Agreement between Marshall, Ranson and Apache II as estimators of morbidity and mortality in acute pancreatitis

Alberto Rodríguez-Varón, MD,[1]   Óscar Mauricio Muñoz-Velandia, MD, MSc, PhD,[2]   Diana Agreda-Rudenko, MD,[3]   Elías García-Consuegra, MD.[4*]

1   Internist physician, gastroenterologist. Professor of Gastroenterology and Internal Medicine, Department of Internal Medicine. Unit of clinical gastroenterology and digestive endoscopy, Pontificia Universidad Javeriana, Hospital Universitario San Ignacio; Bogotá, Colombia.
2   Practitioner and clinical epidemiologist. Professor of Internal Medicine, Department of Internal Medicine, Pontificia Universidad Javeriana, Hospital Universitario San Ignacio; Bogotá, Colombia.
3   Internist physician. Department of Internal Medicine, Pontificia Universidad Javeriana, Hospital Universitario San Ignacio; Bogotá, Colombia.
4   Resident of Internal Medicine, Department of Internal Medicine, Hospital Universitario San Ignacio, Pontificia Universidad Javeriana; Bogotá, Colombia.

*Correspondence: Elías García-Consuegra, MD
Elias_gc10@hotmail.com

Abstract

Introduction: Different scales to estimate the risk of morbidity and mortality in patients with pancreatitis are currently in use in Colombia, which leads to uncertainty when classifying and treating these patients. **Objective:** This study seeks to analyze agreement between the most used scales to estimate the risk of patient morbidity and mortality in a population treated at 2,670 meters above sea level (m.a.s.l.). **Materials and methods:** Two hundred patients between 18 and 65 years old, diagnosed with acute pancreatitis, were evaluated and treated at the Hospital Universitario San Ignacio, Bogotá (Colombia). Three risk scales were used for the estimations. Scores ≥ 8 in the APACHE II system, ≥ 2 in the Modified Marshall Score, or 3 or more positive Ranson criteria were classified as pancreatitis with severity prognostic marker. Agreement between the results was determined using the Kappa coefficient. **Results:** According to the Marshall score, 45.5% of the cases were pancreatitis with predicted severity, while APACHE II and Ranson yielded scores of 39.5% and 38.5%, respectively. The Kappa coefficient showed weak agreement between APACHE II and Ranson (Kappa=0.201; 95%CI 0.05-0.34), poor agreement between Ranson and Marshall (Kappa=0.18; 95%CI 0.04-0.32), and moderate agreement between APACHE II and Marshall (Kappa=0.42; 95%CI 0.28-0.56). **Conclusions:** There is poor agreement between the pancreatitis severity scoring systems used in Colombia, so they cannot be interpreted as clinically equivalent. The data from this study demonstrate the need to validate the scales in Colombia and Latin America. They also suggest that the Marshall scale overestimates the risk in cities above 2,000 m.a.s.l.

INTRODUCTION

Early detection of organ dysfunction or multiorgan failure is very important in the initial assessment of patients with acute pancreatitis. Therefore, it should be considered that these patients benefit from more invasive diagnostic and therapeutic interventions aimed at modifying the course of the disease and reducing its morbidity and mortality (1).

To identify patients at risk of multiorgan failure, several scales have been created to classify them based on clinical and laboratory data. APACHE II (2), Ranson (3) and Marshall (4) are among the most used scales. Each has advantages and disadvantages. APACHE II, for example, is dynamic and allows risks to be staged according to the patient's evolution, but it is expensive and requires multiple laboratory tests. In the meantime, Ranson criteria has been used for a long time

and a large number of physicians are familiar with it; however, it requires a complete assessment after 48 hours for the score to be calculated, so results are available late. And the Marshall score allows assessing different systems (renal, cardiovascular, and respiratory) and is easy to calculate; however, it has not been validated in Colombia.

In this context, these scales are used simultaneously, which means that patients can be classified differently. This generates uncertainty in physicians regarding which risk estimate should be used to provide the best possible management to patients.

To date, whether the different scales that evaluate the risk of morbidity and mortality in acute pancreatitis are clinically equivalent in Colombian population or in patients living at heights 2000 masl, where normal oxygen blood pressure may be significantly lower (69 mm Hg) or not has been studied. This study assesses the concordance between these three scales under such conditions.

METHODS

The study population consisted of patients between 18 and 65 years old diagnosed with acute pancreatitis and who were treated at Hospital Universitario San Ignacio in Bogotá, Colombia, between 2012 and 2016. Patients with a history of chronic pancreatitis and in which hospital monitoring did not exceed 48 hours were excluded, as it was not possible to calculate all risk scores in such patients. The research protocol was approved by the Ethics Committee of the Hospital Universitario San Ignacio and Pontificia Universidad Javeriana.

Follow-up data for these patients were systematically collected from information recorded in the institutional electronic medical record system. Demographic and clinical variables, as well as physical examination findings and biochemical parameters were included. Measurement techniques used for tests processing at the institutional clinical laboratory were the same throughout the study. The calculation of the risk scores was based directly on laboratory reports and clinical characteristics data reported for each moment of time, rather than on calculations described in the medical records of patients by their treating physicians.

In each patient, three risk scores were calculated and the corresponding clinical manifestations on admission were considered. The scales used were APACHE II (2), Ranson (3) and Marshall (4). Based on the results, patients were staged for pancreatitis with or without severity prediction. In this regard, severity-predicted pancreatitis was defined as an APACHE II score ≥8, 3 or more Ranson positive criteria, or a modified Marshall score ≥2.

In addition, information regarding outcomes such as mortality, acute kidney injury, respiratory failure, presence of pancreatic necrosis, acute peri-pancreatic fluid collections, pancreatic pseudocysts, need for surgical intervention, and days of hospital stay, was collected.

Continuous variables are expressed as means and standard deviations (SD) if they follow a normal distribution, and as medians and interquartile ranges (IQR) if they do not. On the other hand, categorical variables are described as percentages. The kappa coefficient was used to analyze the concordance between the different classification methods. The interpretation of this coefficient as an evaluator of concordance strength is as follows: 0.01-0.20 (poor), 0.21-0.40 (weak), 0.41-0.60 (moderate), 0.61-0.80 (good), 0.81-0.90 (very good), and 0.9-1.00 (almost perfect) (5). An alpha significance level of 0.05 was set. Calculations were performed using Stata 15® software.

RESULTS

A total of 200 patients were included and their demographic characteristics are described in **Table 1**. 20.5% had complications during their hospital stay, either necrosis, acute peri-pancreatic fluid collections, pancreatic pseudocysts, or need for surgical intervention. Three patients died during follow-up.

The percentage of patients classified under pancreatitis with severity prediction, for each risk scale, is shown in **Table 2**. The Marshall score ranked the highest number of patients under pancreatitis with severity prediction (45.5%), while only 13.5% of patients scored >12 in the APACHE II scale.

On the other hand, the kappa coefficient showed a weak concordance (kappa: 0.201; 95% CI: 0.05-0.34) Concordant patients and classified as pancreatitis with severity prediction were 40 (20%), while 76 (38%) were discordant cases, of which, 37 (18.5%) were classified under the pancreatitis with severity prediction category based on the Ranson criteria, but not by APACHE II, and 39 (19.5%) were classified under this category according to the APACHE II scale, but not by the Ranson criteria. The assessment of the concordance between APACHE II and Ranson is described below (**Table 3**).

In turn, concordance between APACHE II and Marshall scales was moderate, with a kappa coefficient of 0.42 (95 % CI: 0.28-0.56) In this case, concordant patients classified as pancreatitis with severity prediction were 57 (28.5%). Of the 56 discordant patients (26.5%), most were classified in the pancreatitis with severity prediction category according to the Marshall score, but not by the APACHE II scale (34 patients, 15.5%) (**Table 4**).

In addition, the concordance between the Ranson criteria and the Marshall score was poor, with a kappa coefficient of 0.18 (95% CI: 0.04-0.32) The concordant patients

Table 1. Demographic characteristics of the patients included in the study

Variable	n = 200
Age (years), mean (SD)	54.4 (19)
Female, n (%)	113 (56.2)
Kidney injury, n (%)	39 (19.4)
Days of hospital stay, n (%)	
<5 days	51 (25.5)
6-10 days	95 (47.5)
11-15 days	20 (10)
>16 days	34 (17)
Mean blood pressure (mm Hg), n (%)	
<65	6 (3)
65-90	85 (42.5)
>90	109 (54.5)
PaO$_2$/FiO$_2$, n (%)	
<100	85 (44.04)
100-300	107 (55.4)
300-500	1 (0.52)
No data	7 (3.5)
Creatinine (mg/dL), n (%)	
<1,5	178 (89)
1,5-3	19 (9.5)
>3	1 (0.5)
Hematocrit on admission (%), n (%)	
<30	3 (1.5)
30-40	30 (15)
40-50	124 (62)
>50	43 (21.5)
Hematocrit at 48 h (%), n (%)	
<30	4 (2)
30-40	87 (43.5)
40-50	82 (41)
>50	7 (3.5)
No data	20 (10)
GOT (U/L), n (%)	
<200	128 (64)
>200	66 (33)
No data	6 (1)
GPT (U/L), n (%)	
<200	137 (68.5)
>200	54 (27)
No data	9 (4.5)
Fluid sequestration (L), n (%)	
<4	87 (43.5)
4-10	80 (40)
>10	30 (15)
No data	3 (1.5)
Presence of comorbidity, n (%)	41 (20.5)
Death, n (%)	3 (1.5)

SD: Standard deviation; n: number; mm Hg: millimeters of mercury; PaO$_2$/FiO$_2$ ratio: ratio of Partial arterial pressure of oxygen and fraction of inspired oxygen; mmol/L: millimoles per liter; U/L: units per liter; GOT: Glutamic-Oxalacetic Transaminase; GPT: Glutamic-Pyruvic Transaminase

Table 2. Pancreatitis categories based on the APACHE II scale, the Ranson criteria, and the Marshall score

Variable	n = 200
APACHE II, n (%)	
No SP	121 (60.5)
With SP	79 (39.5)
Ranson, n (%)	
No SP	123 (61.5)
With SP	77 (38.5)
Marshall, n (%)	
No SP	109 (54.5)
With SP	91 (45.5)

APACHE II: Acute Physiology and Chronic Health Disease Classification System II; SP: severity prediction

classified under the pancreatitis with severity prediction category were 44 (22% of the total), while the 80 discordant cases (40%) corresponded to 47 cases classified under this category according to the Marshall score, but not by the Ranson criteria (23.5%). On the contrary, 33 were as pancreatitis with severity prediction cases based on the Ranson criteria, but they were not according to the Marshall score (16.5%) (**Table 5**).

DISCUSSION

This is the first study conducted in Colombian population -and also in patients living at altitudes above 2000 masl- that compares the results obtained by classifying the severity of acute pancreatitis with the three most used risk scales. Based on the results obtained, agreement between these scales is low. Therefore, these results should not be interpreted as clinically equivalent.

In our study most patients were women between 46 and 65 years old. Overall mortality was 1.5%, which is slightly lower than what has been described in the literature (2-3%) (6). However, this may be explained by the fact only patients under the age of 65 were included.

By evaluating each scale individually, it is observed that according to the Marshall score, 45.5% of patients would be classified under the pancreatitis with severity prediction category, whereas according to the APACHE II scale and the Ranson criteria, only 39.5% and 38.5% would be included in this groups. It is noteworthy that in our study the proportion of patients with severe pancreatitis is higher than that what has been reported in similar studies, where on average, 8 to 15% patients had severe pancreatitis (7).

This finding might be explained by the fact that the institution where our study was conducted is a reference hospital, and that only patients with a minimum 48-hour follow-up

Table 3. Concordance between the APACHE II scale and the Ranson criteria

		Ranson		Total
		Pancreatitis without SP	Pancreatitis with SP	
APACHE II	Pancreatitis without SP n (% of total)	84 (42)	37 (18.5)	121 (60.5)
	Pancreatitis with SP n (% of total)	39 (19.5)	40 (20)	79 (39.5)
	Total: n (% of total)	123 (61.5)	77 (38.5)	200 (100)

Kappa: 0.201; 95%CI: 0.05-0.34. APACHE II: Acute Physiology and Chronic Health Disease Classification System II; SP: severity prediction

Table 4. Concordance between the APACHE II scale and the Marshall score

		Marshall		Total
		Pancreatitis without SP	Pancreatitis with SP	
APACHE II	Pancreatitis without SP n (% of total)	87 (43.5)	34 (15.5)	121 (60.5)
	Pancreatitis with SP n (% of total)	22 (11)	57 (28.5)	79 (39.5)
	Total: n (% of total)	109 (54.5)	91 (45.5)	200 (100)

Kappa: 0.42; 95%CI: 0.28-0.56. APACHE II: Acute Physiology and Chronic Health Disease Classification System II; SP: severity prediction

Table 5. Concordance between the Ranson criteria and the Marshall score

		Marshall		Total
		Pancreatitis without SP	Pancreatitis with SP	
Ranson	Pancreatitis without SP n (% of total)	76 (38)	47 (23.5)	123 (61.5)
	Pancreatitis with SP n (% of total)	33 (16.5)	44 (22)	77 (38.5)
	Total: n (% of total)	109 (54.5)	91 (45.5)	200 (100)

Kappa: 0.18; 95% CI: 0.04-0.32. SP: severity prediction

process were include, which could have resulted in patients with mild pancreatitis being poorly represented. In the case of the Marshall score, the high proportion of patients classified as severe pancreatitis cases might be related to the fact that our institution is located at 2670 masl, which might have caused patients to be often categorized as severe, based on the score given by low arterial oxygen pressure levels.

This suggests that a different cut-off point should be sought for oxygen blood pressure in patients treated in cities above 2000 masl, and that the Marshall score should be validated under such conditions before being used.

The difference in the number of patients classified under the pancreatitis with severity prediction group reported in the present study, shows that many patients classified in this category according to this scale, would not receive the same diagnosis if any the other two scales were to be used. This makes it impossible for the treating physician to define which of the classifications is more appropriate to

assess the patient's risk. Therefore, establishing an adequate treatment becomes a challenge.

In addition, in our study a poor agreement between the Ranson criteria, the APACHE II scale, and the Marshall score was found. However, a slight agreement was found between APACHE and Marshall. In general, it can be concluded that concordance between the three scales is low, and so they cannot be interpreted as clinically equivalent in patients with pancreatitis.

This finding leads to important clinical implications for decision-making during the provision of emergency health services. Clinical practice guidelines suggest that, in order to define the disease as pancreatitis with severity prediction, a score >8 points in the APACHE II scale is required, as well as 3 or more positive Ranson criteria, or a score ≥ 2 in the modified Marshall scale.

However, if these parameters were to be accepted, physicians could face two clinical scenarios. In the first one, the treating physician decides to perform diagnostic and therapeutic procedures, such as the admission and management in the intensive care unit for patients classified, according to any of these three scales, as having pancreatitis with severity prediction. This strategy, characterized by high sensitivity but low specificity, would generate a useless expense for the health system, given that unnecessary interventions would be made in some of these patients.

In the second case, on the contrary, the physician decides to act based on the results of one of the scales but ignores the other two. If this were done, a significant number of patients who would benefit from additional interventions could be left out. This situation could lead to increased mortality and complications rates.

It should be noted that the findings obtained here were already expected by us after analyzing the diagnostic performance of these scales reported by other authors. In the case of the Ranson criteria, a sensitivity of 63%, a specificity of 76%, a positive predictive value of 79%, and a negative predictive value of 92% have been described. Meanwhile, for the APACHE II scale a sensitivity of 82%, a specificity of 86%, a positive predictive value of 81%, and a negative predictive value of 98% have been reported (8).

However, the findings of some studies conducted in Latin America differ from the data reported in this paper. In Peru, Ponce (9) conducted an observational study with 77 patients to assess the concordance of three severity prognostic scales (APACHE II, Ranson, and BISAP), where an almost perfect agreement between the scales was found, with concordance percentages above 90%, and particularly strong agreement between BISAP and APACHE II. Thus, in said study, the use any of these scales is recommended,

bearing in mind that the easiest to apply for the treating physician should be the first option.

In 2005, Rosas et al. conducted a similar study in Mexico (10) comparing the APACHE II, Osborne, and Ranson scales, other laboratory tests of severity (hematocrit, serum calcium and base deficit) and the Balthazar computed tomographic severity index. These authors concluded that both, sensitivity, and specificity found for each scale were similar to those reported in the literature worldwide. However, this work also emphasizes the need to continue conducting studies similar to the one carried out here, as well as external validation, considering that so far, no research has assessed these three scales, not only in Colombian population, but in Latin America in general.

On the other hand, divergent results may be associated with differences in health systems. Similarly, there are differences at the time studies are conducted or between populations regarding their biological characteristics. Therefore, it would be necessary to validate the different scales in each country and, probably, in different contexts within a same country (11).

Some limitations of the study include the relatively small sample size and the low number of deaths or complication events. Therefore, at the time it is not possible to validate these scales in Colombian population. Studies with significantly larger sample sizes will be needed to achieve this goal.

A second consequence of the limited sample size is that CIs are relatively large. However, conclusions obtained on the basis of the concordance analysis would be similar at both ends of the ranges, that is, interpretations of the scale results cannot be assumed to be clinically equivalent.

In addition, we did not include patients older than 65 years, so our results cannot be applied to this population. Further studies are required to assess whether the concordance between the different scales changes significantly.

An additional limitation is the selection bias as a result of the underrepresented group of patients with pancreatitis without prediction of severity in our study population. This suggests that our data may not apply to institutions with a lower level of complexity.

Also, our results also show that the agreement between the APACHE II scale, the Ranson criteria, and the Marshall score is low and, therefore, they should be interpreted as clinically equivalent in Colombia. They also suggest that a lower cut-off point should be considered in oxygen blood pressure levels, when using the Marshall score in cities above 2000 masl to avoid misclassifying a high proportion of patients as having severe pancreatitis.

Since there are no validation studies that allow us to define which scale is most appropriate in our context, we emphasize that, apart from the score obtained in the cho-

sen predictive severity scale, the treating physician should always assess the independent risk factors of each patient and individualize the case, which requires the use of laboratory tests and additional imaging studies.

In our opinion, the strategy of classifying pancreatitis cases under the of pancreatitis with severity prediction category is preferable if a positive score is obtained in any of the scales described here. After all, patients' safety comes first than health care costs.

Conflict of interest

None declared by the authors.

Funding sources

None declared by the authors.

REFERENCES

1. Huerta-Mercado J. Tratamiento médico de la pancreatitis aguda. Rev Med Hered. 2013;24(3):231-236. https://doi.org/10.20453/rmh.v24i3.320
2. Larvin M, McMahon MJ. APACHE-II score for assessment and monitoring of acute pancreatitis. Lancet. 1989;2(8656):201-205. http://doi.org/10.1016/s0140-6736(89)90381-4
3. Ranson JH, Spencer FC. Prevention, diagnosis, and treatment of pancreatic abscess. Surgery. 1977;82(1):99-106.
4. Banks PA, Bollen TL, Dervenis C, Gooszen HG, Johnson CD, Sarr MG, Tsiotos GG, Vege SS; Acute Pancreatitis Classification Working Group. Classification of acute pancreatitis--2012: revision of the Atlanta classification and definitions by international consensus. Gut. 2013;62(1):102-11. http://doi.org/10.1136/gutjnl-2012-302779
5. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-174. http://doi.org/10.2307/2529310
6. Werner J, Feuerbach S, Uhl W, Büchler MW. Management of acute pancreatitis: from surgery to interventional intensive care. Gut. 2005;54(3):426-436. http://doi.org/10.1136/gut.2003.035907
7. Everhart J. Pancreatitis. In: The burden of digestive diseases in the United States. US Department of Health and Human Services. 2004. p. 119-122.
8. Wu BU, Johannes RS, Sun X, Tabak Y, Conwell DL, Banks PA. The early prediction of mortality in acute pancreatitis: a large population-based study. Gut. 2008;57(12):1698-1703. http://doi.org/10.1136/gut.2008.152702
9. Ponce-Monar CR. Concordancia que existe entre tres escalas de predictores de severidad en pancreatitis aguda en pacientes atendidos en el servicio de medicina interna en el Hospital María Auxiliadora en el periodo 2016. Lima: Universidad Privada San Juan Bautista; 2017.
10. Rosas FM, Gaxiola WR, Ibáñez GO, Vargas TE, Meza VM, Calvo IJ. Evaluación de las escalas y factores pronóstico en pancreatitis aguda grave. Cir Gen. 2005;27(2):137-143.
11. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605. http://doi.org/10.1136/bmj.b605