analysis within a specific disease category. The equations are statistically independent of one another.

Risk Estimate Equations for Hospital Mortality

An equation combining the explanatory power of the APACHE III score with major disease category and prior patient location permits the calculation of hospital death risk estimates for ICU patients admitted under circumstances similar to those in this study.

The relationship between disease classification and risk prediction is further illustrated in Figure 6, where predicted risk categories for four major disease is plotted against APACHE III score. When the APACHE III score is either low (≤20) or high (≥140), the relative importance of disease is small. Within the middle range of scores, however, variations in disease classification are associated importantly with variations in risk predictions for the same level of APACHE III score.

The overall explanatory power of estimates calculated from the data obtained during the first day of ICU stay is evidenced by the total r^2 of 0.41 and ROC of 0.90 (Fig 7). Overall correct classification on the first day at a 0.50 predicted risk was 88.2 percent (Fig 7). These values are significantly better than the overall explanatory power available from APACHE II (ROC = 0.85 and a correct classification at a 0.50 risk level of 85.5 percent).[6] When the initial or first-day APACHE III equation is applied across the individual ICUs, it accounts for most of the large variation in observed hospital mortality rates (6 percent to 42 percent) from these units ($r^2 = 0.90$ $p < 0.0001$).

When we investigated the use of APACHE III scores on the first and subsequent days to estimate mortality risk over time for individual patients, we discovered that the use of initial and latest-day scores achieved

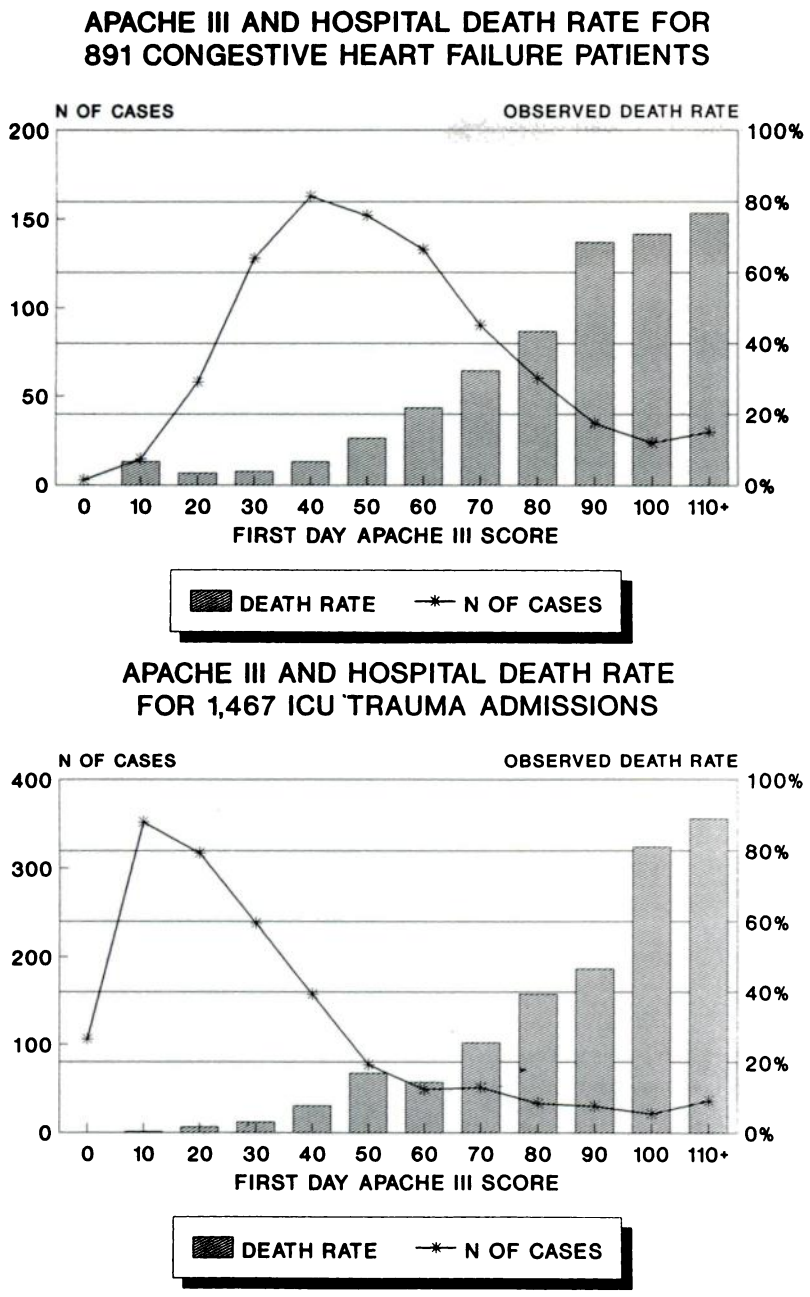


FIGURE 5. Relationship between first-day APACHE III score and risk of hospital death for patients with congestive heart failure *(top)* and trauma *(bottom)*.

maximum explanatory power. In Figure 8, we have indicated the daily risk of hospital mortality over time for two septic shock patients drawn from one of the 40 hospitals. Each daily risk estimate (after the initial day) is derived from the initial day's and most recent day's APACHE III score, along with major disease category and patient location variables.

DISCUSSION

As anticipated, acute physiologic abnormalities accounted for the largest proportion of APACHE III's total explanatory power (Appendix A). In developing new weights, we discovered that the impact of physiologic abnormalities on hospital mortality had been underestimated in APACHE II. For example, the relationship between blood pressure and outcome was more complex than originally hypothesized with greater risk associated with hypotension than hypertension, a relationship previously reported by others.[22] We assign zero weight to missing physiologic data because inferred weights would artificially inflate risk estimates and because other techniques suggested that normal imputed values were most appropriate. Because a continuous weighting approach did not improve overall explanatory power enough to warrant the difficulty of its application, APACHE III uses specific physiologic cut points that permit direct calculation of the APACHE III score (see Appendix A for details).

When we evaluated the incremental contribution of

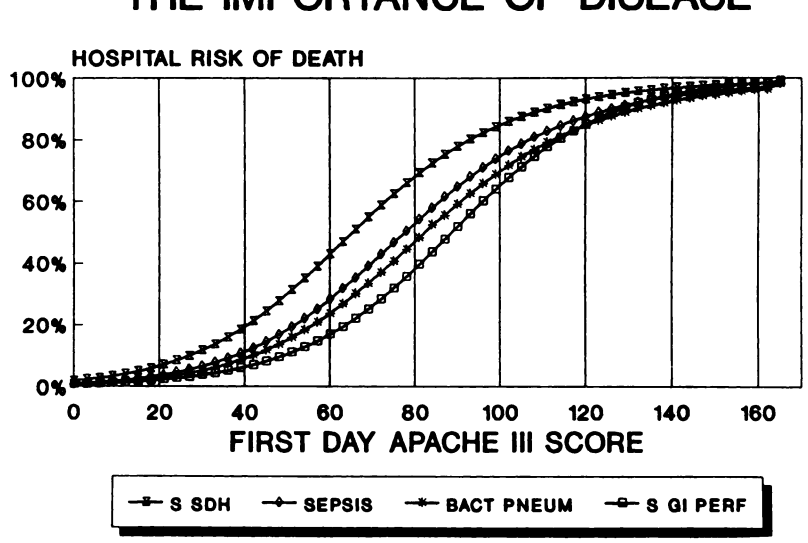


FIGURE 6. Relationship between APACHE III score and predicted risk of hospital death for patients with postoperative subdural hematomas (S SDH), sepsis (other than urinary tract), bacterial pneumonia (BACT PNEUM), and postoperative gastrointestinal perforation (S GI PERF).

comorbidities and functional status limitations to short-term outcome, we discovered that those comorbid conditions that influence the patient's immunologic status were the only ones that met our statistical requirements for inclusion. This is not surprising, since infection is commonly associated with both hospital and ICU mortality.[26] These comorbid conditions are also influential in increasing explanatory power for patients classified as undergoing emergency surgery.

Finally, we added a new variable to the APACHE III equation, the location of the patient prior to intensive care treatment; this variable is aimed at capturing the impact of selection bias on outcome prediction.[24,25] Among nonoperative patients, those who are ICU readmissions, transfers from other units, and admissions from the hospital wards have marginally increased risks of death relative to patients admitted directly to the ICU from the emergency room (Appendix A).

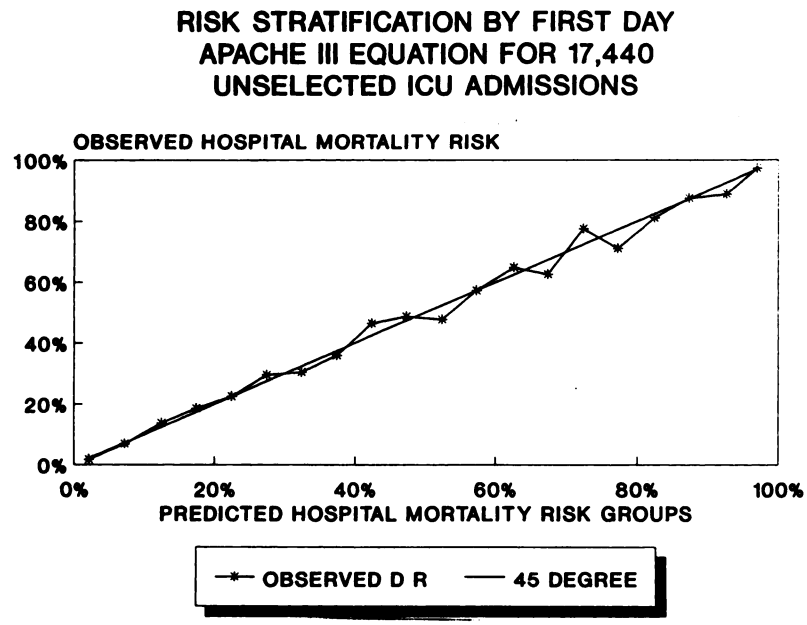
The APACHE III system consists of two options: (1) an APACHE III Score and (2) a series of predictive equations linked to a reference patient data base. The APACHE III score may be used alone within a single disease category or any other independently defined patient group to perform relative risk stratification, as illustrated in Figure 5. Predictive equations link the APACHE III score to our reference data base by the use of separate variables for patient location and disease classification and can produce risk estimates of hospital death for individual ICU patients at various times during their ICU stay (Fig 7 and 8). In both applications, the use of the score with precise measurement of other patient characteristics reduces the amount of unexplained variation in hospital death rates due to previously unmeasured patient characteristics (see Appendix B for examples of application).

Clinical Research Applications

A fundamental challenge for clinical trials involving acutely ill patients is that the treatment and control groups should be at an equivalent baseline risk. Even when patients are randomly assigned, prior risk estimation may detect risk differences that occurred despite randomization. The use of APACHE III also provides a continuous pretreatment risk measure. This is an improvement over categorical risk categories and can reduce the number of patients or observations needed to attain statistical significance between arms of a randomized trial.[27] Within each of 78 diseases listed in Table 3, an increased APACHE III score was statistically significantly associated with an increased risk of death. Therefore, in all 78 of these disease categories, the statistical power and precision of trials of experimental therapies where short-term death is the end point could be improved by using a severity measure such as APACHE III.[28]

The APACHE III score can also be directly calculated and used within disease categories or independently defined patient groups to provide severity stratification either before or after randomization. This application is similar to the recent use of the APACHE II score to assess risk in studies of severely ill hospitalized patients with infection.[28,29] Since patients in such studies may be selected on the basis of criteria different from those of the APACHE III study, the risk levels associated with specific APACHE III scores in this study may not calibrate precisely with those for patients meeting different selection criteria.[18]

The same principle applies to disease classification. For example, the risks displayed in Figure 5, bottom, for trauma patients are those for the four trauma subcategories listed in Table 3 (*ie*, head and multiple trauma, both nonoperative and postoperative). Each



CLASSIFICATION TABLES

A. AT .10 PREDICTED RISK OF HOSPITAL DEATH

	PREDICTED ALIVE	PREDICTED DEAD	TOTAL
TRUE ALIVE	10,283	4,163	14,416
DEAD	330	2,694	3,024
TOTAL	10,583	6,857	17,440

% CORRECT CLASSIFICATION	74.2%
SENSITIVITY	89.1
SPECIFICITY	71.1
PREDICTIVE VALUE POSITIVE	39.3
PREDICITIVE VALUE NEGATIVE	96.9

B. AT .50 PREDICTED RISK OF HOSPITAL DEATH

	PREDICTED ALIVE	PREDICTED DEAD	TOTAL
TRUE ALIVE	13,843	573	14,416
DEAD	1,499	1,525	3,024
TOTAL	15,342	2,098	17,440

% CORRECT CLASSIFICATION	88.1%
SENSITIVITY	50.4
SPECIFICITY	96.3
PREDICTIVE VALUE POSITIVE	72.7
PREDICITIVE VALUE NEGATIVE	90.2

C. AT .90 PREDICTED RISK OF HOSPITAL DEATH

	PREDICTED ALIVE	PREDICTED DEAD	TOTAL
TRUE ALIVE	14,384	32	14,416
DEAD	2,632	392	3,024
TOTAL	17,016	424	17,440

% CORRECT CLASSIFICATION	84.7%
SENSITIVITY	13.0
SPECIFICITY	99.8
PREDICTIVE VALUE POSITIVE	92.5
PREDICITIVE VALUE NEGATIVE	84.5

FIGURE 7. *Top,* Risk stratification by first-day APACHE III equation and assessment of predictive accuracy for 17,440 unselected ICU admissions. *Bottom,* Classification accuracy at 0.10, 0.50, and 0.90 predicted risk of hospital death.

of these subcategories has a different calibration level between the APACHE III score and outcome. These variations in baseline risk are reflected in the unadjusted hospital death rates in Table 3. These variations indicate that it is essential to specify precisely the patient selection and disease classification, defined as the primary reason for ICU admission, with the results of APACHE III scoring.

Applications in Evaluating ICU Outcome

To evaluate outcome for a multidiagnostic group of ICU patients, the APACHE III score must be used in combination with an APACHE III disease classification and patient location weighting. When applied to the 42 individual ICUs in this study, in which the observed hospital death rates varied from 6 percent to 42 percent, the APACHE III equation accounted for a substantial proportion of the variation in death rates ($r^2 = 0.90$). This confirms the findings in previous studies suggesting that the majority of variation in observed death rates across ICUs is related to variations in patient characteristics.[30]

Intensive care units could compare their mortality experience with this reference data base by using a patient-by-patient measurement of risk in order to compare the predicted mortality rate with the actual mortality rate. The difference between predicted and actual death rates is one measure of quality of care. This technique has proved useful in a variety of studies comparing the mortality experience of ICUs and investigating the incremental impact of specific treatment and of structural, process, or organizational changes on patient outcome.[31-33] Because bed availa-

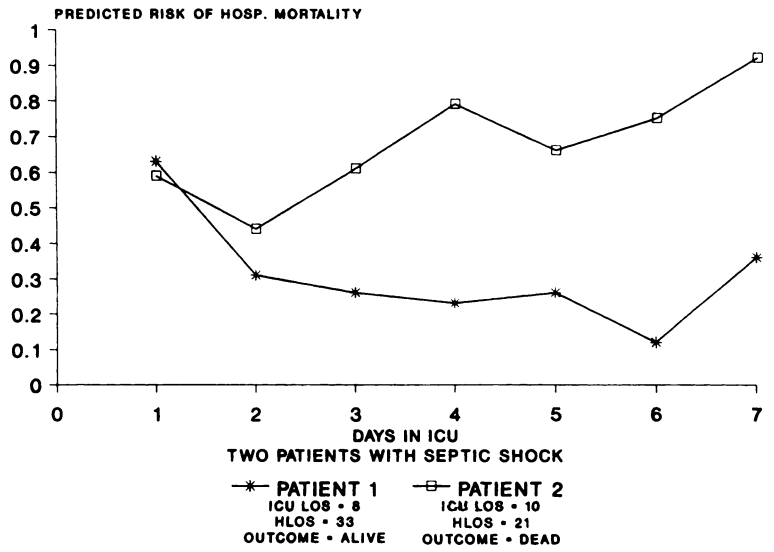


FIGURE 8. Daily risk estimates for two septic shock patients from the ICU with the highest average APACHE III score for all admissions. LOS = length of stay, expressed in days; HLOS = hospital LOS.

bility and screening for admission may vary substantially across different ICUs, comparisons among units in different hospitals may have to account for these variations in patient selection and discharge practices.[24,25]

Potential Applications in Clinical Decision Making

Prognostic efforts will never be able to predict outcome with 100 percent specificity. All clinical decision making, however, uses past experience to guide future decisions. Prognostic systems that are reliable (*eg*, provide identical estimates for an individual patient independent of observer) and well calibrated (*eg*, accurate throughout the range of risk) can ensure that the experiences of the past are taken into consideration in an unbiased manner.[2,6]

The overall explanatory power of the APACHE III system on the initial day of ICU treatment ($r^2 = 0.41$ and ROC = 0.90) compares well to that of previous versions of APACHE[15,20] and to that of other prognostic systems. Overall correct classification using a 0.50 decision rule on the initial day of ICU treatment was 79.1 percent with the Mortality Prediction Model,[34] compared with 88.1 percent with APACHE III (Fig 7, bottom). Because we established weights for the age, chronic health, and disease components of APACHE III using our entire patient data base, some degradation in predictive performance may occur with prospective application. We anticipate degradation will be minimal, however, because most of the explanatory power resides in the derived physiologic weights, which forecast well to independent data (Fig 4).

APACHE III outcome predictions (ROC = 0.90) compare favorably with those of physician judgment. Recent reports analyzing the accuracy of physicians' estimates of hospital death have recorded overall explanatory power measured in terms of ROC areas from 0.84 to 0.89.[35-37] A recent analysis that combined judgments from a variety of clinicians recorded an ROC of 0.85.[38] In a direct comparison of the discriminatory ability of APACHE II with a combination of clinicians' estimates, physicians correctly placed a slightly greater number of patients above the 90 percent risk level. The APACHE II predictions were more discriminating in identifying patients whose risk of death on ICU admission was 10 percent or less.[37]

Objective prognostic estimates derived from APACHE III have at least three potential advantages compared with clinical judgment. First, they should be more reliable than individual estimates because they are based on reproducible data. Second, the data base supporting the risk estimate is substantially larger than any one clinician's experience, thereby providing additional credibility to the prediction. Third, the risk estimates are based solely on the patient's response to treatment, not the order in which he or she presents for care or other commonly used heuristic variables.[39]

Evaluations of the potential usefulness of objective prognostic estimates are most likely to begin with a patient's response to therapy in the form of risk estimates over time (Fig 8). In many cases, these daily risk assessments will confirm uncertainty regarding the patient's ultimate ability to benefit from treatment. The increased reliability and confidence inherent in objective estimates might reduce the potential for error.[6] In a few cases, risk estimates could support clinical judgment that continuing current therapeutic efforts would be futile.[8] A recent pilot study by Knaus et al[40] suggested that probability estimates similar to those available with APACHE III might assist in such decision making. If our technical ability to provide intensive care expands while our financial capabilities

become more restricted, the capability of evaluating competing patients' requirements or their abilities to benefit from intensive care could become more important.[39,41] We emphasize, however, that with the exception of the study by Knaus et al,[40] no formal evaluation of the usefulness of such estimates has yet been completed. The incremental value of objective risk estimates, therefore, remains unknown.

Limitations

Investigators and clinicians using APACHE III must consider the current limitations of the system and its proper application. There were at least three misunderstandings of APACHE II. First, was the use of the score without consideration for disease. The APACHE III score can be used alone only within homogeneous disease categories and then for severity stratification, not risk prediction. The second misunderstanding was the use of a predictive equation calibrated by selection for ICU treatment on a patient sample selected by different criteria.[16,17] We now emphasize that the first-day APACHE III equation is calibrated on all patients selected for ICU admission, not any additional selection criteria, such as the need for mechanical ventilation.[16] The third misunderstanding was the use of a first-day equation at other times during the ICU stay.[18] To address the need for predictions based on treatment response, we developed equations that use the patient's initial and updated APACHE III scores to calculate individual risk estimates over time (Fig 8). APACHE III equations after the initial 24-h period use updated physiologic data and assign somewhat different weights to disease, chronic health, and age components as their implications change over time.

We urge the potential user, however, to appreciate that for patients with rare conditions or with unusual presentations of common conditions, risk may not be accurately estimated by this or other analytic techniques. In determining the confidence one should have in any risk estimate, a review of the total number of patients within a specific disease classification is needed, along with careful scrutiny of selection and other potential confounding biases. Predictive estimates must always be placed in an appropriate clinical context: a risk estimate for a critically ill patient whose clinical status is rapidly changing is likely to change; and the availability of new, untried therapeutic options should be taken into account.

Future Studies

Empirically derived risk estimates have not been a traditional component of medical research.[41] Discovering how best to use these risk estimates to improve the quality of patient care, the precision of clinical research, and patient outcome evaluation will require further investigation. We will be reporting in future analyses the potential use of the APACHE methodology to evaluate the quality of ICU care and to predict outcomes other than hospital death, such as the patient's requirement for unique ICU therapies, or anticipated length of ICU and hospital stay. Since the constructs that make up the APACHE III system (disease, severity, age, and comorbidities) are also closely related to a patient's therapeutic requirements, the APACHE III methodology should be useful in forecasting nursing and other care requirements, as well as the patient's potential need for unique ICU treatment. The most efficient and accurate method for evaluating patients' physiologic responses to therapy by collecting physiologic data over time also needs further investigation. The use of the initial and latest measurement is only one of a variety of analytic approaches that could be used.[12,14,42]

Comparisons of objective risk estimates to clinical judgment are needed both to determine the most appropriate role for these estimates in clinical decisions and to provide a guide to future improvements.[43] Physicians now infrequently use formal probabilistic reasoning or quantitative data to guide decision making.[44] Formal evaluations of the incremental impact, acceptance, and value of objective risk estimates are needed. Especially important to explore is whether feedback of empiric probabilities to clinicians would result in an improvement in the accuracy of their subjective probability estimates or their decision making. Fortunately, one such large-scale study is under way.[45]

Finally, the need to place these empirically derived prognostic estimates into a larger decision-making framework, one that explicitly acknowledges the fundamental roles of patient's preferences and values in clinical decision making, is essential.[8]

Conclusion

Every day, clinicians and physicians engaged in clinical research make complex decisions regarding the scope and intensity of treatment or the potential value of new therapies that might be supported or enhanced by an accurate and objective measurement of patient risk. Indeed, many of the most important questions concerning the quality and appropriateness of advanced medical care cannot be fully addressed until patient risk is accurately assessed and reliably recorded. The completion of the APACHE III prognostic system is an attempt to provide objective probability estimates for critically ill hospitalized patients treated in ICUs.

Financial Disclosure

All the authors certify that affiliations with or involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in this article are disclosed as follows: Drs Knaus, Wagner, and Zimmerman and Ms Draper are each

founders and minority equity shareholders of APACHE Medical Systems, Inc (AMS), a for-profit Delaware-based corporation that funded, in part, the research for the APACHE III study. AMS markets a software-based clinical information system based upon some of the concepts described in this article. Neither Dr Knaus, Dr Wagner, nor Dr Zimmerman, as full-time employees of George Washington University, is permitted to receive any direct financial payment from AMS and cannot participate in any business or policy decisions. Ms Draper is a full-time employee of AMS and sits on the Board of Directors. Dr Sirio has performed consultative services for AMS and has received *per diem* payments. None of the other authors has any formal affiliation with or receives any financial or other consideration from AMS.

ACKNOWLEDGMENTS: We would like to acknowledge the institutions that participated in the APACHE III data collection: St Mary's Hospital, East St Louis; White Memorial Hospital, Los Angeles; United Hospital, Clarksboro, WV; St Lukes-Roosevelt, New York; Cooper Medical Center, Camden, NJ; Monongahela Valley Hospital, Monongahela, Pa; Easton Hospital, Easton, Md; Burlington Medical Center, Burlington, Ia; Union Hospital, Union, NJ; St Lukes Hospital, Newburgh, NY; Wyandotte Hospital/ Medical Center, Wyandotte, Mich; Daniel Freeman Medical Center, Inglewood, Calif; Terre Haute Regional Medical Center, Terre Haute, Ind; Flower Memorial Hospital, Sylvania, Ohio; Richland Memorial Hospital, Columbia, SC; Mount Auburn Hospital, Cambridge, Mass; University Hospital, Denver, Colo; Northeast Georgia Medical Center, Gainesville, Ga; Bayfront Medical Center, St Petersburg, Fla; Craven Regional Medical Center, New Bern, NC; Southwest Florida Regional Medical Center, Fort Meyers, Fla; Mercy San Juan Hospital, Carmichael, Calif; Kaiser Foundation Hospital, Los Angeles; Mesa Lutheran Hospital, Mesa, Ariz; St Lukes Hospital, Kansas City, Mo; Presbyterian Intercommunity Hospital, Whittier, Calif; Stanford University Medical Center, Stanford, Calif; Winchester Medical Center, Winchester, Va; Hennepin County Medical Center, Minneapolis; William Beaumont Medical Center, Royal Oak, Mich; C.V.P.H. Medical Center, Plattsburgh, NY; Mayo Clinic, Rochester, Minn; Cleveland Clinic, Cleveland; North Carolina Baptist Hospital, Winston-Salem, NC; Barnes Hospital, St Louis; Catherine McCauley Health Center, Ann Arbor, Mich; Lexington Medical Center, West Columbia, SC; Fairfax Hospital, Falls Church, Va; George Washington University Medical Center, Washington, DC.

Appendix A: Technical Issues

Timing of Measurement and Weighting of Physiologic Variables

In the process of investigating the optimal approach to selecting and weighting physiologic coefficients for APACHE III, there were a number of specific technical issues considered. The first was whether the patient's initial or admission value was preferred over our previous practice of using the worst physiologic value over 24 h. For this reason, we collected data on both.[15] The first value obtained was the *admission value* (*ie*, the first value in the initial hour of ICU treatment). If no first-hour measurement was available, the data collector looked for values obtained 1 h prior to admission before recording it as missing. The next value obtained was the *worst over initial 23 h value* (*ie*, the most abnormal reading during the remaining initial 23 h of ICU treatment). We designed strict rules to define the most deranged value for the rare instances when there were abnormalities on both sides of a defined normal range. If no data were available for a particular physiologic variable either at admission or during the initial 23-h time period, we recorded the value as missing. Following data collection, we derived a *worse over 24 h value* for each physiologic measure using either the admission or worst over initial 23 h determinations (whichever was more abnormal). Finally, *worst over 24 h values* were obtained for each subsequent 24 h of ICU stay up to seven days.

When mean overall APACHE II scores between data abstractors were compared, no consistent pattern was found in which one abstractor scored higher or lower than another. For each of the three readings (admission, worst over 23 h, worst over 24 h), the absolute difference between the mean scores was not statistically significantly different from zero (n = 196).

Overall, data availability (the proportion of missing values) favored the worst value over the initial 24 h, as did maximum explanatory power.

The second consideration was how to determine the exact weight provided to each physiologic measurement. The physiologic variables were divided into clinically appropriate ranges based partly on cell size and partly on clinical judgment. They were then incorporated into the analysis as a series of separate predictor variables for each range.

The initial results from these analyses were compared with basic clinical and physiologic relationships. Where discrepancies existed (*eg*, a mean blood pressure of 60 mm Hg, assigned a lower coefficient, indicating a lower risk of death, than a mean pressure of 70 mm Hg), we adjusted the ranges. Most of these variations were due to small sample sizes in the originally designated ranges. In a few cases where the results of the analyses remained incompatible with established physiologic patterns, we adjusted the estimated weights by using clinical judgment. Patterns of weights were also checked using restricted cubic splines function.[21]

Cubic splines analysis is a statistical smoothing technique that allows assignment of a continuous varying weight to a physiologic variable. Because there are no discrete threshold values for all physiologic measurements, the use of continuous weighting is attractive compared to using specific cut points, although it would have eliminated the ability to hand-score APACHE III. In this data base, however, the use of cubic splines did not substantially increase total explanatory power.

Our final method of selecting weights for the physiologic variables was still mainly empiric (*ie*, deriving weights from a random half of the data base and validating it in the validation half). As emphasized above, however, when these empiric weights conflicted with known physiologic relationships, they were adjusted.

Disease-Specific Weighting of Physiologic Variables

Next, we explored the possibility that the derived weights for physiologic variables would be substantially different if examined within a specific disease category, as has been suggested by others.[22,23] We examined the weights assigned to specific target physiologic variables in congestive heart failure, the one well-defined disease category with a large number of patients (891) and a substantial death rate (21 percent). We first estimated weights for the target variable (blood pressure), using only patients within the specific disease category, and then compared these weights to ones obtained by using all patients in the estimation data file *except* those in the specific disease categories.

In this analysis, overall explanatory power was not improved by providing disease-specific weighting for individual physiologic variables. It is possible, however, that for some

diseases, use of specific combinations of physiologic variables may enhance predictive ability. It is important to emphasize that when the outcome is dichotomous (*eg*, alive or dead), one must have a very large number of patients within homogeneous disease categories to detect and statistically validate significant advantages in various approaches to severity weighting.

Missing Physiologic Values

We examined our practice of assigning a weight of zero to missing physiologic variables. We did this by examining the pattern of missing values and by using dummy variables to estimate the most appropriate weight to impute to a missing physiologic value. Variations in laboratory-test ordering practices across patients and the various hospitals meant that data availability varied for the physiologic variables. Ninety-nine percent of patients had complete information on all four vital signs (heart rate, respiratory rate, blood pressure, and urine output) during their initial 24 h in the ICU. Serum sodium, serum potassium, and hematocrit values were available for 85 percent and arterial blood gas measurements for 65 percent of patients during the initial 24 h. Analysis indicated that the proportion of missing values was directly related to physiologic stability as determined by vital sign data. Patients with normal or near normal vital signs had the largest proportion of missing laboratory tests, as determined by the mean number of missing physiologic variables per patient. Based on estimation fill-in values with dummy variables, missing physiologic variables were assumed normal and assigned zero weights.

Weighting of Comorbid Conditions

We also based the initial weighting of the 34 candidate variables measuring physiologic reserve and comorbidity on a regression analysis on the estimation data file. The magnitude and direction of the influence of each comorbid chronic disease variable on mortality (a coefficient of 0.05) and the overall statistical significance of the influence (T ratio >2) were considered in determining which variables to include in APACHE III. We also explored potential interactions among chronic disease variables and confounding of acute physiology measurements by chronic conditions (*eg*, a high serum creatinine level and low urine output by chronic conditions (*eg*, a high serum creatinine level and low urine output by chronic hemodialysis or peritoneal dialysis).

Relative Importance of Components of APACHE III Score and Equations

To illustrate the relative importance of the different components of APACHE III and additional information in explaining interpatient differences in risk of death, we reestimated multivariate logistic regression equations, leaving out different groups of variables that measure global concepts. In the aggregate, the complete equation has a global χ^2 of 6,426 ($p<0.00001$). The portion of the global χ^2 that is uniquely associated with specific variables is a measure of the unique importance of that factor in explaining risk of death. When the acute physiology score is deleted from the equation, the total χ^2 drops to 3,396, indicating that 47.2 percent of total interpatient explanatory power is uniquely captured by the acute physiology score. The other factors account for the following portions of unique explanatory power, with the remainder (100 percent − the sum of the parts) allocated jointly to all of the explanatory factors together: chronic health, 2 percent; age, 3 percent; disease, 6 percent; patient origin, 1 percent.

When an equation that uses the acute physiology score alone is estimated, the χ^2 is 5,501. This indicates that the upper limit on the relative importance of acute physiologic abnormalities is 86 percent of the total χ^2.

Timing of ICU Admission

In our review of individual patient trends (Fig 8), occasional discrepancies between initial risk estimates and final patient outcome were often due to ICU admission and initial APACHE assessment late in the course or treatment of an illness. The use of a patient location variable (*eg*, emergency room versus ICU readmission) in APACHE III reduces the impact of such differences. There will be occasional patients, however, for whom even this control for lead-time bias may be inadequate. For example, a patient is admitted to a coronary care unit with an acute myocardial infarction and acute mitral value insufficiency due to a ruptured papillary muscle. After three days of medical therapy for shock and congestive heart failure, an emergency valve replacement is followed by admission to a surgical ICU. Using the APACHE III system, this patient is classified as an emergency postoperative valve replacement. The first-day risk of death in the surgical ICU might not reflect the risk implied by the prior acute myocardial infarction and shock in the coronary care unit. Because the number of such patients is very small, it is not possible to estimate the incremental risk associated with these variations in the onset of intensive therapy. Improved estimation of these risks will have to await further research and the collection of larger reference data bases. The time delay between emergency room and ICU admission did not increase overall explanatory power and is not included in the final system.

Differences in patient selection for ICU therapy, which are not accounted for by APACHE III, also represent an important limitation in explaining patient outcome variations at the hospital and ICU level. Future interinstitutional comparisons, therefore, may have to account for these variations by including selected institutional characteristics as control variables or by limiting comparisons to hospitals with similar triage pressures. Further details on these and other specific analytic issues are available from the authors and will be the subject of subsequent publications.

Appendix B: Guide to APACHE III Risk Prediction

The procedure for estimating the hospital mortality risk of adult medical and surgical patients admitted to an ICU on their initial day of ICU treatment entails the following steps:

Step 1

Choose the single most important reason for ICU treatment from the listing of nonoperative and operative major disease categories (Table 3). The disease category may *not* be identical to the primary hospital diagnosis (*eg*, for a

patient with acute leukemia admitted for respiratory failure due to aspiration pneumonia, the major disease category for APACHE III scoring is aspiration pneumonia; the acute leukemia qualifies as a comorbidity [hematologic malignancy] in the APACHE III score). Always place the patient in the most specific diagnostic category possible (*eg*, a postoperative patient with a gastrointestinal (GI) malignancy who is admitted immediately following an abdominal exploration that revealed an obstructed colon should be classified as GI obstruction, rather than GI cancer). If the patient does not qualify for any of the specific categories, use the most appropriate general category for the primary organ system affected. All patients admitted to the ICU from the recovery room or operating room are considered postoperative, and you must choose from one of the postoperative diagnoses. All other patients are placed into one of the nonoperative diagnostic categories.

Step 2

If the patient is a nonoperative admission, indicate where the patient was being treated immediately prior to ICU admission. Treatment locations include emergency room, hospital floor, other hopsital, other ICU, and readmission to the same ICU. If the patient was in a regular hospital room, the correct variable is "floor"; if they were in a specialized ICU, it is "other ICU." If the patient was previously treated in the same ICU during this hospitalization, the designation is "ICU readmission."

If the patient was admitted to the ICU immediately following surgery, was the surgery performed on an emergent basis? Emergency surgery is defined as surgery required immediately to correct a life-threatening condition.

Step 3

Calculate the patient's APACHE III score by recording and summing points for the 17 potential physiologic measurements, age, and the chronic health evaluation. If multiple comorbid conditions are present, score only the one condition with the highest risk points (scoring is not performed with elective postoperative patients). Points for the 17 potential physiologic measurements reflect the worst (most abnormal) value during the initial 24 h of ICU treatment only. Daily updates of physiologic measurements also represent the most abnormal value within subsequent 24-h periods.

If a physiologic measurement is not obtained during this initial 24-h period, no risk points are assigned. The most abnormal arterial blood gas measurement is the one associated with the widest $P(A\text{-}a)O_2$ or the lowest PaO_2. If a patient is heavily sedated and/or paralyzed, so that his neurologic status cannot be evaluated, and no reliable evaluation prior to sedation is available, the neurologic status should be recorded as normal.

Step 4

Take the coefficients for the patient's major disease category and treatment location prior to ICU admission, together with the total APACHE III score, and use them as part of the first-day APACHE III risk equation to calculate a predicted risk of hospital mortality. Copies of regression coefficients and detailed definitions for the 78 diagnostic categories and other components of the APACHE III equation are available for research and independent confirmation from the authors.

Example 1: Nonoperative Admission

A 56-year-old woman with acute leukemia is admitted to the ICU from her hospital room following an episode of aspiration pneumonia. She had not been treated in the ICU during this hospitalization.

Major Disease Category: Aspiration pneumonia (nonoperative)

Treatment Location: Hospital room

APACHE III Scoring:

Age (56 years)	= 5 points
Chronic health (leukemia)	= 10 points
Acute physiologic abnormalities (most abnormal within initial 24 h):	
Pulse rate (125 beats/min)	= 7 points
Mean blood pressure (75 mm Hg)	= 6 points
Temperature (39.8°C)	= 0 points
Respiratory Rate (36/min)	= 9 points
$PaO_2/P(A\text{-}a)O_2$ ($PaO_2 = 68$ mm Hg; $FIO_2 = 0.70$, mechanically ventilated; therefore, calculate $P(A\text{-}a)O_2$: $PCO_2 = 26$ mm Hg; $P(A\text{-}a)O_2 = 433$)	= 11 points
Hematocrit (24%)	= 3 points
White blood cell count (1,200/cu mm)	= 5 points
Creatinine (2.2 mg/dl)	= 7 points
Urine output (1,200 ml/h)	= 5 points
Blood urea nitrogen (85 mg/dl)	= 12 points
Sodium (136 mmol/L)	= 0 points
Albumin (2.4 g/dl)	= 6 points
Bilirubin (3.3 mg/dl)	= 6 points
Glucose (246 mg/dl)	= 3 points
Acid-base (pH = 7.24; $PCO_2 = 26$ mm Hg)	= 9 points
Neurologic (opens eyes, confused/converses, obeys verbal command)	= 3 points
Total	= 107 points

APACHE III first-day hospital risk equation: Aspiration pneumonia (−4.5575) + Hospital floor admission (+.2744) + APACHE III Score (107 × .0537 = 5.7459) = +1.4628 log odds of death. The natural antilogarithm of +1.4628 = 4.318 = r/(1-r). Solving for r = .8120, risk of hospital mortality = 81.2 percent.

Example 2: Operative Admission

A 79-year-old man is admitted to the ICU from the recovery room following an exploratory laparotomy that revealed an obstructed colon secondary to colon cancer.

Major Disease Category: GI obstruction (postoperative)

Treatment Location: Recovery room (surgery was performed on emergency basis)

APACHE III Scoring:

Age (79 years)	= 17 points
Chronic health (colon cancer, not metastatic)	= 0 points
Acute physiologic abnormalities	

Pulse rate (110 beats/min)	= 5 points
Mean blood pressure (82 mm Hg)	= 0 points
Temperature (37.5°C)	= 0 points
Respiratory rate (10/min ventilated)	= 0 points
$PaO_2/P(A\text{-}a)O_2$ ($PaO_2 = 95$ mm Hg; $FIO_2 = 0.40$, mechanically ventilated; since $FIO_2 < 0.5$, use PaO_2.	= 0 points
Hematocrit (32%)	= 3 points
White blood cell count (12,000/cu mm)	= 0 points
Creatinine (1.8 mg/dl)	= 4 points
Urine output (1,800 ml/h)	= 4 points
Blood Urea Nitrogen (22 mg/dl)	= 7 points
Sodium (142 mmol/L)	= 0 points
Albumin (2.8 g/dl)	= 0 points
Bilirubin (2.5 mg/dl)	= 5 points
Glucose (190 mg/dl)	= 0 points
Acid-base (pH 7.42; $PCO_2 = 38$ mm Hg)	= 0 points
Neurologic (intubated but opens eyes, and obeys commands)	= 0 points
Total	= 45 points

APACHE III first-day hospital risk equation: GI postoperative obstruction (−4.6974) + Emergency surgery (+.0752) + APACHE III Score (45 × .0537 = +2.4167) = −2.2051 = log odds of death. The natural antilogarithm of −2.2051 = 0.1102 = r/(1-r). Solving for r = .0992. Risk of hospital mortality = 9.92 percent.

Comment

Predictions after the initial day contain an additional variable that updates the patient's APACHE III score for changes in physiologic status, and the relative contributions of age and comorbidities are also slightly different on subsequent days. Further explication of the equations is beyond the scope of this article. Research teams should contact the authors to discuss replication or collateral experimental approaches. Within accepted experimental and ethical protocols, the authors plan to make available additional materials as they are developed.

References

1 Feinstein AR. An additional basic science for clinical medicine: I. The constraining fundamental paradigms. Ann Intern Med 1983; 99:939-97

2 Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules: applications and methological standards. N Engl J Med 1985; 313:793-99

3 Silverstein MD. Prediction instruments and clinical judgement in critical care. JAMA 1988; 260:1758-59

4 Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgement. Science 1989; 245:168-74

5 Seneff M, Knaus WA. Predicting patient outcome from intensive care: a guide to APACHE, MPM, SAPS, PRISM, and other prognostic scoring systems. J Intensive Care Med 1990; 5:33-52

6 Kalb PE, Miller DH. Utilization strategies for intensive care units. JAMA 1989; 261:2389-95

7 Detsky AS, Stricker SC, Malley AG, Thibault GE. Prognosis, survival, and the expenditure of hospital resources for patients in an intensive care unit. N Engl J Med 1981; 305:667-72

8 Schneiderman LJ, Jecker NS, Jonsen AR. Medical futility: its meaning and ethical implications. Ann Intern Med 1990; 112:948-54

9 NIH Consensus Development Conference on Critical Care Medicine. Crit Care Med 1983; 14:466-72

10 Zimmerman JE, Knaus WA, Sharpe MD, Anderson AS, Draper EA, Wagner DP. The use and implications of do not resuscitate orders in intensive care units. JAMA 1986; 255:351-56

11 Chang RWS. Individual outcome prediction models for intensive care units. Lancet 1989; 2:143-46

12 Shoemaker WC, Appel PL, Waxman K. Clinical trial of survivors' cardiorespiratory patterns as therapeutic goals in critically ill post-operative patients. Crit Care Med 1982; 10:398-406

13 Forrester JS, Dianost G, Chatterjee K, Swan HJC. Medical therapy of acute myocardial infarction by application of hemodynamic subsets: part I. N Engl J Med 1976; 295:1356-62

14 Levy DE, Caronna JJ, Singer BH, Lapinski RH, Frydman H, Plum F. Predicting outcome from hypoxic-ischemic coma. JAMA 1985; 253:1420-24

15 Zimmerman JE, ed. The APACHE III study design: analytic plan for evaluation of severity and outcome. Crit Care Med 1989; 17(suppl):S169-221

16 Fedullo AJ, Swinburne AJ, Wah GW, Bisby KR. APACHE II score and mortality in respiratory failure due to cardiogenic pulmonary edema. Crit Care Med 1988; 16:1218-21

17 Hopefl AW, Taaffe CL, Herrmann VM. Failure of APACHE II alone as a predictor of mortality in patients receiving total parenteral nutrition. Crit Care Med 1989; 17:414-17

18 Cerra FB, Negro F, Abrams J. APACHE II score does not predict multiple organ failure or mortality in postoperative surgical patients. Arch Surg 1990; 125:519-22

19 Damiano AM, Bergner M, Draper EA, Knaus WA, Wagner DP. Reliability of a measure of severity of illness: acute physiology and chronic health evaluation: II. J Clin Epidemiol (in press)

20 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease classification system. Crit Care Med 1985; 13:818-828

21 Durrleman S, Simon R. Flexible regression models with cubic splines. Stat Med 1989; 8:551-61

22 Daley J, Jencks S, Draper D, Lenhart G, Thomas N, Walker J. Predicting hospital-associated mortality for Medicare patients: a method for patients with stroke, pneumonia, acute myocardial infarction, and congestive heart failure. JAMA 1988; 260:3617-24

23 Jencks S, Daley J, Draper D, Thomas N, Lenhart G, Walker J. Interpreting hospital mortality data: the role of clinical risk adjustment. JAMA 1988; 260:3611-24

24 Dragsted L, Jorgensen J, Jensen NH, Bansing E, Jacobsen S, Knaus WA, et al. Interhospital comparisons of patient outcome from intensive care: importance of lead-time bias. Crit Care Med 1989; 17:418-22

25 Escarce JJ, Kelley MA. Admission source to the medical intensive care unit predicts hospital death independent of APACHE II score. JAMA 1990; 264:2389-94

26 Bone RC, Fisher CJ, Clemmer TP, Slotman GJ, Metz CA, Balk RA, et al. Sepsis syndrome: a valid clinical entity. Crit Care Med 1989; 17:389-93

27 Knaus WA, Wagner DP. Multiple systems organ failure: epidemiology and prognosis. Crit Care Clin 1989; 5:221-32

28 Solomkin JS, Dellinger EP, Christou NU, Busuttil RW. Results of a multicenter trial comparing imipenem/cilastatin to tobramycin/clindamycin for intra-abdominal infections. Ann Surg 1990; 212:581-91

29 Ziegler E, Fisher C, Sprung C, Straube R, Sadoff J, Foulke GE, et al. Treatment of gram-negative bacteremia and septic shock with ha-ia human monoclonal antibody against endotoxin. N Engl J Med 1991; 394:429-36

30 Pollack MM, Ruttimann UE, Geston PR. Accurate prediction of outcome of pediatric intensive care: a new quantitative method. N Engl J Med 1987; 316:134-37

31 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An

evaluation of outcome from intensive care in major medical centers. Ann Intern Med 1986; 104:410-18
32 Knaus WA, LeGall TR, Wagner DP, Draper EA, Loirat P, Campos RA, et al. A comparison of intensive care in the U.S.A. and France. Lancet 1982; 2:642-44
33 Reynolds HN, Haupt MT, Thill-Baharizian MC, Carlson RW. Impact of critical care physician staffing on patients with septic shock in a university hospital medical intensive care unit. JAMA 1988; 260:3446-50
34 Lemeshow S, Teres D, Avrunin SJ, Gage RW. Refining intensive care unit outcome prediction by using changing probabilities of mortality. Crit Care Med 1988; 16:470-77
35 Kruse JA, Thill-Baharizian MC, Carlson RW. Comparison of clinical assessment with APACHE II for predicting mortality risk in patients admitted to a medical intensive care unit. JAMA 1988; 260:1734-42
36 Brannen AL, Godfrey LJ, Goetter WE. Prediction of outcome from critical illness: a comparison of clinical judgement with a prediction rule. Arch Intern Med 1989; 149:1083-86
37 McClish DK, Powell SH. How well can physicians estimate mortality in a medical intensive care unit? Med Decis Making 1989; 9:125-32
38 Poses RM, Bekes C, Winkler RL, Scott WE, Copare FJ. Are two (inexperienced) heads better than one (experienced) head? averaging house officers prognostic judgement for critically ill patients. Arch Intern Med 1990; 150:1874-78
39 Englehardt HT, Rie MA. Intensive care units, scarce resources, and conflicting principles of justice. JAMA 1986; 255:1159-64
40 Knaus WA, Rauss A, Alperovitch A, LeGall JR, Loirat P, Patois E, et al. Do objective estimates of chances for survival influence decisions to withhold or withdraw teatment? Med Decis Making 1990; 10:163-71
41 Knaus WA, Wagner DP, Lynn D. Short-term mortality estimates for critically ill hospitalized adults: science and ethics. Science (in press)
42 Chang RWS, Jacobs S, Lee B. Predicting outcome among intensive care unit patients using computerized trend analysis of daily APACHE II scores corrected for organ system failure. Intensive Care Med 1988; 14:558-66
43 Moskowitz AJ, Kuipers BT, Kassirer JP. Dealing with uncertainty, risks and trade-offs in clinical decisions: a cognitive science approach. Ann Intern Med 1988; 108:435-49
44 Detsky AS, Redelmeier D, Abrams HB. What's wrong with decision analysis? can the left brain influence the right? J Chronic Dis 1987; 40:813-16
45 Murphy DJ, Cluff LE, eds. SUPPORT: study to understand prognoses and preferences for outcomes and risks of treatments. J Clin Epidemiol 1990; 43(suppl):1S-123S