EDITORIAL

Is calcium the solution to the difficult task of predicting severity in acute pancreatitis?

¿Es el calcio la solución a la difícil tarea de pronosticar gravedad en pancreatitis aguda?

The study by Gutierrez-Jiménez et al. evaluates the predictive value of total serum calcium and albumin-corrected calcium in acute pancreatitis. The authors concluded that these parameters, when determined within the first 24 h of the onset of an acute pancreatitis episode, are useful predictors of severity, even comparable or superior to the Ranson and APACHE-II scoring systems.

This study reports that serum calcium measurement has a predictive usefulness similar to that of other scales, in addition to being low-cost, as well as easily obtained and available, but it is important to point out that some of the comparisons were made using Ranson and APACHE-II cut-off points that are different from those now established as severity predictors; using a Ranson’s score > 5 or an APACHE 7 instead of a 3 and 8, respectively, as the definition of severity, increases the risk of underestimating and/or overestimating the true severity of the event.

Furthermore, the authors did not use the Bedside Index for Severity in Acute Pancreatitis (BISAP), which has shown itself to have a prognostic performance for severity and mortality in acute pancreatitis that is comparable or superior to the Ranson and APACHE-II, with areas under the curve (AUC) between 0.81-0.82, 0.94, and 0.78-0.83, respectively, in addition to being easily to calculate. Had the BISAP, which possibly will be the most widely used methodology for predicting severity in acute pancreatitis in the coming years, been included in the comparisons, it would have increased the value of the results.

When the usefulness of any predictive method is analyzed, it is common that the evaluation takes into consideration sensitivity and specificity values, positive predictive value (PPV), and negative predictive value (NPV). However, the most useful and relevant information is obtained by calculating the AUC and the likelihood ratios (LR).

Reviewing the curves corresponding to the total calcium and corrected calcium at first glance (the figures are not reported), the AUCs do not appear to be better than those obtained with the Ranson and APACHE-II scores. With respect to the LR (especially the positive ones), their predictive value is regarded as greater, if they are closer to or above 10. The study results suggest that albumin-corrected calcium (LR + 6.4) in the severe AP patient group has an acceptable predictive value, but these values decrease (LR + 2.47 for a corrected calcium cut-off point of 7.9 mg/dl or less) when those patients classified as having moderately severe disease are included in the severe AP group.

Even when regarding the positive LR obtained by including only severe AP patients as an adequate predictive value, and as an even better one when compared with those obtained by the Ranson and APACHE-II scores, it should be kept in mind that these values depend on the pre-test probability; in other words, on the prevalence of the adverse event to be predicted, which in the case of the severe AP patients of the study is 8%, and when the moderately severe ones are included it is 27%; this means a post-test probability similar to or even below that obtained by the Ranson and APACHE-II scores based on the observed results.

These figures can be explained by a small sample size and a low percentage of patients with severe AP that, in turn, are responsible for the confidence intervals of the calculated odds ratios. These intervals, besides crossing the unit in some cases (which takes away value from the result), are also wide.

Another point to consider, and that the authors mention in the discussion, is the use of the revised Atlanta classification for defining severity; it introduces the category of moderately severe, which decreases the prevalence of the abovementioned severe cases and leaves aside a group whose behavior is uncertain and not previously taken into consideration by the existing prognostic scales. A recent report classified 256 cases of AP according to the previous assessment.
Atlanta classification published in 1993, and found that 49% of the cases were mild and 51% were severe; this same population was reclassified using the revised Atlanta definitions and 49% of the cases remained mild, but 25% were moderately severe and 25% were severe. The AUC for identifying mortality in each of the classifications was 0.76 and 0.89, respectively. It is important to emphasize that this study was conducted at an AP referral center and therefore the severity percentages are higher than those usually reported and expected. Nevertheless, it exemplifies the lack of knowledge and poor information on the cases classified as moderately severe, whose real prevalence is unknown. With the information obtained from these 2 reports, the prevalence of moderately severe cases varies from 17 to 25%, which undoubtedly modifies the predictive performance of the scores used.

Considering this specific group within the severe cases, as Gutierrez-Jiménez et al. did, is the closest approximation to the reality of the dynamic process that is observed in the manifestations of AP, but it possibly overestimates the prevalence of severity.

The word “prediction” implies anticipating a determined outcome. The purpose of this prediction is to implement measures that help prevent a given outcome. In the clinical field, the best scale or predictive marker is the one that is inexpensive, universally accessible and reproducible, rapidly obtained, and highly accurate.

From the perspective of biostatistics, the best predictive test is the one with a high positive LR, or that is associated with, or increases, post-test probability.

This can be achieved by increasing the pre-test probability or by finding a marker or scale with a high predictive value that up to now, including total calcium and albumin-corrected calcium, does not exist. All the markers and scales used for evaluating AP severity are far from being perfect or ideal, and therefore the search for a better biomarker or predictive scale is both necessary and justified.

Although the results of the study by Gutierrez-Jiménez et al. are limited, they are nevertheless provocative and interesting. But before they can be accepted or rejected they should be validated and reproduced in other populations and studies that include a greater number of cases.

Financial disclosure
No financial support was received in relation to this article.

Conflict of interest
The author declares that there is no conflict of interest.

References

M. Pelaez-Luna
Associate Professor of Medicine, Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City, Mexico
Departamento de Gastroenterología, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico
E-mail address: mariopl@prodigy.net.mx

6 February 2014 10 February 2014