



The early prediction of mortality in acute pancreatitis: a large population-based study

B U Wu,¹ R S Johannes,^{1,2} X Sun,² Y Tabak,² D L Conwell,¹ P A Banks¹

See Commentary, p 1645

¹ Brigham and Women's Hospital, Division of Gastroenterology, Center for Pancreatic Disease, Harvard Medical School, Boston Massachusetts, USA; ² Cardinal Health, Marlborough, Massachusetts, USA

Correspondence to: Dr B U Wu, Division of Gastroenterology, Endoscopy Suite, Brigham & Women's Hospital, 75 Francis Street, Boston, MA 02115, USA; buwu@partners.org

Revised 18 April 2008
Accepted 13 May 2008
Published Online First
2 June 2008

ABSTRACT

Background: Identification of patients at risk for mortality early in the course of acute pancreatitis (AP) is an important step in improving outcome.

Methods: Using Classification and Regression Tree (CART) analysis, a clinical scoring system was developed for prediction of in-hospital mortality in AP. The scoring system was derived on data collected from 17 992 cases of AP from 212 hospitals in 2000–2001. The new scoring system was validated on data collected from 18 256 AP cases from 177 hospitals in 2004–2005. The accuracy of the scoring system for prediction of mortality was measured by the area under the receiver operating characteristic curve (AUC). The performance of the new scoring system was further validated by comparing its predictive accuracy with that of Acute Physiology and Chronic Health Examination (APACHE) II.

Results: CART analysis identified five variables for prediction of in-hospital mortality. One point is assigned for the presence of each of the following during the first 24 h: blood urea nitrogen (BUN) >25 mg/dl; impaired mental status; systemic inflammatory response syndrome (SIRS); age >60 years; or the presence of a pleural effusion (BISAP). Mortality ranged from >20% in the highest risk group to <1% in the lowest risk group. In the validation cohort, the BISAP AUC was 0.82 (95% CI 0.79 to 0.84) versus APACHE II AUC of 0.83 (95% CI 0.80 to 0.85).

Conclusions: A new mortality-based prognostic scoring system for use in AP has been derived and validated. The BISAP is a simple and accurate method for the early identification of patients at increased risk for in-hospital mortality.

Acute pancreatitis (AP) is a disease with a substantial burden on the US healthcare system. Recent data indicate a rise in the absolute number as well as the rate of emergency room visits, hospital admissions and direct healthcare costs for AP in the USA (210 000 admissions in 2002; 4.6 of every 1000 hospitalisations from 1988 to 2003, annual direct costs in excess of US\$2 billion).^{1–3} With an overall mortality rate of 2–5%,^{3–4} a reliable method of risk stratification for AP is of significant clinical importance.

Current methods of risk stratification in AP have important limitations. The Ranson⁵ and modified Glasgow score⁶ contain data not routinely collected at the time of hospitalisation. In addition, both require 48 h to complete, missing a potentially valuable early therapeutic window.^{5–7} The most commonly utilised prediction scoring system for clinical research studies in AP is the Acute Physiology and Chronic Health Examination (APACHE) II.^{8–9} However, the APACHE II was

originally developed as an intensive care instrument and requires the collection of a large number of parameters, some of which may not be relevant to prognosis in AP.

The purpose of this study was to develop a simple and accurate clinical scoring system for stratifying patients according to their risk of in-hospital mortality. To develop a clinical tool useful early in the disease course, we examined data collected within the first 24 h of hospitalisation. We used data collected from a large population-based cohort study in both the derivation and validation of the scoring system. For further validation, we compared the accuracy of the new scoring system with that of the APACHE II for prediction of mortality.

METHODS

Patient population and data collection

The current study was approved by the Brigham & Women's Hospital Institutional Review Board. Patient data were generated from the Cardinal Health Clinical Outcomes Research Database (Cardinal Health Clinical Research Services, Cardinal Health, Marlborough, Massachusetts, USA). This large population data set has supported public reporting of hospital performance in Pennsylvania and elsewhere for purposes of quality improvement for >20 years. Details of the data collection and abstraction process for the Cardinal database have been described previously.^{10–13} The database contains information on patient demographics, vital signs, laboratory values, co-morbidities, physical exam findings as well as procedure and diagnosis codes. Unlike previous versions of the database, all laboratory and vital sign data are now recorded as continuous values.

The derivation cohort consisted of all cases in the Cardinal Health Research Database with principal diagnosis (from the International Classification of Diseases, ninth revision, clinical modification) ICD9-CM 577.0 (AP) from January 2000 to December 2001. The validation cohort included all patients with the principal diagnosis of AP admitted from January 2004 to September 2005.

Assessment of risk factors for mortality in AP

We considered the following candidate risk factors in model development:

- ▶ Individual Ranson signs⁵: age, white blood cell (WBC) count, glucose, aspartate aminotransferase (AST), blood urea nitrogen (BUN), lactate dehydrogenase (LDH) and serum calcium.

- ▶ Pleural effusion^{14 15} (on chest radiography or CT).
- ▶ The systemic inflammatory response syndrome (SIRS)^{16 17} defined by the presence of ≥ 2 of the following criteria:
 - pulse >90 beats/min
 - respirations >20 /min or $\text{PaCO}_2 <32$ mmHg
 - temperature $>38^\circ\text{C}$ or $<36^\circ\text{C}$
 - WBC count $>12\,000$ or <4000 cells/mm³ or $>10\%$ immature neutrophils (bands)
- ▶ Haemoconcentration (haemoglobin was included as a continuous variable)¹⁸
- ▶ Atlanta Symposium criteria for organ failure¹⁹: systolic blood pressure, creatinine, partial pressure of arterial oxygen (PaO_2).
- ▶ Altered mental status²⁰: defined as any record of disorientation, lethargy somnolence, coma or stupor in the medical record.

In order to develop a model with widespread applicability, we limited potential laboratory and vital sign parameters to those with collection rates of $\geq 85\%$. SIRS was collected as a dichotomous composite variable. Physical exam findings and co-morbidities were recorded at admission. Data for laboratory tests and vital signs were recorded for the first 24 h admission period including emergency department values. The worst (most extreme) value for vital signs and laboratory data within the first 24 h of admission were utilised for model development. All laboratory and vital sign parameters were included as continuous variables with thresholds determined by Classification and Regression Tree (CART) analysis.

Model derivation

We used CART analysis to identify factors for use in the new clinical prediction rule. CART is a non-parametric, empiric statistical method²¹ that has been increasingly utilised for clinical applications across a number of disease groups^{22–24} but not as yet for clinical prediction in AP. Patients are classified into two groups at each stage of analysis based on classification variables. The optimum split point for each variable is determined by a statistical search algorithm. Patients are grouped into nodes by cut-off points for classification variables. The process is reiterated for subsequent classification variables. Tree building is carried forward until a pruning process determines the optimum tree size without overfitting the data.

In order to avoid model overtraining, we used 10-fold cross-validation in tree development. In addition, we specified a 10-fold misclassification cost such that misclassifying a true death was 10 times worse than misclassification of a case that ultimately survived. We used the Gini index as the splitting rule in tree building. Missing data were incorporated into the tree building process through use of surrogate splits.

A new prediction rule was subsequently generated from the parameters identified in the CART analysis. Specifically, a scoring system was created in which one point was assigned for the presence of each parameter identified in the CART prediction algorithm. We then calculated scores for each case in the derivation cohort and compared this with their observed outcome.

Model validation

To validate performance of the new scoring system on a separate group of patients, we tested its ability to predict in-hospital mortality in the validation cohort. We calculated scores for each case and compared this with observed mortality. From this analysis we calculated the area under the receiver operating characteristic curve (AUC) as a measure of discrimination

accuracy. In order to assess model calibration, we compared observed mortality by point score in both the derivation and validation cohorts. In testing the new scoring system on the validation cohort we treated missing data as normal (reference range) values.

To validate the model further, we compared its performance with that of the APACHE II. We anticipated that a large number of patients would not have complete data for calculation of an APACHE II score. Therefore, in generating APACHE II scores, we treated missing data as normal (reference range) values. Comparison of model AUC with the APACHE II was performed using the method described by De Long *et al.*²⁵

Subgroup analysis

The ability to identify patients at increased risk of mortality from AP prior to the onset of overt organ failure is of significant clinical importance. Patients who present with evidence of organ failure within the first 24 h of hospitalisation have already declared themselves as being at increased risk for experiencing persistent organ failure and death.²⁶ Numerous organ failure scoring systems exist for use in critical care settings that measure the extent of organ failure. An important application of a new scoring system in AP is to identify patients at risk for mortality prior to the onset of organ failure. We were interested in determining how well the new scoring system could predict mortality among patients without evidence of early organ failure during the first 24 h of hospitalisation. Therefore, we performed a subgroup analysis in which we applied the new prediction rule exclusively to patients without evidence of early organ failure by Atlanta criteria (creatinine >2.0 mg/dl, systolic blood pressure <90 mmHg or $\text{PaO}_2 <60$ mmHg on arterial blood gas).

Statistical analysis

CART analysis was performed using the CART statistical software package (CART Professional Extended Edition version 6.0, Salford Systems, California Statistical Software, San Diego, California, USA). Additional statistical analysis was performed in SAS statistical software version 9.1 (SAS Institute, Cary, North Carolina, USA). All reported p values are two sided. We used the Bonferroni method of adjustment for multiple testing when examining differences in mortality between risk groups.

RESULTS

Patient characteristics

In the derivation cohort, there were 17 922 cases of AP identified from 212 hospitals. Median age was 53 years, and 50.5% were men. In the validation cohort there were 18 256 cases of AP identified from 177 hospitals. Median age was 53 years, and 49.4% were men.

There were a total of 335 (1.9%) deaths in the derivation cohort and 234 deaths (1.28%) in the validation cohort. There was a significant reduction in overall mortality between 2000–2001 and 2004–2005 (χ^2 $p < 0.001$). Distributions for demographic and clinical features between the two study populations are depicted in table 1. The serum calcium, LDH, PaO_2 and AST measurements were excluded from further consideration in developing the new scoring system due to their failure to meet the pre-specified 85% collection rate threshold.

CART analysis

Using CART analysis we identified five variables as most efficient in stratifying patients according to risk of mortality:

Table 1 Demographic and clinical characteristics of the two study populations

	Derivation cohort January 2000–December 2001	Validation cohort January 2004–September 2005
Hospital demographics		
Hospitals	212	177
Teaching hospitals	49.9%	39.7%
Bed size		
<100 beds	10%	27%
100–300 beds	58%	50%
>300 beds	32%	23%
Patient demographics and co-morbidity		
Acute pancreatitis cases (n)	17922	18256
Median age in years (range)	53 (18–106)	53 (18–108)
Men	50.5%	49.4%
Gallstones	25.4%	23.8%
Alcohol history (dependency or abuse)	22.4%	21.1%
Chronic pancreatitis	7.0%	8.5%
Congestive heart failure	7.04%	6.5%
Chronic kidney disease	5.1%	5.7%
Chronic obstructive pulmonary disease	12.0%	10.4%
Clinical parameters		
SIRS present*	48.96%	50.3%
WBC count $\times 10^3$ (median, IQR)	10.2 (7.5–13.7)	9.9 (7.3–13.3)
Temperature, °C (median, IQR)	36.1 (35.6 to 36.7)	36.4 (36.1–37.0)
Pulse, beats/min (median, IQR)	70 (60–100)	85 (70–100)
Respirations per minute (median, IQR)	10 (10–20)	18 (10–20)
Systolic blood pressure, mmHg (median, IQR)*	140 (130–160)	134 (118–160)
Haemoglobin, mg/dl (median, IQR)*	13.8 (12.4–15.1)	13.8 (12.4–15.1)
Blood urea nitrogen, mg/dl (median, IQR)*	15 (10–21)	14 (10–21)
Creatinine, mg/dl (median, IQR)*	0.9 (0.8–1.2)	0.9 (0.8–1.2)
Glucose, mg/dl (median, IQR)*	128 (105–170)	123 (102–163)
Pleural effusion*	4.1%	4.4%
Altered mental status*	7.4%	6.7%

Proportions listed in this table reflect the number of patients with each finding divided by the total number of patients in each cohort. Vital signs and laboratory test values were the most extreme within the first 24 h of admission.

*Candidate variables entered into CART analysis.

CART, Classification and Regression Tree; IQR, interquartile range; SIRS, systemic inflammatory response syndrome; WBC, white blood cell.

BUN >25 mg/dl, impaired mental status, SIRS, age >60 years and pleural effusion (BISAP). The CART tree is depicted in fig 1. BUN was identified as the most efficient first splitting variable. Age and SIRS further discriminated between high- and low-risk cases. The remaining parameters (mental status and pleural effusion) served to differentiate intermediate risk patients further.

Validation of scoring system

The five variables from the CART were incorporated into a new scoring system in which the presence of each variable contributes one point to an overall 5-point score. After calculating scores for patients in the derivation cohort, the BISAP score AUC was 0.83 for prediction of in-hospital mortality.

In the validation cohort there were 17 350 (96.8%) cases with complete laboratory and vital sign data for the five parameters included in the scoring system. There were 213 (1.2%) deaths among these patients. After calculating BISAP scores for patients in the validation cohort, the AUC for prediction of in-hospital mortality was 0.82.

Table 2 depicts observed mortality stratified by BISAP point score in both cohorts of patients. Also depicted in table 2 is the frequency of patients within each score category. Using the 5-point scoring system, patients could be reliably classified within 24 h of admission into distinct risk groups for mortality. There

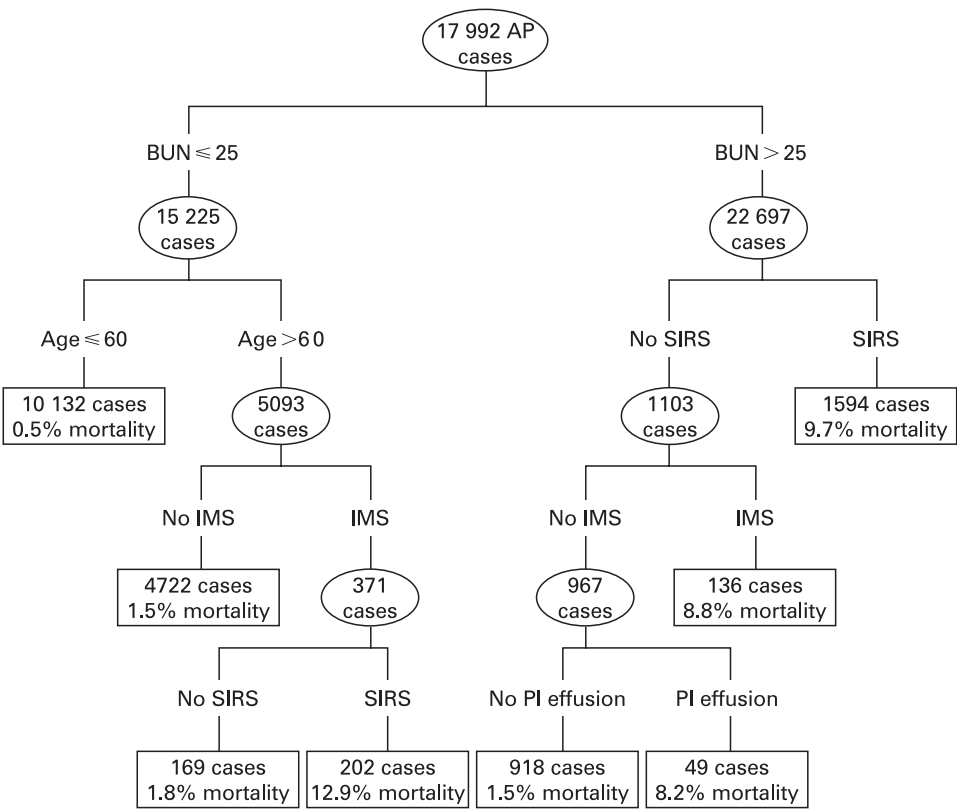
was a significant trend towards higher mortality with increasing BISAP score (Cochrane–Armitage trend test $p < 0.001$). In addition, significant differences existed between risk groups (χ^2 $p < 0.001$ overall, pairwise χ^2 Bonferroni-adjusted $p < 0.001$). Below average mortality was observed in patients with <2 points (<1.0% mortality). Patients with a score of 2 had increased mortality (2%). Mortality continued to rise sharply with BISAP scores of ≥ 3 (5–20%).

In both the derivation and validation cohort, the majority of patients (~60%) presented with BISAP scores <2 and were at very low risk for mortality (<1.0%). The new scoring system was able to identify subgroups of patients (those with scores of ≥ 3) with substantially increased risk of dying in the course of their hospitalisation.

APACHE II

There were 405 (2.2%) patients with complete data for APACHE II. There were 40 (9.9%) deaths among these patients. After imputation, APACHE II scores were calculated for each patient as previously described. The median calculated APACHE II score for the 2004–2005 AP population was 7. The receiver operating characteristic (ROC) curve for the APACHE II score is shown in fig 2. Calculated AUC was 0.83 (95% CI 0.80 to 0.85). For purposes of comparison, a similar ROC curve was plotted for the BISAP score on the 2004–2005 population. The BISAP AUC was 0.82 (95% CI 0.79 to 0.84) in the validation cohort.

Figure 1 Classification and Regression Tree analysis of risk factors for mortality in the derivation cohort. Mortality rates listed are crude (unadjusted) values. AP, acute pancreatitis; BUN, blood urea nitrogen; IMS, impaired mental status; PI effusion, pleural effusion; SIRS, systemic inflammatory response syndrome.



Significance testing for differences in AUC between BISAP and APACHE II yielded a χ^2 p value of 0.2, indicating no significant difference in predictive accuracy.

Subgroup analysis (classification prior to onset of organ failure)
There were 1753 patients with early organ failure in the validation cohort. Among these patients there were 98 (5.5%) deaths. After we excluded cases with early organ failure, there were 16 503 cases remaining in the subgroup analysis. Among these cases there were 136 (0.8%) deaths. Fifty-eight percent of patients that died did not have evidence of organ failure by Atlanta criteria within the first 24 h of hospitalisation. Table 3 depicts the observed mortality by BISAP score among patients without evidence of early organ failure. The scoring system was once again able reliably to identify patients at increased risk of mortality in this subgroup analysis. The model's AUC for prediction of in-hospital mortality in the subgroup analysis was 0.79 versus APACHE II AUC of 0.78 ($\chi^2 = 0.45$, $p = 0.5$).

DISCUSSION
We have derived and validated the first population-based prognostic scoring system for use in AP. Using BUN, impaired mental status, SIRS, age and pleural effusion (BISAP), we were able to stratify patients within the first 24 h of hospitalisation into distinct risk groups for in-hospital mortality.
In the subgroup analysis we examined the ability of the new prediction rule to identify patients at increased risk of mortality prior to the onset of organ failure. Specifically, we excluded patients with evidence of early organ failure by Atlanta criteria (within the first 24 h). Although patients without early organ failure had a low mortality (0.8%), 58% of the patients that ultimately died came from this subgroup. Among these patients, the prediction rule was still able to identify patients with substantially increased mortality.
The ability to risk-stratify patients early in their disease course has several important implications. First, early identification of high-risk patients may alert doctors to institute aggressive resuscitation efforts and to consider specialty care

Table 2 Observed mortality by BISAP point score in the derivation and validation cohorts

BISAP score	Derivation cohort, n = 17 922		Validation cohort, n = 18 256	
	Number of cases	Observed mortality	Number of cases	Observed mortality
0	5121	0.2%	4912	0.1%
1	7206	0.7%	7722	0.5%
2	3829	2.1%	3941	1.9%
3	1390	8.3%	1349	5.3%
4	331	19.3%	292	12.7%
5	45	26.7%	40	22.5%

Overall χ^2 $p < 0.001$ for each study group. Within study group χ^2 $p < 0.001$ for differences in mortality between individual risk (score) categories in pairwise tests of significance.
BISAP, blood urea nitrogen, impaired mental status, systemic inflammatory response syndrome, age and pleural effusion.

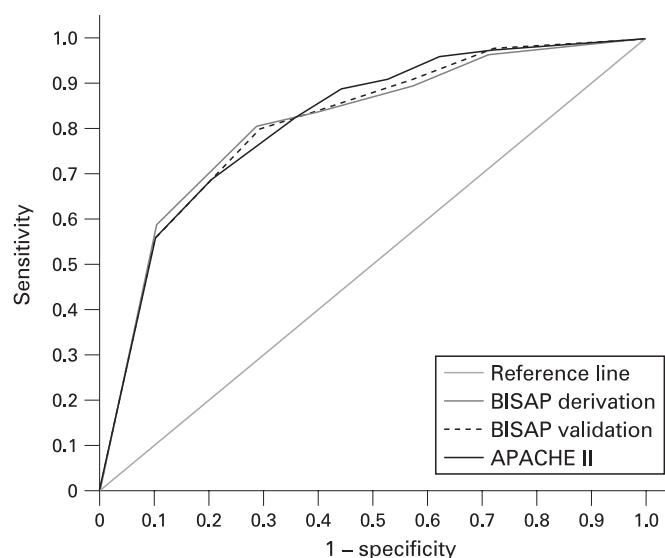


Figure 2 Receiver operator characteristic curves. BISAP (blood urea nitrogen, impaired mental status, systemic inflammatory response syndrome, age and pleural effusion.) derivation (area under the curve (AUC) = 0.83). BISAP validation (AUC = 0.82). Acute Physiology and Chronic Health Examination (APACHE) II (AUC = 0.83).

referral. Second, a severity index provides standardised criteria for enrolment of subjects into future clinical studies. In addition, a population-based system of risk stratification provides an instrument for additional outcomes research. For example, identification of factors associated with death among patients with low BISAP scores may help to lead to improvements in future management strategies in AP.

The primary advantage of BISAP is simplicity. The presence of each variable contributes one point to a total 5-point score. There is no need for additional computation. In addition, each of the parameters can be easily obtained early in the course of a general hospital admission. The only subjective parameter in the new scoring system is the assessment of mental status. Although an uncommon finding, the presence of an altered mental state was a significant predictor of mortality in both populations. Although the Glasgow Coma Score is used as part of the calculation of an APACHE II score as well as the Multiple Organ Failure Score,^{20,26} we simplified determination of this parameter by developing the model in such a way that any evidence of disorientation or further disturbance in mental status qualifies as a positive finding. Although SIRS is a composite parameter that involves the use of four criteria, evaluation of the systemic inflammatory response has become increasingly widespread in clinical practice and has also been demonstrated to have prognostic value in AP.^{16,17}

The use of population-based data in this study has several advantages. First, the large number of cases provided sufficient power to focus on mortality as an end point. Second, the derivation and validation of this new scoring system utilised data collected from a large number of hospitals. Patient information was collected from community hospitals, tertiary care centres, teaching and non-teaching institutions. As a result, the performance of BISAP in our study reflects the combined experience from a variety of treatment settings.

While patient characteristics were similar between the two study populations, the validation cohort differed from the derivation cohort in several important aspects. These differences included a shift in hospital demographics and a decreased overall

Table 3 Subgroup analysis of the validation cohort excluding cases with evidence of organ failure within first 24 h of hospitalisation

BISAP score	n = 16503	
	Number of cases	Observed mortality
0	4796	0.1%
1	7287	0.4%
2	3307	1.6%
3	916	3.6%
4	176	7.4%
5	21	9.5%

Overall and pairwise χ^2 $p < 0.001$.

BISAP, blood urea nitrogen, impaired mental status, systemic inflammatory response syndrome, age and pleural effusion.

mortality (1.3% validation cohort vs 1.9% derivation cohort, t test $p < 0.0001$). We validated the new scoring system in the more recent population of patients who may have benefited from improvements in critical care and the management of severe AP, including changes in the management of sterile as well as infected necrosis.⁴ These changes may have contributed to the reduced mortality observed among patients with higher BISAP scores in the validation cohort.

The early identification of patients at risk for adverse outcome from AP has been an area of active investigation for many years.^{16,28–39} Previous studies have attempted either to develop prognostic scoring systems or to identify individual risk factors for severe disease. Some of these studies have included mortality as an end point.^{16,28,30–32,35–37} Among recently proposed prognostic scoring systems, three have used data collected within the first 24 h of hospitalisation.^{28,31,40} Because all of these scoring systems were based on data from high-risk patient populations, their ability to predict in-hospital mortality among patients with varying disease severity is unknown.

Among studies of individual or combinations of risk factors in AP, several have focused on the first 24 h admission period.^{32,35,37,39,41,42} These smaller cohort studies identified age,³⁵ obesity,^{30,32} glucose,^{37,42} serum creatinine,³⁷ BUN⁴² and organ failure^{41,43} as admission parameters associated with increased mortality. Of these parameters, we were able to evaluate age, glucose, BUN and organ failure (in terms of hypotension, elevated creatinine and hypoxia) within the first 24 h.

To evaluate the performance of the new prediction rule further, we compared its predictive accuracy with that of APACHE II. Although more recent versions of the APACHE system have been developed, the APACHE II remains the most widely accepted method for risk stratification in AP^{4,8,9,44,45} and was therefore chosen as the reference standard. Its major limitations are complexity and reliance upon parameters not routinely collected during a general hospital admission. For example, in our validation cohort, only 2.2% of cases had complete APACHE II data versus 96.8% for the laboratory and vital sign parameters included in the BISAP score. Moreover, cases with a complete APACHE II score had a mortality rate of 9.9%, which was eight times greater than that of the general population. This increased mortality most probably reflects a form of selection bias wherein more severely ill patients are the ones most likely to have the exhaustive data collection required for completion of an APACHE II score. Nevertheless, the BISAP score was able to achieve a similar level of predictive accuracy to the more complex APACHE II score, with far fewer variables.

There were several potential limitations to the present study. We relied upon ICD-9 data for diagnosis rather than the more

strict Atlanta Symposium criteria.¹⁹ As a result, milder causes of abdominal pain may have been misclassified as AP. Nevertheless, observed mortality in both cohorts of patients was consistent with data from recent studies examining trends in AP.^{3, 4, 46, 47} A second limitation was limited information regarding aetiology, obesity or initial versus recurrent episode of AP, each of which may have prognostic value in AP.^{30, 32, 48, 49}

In summary, we have derived and validated a prognostic scoring system for use in AP. The BISAP score stratifies patients within the first 24 h of admission according to their risk of in-hospital mortality and was able to identify patients at increased risk of mortality prior to the onset of organ failure. The ability to risk-stratify patients early in their course is a major step to improving future management strategies in acute pancreatitis.

Acknowledgements: The authors would like to thank Linda Hyde, Karen Derby and Stephen Kurtz of Cardinal Health for their assistance with data management. In addition, we would like to thank Dr Earl F Cook of the Harvard School of Public Health for assistance with study design and data analysis.

Competing interests: None.

Ethics approval: The study was approved by the Brigham & Women's Hospital Institutional Review Board.

REFERENCES

- Fagenholz PJ, Fernandez-del Castillo C, Harris NS, et al. Direct medical costs of acute pancreatitis hospitalizations in the United States. *Pancreas* 2007;**35**:302–7.
- Fagenholz PJ, Fernandez-del Castillo C, Harris NS, et al. National study of United States emergency department visits for acute pancreatitis, 1993–2003. *BMC Emerg Med* 2007;**7**:1.
- Fagenholz PJ, Castillo CF, Harris NS, et al. Increasing United States hospital admissions for acute pancreatitis, 1988–2003. *Ann Epidemiol* 2007;**17**:491–7.
- Banks PA, Freeman ML. Practice guidelines in acute pancreatitis. *Am J Gastroenterol* 2006;**101**:2379–400.
- Ranson JH, Rifkind KM, Roses DF, et al. Objective early identification of severe acute pancreatitis. *Am J Gastroenterol* 1974;**61**:443–51.
- Blamey SL, Imrie CW, O'Neill J, et al. Prognostic factors in acute pancreatitis. *Gut* 1984;**25**:1340–6.
- Ranson JH, Pasternack BS. Statistical methods for quantifying the severity of clinical acute pancreatitis. *J Surg Res* 1977;**22**:79–91.
- Yeung YP, Lam BY, Yip AW. APACHE system is better than Ranson system in the prediction of severity of acute pancreatitis. *Hepatobiliary Pancreat Dis Int* 2006;**5**:294–9.
- Larvin M, McMahon MJ. APACHE-II score for assessment and monitoring of acute pancreatitis. *Lancet* 1989;**2**:201–5.
- Iezzoni LI, Hotchkiss EK, Ash AS, et al. MedisGroups data bases. The impact of data collection guidelines on predicting in-hospital mortality. *Med Care* 1993;**31**:277–83.
- Tabak YP, Johannes RS, Silber JH. Using automated clinical data for risk adjustment: development and validation of six disease-specific mortality predictive models for pay-for-performance. *Med Care* 2007;**45**:789–805.
- Iezzoni LI, Moskowitz MA. A clinical assessment of MedisGroups. *JAMA* 1988;**260**:3159–63.
- Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997;**336**:243–50.
- Heller SJ, Noordhoek E, Tenner SM, et al. Pleural effusion as a predictor of severity in acute pancreatitis. *Pancreas* 1997;**15**:222–5.
- Belfar HL, Radecki PD, Friedman AC, et al. Pancreatitis presenting as pleural effusions: computed tomography demonstration of pleural space extension of pancreatitis exudate. *J Comput Tomogr* 1987;**11**:184–7.
- Mofidi R, Duff MD, Wigmore SJ, et al. Association between early systemic inflammatory response, severity of multiorgan dysfunction and death in acute pancreatitis. *Br J Surg* 2006;**93**:738–44.
- Buter A, Imrie CW, Carter CR, et al. Dynamic nature of early organ dysfunction determines outcome in acute pancreatitis. *Br J Surg* 2002;**89**:298–302.
- Brown A, Orav J, Banks PA. Hemoconcentration is an early marker for organ failure and necrotizing pancreatitis. *Pancreas* 2000;**20**:367–72.
- Bradley EL 3rd. A clinically based classification system for acute pancreatitis. Summary of the International Symposium on Acute Pancreatitis, Atlanta, Ga, September 11 through 13, 1992. *Arch Surg* 1993;**128**:586–90.
- Tran DD, Cuesta MA. Evaluation of severity in patients with acute pancreatitis. *Am J Gastroenterol* 1992;**87**:604–8.
- Breiman L. *Classification and regression trees*. Belmont, California: Wadsworth International Group, 1984.
- Fonarow GC, Adams KF Jr, Abraham WT, et al. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *JAMA* 2005;**293**:572–80.
- Miller PD, Barlas S, Brennenman SK, et al. An approach to identifying osteopenic women at increased short-term risk of fracture. *Arch Intern Med* 2004;**164**:1113–20.
- Takahashi O, Cook EF, Nakamura T, et al. Risk stratification for in-hospital mortality in spontaneous intracerebral haemorrhage: a Classification and Regression Tree analysis. *QJM* 2006;**99**:743–50.
- 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**:837–45.
- Johnson CD, Abu-Hilal M. Persistent organ failure during the first week as a marker of fatal outcome in acute pancreatitis. *Gut* 2004;**53**:1340–4.
- Knaus WA, Draper EA, Wagner DP, et al. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;**13**:818–29.
- Ueda T, Takeyama Y, Yasuda T, et al. Simple scoring system for the prediction of the prognosis of severe acute pancreatitis. *Surgery* 2007;**141**:51–8.
- Rau BM, Bothe A, Kron M, et al. Role of early multisystem organ failure as major risk factor for pancreatic infections and death in severe acute pancreatitis. *Clin Gastroenterol Hepatol* 2006;**4**:1053–61.
- Papachristou GI, Papachristou DJ, Avula H, et al. Obesity increases the severity of acute pancreatitis: performance of APACHE-O score and correlation with the inflammatory response. *Pancreatol* 2006;**6**:279–85.
- Spitzer AL, Barcia AM, Schell MT, et al. Applying Ockham's razor to pancreatitis prognostication: a four-variable predictive model. *Ann Surg* 2006;**243**:380–8.
- Martinez J, Johnson CD, Sanchez-Paya J, et al. Obesity is a definitive risk factor of severity and mortality in acute pancreatitis: an updated meta-analysis. *Pancreatol* 2006;**6**:206–9.
- Mentula P, Kylanpaa ML, Kemppainen E, et al. Early prediction of organ failure by combined markers in patients with acute pancreatitis. *Br J Surg* 2005;**92**:68–75.
- Gan SI, Romagnuolo J. Admission hematocrit: a simple, useful and early predictor of severe pancreatitis. *Dig Dis Sci* 2004;**49**:1946–52.
- Company L, Saez J, Martinez J, et al. Factors predicting mortality in severe acute pancreatitis. *Pancreatol* 2003;**3**:144–8.
- Halonen KI, Leppaniemi AK, Lundin JE, et al. Predicting fatal outcome in the early phase of severe acute pancreatitis by using novel prognostic models. *Pancreatol* 2003;**3**:309–15.
- Blum T, Maisonneuve P, Lowenfels AB, et al. Fatal outcome in acute pancreatitis: its occurrence and early prediction. *Pancreatol* 2001;**1**:237–41.
- Losanoff JE, Asparouhov OK, Jones JW. Multiple factor scoring system for risk assessment of acute pancreatitis. *J Surg Res* 2001;**101**:73–8.
- Halonen KI, Leppaniemi AK, Puolakkainen PA, et al. Severe acute pancreatitis: prognostic factors in 270 consecutive patients. *Pancreas* 2000;**21**:266–71.
- Harrison DA, D'Amico G, Singer M. The Pancreatitis Outcome Prediction (POP) Score: a new prognostic index for patients with severe acute pancreatitis. *Crit Care Med* 2007;**35**:1703–8.
- Isenmann R, Rau B, Beger HG. Early severe acute pancreatitis: characteristics of a new subgroup. *Pancreas* 2001;**22**:274–8.
- Fan ST, Choi TK, Lai EC, et al. Prediction of severity of acute pancreatitis: an alternative approach. *Gut* 1989;**30**:1591–5.
- Perez A, Whang EE, Brooks DC, et al. Is severity of necrotizing pancreatitis increased in extended necrosis and infected necrosis? *Pancreas* 2002;**25**:229–33.
- Taylor SL, Morgan DL, Denson KD, et al. A comparison of the Ranson, Glasgow, and APACHE II scoring systems to a multiple organ system score in predicting patient outcome in pancreatitis. *Am J Surg* 2005;**189**:219–22.
- Robert JH, Frossard JL, Mermillod B, et al. Early prediction of acute pancreatitis: prospective study comparing computed tomography scans, Ranson, Glasgow, Acute Physiology and Chronic Health Evaluation II scores, and various serum markers. *World J Surg* 2002;**26**:612–9.
- Yadav D, Lowenfels AB. Trends in the epidemiology of the first attack of acute pancreatitis: a systematic review. *Pancreas* 2006;**33**:323–30.
- Frey CF, Zhou H, Harvey DJ, et al. The incidence and case-fatality rates of acute biliary, alcoholic, and idiopathic pancreatitis in California, 1994–2001. *Pancreas* 2006;**33**:336–44.
- Pezzilli R, Billi P, Morselli-Labate AM. Severity of acute pancreatitis: relationship with etiology, sex and age. *Hepato-gastroenterology* 1998;**45**:1859–64.
- Lankisch PG, Assmus C, Pflüthofer D, et al. Which etiology causes the most severe acute pancreatitis? *Int J Pancreatol* 1999;**26**:55–7.