APACHE II: a severity of disease classification system

Article *in* Critical Care Medicine · November 1985

DOI: 10.1097/00003465-198603000-00013 · Source: PubMed

4 authors, including:

William A. Knaus
University of Virginia
334 PUBLICATIONS **69,816** CITATIONS

0090-3493/85/1309-0818$02.00/0
CRITICAL CARE MEDICINE
Copyright © 1985 by The Williams & Wilkins Co.

Vol. 13, No. 10
Printed in U.S.A.

APACHE II: A severity of disease classification system

WILLIAM A. KNAUS, MD; ELIZABETH A. DRAPER, MS; DOUGLAS P. WAGNER, PHD; JACK E. ZIMMERMAN, MD

This paper presents the form and validation results of APACHE II, a severity of disease classification system. APACHE II uses a point score based upon initial values of 12 routine physiologic measurements, age, and previous health status to provide a general measure of severity of disease. An increasing score (range 0 to 71) was closely correlated with the subsequent risk of hospital death for 5815 intensive care admissions from 13 hospitals. This relationship was also found for many common diseases.

When APACHE II scores are combined with an accurate description of disease, they can prognostically stratify acutely ill patients and assist investigators comparing the success of new or differing forms of therapy. This scoring index can be used to evaluate the use of hospital resources and compare the efficacy of intensive care in different hospitals or over time.

In the past three decades, American medicine, especially hospital medicine, has undergone dramatic changes.[1] A profession that previously emphasized diagnosis and close observation now places therapeutic concerns first. We now undertake complex multidisciplinary treatment of the advanced stages of disease and much of this care is concentrated in ICUs. The rapidity of this change and the large and still growing investment in these high cost services have prompted demands for better evidence of the indications and benefit of intensive care.[2] One of the best ways to do this is to have precise estimates of outcome. Yet, as recently emphasized by Feinstein,[3] the science of prediction, and specifically risk stratification, has not kept pace with our new therapeutic abilities.

This report presents the results of a nationwide effort to validate APACHE II, a severity of disease classification system that uses basic physiologic principles to stratify acutely ill patients prognostically by risk of death. We describe the development, rationale, and data collection requirements for APACHE II, the results supporting its validity, and the potential uses of APACHE II.

From the ICU Research Unit, Departments of Anesthesiology and Computer Medicine, The George Washington University Medical Center, Washington, DC.

This study was supported by NCHSR grant no. 04857 and The Robert Wood Johnson Foundation grant no. 8498.

Address requests for reprints to: William A. Knaus, MD, ICU Research Unit, The George Washington University Medical Center, 2300 K Street, NW, Washington, DC 20037.

PREVIOUS RESEARCH

The APACHE II classification system is a revised version of a prototype system, APACHE (acute physiology and chronic health evaluation).[4] The basis for APACHE's development was the hypothesis that the severity of acute disease can be measured by quantifying the degree of abnormality of multiple physiologic variables. We used this approach because we believe that one of intensive care's major functions is to detect and treat life-threatening acute physiologic derangements, and that a severity classification system must be based on objective physiologic measurements and be as independent of therapy as possible.[5] Finally, the index should be valid for a wide range of diagnoses, easy to use, and based upon data available in most hospitals.

We used a variation of the nominal group process to choose and weight physiologic variables. This process followed closely the suggestions of Gustafson et al.[6] in regard to proper construction of severity scales, and took advantage of the long-established principle of homeostasis. The original APACHE system provides weightings for 34 potential physiologic measures, the sum of which yields an acute physiology score (APS).

This weighting system is based on a scale of 0 to 4, as illustrated by the following weights assigned to variations in serum pH:

Weighted Score	pH Range
+4	<7.15
+3	7.15–7.24
+2	7.25–7.32
0	7.33–7.49
+1	7.50–7.59
+3	7.6–7.69
+4	7.7 or>

The APS is determined from the most deranged (worst) physiologic value, e.g., the lowest BP or the highest respiratory rate, during the initial 24 h after ICU admission. The 24-h time period ensures that all pertinent physiologic values are available, and clinical judgment ensures that each value is legitimate. Because

severe chronic disease significantly reduces the probability of survival during acute illness, the original APACHE system incorporates a four-letter (A, B, C, and D) designation corresponding to a spectrum ranging from excellent health (A) to severe chronic organ system insufficiency (D).[4]

APACHE is a reliable and useful means of classifying ICU patients. Increases in APS are associated with increased risk of subsequent hospital death.[4] APACHE has also proved useful in evaluating outcome from intensive care and in comparing the success of different treatment programs.[7,8] However, the original APACHE system is complex and needed formal multi-institutional validation.[9] The APACHE II system is the result of our efforts to simplify and present a more clinically useful yet statistically accurate and valid patient classification system.

DEVELOPMENT OF APACHE II

Using clinical judgment and documented physiologic relationships to choose variables and assign weights remains the essence of APACHE II. The number of physiologic measurements, however, has been reduced from 34 to 12. Infrequently measured physiologic variables such as serum osmolarity, lactic acid level, and skin testing for anergy were deleted, as were potentially redundant variables. Thus, serum BUN was replaced by the more specific serum creatinine value and serum pH was retained in preference to bicarbonate.

Subsequent reductions were accomplished by establishing a minimum set of clinically essential variables and then carefully evaluating the role of additional physiologic measurements with regard to their impact on survival. We deleted each measurement based upon clinical judgment, and then evaluated that decision using a multivariate comparison of the original APACHE system with each proposed revision. The total R^2 and correct classification rate for hospital mortality were used as standards. The smallest number of variables that reflected physiologic derangement for all vital organ systems as well as maintained statistical precision was 12 (Fig. 1).

During this reduction process, we learned that many variables crucial in patient care, such as serum glucose, albumin, CVP and urinary output resulted in little increase in explanatory power. This could be because some of these variables, such as CVP measurement, are more sensitive to variations in therapeutic decisions than severity of disease. Also, most patients with abnormal serum albumin values, for example, frequently have other abnormal values captured within the basic 12 measurements.

Some of the thresholds and weights for the physiologic variables have been changed. Analysis of previously collected data as well as research by others suggested that the Glasgow coma score, our only measure of neurologic function, should be more heavily weighted relative to the other measures.[10] Because loss of renal function is also a very poor prognostic sign, we increased the threshold at which serum creatinine contributes to the total score, and for all cases of acute renal failure, doubled its weighting.[11] Finally, because the equation for computing the alveolar-arterial O_2 gradient ($P[A\text{-}a]O_2$) is heavily dependent on inspired O_2 (FIO_2) levels, we developed a direct weighting for all PaO_2 values when the FIO_2 is less than 0.50.

The weights for the nine remaining physiologic variables used in APACHE II are the same as in the original APACHE system. The recorded value is still based on the most deranged reading during each patient's initial 24 h in an ICU. Unlike APACHE, however, measurement of all 12 physiologic values is mandatory when using APACHE II. This eliminates the problem of missing values and concerns about the assumption that an unmeasured variable was normal.[9,12] Although arterial blood gas measurements may be inappropriate for some patients, exclusion of these values is not encouraged and should only be done when clinical judgment strongly suggests the results would be within normal limits.

Because age and severe chronic health problems reflect diminished physiologic reserve, they have been directly incorporated into APACHE II. Chronologic age is a well-documented risk factor for death from acute illness, that is independent of the severity of disease.[13,14] The weights assigned to age in APACHE II are based on their relative impact within this validation.

We discovered that when we controlled for acute physiologic derangement and age, three of the four chronic health classifications (B,C, and D) were associated with higher death rates. However, only the most severe chronic organ system insufficiency or immunocompromised state (class D) markedly influenced outcome.[13] We also discovered that nonoperative and emergency surgery admissions had a substantially higher risk for death from their prior organ system insufficiency than elective surgical admissions. (Surgery or postoperative patients are those admitted to the ICU directly from the operating or recovery room. All others are nonoperative.) This was probably because patients with the most severe chronic conditions are not considered to be candidates for elective surgery. Therefore, nonoperative or emergency operative admissions with a severe chronic organ system dysfunction are given an additional five points, while similar elective surgical admissions are only given two points.

The maximum possible APACHE II score is 71. In experience to date, no patient has exceeded 55.

VALIDATION METHODS

We contend that the most specific standard for judging the validity of a severity of disease classification

THE APACHE II SEVERITY OF DISEASE CLASSIFICATION SYSTEM

PHYSIOLOGIC VARIABLE	HIGH ABNORMAL RANGE					LOW ABNORMAL RANGE			
	+4	+3	+2	+1	0	+1	+2	+3	+4
TEMPERATURE — rectal (°C)	≥41°	39°-40.9°		38.5°-38.9°	36°-38.4°	34°-35.9°	32°-33.9°	30°-31.9°	≤29.9°
MEAN ARTERIAL PRESSURE — mm Hg	≥160	130-159	110-129		70-109		50-69		≤49
HEART RATE (ventricular response)	≥180	140-179	110-139		70-109		55-69	40-54	≤39
RESPIRATORY RATE — (non-ventilated or ventilated)	≥50	35-49		25-34	12-24	10-11	6-9		≤5
OXYGENATION: A-aDO_2 or PaO_2 (mm Hg) a. FIO_2 ≥ 0.5 record A-aDO_2	≥500	350-499	200-349		<200				
b. FIO_2 < 0.5 record only PaO_2					PO_2 > 70	PO_2 61-70		PO_2 55-60	PO_2 < 55
ARTERIAL pH	≥7.7	7.6-7.69		7.5-7.59	7.33-7.49		7.25-7.32	7.15-7.24	<7.15
SERUM SODIUM (mMol/L)	≥180	160-179	155-159	150-154	130-149		120-129	111-119	≤110
SERUM POTASSIUM (mMol/L)	≥7	6-6.9		5.5-5.9	3.5-5.4	3-3.4	2.5-2.9		<2.5
SERUM CREATININE (mg/100 ml) (Double point score for **acute** renal failure)	≥3.5	2-3.4	1.5-1.9		0.6-1.4		<0.6		
HEMATOCRIT (%)	≥60		50-59.9	46-49.9	30-45.9		20-29.9		<20
WHITE BLOOD COUNT (total/mm3) (in 1,000s)	≥40		20-39.9	15-19.9	3-14.9		1-2.9		<1
GLASGOW COMA SCORE (GCS): Score = 15 minus actual GCS									
[A] Total ACUTE PHYSIOLOGY SCORE (APS): Sum of the 12 individual variable points									
Serum HCO_3 (venous-mMol/L) [Not preferred, use if no ABGs]	≥52	41-51.9		32-40.9	22-31.9		18-21.9	15-17.9	<15

[B] AGE POINTS: Assign points to age as follows:

AGE(yrs)	Points
≤44	0
45-54	2
55-64	3
65-74	5
≥75	6

[C] CHRONIC HEALTH POINTS

If the patient has a history of severe organ system insufficiency or is immuno-compromised assign points as follows:

a. for nonoperative or emergency postoperative patients — 5 points
or
b. for elective postoperative patients — 2 points

DEFINITIONS

Organ Insufficiency or immuno-compromised state must have been evident **prior** to this hospital admission and conform to the following criteria:

LIVER: Biopsy proven cirrhosis and documented portal hypertension; episodes of past upper GI bleeding attributed to portal hypertension; or prior episodes of hepatic failure/encephalopathy/coma.

CARDIOVASCULAR: New York Heart Association Class IV.

RESPIRATORY: Chronic restrictive, obstructive, or vascular disease resulting in severe exercise restriction, i.e., unable to climb stairs or perform household duties; or documented chronic hypoxia, hypercapnia, secondary polycythemia, severe pulmonary hypertension (>40mmHg), or respirator dependency.

RENAL: Receiving chronic dialysis.

IMMUNO-COMPROMISED: The patient has received therapy that suppresses resistance to infection, e.g., immuno-suppression, chemotherapy, radiation, long term or recent high dose steroids, or has a disease that is sufficiently advanced to suppress resistance to infection, e.g., leukemia, lymphoma, AIDS.

APACHE II SCORE

Sum of [A] + [B] + [C] :

[A] APS points _____

[B] Age points _____

[C] Chronic Health points _____

Total APACHE II _____

FIG. 1. The APACHE II severity of disease classification system.

system is hospital mortality. Hospital mortality can be accurately measured, and although not sensitive to the important question of quality of survival, mortality is an objective and reasonable starting point for evaluation. We therefore evaluated the validity of APACHE II by testing its association with hospital mortality in a large number of unselected but carefully described ICU admissions from 13 hospitals.

The 13 hospitals and the characteristics of the units surveyed for this evaluation are listed in alphabetical order in Table 1. With the exception of the George Washington University Medical Center (GWUMC), where data were collected for all ICU admissions during 1979 through 1981, the remaining 12 hospitals collected data on a minimum of 200 unselected consecutive ICU admissions during 1982. The hospitals were chosen because of their willingness to participate and, in most cases, to support supervised data collection independently. The number of patients per hospital reflects the amount of labor available for data collection—not the overall ICU utilization rate.

One of us (E.A.D.) visited each of the hospitals to initiate and train the data collectors. In most hospitals, these were experienced ICU nurses, but also included medical records personnel, medical corpsmen, and physicians. All data were checked for transcription errors, completeness, and internal consistency. Interobserver reliability testing by others (DL Jackson, personal communication) revealed 96% agreement for all physiologic data. Agreement on preadmission data was somewhat less, but disagreements were minor.

For each patient, the data collected included age, diagnosis, indication for ICU admission, surgical status, preadmission history, APACHE classifications on each ICU day, and outcome at ICU and hospital discharge. Copies of most patients' discharge face sheets (H-ICDA) were also obtained.

At ICU admission, the patients were assigned to specific diagnostic categories according to the *one* principal reason for their admission. Some of the most frequent and important of these diagnostic categories appear in Tables 2 and 3, and a full listing is available in the Appendix and Table 6. Patients without one of these principal diagnoses were designated as admitted

TABLE 1. Description of 13 hospitals participating in APACHE II validation

Hospital (State)	Total Number of Hospital Beds	Total Number of Adult ICU Beds	Number of Adult ICU Beds Studied	Type of ICU(s) Studied
Cooper Medical Center (NJ)	522	14	14	Mixed medical/surgical
George Washington University Medical Center (DC)	511	24	16	Mixed medical/surgical
Medical College of Georgia (GA)	706	21	6	Medical
Johns Hopkins University (MD)	1025	36	7	Medical
Maine Medical Center (ME)	533	32	20	Mixed medical/surgical
University of Maryland Hospital (MD)	729	31	10	Surgical
Massachusetts General Hospital (MA)	1092	90	20	Surgical (2 units)
Polyclinic Medical Center (PA)	556	14	6	Mixed medical/surgical
St. Francis Hospital (OK)	802	40	16	Mixed medical/surgical
South Shore Hospital (MA)	280	28	16	Surgical and mixed medical/surgical (2 units)
Stanford University Hospital (CA)	633	65	57	Surgical, medical, cardiac surgery (3 units)
University of Virginia Medical Center (VA)	683	44	16	Surgical
University of Wisconsin Hospital (WI)	548	36	32	Surgical, medical, mixed medical/surgical (3 units)

under one of five mutually exclusive categories of organ system failure or insufficiency: neurologic, cardiovascular, respiratory, gastrointestinal, or renal/metabolic. These five categories are not as precise as specific disease but do permit the system to be used to study entire ICU populations.

Table 2 summarizes demographic data as well as the most frequent diagnostic and organ system indication for admission at each of the 13 hospitals. This and all subsequent description of hospitals do not correspond to the alphabetical listing in Table 1, because utilization and outcome information from each hospital is confidential.

RESULTS

In this validation study, all 12 physiologic measurements were available for 87% of the 5815 ICU admissions. The most frequent exceptions were serum creatinine and arterial blood gas values (10% of all patients). Almost all patients with missing values were admitted to the ICU for monitoring, and the results of arterial blood gases and serum creatinine were not considered essential for patient care.

Figure 2 illustrates the distribution of APACHE II scores according to medical or surgical status. Patients were admitted at all levels of severity. Postoperative admissions were concentrated in the midrange of severity, reflecting more uniformity in their physiologic derangements immediately after surgery. Medical patients had more widely distributed scores because they were all emergency admissions with a greater variety of acute insults. The variation between nonoperative admissions and patients admitted to the ICU directly after surgery illustrates the importance of combining severity classification with precise clinical diagnosis and other relevant information.

Figure 3 illustrates the direct relationship between APACHE II scores and observed hospital death rates. For each five-point increase in APACHE II, there was a significant increase in death rate. For example, the 1.9% death rate for patients with 0 to 4 points was significantly (chi-square = 5.28, $p = .02$) lower than the 3.9% death rate for patients with 5 to 9 points. At the other end of the spectrum, the 73% death rate for patients with 30 to 34 points was significantly lower than the 84% death rate for patients with 35 or more points (chi-square = 7.5, $p = .01$). A five-point increase in APACHE II also significantly ($p < .0001$) increased death rate in the intermediate ranges of severity.

TABLE 2. Description of ICU admissions to 13 hospitals participating in APACHE II validation

Hospital	Patients over 65 yr	Patients with Chronic Failing Health	Patients with Cancer	Postoperative Patients	Most Frequent Postoperative Admission System and Disease Group (%)	Most Frequent Nonoperative Admission System and Disease Group (%)
1	41%	15%	3%	50%	Cardiovascular (86) Coronary artery disease	Cardiovascular (41) Sepsis
2	43%	21%	8%	43%	Cardiovascular (57) Coronary artery disease	Cardiovascular (34) Postarrest
3	36%	20%	2%	66%	Cardiovascular (70) Coronary artery disease	Neurologic (37) Intracranial hemorrhage
4	28%	14%	4%	86%	Cardiovascular (67) Coronary artery disease	Cardiovascular (38) Trauma
5	35%	29%	8%	75%	Cardiovascular (77) Coronary artery disease	Cardiovascular (45) Sepsis
6	39%	5%	1%	46%	Cardiovascular (34) Peripheral vascular	Cardiovascular (33) Trauma
7	30%	17%	8%	2%	(Not applicable)	Respiratory (37) Pneumonia
8	30%	13%	8%	70%	Gastrointestinal (32) GI neoplasm	Cardiovascular (33) Peripheral vascular
9	37%	20%	11%	58%	Cardiovascular (48) Coronary artery disease	Cardiovascular (29) Sepsis
10	34%	23%	5%	54%	Cardiovascular (55) Coronary artery disease	Cardiovascular (48) Coronary artery disease
11	24%	12%	3%	13%	Cardiovascular (38) Peripheral vascular	Cardiovascular (57) Coronary artery disease
12	54%	24%	8%	42%	Cardiovascular (32) Peripheral vascular	Cardiovascular (32) Trauma
13	42%	14%	6%	89%	Cardiovascular (56) Peripheral vascular	Cardiovascular (36) Postarrest

TABLE 3. A comparison of the statistical accuracy of APACHE II with APACHE using logistic regression analysis. The dependent variable is hospital death. The independent variables are severity of disease (APS-12 and APS-34), age, chronic health status, surgical status, and major diagnoses. All patients from the 13 hospitals are included, except those with CABG (N = 5030, 993 deceased).

Variables	APACHE II	Original APACHE
APS-12[a]	0.138[b] (529.28)	
APS-34		0.111[b] (494.22)
Age groups[a,c]		
45–54	0.436[d] (7.36)	0.394[e] (6.21)
55–64	0.621[b] (18.78)	0.573[b] (16.60)
65–74	1.113[b] (65.74)	1.071[b] (61.95)
≥75	1.323[b] (82.28)	1.297[b] (81.08)
Chronic health status[a,c]	0.739[b] (49.27)	0.692[b] (43.75)
Surgical status[c]		
Nonoperative	0.860[b] (48.75)	0.901[b] (55.40)
Emergency operative	0.698[b] (22.12)	0.729[b] (24.18)
ICU admission diagnoses[c]		
Postarrest	0.125 (0.27)	0.325 (1.84)
Septic shock	0.107 (0.17)	0.021 (0.01)
Intracranial bleeding	0.307 (1.56)	0.656[b] (7.39)
GI bleeding	0.122 (0.24)	−0.064 (0.07)
Multiple trauma	−0.697[e] (4.96)	−0.568 (3.30)
Other neurologic	−0.558[e] (5.32)	−0.514[e] (4.79)
Other respiratory	−0.404 (3.38)	−0.400 (3.34)
Other cardiovascular	−0.507[e] (5.80)	−0.534[e] (6.46)
Other GI	0.359 (2.31)	0.269 (1.24)
Renal	−0.955[d] (7.96)	−0.935[d] (7.27)
Metabolic and drug overdose	−2.018[b] (34.82)	−1.907[b] (24.31)
Intercept	−4.381[b] (303.56)	−4.397[b] (305.24)
Measures of aggregate explanatory power:		
Model chi-square	1634.5[b]	1537.1[b]
% cases correctly classified (<.5>)	85.5%	85.7%
R-squared	.319	.310
Rank correlation between outcome and predicted probability	.739	.730
Area under ROC curve	.863	.851

[a] Contained in APACHE II score.

[b] $p < .001$ (partial chi-square statistics are in parentheses under each coefficient).

[c] The reference group is composed of elective surgery patients admitted to the ICU with an admission diagnosis of respiratory infection, under age 45, and with no chronic organ system failure. All of the independent variables other than the acute physiology score are categorical, with the coefficients measuring the impact of being in the named category, relative to the reference category. For example, the coefficient on age 65 to 74 measures the increased risk of death for that age, relative to the reference age group, under 45 (hospital dead = 1, alive = 0).

[d] $p < .01$ (partial chi-square statistics are in parentheses under each coefficient).

[e] $p < .05$ (partial chi-square statistics are in parentheses under each coefficient).

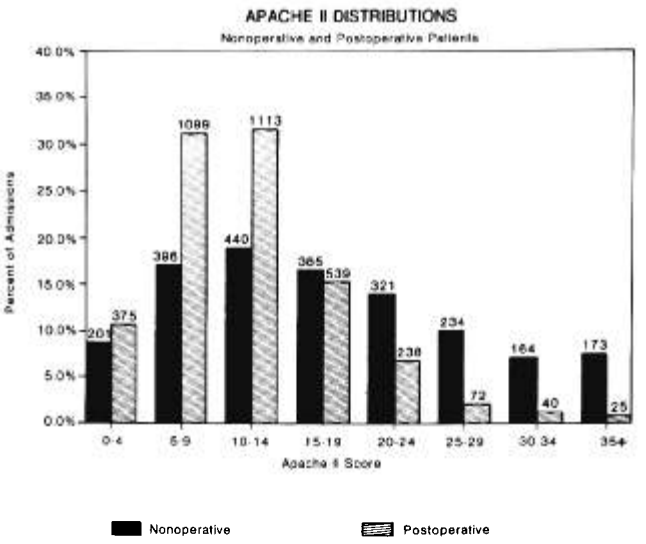

FIG. 2. The distribution of APACHE II scores among 5815 ICU admissions.

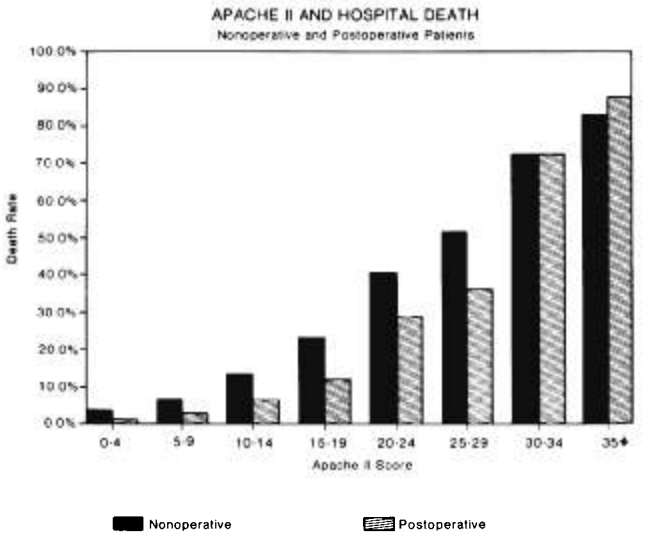

FIG. 3. The relationship between APACHE II scores and hospital mortality among 5815 ICU admissions.

The overall risk of hospital death varied according to the disease (Fig. 4). For example, patients with congestive heart failure admitted with APACHE II scores of 10 to 19 had a lower observed hospital death rate than septic shock patients with similar scores (13% vs. 26%, respectively). The only exception was septic shock patients with scores less than 9, because this group had only six cases and two deaths.

In addition to direct cross-tabulations, we also used multivariate techniques to illustrate the validity of APACHE II. These allowed us to compare the power of APACHE II with the original APACHE classification system, examine the relative importance of the components of APACHE II, and compare the relative importance of diagnoses with the APACHE II system. Table 3 reports a multiple logistic regression equation

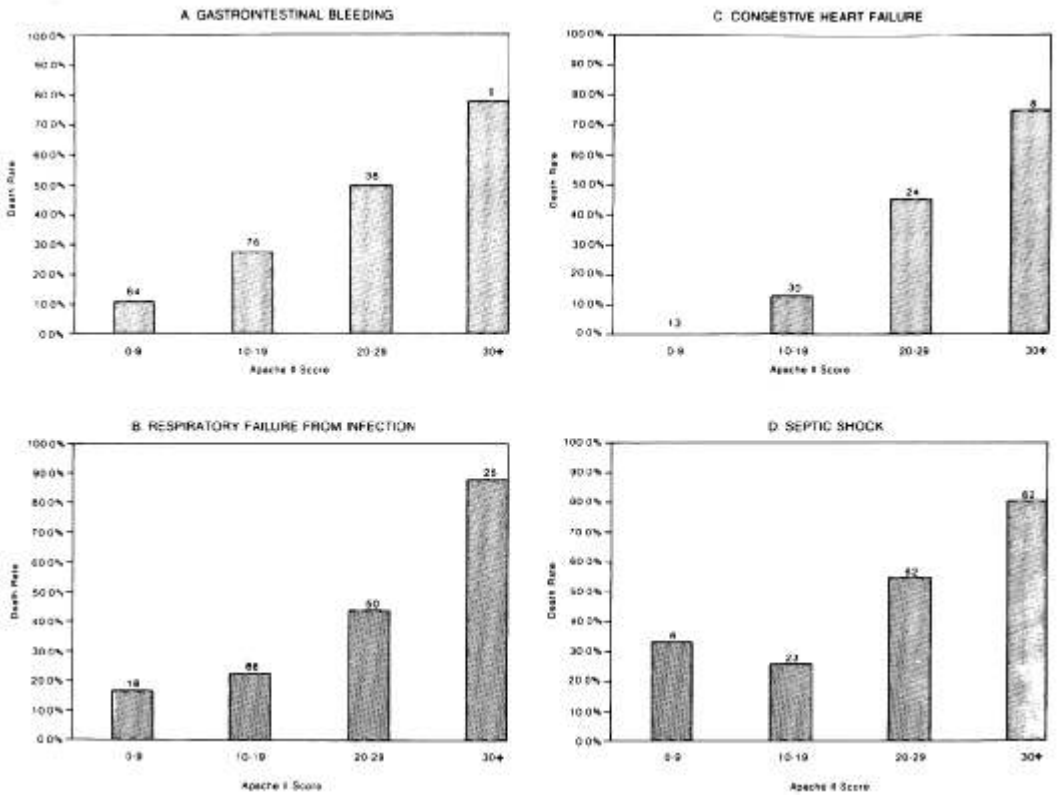

FIG. 4. The relationship of APACHE II scores to hospital mortality within four nonoperative diagnostic categories.

appropriate for analysis of a dichotomous outcome such as mortality prediction.[15] This equation is of the form: $\ln (R/1-R) = A + B_iX_i$; where R is the risk of death, (R/1−R) is the odds ratio, A is the estimated intercept and B is the estimated coefficient for each independent (i) variable such as physiologic derangement, age group, severe chronic health impairment, and diagnostic category. The derivation is detailed in the Appendix. As shown by the summary statistics at the bottom of Table 3, the 12-variable APS had a slightly better aggregate explanatory power than did the 34-variable score. It is important to emphasize that the form of the equation in Table 3 was not obtained with stepwise or other data search techniques.

The aggregate analysis in Table 3 excludes postcoronary artery bypass graft (CABG) patients. Although APACHE II scoring works for CABG admissions, these patients represented a large group (N = 785) whose surgical and anesthetic management resulted in high scores at ICU admission (average 12.4) but very low hospital death rates (1.5%). Because the implications of their physiologic derangement are so different from the majority of ICU admissions, it is unwise to trust the linearity assumptions of multivariate logistic regression to adjust for these differences. Including this large group of CABG patients resulted in little change in the equation but would have made the resulting predictions slightly less accurate for the majority of ICU admissions.

Individual estimated death rates obtained from equation 1 in Table 3 were used with a decision criterion of .50 to derive a classification matrix (Table 4). A decision criterion of .50 means that every patient with a risk greater than .50 is predicted to die. The overall correct classification rate was 86%. This information can also be displayed by a receiver operator characteristic (ROC) curve (Fig. 5, Point A). Because each point on this ROC curve corresponds to a different classification table, we selected three additional points (B, C, and D) from Figure 5 which correspond to classification matrixes in Table 5. This illustrates the trade-off between the specificity and sensitivity of the predictions as the decision rule varies.[16] Thus, as severity and APACHE II scores increase (at .70, .80, and .90 predicted risk of death), the false-positive rate (predicted to die but lived) decreases.

TABLE 4. Classification for a .50 predicted risk (Point *A* on Fig. 5)[a]

	Predicted		
	Alive	Dead	Total
True			
Alive	3833	204	4037
Dead	526	467	993
Total	4359	671	5030

[a] Sensitivity: 47.0%; specificity: 94.9%; correct: 85.5%; predictive value positive: 69.6%; predictive value negative: 87.9%.

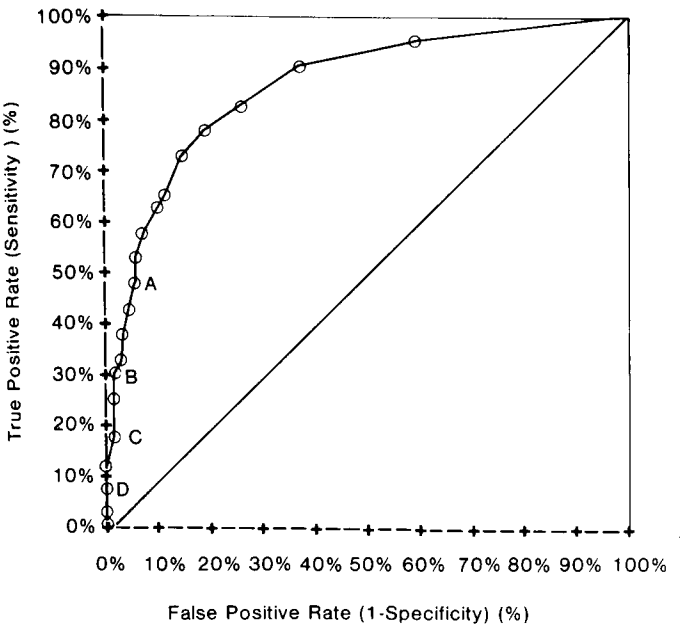

FIG. 5. ROC curve demonstrating predictive ability of APACHE II, based on 5030 ICU admissions to 13 hospitals. Points *A*, *B*, *C*, and *D* correspond to decision criteria of .50, .70, .80, and .90, respectively, for predicted risk of death. The diagonal line indicates an index that operates no better than chance.

TABLE 5. Classifications for alternative risks

.70 Predicted risk (Point *B* on Fig. 5)[a]

	Predicted		
	Alive	Dead	Total
True			
Alive	3974	63	4037
Dead	702	291	993
Total	4676	354	5030

.80 Predicted risk (Point *C* on Fig. 5)[b]

	Predicted		
	Alive	Dead	Total
True			
Alive	4006	31	4037
Dead	808	185	993
Total	4814	216	5030

.90 Predicted risk (Point *D* on Fig. 5)[c]

	Predicted		
	Alive	Dead	Total
True			
Alive	4030	7	4037
Dead	920	73	993
Total	4950	80	5030

[a] Sensitivity: 29.3%; specificity: 98.4%; correct: 84.8%; predictive value positive: 82.2%; predictive value negative: 85.0%.

[b] Sensitivity: 18.6%; specificity: 99.2%; correct: 83.3%; predictive value positive: 85.6%; predictive value negative: 83.2%.

[c] Sensitivity: 7.3%; specificity: 99.8%; correct: 81.5%; predictive value positive: 91.1%; predictive value negative: 81.4%.

DISCUSSION

When APACHE II scores were evaluated for their ability to stratify groups of ICU admissions prognostically, the results were very strong and stable, even for relatively small groups of patients within specific disease categories. We believe that the APACHE II system is able to stratify a wide variety of patients prognostically because of the strong and consistent underlying relationship between acute physiologic derangement and the risk of death during acute illness.[17-21]

Preliminary analyses of this sample suggest that APACHE II performs equally well in 34 disease categories. Variations in death rates by disease reflect the nature of the underlying process.[17] This is why it is crucial to combine the APACHE II score with a precise description of disease, especially for those diseases with a good overall prognosis (as indicated by a very negative coefficient, such as acute asthma or diabetic ketoacidosis) and those with a poor prognosis (corresponding to a large positive coefficient, such as septic shock) (Appendix and Table 6). Disease-specific mortality predictions should be derived from at least 50 patients in each diagnostic category, with at least 20 patients in the least-frequent outcome category. When such results are combined with details on therapeutic approach, diagnostic-specific nomograms can be developed relating APACHE II scores to outcome for individual diseases.[22] In the future, as we obtain additional information on APACHE II in specific diagnostic categories, we also will be updating the coefficients provided in the Appendix.

Our data further indicate that classification would be more appropriate if done at an early point in time, such as in the emergency room or at the time of ICU admission. This would make the severity classification more independent from treatment. When we tested the association between admission and worst-value APACHE II scores on a subset of GWUMC patients, in 88% of the physiologic measurements the worst value over 24 h was the ICU admission value. Also, 81% of APS scores changed less than five points when using admission values only. However, although APACHE II scores based on admission values were close to those obtained using worst values over the initial 24 h, they were not identical. Therefore, we are in the process of further comparing initial values and worst 24-h values. Until this is completed, investigators must still use the worst value over 24 h. We recommend, however, that data collection also include admission physiologic values, because we will eventually be adopting that approach.

It should be emphasized that first-day APACHE II scores do not perfectly predict death rates for individual patients. However, our data indicate that misclassifi-

TABLE 6. Number of patients by diagnostic category and observed deaths

	No. of Patients	No. of Deaths (%)
Nonoperative Patients		
Respiratory failure or insufficiency from:		
Asthma/allergy	32	2 (6)
Chronic obstructive pulmonary disease	50	15 (30)
Pulmonary edema (noncardiogenic)	65	24 (37)
Postrespiratory arrest	89	33 (37)
Aspirations/poisoning/toxic	38	14 (37)
Pulmonary embolus	32	7 (22)
Infection	159	62 (39)
Neoplasm	31	17 (55)
Cardiovascular failure or insufficiency from:		
Hypertension	43	3 (7)
Rhythm disturbance	56	4 (7)
Congestive heart failure	75	21 (28)
Hemorrhagic shock/hypovolemia	38	19 (50)
Coronary artery disease	110	17 (15)
Sepsis	180	104 (58)
Postcardiac arrest	155	103 (66)
Cardiogenic shock	24	8 (33)
Dissecting thoracic/abdominal aneurysm	58	18 (31)
Trauma:		
Multiple trauma	123	8 (7)
Head trauma	69	8 (12)
Neurologic:		
Seizure disorder	51	13 (25)
ICH/SDH/SAH	126	63 (50)
Other:		
Drug overdose	153	2 (1)
Diabetic ketoacidosis	100	13 (13)
Gastrointestinal bleeding	187	54 (29)
If not in one of the specific groups above, then which major vital organ system was the principal reason for admission?		
Metabolic/Renal	72	18 (25)
Respiratory	63	7 (11)
Neurologic	67	16 (24)
Cardiovascular	54	17 (31)
Gastrointestinal	66	26 (39)
Postoperative patients		
Multiple trauma	121	11 (9)
Admission due to chronic cardiovascular disease	90	6 (7)
Peripheral vascular surgery	494	31 (6)
Heart valve surgery	225	18 (8)
Craniotomy for neoplasm	301	15 (5)
Renal surgery for neoplasm	37	2 (5)
Renal transplant	47	5 (11)
Head trauma	51	7 (14)
Thoracic surgery for neoplasm	221	13 (6)
Craniotomy for ICH/SDH/SAH	147	29 (20)
Laminectomy and other spinal cord surgery	57	4 (7)
Hemorrhagic shock/hypovolemia	69	15 (22)
Gastrointestinal bleeding	47	8 (17)
GI surgery for neoplasm	153	22 (14)
Respiratory insufficiency after surgery	52	10 (19)
Gastrointestinal perforation/obstruction	106	36 (34)
For postoperative patients admitted to the ICU for sepsis or postarrest, use the corresponding weights for nonoperative patients.		
If not in one of the above, which major vital organ system led to ICU admission postsurgery?		
Neurologic	82	7 (9)
Cardiovascular	88	10 (11)
Respiratory	127	9 (7)
Gastrointestinal	90	13 (14)
Metabolic/Renal	36	6 (36)
Total	5030	993 (20)

cation rates become smaller as the probability of death increases (Tables 4 and 5, Fig. 5).

APPLICATIONS

We believe the ability to classify patient groups according to severity of illness will provide researchers with a new tool for improving the treatment of critically ill patients. APACHE II can be very useful in clinical trials or in nonrandomized or multi-institutional studies of therapeutic efficacy. By providing a measure of severity of disease, APACHE II scores will help investigators determine whether control and treatment groups are similar. In addition, using the Appendix, an expected death rate based on APACHE II can be compared to actual death rate as a test of therapeutic efficacy.[8]

In this regard, it is important to note that the statistical precision of APACHE II predictions is comparable to that found with the burn index.[23] The use of the burn index has made it possible for investigators to demonstrate an overall improvement in the quality of burn care during the last decade. Similar comparisons would be possible for intensive care using APACHE II data collected over time. Like the Glasgow coma score, APACHE II should also be able to help determine whether new therapeutic interventions really benefit severely ill patients.[24]

In studies of specific disease groups, APACHE II scores can only be expected to provide a minimal description of severity of disease. Investigators may also want to use additional indicators of severity, such as serum albumin and anergy testing for nutritional studies, or pulmonary mechanics for respiratory surveys. The importance of APACHE II is that it combines in one summary measure the risk factors of physiologic derangement, age, and poor chronic health status. This is an improvement over the comparison of mean values which do not take into account comorbidity, interaction of variables from different organ systems, or important physiologic thresholds.[25]

The original APACHE system demonstrates that the degree of physiologic derangement correlates closely with the need for admission and continued stay in an ICU for low-risk monitored patients.[26] Because it is less complex and still relatively independent of therapeutic decisions, the APACHE II system should be even more useful for such questions or for determining the relative benefit of an invasive procedure. For specific research questions, we suggest using only the 12 physiologic variables without adding points for age and a chronic disease. While advanced age and severe chronic disease are risk factors for death from an acute illness, they may not be needed for risk stratification in studies where the end-point is not hospital mortality.

Although the original APACHE system was not primarily developed to be used for individual patient treatment decisions, APACHE II can provide the clinician with a systematic evaluation and an improved understanding of how an individual patient's severity of disease influences outcome. We believe that outcome data from very carefully described groups of patients can be used for making clinical decisions.[27] For example, it is important for a clinician to know the expected death rate for a group of respiratory failure patients scheduled for treatment with a new drug or ventilator. It is also useful to know that, in this study, there were no survivors among 24 septic shock patients with APACHE II scores greater than 40. Such information, when integrated with a particular patient's overall clinical course, can provide a useful element for good clinical decision-making.[28] Similar physiologic data collected over time can be even more precise, and this is an important area for future research.

We encourage persons using APACHE II to review carefully Table 6. This table lists the number of patients from our research in each diagnostic category, along with the group death rate. Diagnostic categories with small numbers of patients and/or low death rates mean the user should be cautious in applying projections to study patients with that disease. The same caution applies to patients whose major diagnosis is not listed and for whom only major organ system stratification is available.

Regardless of the number of patients studied, however, prognostic estimates are still only estimates. Providing intensive medical care to individuals will always require experienced clinical judgment and careful integration of objective data with other relevant information, such as the individual reaction and the personal wishes of the patient.

ACKNOWLEDGMENTS

This study and the final version of APACHE II would not have been possible without the substantial and invaluable contribution of time and talent from the following people involved in data acquisition: Carolyn Bekes, MD, George Kuhn, RN, and W. Eric Scott, MD (Cooper Medical Center, Camden, NJ); Peter E. Dans, MD, Jeanne Keruly, RN, and Warren R. Summer, MD (John Hopkins Hospital, Baltimore, MD); Susan Brown, RN, Paul Cox, MD, and Phyllis Ogrodnik, RN (Maine Medical Center, Portland, ME); David J. Cullen, MD, and Roberta Keene, RN (Massachusetts General Hospital, Boston, MA); Robert Bevis, MD, and David Mize, MD (Medical College of Georgia, Augusta, GA); Robert C. Gilroy, MD, and Carey Goodrich, RRT (Polyclinic Medical Center, Harrisburg, PA); Helen Epstein, RRA, Gerald E. Gustafson, MD, and Barbara Reynolds, RN (St. Francis Hospital, Tulsa, OK); Jean Henderson, RN, MS, James M. Klick, MD, and Kenneth W. Travis, MD (South Shore Hospital, South Weymouth, MA); Judith Moran, RN, PhD, Michael Rosenthal, MD, and James F. Silverman, MD (Stanford University Hospital, Stanford, CA); Mary McKinley, RN, and Baekhyo Shin, MD (University of Maryland Hospital, Baltimore, MD); Mike Flannagan and John W. Hoyt, MD (University of Virginia Medical Center, Charlottesville, VA); Jean Grube, RN, Mary Kay Kohles, BSN, MSW, and Dennis G. Maki, MD (University of Wisconsin Hospital and Clinics, Madison, WI). We are also indebted to the following individuals for their invaluable research assistance: David

Abrams, Gladys Campbell, Agnes Courtney-Jenkins, Netta Fedor, Terri Kuznicki, Diane Lawrence Reba, Jerry Tietelbaum, Tom Wineland, Jean Wolff, and Lauri Yablick.

REFERENCES

1. Russell LB: Technology in hospitals: Medical advances and their diffusion. Washington, DC, The Brookings Institution, 1979
2. Relman AS: Intensive care units: Who needs them? *N Engl J Med* 1980; 302:965
3. Feinstein AR: An additional basic science for clinical medicine. I: The constraining fundamental paradigms. *Ann Intern Med* 1983; 99:393
4. Knaus WA, Zimmerman JE, Wagner DP, et al: APACHE—acute physiology and chronic health evaluation: A physiologically based classification system. *Crit Care Med* 1981; 9:591
5. Scheffler RM, Knaus WA, Wagner DP, et al: Severity of illness and the relationship between intensive care and survival. *Am J Public Health* 1982; 72:449
6. Gustafson DH, Fryback D, Rose J, et al: Decision: Theoretical methodology for severity index development. Madison, WI, Center for Health Systems Research and Analysis, University of Wisconsin, 1981
7. Knaus WA, Draper EA, Wagner DP, et al: Evaluating outcome from intensive care: A preliminary multihospital comparison. *Crit Care Med* 1982; 10:491
8. Knaus WA, Le Gall JR, Wagner DP, et al: A comparison of intensive care in the U.S.A. and France. *Lancet* 1982; ii:642
9. Champion HR, Sacco WJ: Measurement of patient illness severity. *Crit Care Med* 1982; 10:552
10. Teres D, Brown RB, Lemeshow S: Predicting mortality of intensive care patients: The importance of coma. *Crit Care Med* 1982; 10:86
11. Sweet SJ, Glenney CU, Fitzgibbons JP, et al: Synergistic effect of acute renal failure and respiratory failure in the surgical intensive care unit. *Am J Surg* 1981; 141:492
12. Li TC, Phillips MC, Shaw L, et al: Staffing in a community hospital intensive care unit. *JAMA* 1984; 252:2023
13. Wagner DP, Knaus WA, Draper EA: Statistical validation of a severity of illness measure. *Am J Public Health* 1983; 73:878
14. Kenney RA: Physiology of Aging: A Synopsis. Chicago, Year Book Medical Publishers, Inc., 1982
15. Chambers EA, Cox DR: Discrimination between alternative binary response models. *Biometrika* 1967; 54:573
16. McNeil BJ, Keeler E, Adelstein SJ: Primer on certain elements of medical decision making. *N Engl J Med* 1975; 293:211
17. Knaus WA, Wagner DP, Draper EA: Relationship between acute physiologic derangement and risk of death. *J Chronic Dis* 1985; 38:295
18. Wilson RF, Gibson D, Percinel AK, et al: Severe alkalosis in critically ill surgical patients. *Arch Surg* 1972; 105:197
19. Winkel P, Afifi AA, Cady LD, et al: Application of statistical techniques for assessment of prognosis in patients with acute circulatory failure (shock). *J Chronic Dis* 1971; 24:61
20. Siegel JH, Goldwyn RM, Friedman HP: Patterns and process in the evaluation of human septic shock. *Surgery* 1971; 70:232
21. Shoemaker WP, Chang P, Czer L: Cardiovascular monitoring in postoperative patients. *Crit Care Med* 1979; 7:237
22. Dellinger EP, Wertz MJ, Meakins JL, et al: Surgical infection stratification system for intra-abdominal infection. *Arch Surg* 1985; 120:21
23. Feller I, Tholen D, Cornell RG: Improvements in burn care, 1965 to 1979. *JAMA* 1980; 244:2074
24. Jennett B, Bond M: Assessment of outcome after severe brain damage. *Lancet* 1975; i:480
25. Knaus WA, Wagner DP, Draper EA: The value of measuring severity of disease in clinical research on acutely ill patients. *J Chronic Dis* 1984; 37:455
26. Wagner DP, Knaus WA, Draper EA: Identification of low-risk monitor patients within a medical-surgical intensive care unit. *Med Care* 1983; 21:425
27. Levy DE, Bates D, Cardonna JJ, et al: Prognosis in non-traumatic coma. *Ann Intern Med* 1981; 94:293
28. Knaus WA, Draper EA, Wagner DP: The use of intensive care: New research initiatives and their implications for national health policy. *Milbank Mem Fund Q* 1983; 61:561

APPENDIX

To compute predicted death rates for groups of acutely ill patients, for each individual compute the risk (R) of hospital death with the following equation; then sum the individual risks and divide by the total number of patients.

$$\ln(R/1\text{-}R) = -3.517 + (\text{APACHE II score} \times 0.146) + (0.603, \text{only if postemergency surgery}) + (\text{Diagnostic category weight, as shown below})$$

Principal Diagnostic Categories Leading to ICU Admission

Nonoperative patients

Respiratory failure or insufficiency from:	
Asthma/allergy	−2.108
COPD	−0.367
Pulmonary edema (noncardiogenic)	−0.251
Postrespiratory arrest	−0.168
Aspiration/poisoning/toxic	−0.142
Pulmonary embolus	−0.128
Infection	0
Neoplasm	0.891
Cardiovascular failure or insufficiency from:	
Hypertension	−1.798
Rhythm disturbance	−1.368
Congestive heart failure	−0.424
Hemorrhagic shock/hypovolemia	0.493
Coronary artery disease	−0.191
Sepsis	0.113
Postcardiac arrest	0.393
Cardiogenic shock	−0.259
Dissecting thoracic/abdominal aneurysm	0.731
Trauma:	
Multiple trauma	−1.228
Head trauma	−0.517
Neurologic:	
Seizure disorder	−0.584
ICH/SDH/SAH	0.723
Other:	
Drug overdose	−3.353
Diabetic ketoacidosis	−1.507
GI bleeding	0.334
If not in one of the specific groups above, then which major vital organ system was the principal reason for admission?	
Metabolic/renal	−0.885
Respiratory	−0.890
Neurologic	−0.759
Cardiovascular	0.470
Gastrointestinal	0.501

'ostoperative patients	
Multiple trauma	−1.684
Admission due to chronic cardiovascular disease	−1.376
Peripheral vascular surgery	−1.315
Heart valve surgery	−1.261
Craniotomy for neoplasm	−1.245
Renal surgery for neoplasm	−1.204
Renal transplant	−1.042
Head trauma	−0.955
Thoracic surgery for neoplasm	−0.802
Craniotomy for ICH/SDH/SAH	−0.788
Laminectomy and other spinal cord surgery	−0.699
Hemorrhagic shock	−0.682
GI bleeding	−0.617
GI surgery for neoplasm	−0.248
Respiratory insufficiency after surgery	−0.140
GI perforation/obstruction	0.060

For postoperative patients admitted to the ICU for sepsis or postarrest, use the corresponding weights for nonoperative patients.

If not in one of the above, which major vital organ system led to ICU admission postsurgery?

Neurologic	−1.150
Cardiovascular	−0.797
Respiratory	−0.610
Gastrointestinal	−0.613
Metabolic/renal	−0.196

For example: A patient admitted with noncardiogenic pulmonary edema (nonoperative) having 15 APACHE II points would have the following estimated risk:

$$\begin{aligned}\ln (R/1-R) &= -3.517 + (15 \times 0.146) \\ &\quad + (0 \times 0.603) - 0.251 \\ &= -3.517 + 2.19 + 0 - 0.251 \\ &= -1.578\end{aligned}$$

Since the exponential of −1.578 is +0.206, then (R/1−R) equals +0.206, and R is 0.17 or 17% estimated risk of hospital death.