

Applying Ockham's Razor to Pancreatitis Prognostication

A Four-Variable Predictive Model

Austin L. Spitzer, MD,*† Anthony M. Barcia, BS,* Michael T. Schell, MD,* Annabel Barber, MD,‡ James Norman, MD,§ James Grendell, MD,|| and Hobart W. Harris, MD, MPH*

Objective: We sought to develop a simple yet accurate prognostic scoring system to determine the severity of acute pancreatitis at admission.

Summary Background Data: Because acute pancreatitis has a variable and frequently unpredictable course, identifying individuals at greatest risk for significant, life-threatening complications and stratifying their care appropriately remain a concern. Previous prognostic scoring systems predict severity reasonably well but are limited by time constraints, are unwieldy to use, or both.

Methods: Data from the international phase III trial of the platelet-activating factor receptor-antagonist Lexipafant were used to develop a 4-variable prognostic model. We then compared the model's ability to predict the severity of acute pancreatitis with the Ranson, Glasgow, and APACHE II systems.

Results: The model (BALI), which included $BUN \geq 25$ mg/dL, $Age \geq 65$ years, $LDH \geq 300$ IU/L, and $IL-6 \geq 300$ pg/mL, measured at admission, was similar to the Ranson, Glasgow, and APACHE II systems in its ability to identify increased mortality from acute pancreatitis. The receiver operating characteristic curve area for the BALI model was $\geq 0.82 \pm 0.03$ (mean \pm SD) versus 0.75 ± 0.04 (Ranson), 0.80 ± 0.03 (Glasgow), and 0.79 ± 0.03 (APACHE II). Furthermore, at a prevalence of 15%, the positive and negative predictive values for increased mortality were similar for all systems.

Conclusion: The prognostic ability of the BALI 4-variable model was similar to the Ranson, Glasgow, and APACHE II systems but is unique in its simplicity and ability to accurately predict disease severity when used at admission or anytime during the first 48 hours of hospitalization.

(*Ann Surg* 2006;243: 380–388)

From the *Department of Surgery, University of California-San Francisco, San Francisco, CA; and †Department of Surgery, University of California-East Bay, East Bay, CA; ‡Department of Surgery, University of Nevada, Las Vegas, NV; §Department of Surgery, University of South Florida, Tampa, FL; and ||Department of Surgery, New York Hospital-Cornell Medical Center, New York, NY.

Reprints: Hobart W. Harris, MD, MPH, Division of General Surgery, University of California-San Francisco, 513 Parnassus Avenue, S-301, San Francisco, CA 94143-0104. E-mail: harrish@surgery.ucsf.edu.

Copyright © 2006 by Lippincott Williams & Wilkins
ISSN: 0003-4932/06/24303-0380

DOI: 10.1097/01.sla.0000202213.22389.36

Although advances in understanding the pathophysiology of acute pancreatitis and treating its complications have improved patient outcomes, the mortality rate for severe pancreatitis remains 8% to 15%.^{1–4} Identifying patients who will require aggressive resuscitation and intensive care measures therefore remains imperative.

The perpetual failure of clinical assessment to accurately predict pancreatitis severity^{5,6} led to the development of prognostic scoring systems based on objective clinical and laboratory data by Ranson et al^{7,8} and Imrie et al.^{9,10} Known as the Ranson and Glasgow systems, they have allowed patients to be stratified according to disease severity and resources to be focused on appropriate treatment regimens; however, both systems use a prognostic score that is based on data collected over 24 to 48 hours, effectively resulting in what many consider a missed opportunity for intervention. The revised Acute Physiologic and Chronic Health Evaluation (APACHE II)^{11,12} scoring system allows patients to be stratified at admission,^{13–15} but its unwieldiness (weighted scoring of 3 sets of variables) continues to limit its use. Consequently, several simple, clinically useful predictors of early severity have been proposed, such as trypsinogen activation peptide (TAP),^{16,17} C-reactive protein (CRP),^{18–20} and interleukin-6 (IL-6),¹⁹ but lack of specificity has limited their widespread clinical application.

We sought to develop a simple model for predicting acute pancreatitis severity. To accomplish this, we analyzed data from the recent Lexipafant^{21–27} pancreatitis trial to develop a system that could be used at admission. We then compared its prognostic reliability with the established systems for predicting acute pancreatitis severity.

METHODS

Data Obtained From the Lexipafant Study

The Lexipafant study was a randomized, multicenter, double-blind, placebo-controlled, phase III trial to determine if an infusion of Lexipafant given within 48 hours of the onset of symptoms of pancreatitis could reduce all-cause mortality within 28 days.²⁸ The trial was conducted in accordance with the declaration of Helsinki,²⁹ the recommendations of the EEC Committee for Proprietary Medicinal Products,³⁰ and under the supervision of the Food and Drug Administration in the United States.

A total of 1518 subjects (56% men; mean \pm SD [SD] age, 61 ± 16 years) were enrolled according to the inclusion criteria; 510 were randomized to receive placebo, 498 to receive 10 mg/24 hours Lexipafant, and 510 to receive 100 mg/24 hours Lexipafant. The mean APACHE II score for each group was approximately 11 at enrollment. Blood samples were taken at 0, 12, 24, and 48 hours from all subjects. Serum samples obtained from the first 450 patients were analyzed for the cytokines IL-6 and IL-8. The 3 groups did not differ significantly in terms of organ dysfunction,^{31,32} sepsis,³³ or death at 90 days.²⁸ We confirmed that there was no statistical association between treatment group and mortality at 90 days (Fisher exact test $P > 0.9$).

Data records were downloaded into ASCII files and converted into SAS version 6.11 data sets (SAS Institute Inc., Cary, NC). Model development and system analyses were done using SAS version 8.2.

Model Development

Because subjects were enrolled on an “intent-to-treat” basis, we included mortality in all subjects ($n = 149$ at 90 days) in our analyses. Because no association was found between the placebo and treatment groups in terms of presentation, subject characteristics, and mortality, we assumed that there was no difference in treatment response or its effect on these variables, and that the entire group therefore consisted of a homogeneous population of individuals with acute pancreatitis as defined by the study inclusion criteria.

An index data set was created as a stratified random subset from the entire population of trial participants. We selected out two thirds of the population based on the 6 combinations of the 2 stratification variables, death (yes/no) and treatment (placebo, 10 mg/24 hours, or 100 mg/24 hours). A subset of those variables significant at $P \leq 0.001$ was analyzed to determine “optimal” cutoff points for each of the variables. Multiple logistic regression analyses were performed to explore the combinations of dichotomous predictors to reduce the codependency of variables. To develop a practical clinical model, we simplified the weight of each variable to be “1.” This approach made any combination of variables equivalent with regard to prognostication. The final values and cutoffs fit on the index set were BUN ≥ 25 mg/dL, age ≥ 65 years, LDH ≥ 300 IU/L, and IL-6 ≥ 300 pg/mL (BALI). Only initial, not peak or random, values were used as variables in the BALI model. As a secondary study, we tried to develop a similar model by using an index set of the placebo-treated patients. Statistical significance was considered at $P \leq 0.001$ as evaluated by Fisher’s 2-tailed exact probability or the Mann-Whitney rank sum test.

Discrimination was defined as a prognostic model’s ability to distinguish high-risk from low-risk individuals, which, in this case, meant 90-day mortality versus survival. Discrimination was quantified by the area under the receiver operating characteristic (ROC) curves^{34–36} and by positive predictive values (PPV) and negative predictive values (NPV) for each system. For PPV and NPV comparisons, a prevalence of 15% mortality was used. The 4 variables of the BALI model were compared with scores of 1, 3, 5, and 7 for

the Ranson and Glasgow systems and with APACHE II interval scores of <10 , 10–15, 15–20, 20–25, and >25 .

The validity of the BALI model was confirmed by comparing it against each system in a pairwise fashion with regard to mortality rate at each score. Only data for individuals who met the criteria of our model and the Ranson, Glasgow, or APACHE II systems were compared. Analysis of the Ranson model was limited to 9 of its 11 available variables because fluid sequestration and base deficit data were not available from subjects in the Lexipafant trial.

The performance of the BALI model was verified using the remaining one third sample (validation set) that had been reserved during the model building process, and its ability to predict severity was subsequently compared with the Ranson, Glasgow, and APACHE II systems. The discrepancy in the number of patients considered for each comparison is due to missing variables from each system. When the Ranson or Glasgow systems were evaluated, data were reviewed with “no missing” data sets or with a missing variable receiving a score of “0.”

RESULTS

Descriptive variables in the index and validation data sets did not differ significantly. Not surprisingly, many of the variables previously used in the Ranson and Glasgow systems were significantly related to mortality at 90 days, including age, BUN, LDH, calcium, and glucose, but hematocrit, white blood cell count (WBC), alanine aminotransferase (ALT), and arterial oxygenation (paO_2) were not, and were therefore excluded from the model (Table 1). Although the percentage change in hematocrit at 48 hours was associated with increased mortality ($P < 0.01$), its contribution was not significant in model development. Fluid sequestration and base deficit, variables previously discarded by Imrie et al⁹ in developing the Glasgow model, were not analyzed. Neither was aspartate aminotransferase (AST), a variable that Blamey et al¹⁰ discarded from a revised version of the Glasgow model because of its lack of significance.

To reduce the effects of codependency of the variables on overall prognostication, we included only BUN ≥ 25 mg/dL, age ≥ 65 years, LDH ≥ 300 IU/L, and IL-6 ≥ 300 pg/mL, as their contribution to the BALI model provided the best prediction of mortality. Only initial, not peak or random, values were used to develop the model parameters. The secondary analysis of placebo-treated patients alone also showed that age ≥ 65 , LDH ≥ 300 , and IL-6 ≥ 300 were 3 of the strongest predictors for increased mortality, although multivariate modeling was not possible because of sample size limitations. Removing BUN from the BALI model further increased the contribution of each remaining variable (odds ratios not shown); however, this necessarily decreased specificity, and ROC comparisons of our model with and without BUN resulted in a similar area (0.84), but no significant improvement ($P > 0.6$). This led us to compare the predictive power of the model with and without the IL-6 variable, which limited the area under the ROC curve (0.81); but again, the difference in areas was not significant ($P > 0.2$).

TABLE 1. Relationship Between Mean Values of Common System Variables and Mortality at Time 0 and 48 Hours

Variable	Time	Death	Mean ± SD	Median 95% CI	P*
IL-6 (pg/mL) (27%)	0	No	465 ± 1100	119, 217	<0.0001
		Yes	3970 ± 16000	424, 962	
	48	No	145 ± 316	0, 21	
		Yes	1630 ± 3400	163, 763	
BUN (mg/dL) (94%)	0	No	24.3 ± 21.8	17.6, 18.5	<0.0001
		Yes	46.6 ± 32.9	32.2, 44.0	
	48	No	18.9 ± 19.3	12.6, 14.0	
		Yes	53.0 ± 45.9	33.9, 46.8	
LDH (IU/L) (67%)	0	No	302 ± 329	230, 250	<0.0001
		Yes	677 ± 847	345, 567	
	48	No	301 ± 249	209, 231	
		Yes	776 ± 965	383, 669	
Albumin (g/dL) (91%)	0	No	3.29 ± .563	3.3, 3.4	<0.0001
		Yes	2.88 ± .747	2.8, 3.1	
	48	No	2.86 ± .543	2.8, 2.9	
		Yes	2.36 ± 0.562	2.2, 2.6	
Glucose (mg/dL) (90%)	0	No	152 ± 81.8	126, 133	<0.0001
		Yes	199 ± 149	155, 189	
	48	No	141 ± 92.6	115, 121	
		Yes	185 ± 99.0	135, 178	
Calcium (mg/dL) (94%)	0	No	8.04 ± 0.904	8.0, 8.4	<0.0001
		Yes	7.30 ± 1.22	7.2, 7.6	
	48	No	7.91 ± 0.836	2.0, 2.0	
		Yes	7.08 ± 1.15	1.7, 1.9	
AST (U/L) (91%)	0	No	123 ± 298	58, 67	<0.027
		Yes	231 ± 652	53, 92	
	48	No	50.0 ± 100	33, 36	
		Yes	243 ± 693	39, 58	
ALT (U/L) (91%)	0	No	136 ± 208	54, 70	0.189
		Yes	141 ± 245	32, 61	
	48	No	67.9 ± 96.3	36, 43	
		Yes	156 ± 406	28, 45	
paO ₂ (mm Hg) (65%)	0	No	84.7 ± 33.1	75.8, 78.7	0.047
		Yes	92.2 ± 39.4	75.7, 84.2	
	48	No	83.2 ± 24.6	76.5, 83.0	
		Yes	88.8 ± 35.6	73.0, 89.7	
WBC (1000 s/mm ³) (100%)	0	No	13.9 ± 6.10	12.8, 13.6	0.144
		Yes	14.7 ± 6.76	12.9, 15.1	
	48	No	10.9 ± 4.53	9.9, 10.5	
		Yes	11.9 ± 5.37	10.4, 12.6	
Hct (%) (100%)	0	No	41.3 ± 6.50	40.8, 41.6	0.866
		Yes	41.0 ± 7.76	40.0, 43.2	
	48	No	36.0 ± 5.26	35.9, 36.6	
		Yes	34.3 ± 6.55	33.8, 36.5	

*Mann-Whitney *U* test.

IL-6 indicates interleukin-6; BUN, blood urea nitrogen; LDH, lactate dehydrogenase; AST, aspartate aminotransferase; ALT, alanine aminotransferase; WBC, white blood cell count; Hct, hematocrit; paO₂, arterial oxygenation; SI unit conversion: BUN mg/dL × 0.357 = mmol/L (SI); albumin g/dL × 10 = g/L (SI); glucose mg/dL × 0.0555 = mmol/L (SI); calcium mg/dL × 0.25 = mmol/L (SI); and paO₂ mm Hg × 0.1333 = kPa (SI). Percentage of individuals with variable indicated in parentheses.

As anticipated, when prognostic scores were compared with mortality rates, the relationship was consistent throughout (Fig. 1). The BALI model demonstrated $\geq 25\%$ mortality with 3 positive variables, and $>50\%$ mortality with 4 positive variables. This relationship was consistent throughout the first 48 hours of admission (data not shown). Similar trends in prognostic scores and mortality rates were demonstrated with the Ranson and Glasgow scoring systems. For the Ranson criteria, ≤ 2 positive variables yielded approximately 4% mortality, 3 or 4 positive variables yielded 10%, 5 or 6 variables yielded 36%, and 7 or 8 variables yielded $>50\%$. For the Glasgow criteria, ≤ 2 positive variables yielded approximately 3% mortality, 3 or 4 positive variables yielded 4%, 5 or 6 variables yielded 19%, and 7 or 8 variables yielded $>46\%$. The APACHE II scores likewise demonstrated lower mortality at lower scores: <10 yielded 2% mortality, whereas scores >20 yielded greater than 50% mortality.

In pairwise comparisons, the ROC curve areas were similar for the BALI model and the established systems (Fig. 2). At

the power of this study, there was no significant difference between the BALI model (with the score determined at admission) and the established systems. This was true even though the comparisons were done using the most predictive APACHE II score obtained over the first 48 hours, and Ranson and Glasgow scores obtained over the first 48 hours.

The area under the ROC curve for the BALI model validation data set was 0.84, which was similar to the area for the validation data sets of the Ranson (0.70), Glasgow (0.77), and APACHE II (0.80) systems (data not shown). When we compared the established systems against our entire data set to ensure greatest predictability, the areas under the ROC curves were only slightly higher. Because each ROC curve was based on different subsets of patients, only a qualitative comparison was possible. When we attempted to improve the prediction of the Ranson analysis with LDH set at 350 versus 700 IU/L, there was no significant difference in area (0.77 versus 0.75). The same was true when the Glasgow system was analyzed without AST (0.79 versus 0.79).¹⁰

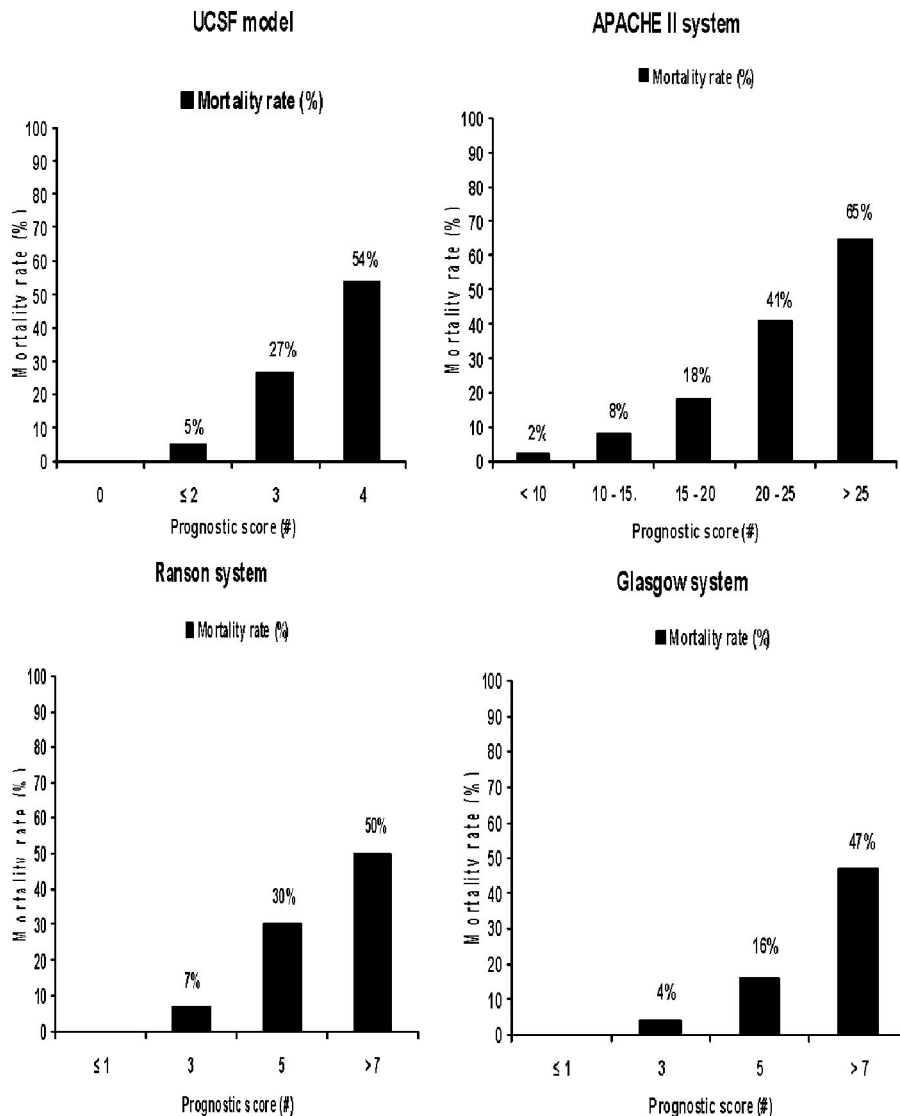


FIGURE 1. Mortality rate versus prognostic score for BALI model versus established systems. y-axis = mortality rate (%); x-axis = prognostic score or range for APACHE II interval score. Number of deaths per sample population: BALI model (n = 40 of 365), APACHE II (149 of 1518), Ranson (79 of 627), and Glasgow (105 of 854) systems.

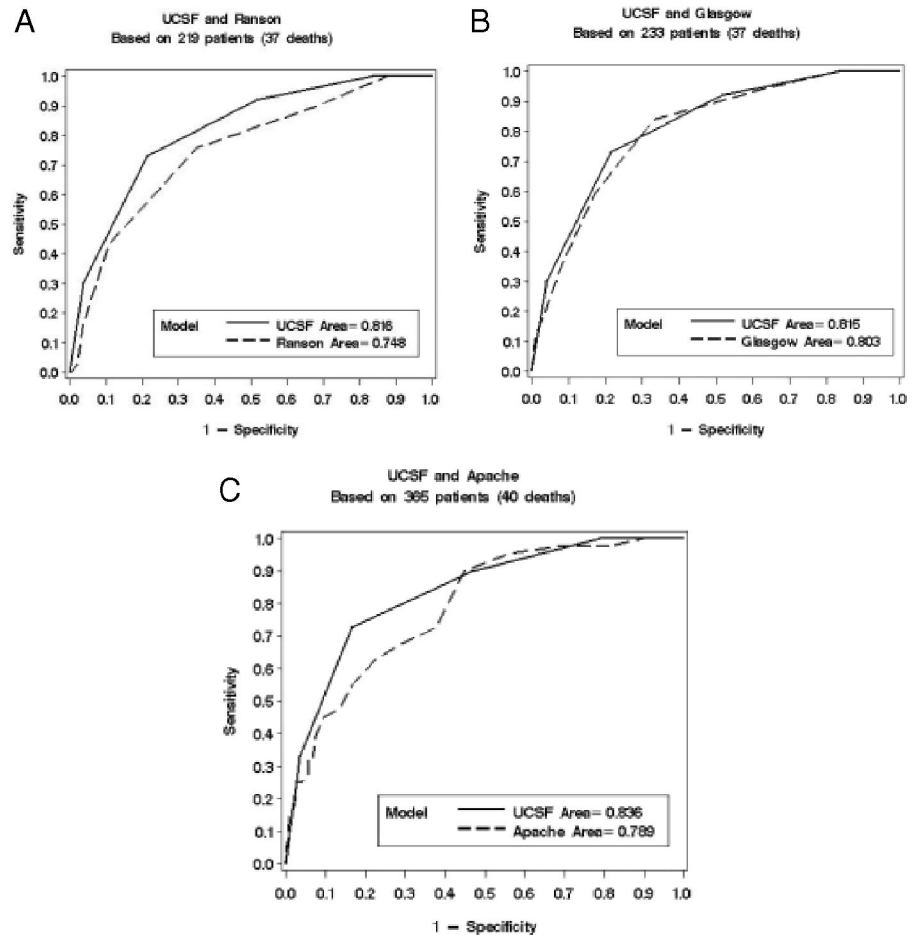


FIGURE 2. Receiver operating characteristic curves for BALI model versus the (A) Ranson, (B) Glasgow, and (C) APACHE II systems, respectively. y-axis = sensitivity; x-axis = 1 – specificity. Areas under the curves (AUC) for the BALI model were: $\geq 0.81 \pm 0.03$ versus 0.75 ± 0.04 , a difference of 0.07 (CI, -0.15 to 0.01 ; $P = 0.08$) (Ranson), 0.80 ± 0.03 ; a difference of 0.01 (CI, -0.08 to 0.06 ; $P = 0.72$) (Glasgow), and 0.79 ± 0.03 ; a difference of 0.05 (CI, -0.12 to 0.03 ; $P = 0.21$) (APACHE II), respectively.

The predictive values of the BALI model and the established systems consistently demonstrated the same patterns for sensitivity, PPV and NPV (Table 2). It is clear that with the BALI model and the Ranson and Glasgow systems, increasing the number of variables results in increased specificity and reduced sensitivity.

DISCUSSION

Utilizing the largest prospective database of individuals with presumed pancreatitis ever established, we developed a 4-variable model (BALI) that predicted increased mortality with similar sensitivity and specificity to the Ranson, Glasgow, and APACHE II scoring systems. Our model was based on a subset of variables from these systems and the more recently investigated cytokine assay for IL-6. Much as the creators of the Glasgow system refined the Ranson model into a simpler clinical tool, we attempted to streamline our model to include only the most predictive variables and to exclude duplicate or codependent variables. The BALI model has a similar advantage to the Ranson and Glasgow systems in that the score of each variable is “1”; therefore, there should not be a significant predictive advantage in one variable versus another and any combination should provide similar predictive power.

Age, BUN, and LDH, which were used in our model and in the Ranson and Glasgow systems, likely represent surrogate indicators of the individual’s physiologic reserve, end-organ function, and general inflammatory response. The immune system changes with the aging process, and age limits an individual’s response to an inflammatory insult.^{37–39} The rate at which BUN increases is influenced by the degree of tissue necrosis, protein catabolism, and the rate at which the kidneys excrete urea nitrogen.^{40,41} Elevated levels of LDH appear to provide a general, albeit nonspecific, indication of inflammation.^{40,41} Although adding the IL-6 assay to age, BUN, and LDH variables increases the discriminatory power of our model, establishing 3 of 3 positive variables without IL-6 data predicted approximately 40% mortality, enough to warrant increased vigilance for clinical deterioration in a patient’s condition.

A 1998 review of the existing scoring systems⁴² found their weakness is that their PPV for severe pancreatitis are only 40% to 60% and their sensitivities are only 60% to 80%, implying that 20% to 40% of patients with severe disease will not be detected before their demise becomes apparent.^{5,15,42,43} However, others have argued that high NPV¹⁷ or both high PPVs and NPVs¹⁵ are preferred when assessing acute pancreatitis severity. The high NPVs of the multifactorial scoring

TABLE 2. Sensitivity, Positive and Negative Predictive Value(s) for BALI Model, Ranson, Glasgow, and APACHE II Scoring Systems at a Mortality Prevalence of 15%

Model	Data Set	Criteria (no.)	Sensitivity (%)	Specificity (%)	PPV	NPV	No. of Deaths*
BALI	Index	1	100	22	0.18	1.0	29
		2	90	52	0.25	0.97	
		3	69	83	0.41	0.94	
		4	38	96	0.61	0.90	
	Validation	1	100	20	0.18	1.0	11
		2	91	56	0.27	0.97	
		3	82	85	0.49	0.96	
		4	18	98	0.65	0.87	
RANSON	Index	1	99	3	0.15	0.95	100
			100	1	0.15	1.0	54
		3	77	46	0.20	0.92	100
			91	36	0.20	0.96	54
		5	35	92	0.45	0.89	100
			44	89	0.43	0.90	54
		7	5	99	0.57	0.86	100
	Validation		9	98	0.49	0.86	54
		1	100	4	0.16	1.0	49
			100	4	0.16	1.0	25
		3	74	46	0.19	0.91	49
			84	34	0.18	0.92	25
		5	35	93	0.47	0.89	49
			44	90	0.44	0.90	25
GLASGOW	Index	1	99	2	0.15	0.93	100
			100	1	0.15	1.0	73
		3	91	25	0.18	0.94	100
			97	17	0.17	0.97	73
		5	72	74	0.33	0.94	100
			81	69	0.32	0.95	73
		7	26	97	0.57	0.88	100
	Validation		32	95	0.53	0.89	73
		1	100	4	0.16	1.0	49
			100	3	0.15	1.0	32
		3	88	27	0.18	0.93	49
			97	19	0.17	0.97	32
		5	61	73	0.29	0.92	49
			75	65	0.28	0.94	32
APACHE II	Index	10	95	43	0.23	0.98	100
		15	57	86	0.42	0.92	
		20	35	97	0.65	0.89	
		25	13	99	0.70	0.87	
	Validation	10	84	48	0.22	0.94	49
		15	57	84	0.39	0.92	
		20	41	94	0.60	0.90	
		25	22	99	0.82	0.88	

*The numbers in the Ranson and Glasgow divisions represent the entire data set and the subset without missing data. PPV indicates positive predictive value; NPV, negative predictive value.

systems have resulted in their application in many clinical practices, where the presence of <3 Ranson signs or APACHE II scores <10 are viewed as indicators of mild disease.^{14,44} The sensitivity, PPV, and NPV of our model and the Ranson and Glasgow systems consistently demonstrated similar results when 3 positive criteria were used (Table 2); however, we found slightly lower PPV.^{4-7,10,15,18,45-48} Furthermore, the APACHE II data PPV in our study are slightly lower than those anticipated from the literature.^{15,48} This discrepancy is likely due to the specificity of the systems; given that nondeaths are the larger and less homogeneous group, the probability of a negative result given a nondeath (specificity) is low. The deaths, in contrast, are likely more homogeneous.

There have been many other criticisms of the existing systems. Because not all of the Ranson criteria are completed during a typical patient evaluation, the system does not help discriminate at presentation. The Glasgow system, while an improvement in terms of available variables, also requires 24 to 48 hours for calculation. Another criticism of both systems is that certain variables are, in essence, duplicates (eg, WBC and LDH are nonspecific measures of inflammation) and do not actually contribute to predictive value. A distinct advantage of APACHE II is that it can be calculated at admission and on a daily basis to monitor changes in prognosis. Severe attacks of pancreatitis have been shown to correspond to increasing scores over the first 48 hours, whereas milder attacks demonstrate decreasing scores over time.¹⁴ On the basis of data from the Lexipafant trial, the success of the APACHE II system in predicting acute pancreatitis severity has been speculated to be due to its ability to identify patients with early organ failure.⁴⁹ The APACHE II system, which reportedly has the best accuracy of the scoring systems, has a sensitivity of predicting a severe attack of pancreatitis in approximately 61% of patients at admission.^{50,51} The BALI model is similar to APACHE II in that it can be calculated at admission and repeatedly during the initial 48 hours of hospitalization, thereby permitting early stratification of patient care and resources. However, we think that the BALI 4-variable model is simpler to use than the APACHE II system, which entails weighted scoring of multiple variables.

Other proposed markers of pancreatitis severity have been based on the knowledge that the earliest events of acute pancreatitis occur in acinar cells when pancreatic enzymes are overproduced and prematurely activated,⁵² an insult that is followed by an inflammatory response mediated by cytokines including IL-1, IL-6, IL-8, tumor necrosis (TNF), and platelet-activating factor.⁵³ These molecules, as well as TAP⁴² and CRP, have all been proposed as potential early markers of severe acute pancreatitis. TAP has shown promise as a single marker because it is specifically related to the onset of acute pancreatitis; however, as with CRP,^{18,47,54} it requires 24 hours to evaluate. Circulating levels of both IL-6 and IL-8 were shown to be higher in patients with complicated pancreatitis and were more predictive of end organ failure, duration of hospital stay, and overall mortality.^{19,55-57} However, at the time of these preliminary studies, the assays were substantially more cumbersome and time dependent.

Our findings were especially similar to those reported by Viedma et al⁵⁸ and Leser et al,⁵⁹ which showed marked elevation of IL-6 levels in patients with fatal outcomes. A recent report by Pooran et al⁶⁰ found significantly higher levels of IL-6, IL-8, and TNF in patients with severe acute pancreatitis than in patients with mild disease, and showed that IL-6 differentiated control from mild disease. Although the mean level for IL-6 (83 pg/mL) was lower than our “optimal” cutoff of 300 pg/mL for IL-6, their definition of severity was not based on mortality and would, by definition, increase the sensitivity of our test at that value at the expense of specificity. Their study also begs the question as to which cytokine marker would be the most appropriate in a predictive model. The brief production of TNF, hepatic clearance, and nonuniformity in measurement reduce its attractiveness as a predictor.⁶¹⁻⁶⁴ In our study, increased IL-8 was associated with increased mortality; however, a significant number of individuals in the Lexipafant database had undetectable levels of IL-8, and the contribution to the predictive model for the number of samples was therefore not as significant as that of IL-6. However, IL-8 may remain elevated longer than IL-6;⁵⁶ conceivably, an elevation in either cytokine would complement our model.

Although clinical testing of cytokines is not yet widely accepted, they can now be quickly and reliably determined at admission, making their use as prognostic markers feasible. For instance, the Milenia IL-6 Quickline Assay, which costs less than \$20 per test, can be performed in the emergency room or at the bedside and provides semi-quantitative results (range: <100 pg/mL, ≥100 pg/mL, ≥300 pg/mL, ≥1000 pg/mL) within 20 minutes.⁶⁵ Additionally, IL-6 could be measured in large batches using multiplex technology at a per unit cost that makes it feasible in a major clinical setting.⁶⁶

Ultimately, no matter how one evaluates a group of patients, the appearance of healthy and diseased populations will overlap to some extent, making it difficult to distinguish between the groups completely, a problem compounded by the spectrum of disease seen in pancreatitis. The accuracy of a diagnostic test is frequently expressed in terms of specificity and sensitivity, implying that every test has only one true positive and one true negative fraction. In reality, the sensitivity and specificity will change if the diagnostic criteria are altered; the ROC curve takes the variability of the criteria into account.^{34,35} In this study, the ROC curves were remarkably similar for the index and validation data sets, when either the entire data set or a subset without missing data was used. However, in creating the subset, any criterion that was missing from the data set was scored with a “0,” a strategy that likely underrepresents the severity of pancreatitis demonstrated by the Ranson and Glasgow systems.

False positives (type I errors) and false negatives (type II errors) can be expected in studies with relatively few patients. The BALI model is based on individuals who had cytokine levels drawn (the initial 450 patients enrolled in the study), and although the mortality in this subset was similar to that for the entire data set, fewer patients in the validation set died ($n = 11$) than in the Ranson, Glasgow, and APACHE II validation sets ($n = 25-49$). However, the PPV and NPV

were similar between the index and validation sets for our model. For example, for 4 of 4 criteria, the PPVs were 61% versus 65% and the NPVs were 90% versus 87%. Therefore, development of the BALI model (on the index set) was not likely adversely affected by overfitting or other data-driven procedures.

Another potential source of error in model development could have come from combining the 3 patient groups (placebo, 10 mg/24 hours, and 100 mg/24 hours Lexipafant) into one population, if their demographics, physiologic variables, or outcomes were different, as was true for the earlier Lexipafant trial.²¹ However, in developing the index data set, we compared the effects of treatment on outcome variables and did not find substantial differences among the groups. Furthermore, our attempt to develop a predictive model based on the placebo group alone, although limited by sample size, showed that once again, age ≥ 65 , LDH ≥ 300 , and IL-6 ≥ 300 were strong predictors of increased mortality.

The inclusion criteria for patient recruitment were another potential source of error in the BALI model. During the Lexipafant trial, investigators were concerned that the APACHE II score used as an inclusion criterion overestimated mortality in acute pancreatitis.⁶⁷ This score likely contributed to greater patient numbers through recruitment of some individuals who did not have pancreatitis. Furthermore, 18 of the 149 patients who died during the trial did not have pancreatitis.⁶⁸ The BALI model was based on all patients who fit the inclusion criteria for the trial and began “therapy” on an “intent-to-treat” basis. Therefore, the BALI model’s validity in predicting mortality from acute pancreatitis is limited by the 18 individuals who ultimately did not have pancreatitis. However, given the BALI model’s predictive value, eliminating these individuals from model development should strengthen the model because it would make the population of individuals who died more homogeneous. Ideally, our model would be compared with the existing systems using only individuals with severe acute pancreatitis as accurately defined by the Atlanta classification. We are now enrolling individuals in a prospective validation of the BALI model at our institution and are using it to predict severity throughout hospitalization. We think that the BALI model, with cautionary interpretation of the IL-6 concentration if available, is a valuable, efficient, and early predictor of severe acute pancreatitis.

CONCLUSION

The Ranson, Glasgow, and APACHE II systems each predicted acute pancreatitis severity, as defined by potential for death within 90 days, but our predictive model (BALI), which included elevated values for just 4 variables (BUN ≥ 25 mg/dL, Age ≥ 65 years, LDH ≥ 300 IU/L, and IL-6 ≥ 300 pg/mL) was equally capable of predicting severity at admission and throughout the initial 48 hours.

ACKNOWLEDGMENTS

The authors thank Pamela Derish for editorial assistance and Patrick Twomey for helpful discussions in preparation of the manuscript.

REFERENCES

1. Corfield AP, Cooper MJ, Williamson RC. Acute pancreatitis: a lethal disease of increasing incidence. *Gut*. 1985;26:724–729.
2. Mann DV, Hershman MJ, Hittinger R, et al. Multicentre audit of death from acute pancreatitis. *Br J Surg*. 1994;81:890–893.
3. McKay CJ, Evans S, Sinclair M, et al. High early mortality rate from acute pancreatitis in Scotland, 1984–1995. *Br J Surg*. 1999;86:1302–1305.
4. Banks PA. Practice guidelines in acute pancreatitis. *Am J Gastroenterol*. 1997;92:377–386.
5. Corfield AP, Cooper MJ, Williamson RC, et al. Prediction of severity in acute pancreatitis: prospective comparison of three prognostic indices. *Lancet*. 1985;2:403–407.
6. McMahon MJ, Playforth MJ, Pickford IR. A comparative study of methods for the prediction of severity of attacks of acute pancreatitis. *Br J Surg*. 1980;67:22–25.
7. Ranson JH, Rifkind KM, Roses DF, et al. Prognostic signs and the role of operative management in acute pancreatitis. *Surg Gynecol Obstet*. 1974;139:69–81.
8. Ranson JH, Rifkind KM, Turner JW. Prognostic signs and nonoperative peritoneal lavage in acute pancreatitis. *Surg Gynecol Obstet*. 1976;143:209–219.
9. Imrie CW, Benjamin IS, Ferguson JC, et al. A single-centre double-blind trial of Trasylol therapy in primary acute pancreatitis. *Br J Surg*. 1978;65:337–341.
10. Blamey SL, Imrie CW, O’Neill J, et al. Prognostic factors in acute pancreatitis. *Gut*. 1984;25:1340–1346.
11. Knaus WA, Zimmerman JE, Wagner DP, et al. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med*. 1981;9:591–597.
12. Knaus WA, Draper EA, Wagner DP, et al. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13:818–829.
13. Chatzicostas C, Roussomoustakaki M, Vlachonikolis IG, et al. Comparison of Ranson, APACHE II and APACHE III scoring systems in acute pancreatitis. *Pancreas*. 2002;25:331–335.
14. Wilson C, Heath DI, Imrie CW. Prediction of outcome in acute pancreatitis: a comparative study of APACHE II, clinical assessment and multiple factor scoring systems. *Br J Surg*. 1990;77:1260–1264.
15. Larvin M, McMahon MJ. APACHE-II score for assessment and monitoring of acute pancreatitis. *Lancet*. 1989;2:201–205.
16. Gudgeon AM, Heath DI, Hurley P, et al. Trypsinogen activation peptides assay in the early prediction of severity of acute pancreatitis. *Lancet*. 1990;335:4–8.
17. Neoptolemos JP, Kemppainen EA, Mayer JM, et al. Early prediction of severity in acute pancreatitis by urinary trypsinogen activation peptide: a multicentre study. *Lancet*. 2000;355:1955–1960.
18. Wilson C, Heads A, Shenkin A, et al. C-reactive protein, antiproteases and complement factors as objective markers of severity in acute pancreatitis. *Br J Surg*. 1989;76:177–181.
19. Heath DI, Cruickshank A, Gudgeon M, et al. Role of interleukin-6 in mediating the acute phase protein response and potential as an early means of severity assessment in acute pancreatitis. *Gut*. 1993;34:41–45.
20. Viedma JA, Perez-Mateo M, Agullo J, et al. Inflammatory response in the early prediction of severity in human acute pancreatitis. *Gut*. 1994;35:822–827.
21. Johnson CD, Kingsnorth AN, Imrie CW, et al. Double blind, randomised, placebo controlled study of a platelet activating factor antagonist, lexipafant, in the treatment and prevention of organ failure in predicted severe acute pancreatitis. *Gut*. 2001;48:62–69.
22. Kingsnorth AN. Early treatment with lexipafant, a platelet activating factor antagonist reduces mortality in acute pancreatitis; a double blind, randomized placebo controlled study [Abstract]. *Gastroenterology*. 1997;112:A453.
23. Kingsnorth AN, Galloway SW, Formela LJ. Randomized, double-blind phase II trial of Lexipafant, a platelet-activating factor antagonist, in human acute pancreatitis. *Br J Surg*. 1995;82:1414–1420.
24. Curran FJM SC, Young CA, et al. Controlled trial of lexipafant in severe acute pancreatitis [Abstract]. *Pancreas*. 1995;11:424.
25. Formela LJ, Wood LM, Whittaker M, et al. Amelioration of experimental acute pancreatitis with a potent platelet-activating factor antagonist. *Br J Surg*. 1994;81:1783–1785.

26. Emanuelli G, Montrucchio G, Dughera L, et al. Role of platelet activating factor in acute pancreatitis induced by lipopolysaccharides in rabbits. *Eur J Pharmacol.* 1994;261:265–272.
27. Anderson BO, Bensard DD, Harken AH. The role of platelet activating factor and its antagonists in shock, sepsis and multiple organ failure. *Surg Gynecol Obstet.* 1991;172:415–424.
28. *Lexipafant Report* (in press). Europe, 2004.
29. *World Medical Association of Declarations.* France: Ferney-Voltaire, 1964.
30. EEC note for guidance: good clinical practice for trials on medicinal products in the European Community. CPMP Working Party on Efficacy of Medicinal Products. *Pharmacol Toxicol.* 1990;67:361–372.
31. Bernard G, Doig G, Hudson L. Quantification of organ failure for clinical trials and clinical practice. *AM J Respir Crit Care Med.* 1995;151:A323.
32. Marshall JC, Cook DJ, Christou NV, et al. Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. *Crit Care Med.* 1995;23:1638–1652.
33. Bone RC, Balk RA, Cerra FB, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis: the ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest.* 1992;101:1644–1655.
34. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143:29–36.
35. Turner DA. An intuitive approach to receiver operating characteristic curve analysis. *J Nucl Med.* 1978;19:213–220.
36. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837–845.
37. Bogden JD, Bendich A, Kemp FW, et al. Daily micronutrient supplements enhance delayed-hypersensitivity skin test responses in older people. *Am J Clin Nutr.* 1994;60:437–447.
38. Gillis S, Kozak R, Durante M, et al. Immunological studies of aging: decreased production of and response to T cell growth factor by lymphocytes from aged humans. *J Clin Invest.* 1981;67:937–942.
39. Inkeles B, Innes JB, Kuntz MM, et al. Immunological studies of aging: III. Cytokinetic basis for the impaired response of lymphocytes from aged humans to plant lectins. *J Exp Med.* 1977;145:1176–1187.
40. *Clinical Laboratory Medicine.* Philadelphia: Lippincott, Williams & Williams, 2000.
41. *Clinical Diagnosis and Management by Laboratory Methods.* Philadelphia: Saunders, 2001.
42. Neoptolemos JP, Raraty M, Finch M, et al. Acute pancreatitis: the substantial human and financial costs. *Gut.* 1998;42:886–891.
43. Leese T, Shaw D. Comparison of three Glasgow multifactor prognostic scoring systems in acute pancreatitis. *Br J Surg.* 1988;75:460–462.
44. Liu TH, Kwong KL, Tamm EP, et al. Acute pancreatitis in intensive care unit patients: value of clinical and radiologic prognosticators at predicting clinical course and outcome. *Crit Care Med.* 2003;31:1026–1030.
45. Agarwal N, Pitchumoni CS. Simplified prognostic criteria in acute pancreatitis. *Pancreas.* 1986;1:69–73.
46. London NJ, Neoptolemos JP, Lavelle J, et al. Contrast-enhanced abdominal computed tomography scanning and prediction of severity of acute pancreatitis: a prospective study. *Br J Surg.* 1989;76:268–272.
47. Mayer AD, McMahon MJ, Bowen M, et al. C reactive protein: an aid to assessment and monitoring of acute pancreatitis. *J Clin Pathol.* 1984;37:207–211.
48. Steinberg WM. Predictors of severity of acute pancreatitis. *Gastroenterol Clin North Am.* 1990;19:849–861.
49. Johnson CD, Abu-Hilal M. Persistent organ failure during the first week as a marker of fatal outcome in acute pancreatitis. *Gut.* 2004;53:1340–1344.
50. Larvin M. Assessment of severity and prognosis in acute pancreatitis. *Eur J Gastroenterol Hepatol.* 1997;9:122–130.
51. Heath D, Alexander D, Wilson C, et al. Which complications of acute pancreatitis are most lethal? A prospective multi-centre clinical study of 719 episode [Abstract]. *Gut.* 1995;36:A478.
52. Saluja AK, Steer MLP. Pathophysiology of pancreatitis: role of cytokines and other mediators of inflammation. *Digestion.* 1999;60(suppl 1):27–33.
53. Tetta C, Mariano F, Buades J, et al. Relevance of platelet-activating factor in inflammation and sepsis: mechanisms and kinetics of removal in extracorporeal treatments. *Am J Kidney Dis.* 1997;30(suppl 4):57–65.
54. Puolakkainen P, Valtonen V, Paananen A, et al. C-reactive protein (CRP) and serum phospholipase A2 in the assessment of the severity of acute pancreatitis. *Gut.* 1987;28:764–771.
55. Galloway SW, Kingsnorth AN. Reduction in circulating levels of CD4-positive lymphocytes in acute pancreatitis: relationship to endotoxin, interleukin 6 and disease severity. *Br J Surg.* 1994;81:312.
56. Pezzilli R, Billi P, Miniero R, et al. Serum interleukin-6, interleukin-8, and beta 2-microglobulin in early assessment of severity of acute pancreatitis. Comparison with serum C-reactive protein. *Dig Dis Sci.* 1995;40:2341–2348.
57. Inagaki T, Hoshino M, Hayakawa T, et al. Interleukin-6 is a useful marker for early prediction of the severity of acute pancreatitis. *Pancreas.* 1997;14:1–8.
58. Viedma JA, Perez-Mateo M, Dominguez JE, et al. Role of interleukin-6 in acute pancreatitis: comparison with C-reactive protein and phospholipase A. *Gut.* 1992;33:1264–1267.
59. Leser HG, Gross V, Scheibenbogen C, et al. Elevation of serum interleukin-6 concentration precedes acute-phase response and reflects severity in acute pancreatitis. *Gastroenterology.* 1991;101:782–785.
60. Pooran N, Indaram A, Singh P, et al. Cytokines (IL-6, IL-8, TNF): early and reliable predictors of severe acute pancreatitis. *J Clin Gastroenterol.* 2003;37:263–266.
61. Engelberts I, Stephens S, Francot GJ, et al. Evidence for different effects of soluble TNF-receptors on various TNF measurements in human biological fluids. *Lancet.* 1991;338:515–516.
62. Grewal HP, Mohey el Din A, Gaber L, et al. Amelioration of the physiologic and biochemical changes of acute pancreatitis using an anti-TNF-alpha polyclonal antibody. *Am J Surg.* 1994;167:214–218; discussion 218–219.
63. McLaughlin PJ, Davies HM, Aikawa A. Tumour-necrosis factor in normal plasma. *Lancet.* 1990;336:1014–1015.
64. van Kessel KP, van Strijp JA, Verhoef J. Inactivation of recombinant human tumor necrosis factor-alpha by proteolytic enzymes released from stimulated human neutrophils. *J Immunol.* 1991;147:3862–3868.
65. Galetto-Lacour A, Zamora SA, Gervaix A. Bedside procalcitonin and C-reactive protein tests in children with fever without localizing signs of infection seen in a referral center. *Pediatrics.* 2003;112:1054–1060.
66. Ray CA, Bowsher RR, Smith WC, et al. Development, validation, and implementation of a multiplex immunoassay for the simultaneous determination of five cytokines in human serum. *J Pharm Biomed Anal.* 2005;36:1037–1044.
67. Zyromski N, Murr MM. Evolving concepts in the pathophysiology of acute pancreatitis. *Surgery.* 2003;133:235–237.
68. McKay CJ, Imrie CW. The continuing challenge of early mortality in acute pancreatitis. *Br J Surg.* 2004;91:1243–1244.