Use of Modified Balthazar Grades for the Early Prediction of Acute Pancreatitis Severity in the Emergency Department

Tao Cheng¹,²,*; Tian-Yong Han¹,²,*; Bo-Fu Liu¹,²; Pan Pan¹,²; Qiang Lai¹,²; Haifang Yu¹,²; Yu Cao¹,²

¹Emergency Department, West China Hospital of Sichuan University, Chengdu, Sichuan, People’s Republic of China; ²Laboratory of Emergency Medicine, West China Hospital of Sichuan University, Chengdu, Sichuan, People’s Republic of China

*These authors contributed equally to this work

Correspondence: Yu Cao, Emergency Department of West China Hospital and Laboratory of Emergency Medicine of West China Hospital, Sichuan University, 37 Guoxue Road, Chengdu, 610041, Sichuan, People’s Republic of China, Tel +86-28-85422288, Email caoyu@wchscu.cn

Objective: To develop a new approach based on Balthazar grades of acute pancreatitis (AP) and to assess this modified method as a tool for the early prediction of AP severity in the emergency department (ED).

Methods: Data pertaining to AP patients ≥18 years old that had undergone computed tomography (CT) scanning within 24 h following ED admission between January 1, 2017 and September 30, 2017 were retrospectively analyzed. Patients were separated into two groups based on the length of time between the onset of their AP symptoms and the completion of CT scanning (Group 1: <72 h; Group 2: ≥72 h). Modified Balthazar grades for these patients were then assessed, with the concordance between these modified grades and the 2012 revised Atlanta classification being assessed using the Kappa (κ) statistic. The modified grade with the largest κ value was evaluated based on performance traits including Harrell’s concordance index (C-index), area under the receiver operating characteristic curve (AUC) analyses, calibration curves, and decision curve analyses (DCA) in comparison with bedside index for severity in AP (BISAP) scores.

Results: In total, 372 patients were included in the present analysis. These patients were regraded according to six methods, with the method yielding the largest κ value consisting of regraded Balthazar grades A–C, D, and E, respectively, corresponding to mild, moderate, and severe AP. The κ values for this method were 0.786 (95% CI, 0.706–0.853) in Group 1 and 0.907 (95% CI, 0.842–0.955) in Group 2, exhibiting nearly complete agreement with the latest Atlanta classification of AP. AUROC values for these modified Balthazar grades when used to predict SAP were significantly higher than those for BISAP scores in Group 1, Group 2, and the overall cohort (P < 0.05). The DCA curves for Group 1, Group 2, and the overall patient cohort exhibited substantial net benefits when using these modified grades across a range of POFs relative to BISAP scores. The calibration curve for this modified approach to predicting POF in AP patients revealed good agreement in this cohort.

Conclusion: Modified Balthazar grades exhibited substantial to near-total agreement with the 2012 revised Atlanta classification of AP patients, and this modified method can thus be used for the early prediction of AP severity in the ED.

Keywords: acute pancreatitis, severity, Balthazar grade, revised Atlanta classification

Introduction

Acute pancreatitis (AP) is an inflammatory and exudative condition of the pancreas that can involve other local or distal organ systems.¹ AP is among the leading gastrointestinal causes of hospitalization, affecting estimated 13–45/100,000 persons worldwide.² The overall incidence of AP is steadily rising, with a median 3.4% annual growth rate.³ Mortality rates for AP patients range from 2% to 20% depending on disease severity.²,⁴ Per the 2012 revised Atlanta classification system, AP is classified as mild AP (MAP), moderate-severe AP (MSAP), or severe AP (SAP) based upon whether or not it manifests with local or systemic complications and organ failure lasting for ≤48 h or >48 h.⁶ MAP patients generally recover within a few days without the need for hospitalization.⁷,⁸ MSAP and SAP patients, in
contrast, can exhibit local and/or systemic complications and potentially fatal multi-organ dysfunction syndrome (MODS) driven by a pronounced inflammatory response, while early hospitalization or intensive care unit (ICU) treatment can improve outcomes in these patients. It is thus important to differentiate between MSAP/SAP patients and MAP patients such that the former can undergo early ICU admission, while the latter can undergo outpatient treatment in order to improve patient prognosis while minimizing the waste of medical resources associated with MAP patient hospitalization.

Predicting AP severity during the early stages of the disease is vital to appropriate patient triage and can markedly improve patient outcomes. As such, a series of scoring systems have been developed to assess pancreatitis severity at presentation or at 48 h, including the Ranson criteria, Glasgow-Imrie score, Acute Physiology and Chronic Health Evaluation II (APACHE II), Simplified Acute Physiology Score (SAPS II), Multiple Organ System Score (MOSS), and Balthazar grade systems. All of these evaluation approaches, however, are primarily derived from the assessment of AP as per the Atlanta criteria or mortality outcomes rather than the 2012 revised Atlanta classification system capable of differentiating MSAP patients from those with MAP or SAP. Moreover, these methods necessitate complex or cumbersome calculations and can necessitate dynamic examination for 24–48 h, thus precluding the early prediction of AP severity. There is thus the urgent need for the development of a rapid and robust approach to accurately predicting AP severity at an early time point.

Introduced by Balthazar in 1994, the Balthazard grading scale can assess AP severity based on the appearance of the pancreas in computed tomography (CT) scans, stratifying AP patients into grades A - E. This system has proven to be superior to many of the other clinical scoring methods discussed above, but it is also subject to many of the same limitations given that its accuracy within 72 h of AP onset is limited. In an effort to guide the early prediction of AP severity in the emergency department (ED), we therefore sought to modify these Balthazar grades and we assessed the efficacy of our modified grading system as a tool for predicting AP severity in the ED.

Materials and Methods
Study Design
This was a single-center retrospective cohort study that was conducted in accordance with the Declaration of Helsinki and received approval from the Human Ethical Committee of West China Hospital of Sichuan University (No.2019-334). As data were obtained retrospectively from an electronic case information system and laboratory test system without the need for any additional analyses, informed consent was not obtained from individual patients.

Study Population
AP patients evaluated in the ED of West China Hospital from January 1, 2017 to September 30, 2017 were retrospectively enrolled in this study if they were ≥18 years old and had undergone CT examination within 24 h following ED admission. Patients were excluded if they had a history of organ failure, were diagnosed with chronic pancreatitis or pancreatic cancer, had undergone prior pancreatic surgery, or exhibited incomplete clinical data.

Patients were separated into two groups based upon whether or not their CT scans had been completed within 72 h of AP symptom onset (Group 1: <72 h; Group 2: ≥72 h). MAP, MSAP, and SAP were diagnosed by adopting the 2012 revision of the Atlanta Classification of AP criteria.

Data Collection
Data collected from the electronic case information system and laboratory test information system of our hospital pertaining to enrolled patients included age, gender, clinical features, laboratory test results, disease severity, time of symptom onset, time of admission, and time of initial CT scan. All patient data was anonymized and encrypted. Patient Balthazar grades were confirmed by two independent radiologists blinded to patient outcomes and other examination data.
Statistical Analysis

SPSS v 25.0 was used to analyze all data. Continuous data are given as means ± standard deviation and were compared via non-parametric Mann–Whitney tests or Student’s t-tests. Categorical data are given as frequencies and percentages and were compared via Fisher’s exact test or \( \chi^2 \) tests. Concordance between the modified grading methods developed herein and the 2012 revised Atlanta classification criteria was assessed using the \( \kappa \) statistic. \( \Delta \) Area under the ROC curve (AUC) values were assessed to gauge the discriminative performance of the modified method. Calibration was assessed using a calibration curve, a decision curve analysis (DCA) was performed to assess the clinical utility of this modified method. \( P < 0.05 \) was the threshold of significance.

Results

Patient Characteristics

In total, 502 AP patients were identified over the study period, of whom 130 were excluded based on the exclusion criteria. The remaining 372 patients were enrolled in this study, with 200 and 172 being, respectively, included in Group 1 and Group 2.

Patient demographic characteristics, clinical features, laboratory findings, and grouping data are compiled in Table 1. The average age of patients in this study was 49.40 ± 15.14 years (range: 14–92), and 231 (62.10%) of them were male. In total, there were 81 (21.77%), 117 (31.45%), and 174 (46.78%) MAP, MSAP, and SAP patients in this study cohort, respectively (Table 1).

| Characteristics               | Mild (n=81) | Moderate Severe (n=117) | Severe (n=174) | P   |
|-------------------------------|------------|-------------------------|----------------|-----|
| Male, n (%)                   | 41 (50.62) | 74 (63.25)              | 116 (66.67)    | 0.081|
| Age, year                     | 52.99±17.25| 48.48±14.20             | 48.24±14.50    | 0.063|
| Fever, °C                     | 36.70±0.53 | 36.82±0.60              | 36.94±0.72     | 0.028|
| HR, beats/min                 | 90.13±18.86| 102.23±23.55            | 112.54±24.12   | <0.001|
| MAP, mmHg                     | 100.03±14.95| 98.58±13.84            | 100.00±16.06   | 0.724|
| RR, /min                      | 20.86±6.24 | 22.90±4.92              | 23.85±5.17     | <0.001|
| PaO\(_2\)/FiO\(_2\)          | 365.72±84.00| 329.18±121.66          | 279.94±115.42  | <0.001|
| WBC, 10\(^9\)/L               | 12.39±5.08 | 13.21±2.12              | 133.85±34.16   | 0.702|
| HCT, %                        | 37.80±5.06 | 38.48±5.53              | 39.35±8.60     | 0.270|
| PLT, 10\(^9\)/L               | 175.86±75.28| 176.8±80.88            | 173.6±89.92    | 0.945|
| TBIL, \(\mu\)mol/L           | 32.38±36.74| 29.67±37.24             | 35.76±60.87    | 0.618|
| BUN, mmol/L                   | 5.01±3.71  | 4.87±2.49               | 6.82±4.41      | <0.001|
| ALB, mmol/L                   | 39.73±5.86 | 36.29±7.07              | 34.07±6.13     | <0.001|
| Cr, mmol/L                    | 66.75±18.32| 76.00±55.68             | 119.79±114.14  | <0.001|
| PT, s                         | 12.67±2.65 | 13.03±1.49              | 13.60±2.08     | <0.001|
| D-Dimer                       | 2.87±5.05  | 5.27±5.70               | 6.21±5.19      | <0.001|
| AMY, U/L                      | 357.34±518.33| 351.84±408.29          | 614.59±698.38  | <0.001|
| LIP, U/L                      | 432.00±534.50| 424.29±478.19          | 756.95±1270.26 | 0.006|
| Ca, mmol/L                    | 2.16±0.19  | 2.05±0.28               | 1.95±0.30      | <0.001|
| LAC, mmol/L                   | 1.71±1.19  | 1.84±0.89               | 2.29±1.55      | 0.002|
| Balthazar grade *             | 2.43±0.87  | 3.91±0.54               | 4.85±0.55      | <0.001|
| Marshall scores               | 1.01±0.95  | 1.41±1.14               | 2.57±1.75      | <0.001|
| APACHE-II                     | 6.01±3.32  | 6.59±3.37               | 8.62±4.15      | <0.001|
| SOFA                          | 2.36±1.75  | 2.76±1.78               | 4.37±2.47      | <0.001|
| BISAP                         | 1.37±0.83  | 1.76±0.89               | 2.18±0.91      | <0.001|
| LOS in hospital†, day         | 11.63±7.19 | 13.25±7.10             | 20.81±15.94    | <0.001|

Notes: *Translated the Balthazar grade scale A to E into a point scale 1 to 5. † LOS in hospital, length of stay in hospital.
Clinical and Laboratory Information
There were significant differences among MAP, MSAP, and SAP patients with respect to indices including fever, respiratory rate (RR), heart rate (HR), white blood cell (WBC) count, oxygenation index (PaO$_2$/FiO$_2$), lactate (LAC), blood urea nitrogen (BUN), albumin (ALB), creatinine (Cr), prothrombin time (PT), D-Dimer, amylase (AMY), lipase (LIP), calcium (Ca), Balthazar grades, Marshall scores, SOFA scores, APACHE II scores, Bedside Index for Severity in Acute Pancreatitis (BISAP) scores, and length of hospitalization ($P < 0.05$; Table 1). In contrast, there were no differences in gender, mean arterial pressure (MAP), hemoglobin (HB), hematocrit (HCT), platelets (PLT), or total bilirubin (TBIL) levels among these groups ($P > 0.05$).

Balthazar CT Grades
With respect to Balthazar grade distributions in this study cohort, 80, 64, 39, 9, and 8 patients in Group 1 were classified as Balthazar grade E, D, C, B, and A, respectively, while in Group 2 there were 85, 46, 22, 9, and 10 patients classified in these respective grade categories (Figure 1). The mean Balthazar grade scores in Group 1, Group 2, and the overall patient cohort were 4.00, 4.09, and 4.04, respectively ($P > 0.05$).

Patient Regrading and κ Value Calculation
We next modified the Balthazar grading criteria to establish six new classification methods corresponding to the three grades of AP severity defined by the 2012 Atlanta classification system, as shown in Table 2. Of these methods, Method 6, which regraded Balthazar scores of A-C, D, and E as MAP, MSAP, and SAP, respectively, exhibited the largest $\kappa$ value of these tested methods, with values of 0.786 (0.706–0.853), 0.907 (0.842–0.955) and 0.841 (0.794–0.885) in Group 1, Group 2, and the overall cohort, respectively, as shown in Table 3, consistent with substantial or near-total agreement with the 2012 revised Atlanta classifications.

Correlations Between Modified Balthazar Grading and AP Severity Parameters
Using regrading method 6 established above, we were able to classify 97, 110, and 167 AP patients in the overall study cohort as having MAP, MSAP, and SAP, respectively, including 56, 64, and 80 patients in Group 1 and 41, 46, and 85 patients in Group 2, respectively (Table 4). There were also significant differences in transient organ dysfunction, long-term organ dysfunction, ICU treatment, and in-hospital mortality among patients with different levels of disease severity established using this method in both the individual groups and the overall study cohort ($P < 0.001$).

Assessment of Modified Balthazar Grade Prognostic Performance
The ROC curves generated for Group 1, Group 2, and the overall cohort revealed that these modified Balthazar grades exhibited good discriminative power as a means of predicting SAP, with respective AUC values of 0.928 (95% CI, 0.887–0.968), 0.968 (95% CI, 0.939–0.996), and 0.944 (95% CI, 0.919–0.970), as compared to respective BISAP
AUC values of 0.707 (95% CI, 0.635–0.779), 0.662 (95% CI, 0.581–0.742), and 0.688 (95% CI, 0.634–0.741), Marshall AUC values of 0.762 (95% CI, 0.695–0.829), 0.735 (95% CI, 0.661–0.810), and 0.750 (95% CI, 0.700–0.799) and APACHE II AUC values of 0.699 (95% CI, 0.626–0.772), 0.665 (95% CI, 0.585–0.745), and 0.690 (95% CI, 0.637–0.744) (Figure 2A–C). These AUROC values for the modified Balthazar grades were significantly higher than those for BISAP scores in Group 1, Group 2, and the overall cohort (P < 0.05). The sensitivity, specificity and

**Table 2** Six Regrading Methods Based on Balthazar Grades of AP

| Balthazar Grades of AP | A | B | C | D | E |
|------------------------|---|---|---|---|---|
| Regrading method 1     | MAP | MSAP | SAP |
| Regrading method 2     | MAP | MSAP | SAP |
| Regrading method 3     | MAP | MSAP | SAP |
| Regrading method 4     | MAP | MSAP | SAP |
| Regrading method 5     | MAP | MSAP | SAP |
| Regrading method 6     | MAP | MSAP | SAP |

**Notes:** Green, MAP = mild acute pancreatitis; yellow, MSAP = moderate severe acute pancreatitis; red, SAP = severe acute pancreatitis.

**Table 3**κ Values Corresponding to Comparisons Between Different Regrading Methods and the 2012 Revised Atlanta Classifications

| Regrading Method | Group 1 (n=202) | Group 2 (n=170) | Total |
|------------------|-----------------|-----------------|-------|
|                  | κ (95% CI)      | κ (95% CI)      | κ (95% CI) |
| Regrading method 1 | 0.563 (0.481–0.648) | 0.685 (0.600–0.769) | 0.620 (0.559–0.676) |
| Regrading method 2 | 0.168 (0.088–0.254) | 0.213 (0.123–0.306) | 0.189 (0.131–0.249) |
| Regrading method 3 | 0.067 (0.013–0.127) | 0.130 (0.057–0.210) | 0.094 (0.051–0.139) |
| Regrading method 4 | 0.622 (0.533–0.708) | 0.772 (0.691–0.850) | 0.691 (0.626–0.749) |
| Regrading method 5 | 0.232 (0.148–0.323) | 0.313 (0.212–0.418) | 0.270 (0.204–0.337) |
| Regrading method 6 | 0.786 (0.706–0.853) | 0.907 (0.842–0.955) | 0.841 (0.794–0.885) |

**Table 4** The Relationship Between Severity Parameter and the Modified Balthazar Grades

| Severity Parameter | The Modified Balthazar Grades (Regrade Method 6) | p |
|-------------------|-----------------------------------------------|---|
|                   | Group 1 (n=56) | Moderate (n=45) | Severity (n=80) | p | Group 2 (n=41) | Moderate (n=46) | Severity (n=85) | p | Total (n=97) | Moderate (n=110) | Severity (n=165) | p |
| Mild              | 43             | 2               | 41             | 0             | 0             | <0.001          | 84             | 2             | 1               | <0.001          | 10             | 97             | 5               | <0.001          |
| Moderate          | 10             | 53              | 0             | 44            | 2             | <0.001          | 10             | 97            | 5               | <0.001          | 10             | 97             | 5               | <0.001          |
| Persistent        | 3              | 9               | 76            | 0             | 2             | <0.001          | 3              | 11            | 159             | <0.001          | 3              | 11             | 159             | <0.001          |
| Need for ICU, n   | 1              | 5               | 25            | 1             | 4             | 35              | 2              | 9             | 60              | <0.001          | 2              | 9             | 60              | <0.001          |
| Death, n          | 0              | 2               | 9             | 0             | 3             | 15              | 0              | 5             | 24              | <0.001          | 0              | 5             | 24              | <0.001          |

**Abbreviations:** ICU, intensive care unit; Death, hospital death.
Youden index of modified Balthazar grades of the overall cohort is 89.77%, 96.43% and 0.862 (85.39%, 96.40% and 0.818 for Group 1 and for 94.32%, 97.62% and 0.919 for Group 2.) DCA curves were next generated to assess the clinical benefits associated with the use of our modified Balthazar grading system, revealing that this modified grading approach exhibited a large net benefit across SAP as compared to BISAP scores when evaluating patients in Group 1, Group 2, and the overall cohort (Figure 2D–F). Calibration curves for this modified Balthazar grading system when used to predict SAP exhibited good agreement with this cohort (Figure 2G–I). In summary, C-index, ROC curve, and calibration curve analyses all indicated that this modified Balthazar grading system was effective as a means of screening AP patients for risk of SAP during the early stages of disease.

**Discussion**

A range of scoring systems and biomarkers have previously been used to predict disease severity in AP patients, but these approaches have been subject to certain limitations. For example, several require complex calculations and an observation period of at least 48 h, limiting their utility in an ED setting. The Balthazar grading system is a CT-based approach to assessing AP severity based on the appearance of the pancreas that was first introduced in 1994, and it has been shown to be superior to many other clinical scoring systems and approaches to gauging AP severity. However, Balthazar grades separate patients into five levels (A-E), thus yielding a complex system that does not directly correspond to the 2012 revised Atlanta classification criteria for AP. Moreover, CT scans are typically performed more than 72 h after AP symptom onset other than in specific diagnostic contexts, making Balthazar scoring impractical as a means of predicting AP severity at an early stage.
The present study is the first to our knowledge to have utilized modified Balthazard grades to predict AP patient disease severity for individuals completing CT scans within 72 h of symptom onset based on the 2012 revised Atlanta classification criteria. Results including \( \kappa \) values, C-index, calibration curves, and ROC curves demonstrated that our modified Balthazar grading strategy was able to predict POF among AP patients more accurately than BISAP scores, consistent with the prognostic value of these scores during the early stages of AP. This modified grading approach may thus guide clinicians in efforts to select patients for early ICU treatment or admission in an effort to improve patient outcomes. SAP entails potentially serious complications and high mortality rates, and patient prognosis can be improved via early admission or ICU treatment. The proportion of SAP was higher in this study relative to prior reports, potentially because West China Hospital is a prestigious and well-known medical center that treats many critically ill patients transferred from other hospitals.

Figure 2 The prognostic performance of modified Balthazar grade in patients with AP. (A–C) ROC curves corresponding to the utility of modified Balthazar grades, Marshall, APACHE II and BISAP scores as predictors of SAP in Group 1, Group 2, and the overall study cohort; (D–F) DCA curves corresponding to the utility of modified Balthazar grades and BISAP scores as predictors of SAP in Group 1, Group 2, and the overall study cohort; (G–I) calibration curves corresponding to modified Balthazar grade as predictors of SAP in Group 1, Group 2, and the overall study cohort.

Abbreviations: ROC, receiver operating characteristic; APACHE, acute physiology and chronic health evaluation; BISAP, bedside index for severity in acute pancreatitis; AUC, area under the curve.
There are certain limitations to this study. For one, the study design was retrospective and is thus susceptible to potential selection bias, and certain clinical or laboratory details corresponding to individual patients may have been poorly documented. Second, this was a single-center study conducted at a relatively prestigious medical center in China that has many critically ill patients transferred from other hospitals, resulting in a higher proportion of patients with severe pancreatitis and thus introducing a degree of selection bias. We were additionally unable to establish the etiology of AP in individual patients in this retrospective analysis owing to the complexity and multifaceted nature of this disease.

**Conclusion**

In summary, the modified Balthazar grading strategy developed herein wherein AP patients with Balthazard grades A-C, D, and E are classified as having mild, moderate, and severe disease, respectively, offers excellent concordance with the 2012 revised Atlanta classification system, suggesting that this method can be reliably used to predict AP severity at an earlier time point in the ED. However, additional large-scale prospective studies will be necessary to establish whether this approach can be reliably used to predict AP patient prognosis in a clinical setting.

**Author Contributions**

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

**Funding**

This work was supported financially by grants from the Key R&D Project of Sichuan Provincial Department of Science and Technology (2021YFS0023), the Science Foundation of Health Commission of Sichuan Province (No. S15065), Technology Innovation Project of Key R & D Support Plans of Chengdu Science and Technology Municipality (2020-YF05-00074-SN) and Project of Beijing medical and health foundation (YWJKJJHYJJ-B184096-Q26).

**Disclosure**

The authors report no conflicts of interest in this work.

**References**

1. Bradley EL. 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(5):586. doi:10.1001/archsurg.1993.01420170122019
2. Yadav D, Lowenfels AB. Trends in the epidemiology of the first attack of acute pancreatitis: a systematic review. *Pancreas.* 2006;33(4):323–330. doi:10.1097/01.mpa.0000236373.31617.52
3. Roberts SE, Morrison-Rees S, John A, Williams JG, Brown TH, Samuel DG. The incidence and aetiology of acute pancreatitis across Europe. *Pancreatology.* 2017;17(2):155–165. doi:10.1016/j.pan.2017.01.005
4. Munigala S, Yadav D. Case-fatality from acute pancreatitis is decreasing but its population mortality shows little change. *Pancreatology.* 2016;16(4):542–550. doi:10.1016/j.pan.2016.04.008
5. Roberts SE, Akbari A, Thorne K, Atkinson M, Evans PA. The incidence of acute pancreatitis: impact of social deprivation, alcohol consumption, seasonal and demographic factors. *Aliment Pharmacol Ther.* 2013;38(5):539–548. doi:10.1111/apt.12408
6. Banks PA, Bollen TL, Dervenis C, et al. Classification of acute pancreatitis-2012: revision of the Atlanta classification and definitions by international consensus. *Gut.* 2013;62(1):102–111. doi:10.1136/gutjnl-2012-302779
7. Ince AT, Senturk H, Singh VK, Yildiz K, Gurakar A. A randomized controlled trial of home monitoring versus hospitalization for mild nonalcoholic acute interstitial pancreatitis: a pilot study. *Pancreatology.* 2014;14(3):174–178. doi:10.1016/j.pan.2014.02.007
8. Kulvatunyou N, Watt J, Friese RS, et al. Management of acute mild gallstone pancreatitis under acute care surgery: should patients be admitted to the surgery or medicine service? *Am J Surg.* 2014;208(6):981–987. doi:10.1016/j.amjsurg.2014.09.003
9. Working Group IAP/APA Acute Pancreatitis Guidelines. IAP/ APA evidence-based guidelines for the management of acute pancreatitis. *Pancreatology.* 2013;13(4):e1–e15. doi:10.1016/j.pan.2013.07.063
10. Ranson JH, Pasternack BS. Statistical methods for quantifying the severity of clinical acute pancreatitis. *J Surg Res.* 1977;22(2):79–91. doi:10.1016/0022-4804(77)90045-2
11. Imrie CW, Benjamin IS, Ferguson JC, et al. A single-centre double-blind trial of trasylool therapy in primary acute pancreatitis. *Br J Surg.* 1978;65(5):337–341. doi:10.1002/bjs.1800650514
12. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. Apache ii: a severity of disease classification system. *Crit Care Med.* 1985;13(10):818–829. doi:10.1097/00003246-198510000-00009
13. Larvin M, McMahon MJ. Apache-ii score for assessment and monitoring of acute pancreatitis. *Lancet*. 1989;2(8656):201–205. doi:10.1016/S0140-6736(89)90381-4

14. Le Gall JR, Loirat P, Alperovitch A, et al. A simplified acute physiology score for ICU patients. *Crit Care Med*. 1984;12(11):975–977. doi:10.1097/00003246-198411000-00012

15. Taylor SL, Morgan DL, Denson KD, Lane MM, Pennington LR. 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(2):219–222. doi:10.1016/j.amjsurg.2004.11.010

16. Balthazar EJ, Ranson JH, Naidich DP, Megibow AJ, Caccavale R, Cooper MM. Acute pancreatitis: prognostic value of CT. *Radiology*. 1985;156(3):767–772. doi:10.1148/radiology.156.3.4023241

17. Li W, Luo S, Zhu Y, et al. Concordance of the Balthazar grade and the Revised Atlanta classification. *Pancreas*. 2018;47(10):1312–1316. doi:10.1097/MPA.0000000000001166

18. Chatzicostas C, Roussomoustakaki M, Vardas E, Romanos J, Kouroumalis EA. Balthazar computed tomography severity index is superior to Ranson criteria and APACHE II and III scoring systems in predicting acute pancreatitis outcome. *J Clin Gastroenterol*. 2003;36(3):253–260. doi:10.1097/00004836-200303000-00013

19. Leung TK, Lee CM, Lin SY, et al. Balthazar computed tomography severity index is superior to Ranson criteria and APACHE II scoring system in predicting acute pancreatitis outcome. *World J Gastroenterol*. 2006;11(38):6049–6052. doi:10.3748/wjg.v11.i38.6049

20. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37–46. doi:10.1177/0013164460020000104

21. Mounzer R, Langmead CJ, Wu BU, et al. Comparison of existing clinical scoring systems to predict persistent organ failure in patients with acute pancreatitis. *Gastroenterology*. 2012;142(7):1476–1482; quiz e15–e16. doi:10.1053/j.gastro.2012.03.005

22. De Waele JJ, Delrue L, Hoste EA, De Vos M, Duyck P, Colardyn FA. Extrapancreatic inflammation on abdominal computed tomography as an early predictor of disease severity in acute pancreatitis: evaluation of a new scoring system. *Pancreas*. 2007;34(2):185–190. doi:10.1097/mpa.0b013e31802d4136

23. Yousef M, McCallion K, Diamond T. Management of severe acute pancreatitis. *Br J Surg*. 2003;90(4):407–420. doi:10.1002/bjs.4179

24. Forsmark CE, Vege SS, Wilcox CM. Acute pancreatitis. *N Engl J Med*. 2016;375(20):1972–1981. doi:10.1056/NEJMra1505202