Combined Serum Amylase and Lipase Determinations for Diagnosis of Suspected Acute Pancreatitis

James P. Corsetti, Christopher Cox, Thadeus J. Schulz, and Dean A. Arvan

Serum amylase and lipase measurements are often used to diagnose acute pancreatitis. This study addresses the question of whether it is advantageous to order serum amylase and lipase tests simultaneously. We evaluated performance of the two tests separately and in combination through a retrospective study of patients for whom both amylase and lipase determinations were ordered. Initial analysis of test performance was conducted with a uniformly applied criterion based on determination of optimal sensitivity–specificity pairs. Individual tests and combinations of tests, including the "AND" and "OR" rules and discriminant functions, were examined. Only the discriminant approach demonstrated better performance than the lipase test alone. This finding was subsequently confirmed by logistic regression analysis. We conclude that ordering both tests simultaneously can be advantageous in diagnosing acute pancreatitis when a bivariate approach is used; however, this must be weighed against the difficulties associated with clinical implementation of such approaches.

Indexing Terms: sensitivity • specificity • predictive value • logistic regression

Serum amylase and lipase determinations are the most commonly ordered laboratory tests for the assessment of acute pancreatitis, with most previous reports showing a higher sensitivity and specificity for lipase (1–7). However, whether or not there is any advantage in ordering both tests together is an important and frequently raised question. From a cost-effectiveness standpoint, an amylase test should be ordered only if it improves sensitivity and (or) specificity in the clinical assessment of patients with suspected acute pancreatitis. If this is not the case, the test for amylase should be discouraged. This could result in significant savings of laboratory resources, depending on the local prevalence of the disease and how often the tests are ordered together.

The question of whether information is gained by ordering both amylase and lipase tests together has been approached in several studies. Results have been mixed, with some reports supporting the practice (2, 8, 9) and others not (5–7). Because simultaneous ordering of these tests was not the primary focus of these studies, the techniques for combining the two tests have been generally limited, such as the logical "AND" and "OR" rules with cutoffs that have been separately optimized for each test. However, additional techniques are available that could be used for combining these tests to optimize performance. This study systematically addresses whether or not the combined serum amylase and lipase measurements are better than the individual tests for suspected acute pancreatitis. We compared individual tests and combinations of tests, including the "AND" and "OR" rules, and linear and quadratic discriminant functions, by using the best value of the simultaneously optimized sensitivity–specificity pair to identify the most accurate test combinations. Results were confirmed with logistic regression analysis.

Materials and Methods

Study Design and Subjects

We performed a retrospective study by retrieval of initial serum amylase and lipase results from the laboratory information system for all patients who had a serum amylase and (or) lipase determination performed over a period of 1 year. Inpatients and our emergency department patients were included. This resulted in 4338 separate initial tests being ordered. Subsequent exclusions were as follows: age <18 years, 309; follow-up tests for some subjects, 474 (initial tests were retained); amylase not done, 41; lipase not done, 843; and renal failure, 103 (blood urea nitrogen or creatinine more than twice the upper reference limit); thus 2568 patients remained with both serum amylase and lipase tests performed. Of these patients, 443 (17.3%) had increased serum amylase and (or) lipase concentrations (upper reference limits: amylase, 110 U/L; lipase, 208 U/L) and 2125 (82.7%) were within the reference limits for both tests. Of the 443 enzyme-positive subjects, a random sample of one-third (148) was assessed for acute pancreatitis by chart review. Because of incomplete information in some medical records, 169 cases had to be reviewed to achieve this result. Diagnosis was based on: presentation and clinical course (71); amylase and lipase concentrations both greater than twice the upper normal reference limit (24) (1, 5); or positive findings in ultrasound (4), laparotomy (3), computerized tomography (2), biopsy (1), or autopsy (1). Of the 148 enzyme-positive cases reviewed, 106 (71.6%) had acute pancreatitis and 42 (28.4%) did not. Of the 2125 enzyme-normal cases, a random sample of one-third (709) was selected and added to the database for further study. These patients were assumed not to have acute pancreatitis, as in the study by Tetraulit (10). Thus, the final database contained 106 patients with acute pancreatitis.
[57 men, ages 19–87 years, mean age 49 ± 17 (SD) years; and 49 women, ages 18–93 years, mean age 55 ± 21 years] and 751 patients without acute pancreatitis (371 men, ages 18–91 years, mean age 46 ± 20 years; and 380 women, ages 18–104 years, mean age 48 ± 22 years).

Laboratory Analysis

Serum amylase and lipase determinations were performed on an Ektachem 700 analyzer (Eastman Kodak Co., Rochester, NY) according to the manufacturer's instructions.

Data Analysis

In the first phase of the analysis, we used optimal sensitivity–specificity pairs to evaluate test performance; thus, a uniformly applicable criterion could be used for each of the rules, including the "AND" and "OR" rules. [The "AND" and "OR" rules cannot be treated with more traditional approaches such as receiver-operating characteristic (ROC) analysis.] Optimal sensitivity–specificity pairs were defined as those that resulted in the closest approach to the point characterizing an ideal test (sensitivity = 1, specificity = 1). We determined optimal sensitivity–specificity pairs for the assessment of acute pancreatitis for amylase alone (univariate case); lipase alone (univariate case); amylase and lipase with the "AND" rule, using the individually optimized amylase and lipase cutoffs (univariate cases); amylase and lipase with the "OR" rule, using the individually optimized amylase and lipase cutoffs (univariate cases); amylase and lipase with the "AND" rule, using simultaneously optimized amylase and lipase cutoffs (bivariate case); amylase and lipase with the "OR" rule, using simultaneously optimized amylase and lipase cutoffs (bivariate case); linear discriminant function; and quadratic discriminant function. The optimized sensitivity–specificity pairs and corresponding cutoffs for all approaches except the discriminant functions were determined by computerized examination of the distance of each sensitivity–specificity pair to the point characterizing an ideal test for all possible cutoffs (i.e., between all adjacent data points). The cutoffs were determined by minimizing the square root of [(1 – sensitivity)² + (1 – specificity)²] (11). The determination of optimal discriminant functions is a discrete optimization problem and was addressed by using graphical methods to eliminate all but a relatively small number of possibilities, which were then examined exhaustively by computer with the closest-approach criterion. This nonparametric approach was used to maintain the closest-approach criterion, which would not be the case for parametric techniques (e.g., logistic regression or classical discriminant analysis). An interesting feature of the nonparametric approach is that the actual cutoffs and discriminant functions, but probably not the estimated sensitivity and specificity, are sensitive to data points in the center of the plot.

In the second phase of the analysis, we used logistic regression analysis to confirm the initial results and to assess more directly the effects of sampling variation. This contrasts with the closest-approach procedure, in which this appraisal is more difficult.

The database was used only to derive the optimized parameters for each of the rules so that the performance of the rules could be compared. The actual clinical implementation of any rule would require further validation with an independent data set. Tests for statistical significance of differences between sensitivities and specificities were performed as described by Galen and Gambino (12). Positive predictive values were calculated assuming the observed prevalence of acute pancreatitis in the final database (12.5%).

Results and Discussion

Chart review of the 148 enzyme-positive cases revealed 74 emergency room visits, 49 medical inpatients, and 25 postsurgery cases, with the following characteristics in the 106 subjects with acute pancreatitis: ethanol use, 28 (26.4%); trauma, 26 (24.5%); biliary tract disease, 10 (9.4%); peptic ulcer disease, 7 (6.6%); drug use, 4 (3.8%); pancreatic duct obstruction, 4 (3.8%); acquired immunodeficiency syndrome, 2 (1.9%); post-endoscopic retrograde cholangiopancreatography, 1 (0.9%); systemic lupus erythematosus, 1 (0.9%); hyperalimentation, 1 (0.9%); and undetermined, 22 (20.8%). Mean (±SD) serum enzyme concentrations for the 106 pancreatitis-positive subjects were 353 ± 767 U/L (range: 30–4590 U/L) for amylase and 2745 ± 6786 U/L (range: 179–37746 U/L) for lipase. The respective values for the 751 pancreatitis-negative subjects were 56 ± 27 U/L (range: 30–354 U/L) and 97 ± 51 U/L (range: 2–379 U/L).

ROC curves for the assessment of acute pancreatitis for the individual (univariate) serum amylase and lipase concentrations are shown in Figure 1. The superiority of lipase over amylase is clearly demonstrated, in agreement with several previous reports (1–7). The lipase ROC curve is virtually identical to that of Werner et al. (5) and Van Lente and Kazmierczak (6), but dem-
The amylase ROC curve is virtually identical to that of Van Lente and Kazmierczak (6) but demonstrates less sensitivity and specificity than that of Werner et al. (5). Such differences may be due to different study designs, populations of patients, and analytical methodologies, but the qualitative result with regard to the superiority of lipase over amylase is the same.

Figure 2 illustrates scatter plots of the log_{10}-transformed lipase vs amylase concentrations, optimal cutoffs, and regions indicative of acute pancreatitis as defined by various combination rules by using the closest-approach criterion. The optimal cutoffs, true positives, false negatives, true negatives, false positives, optimal sensitivity–specificity pair, and distance of the optimal sensitivity–specificity pair from that of an ideal test (1, 1) are given in Table 1. Although lipase alone is most frequently cited (5-7) as the best test for the assessment of acute pancreatitis, Table 1 shows distances less than that for lipase alone (0.0420)—and more closely approaching an ideal test—for the “OR” rule with simultaneously optimized cutoffs (0.0373), the linear discriminant rule (0.0305), and the quadratic discriminant rule (0.0221). Statistical tests of significant differences among the three rules were conducted separately for the optimal sensitivity and specificity and were compared with lipase alone. Significant differences were found...
only in the specificities of the linear ($P = 0.0044$) and quadratic ($P = 0.0069$) discriminant rules vs lipase alone. It is not entirely correct to perform significance tests on the same data that were used to develop the classification rules because both sensitivity and specificity are typically overestimated with this procedure; however, the reported $P$ values are small enough to assure reasonable confidence in the results.

The optimization criterion used in this work, including linear and quadratic discriminant approaches, involved minimization of the quantity $[(1 - sensitivity)^2 + (1 - specificity)^2]$, which is equivalent to choosing a sensitivity–specificity pair on a ROC curve that most closely approaches the point characterizing an ideal test (sensitivity $= 1$, specificity $= 1$). The criterion is not equivalent to that used in classical parametric discriminant analysis or logistic regression. The closest-approach criterion is one of several standard criteria for evaluating test performance based on a specific cutoff (11). It attributes equal weight to both sensitivity and specificity. By itself, this criterion does not account for any clinical impact or costs to the health care system associated with misclassifications of patients (false-positive and false-negative rates). However, it remains an excellent criterion for the assessment of the discriminatory ability of clinical tests. An alternative, well-accepted parameter for comparing performance of clinical tests is the area under the ROC curve (13); however, there are problems related to overlap of curves and decreased specific clinical relevance (13, 14) caused by not defining an operating point on the curve that is optimal for a given clinical situation. In any case, area under the ROC curve was not used for the evaluation of test performance in the first phase of this study because it could not be applied uniformly to all test combinations. ROC analysis of the “AND” and “OR” rules results in collections of points on the plot instead of a single curve. It is important to be able to evaluate the “AND” and “OR” rules because they can be clinically implemented relatively easily. It is also important to be able to compare these results to earlier studies.

To derive additional information about the clinical implications of these results, we calculated positive predictive values for lipase alone and for the linear and quadratic discriminant functions by using the prevalence of acute pancreatitis in the final database. This calculation should provide a reasonable estimate of prevalence because the only confounding factor is the exclusion from the final database of 843 cases in which only an amylase test was ordered and of 41 cases in which only a lipase test was ordered. There is no reason to believe that the prevalence of acute pancreatitis in the amylase-only group is significantly different from that in the final database. In our experience, such cases are probably more reflective of differences in individual physicians’ ordering practices. The same can be said of the lipase-only cases. Positive predictive values for lipase alone and for the linear and quadratic discriminant functions were 78.8%, 85.2%, and 87.5%, respectively. The observed higher predictive values with discriminant functions are a result of greater discrimination in the pancreatitis-negative population, leading to fewer false positives. A higher predictive value is especially useful in assessment of acute pancreatitis by tending to decrease needless or inappropriate treatment of such patients, thus avoiding serious clinical consequences in the false-positive patients as described by Galen and Gambino (15) and Robertson and Zweig (16).

With the closest-approach procedure, the assessment of the effect of sampling variation on the derived parameters is not simple, making more rigorous statistical determination of differences in the various rules difficult. For this reason, logistic regression analysis was performed in the subsequent phase to confirm that amylase and lipase measurements together resulted in better performance than lipase alone. Logistic regression analysis showed that both amylase ($P <0.0001$) and lipase concentrations ($P <0.0001$) were highly significant predictors of disease. Furthermore, because both variables were included in the model, it is clear that each test provides predictive information that the other
does not. Thus, the performance of both tests together is statistically better than lipase alone. The discriminant function resulting from logistic regression for the log$_{10}$-transformed data was lipase + (0.2539 amylase) > 2.847 (addition of second-order terms was not statistically significant), resulting in: sensitivity, 89.6%; specificity, 99.2%; true positive, 95; false negative, 11; true negative, 745; and false positive, 6. ROC curves for lipase alone and the discriminant function are presented in Figure 3. The logistic regression results clearly indicate that the addition of amylase is significant and is consistent with the finding that the curve corresponding to the discriminant function demonstrates equal or better characteristics than that of lipase alone. On the other hand, it is obvious that the curves are close to each other. Thus, the addition of amylase may result in only marginally better clinical performance, as suggested by the nearly identical areas under the curves.

Results of previous studies of whether ordering amylase and lipase tests together is advantageous have been mixed. Our finding that ordering both tests simultaneously may be warranted is in contrast with those of several previous reports (5–7). This may be due to a number of factors. In the report of Werner et al. (5), the only test combinations examined were the "AND" and "OR" rules with individually optimized cutoffs. Our results confirm that such an approach does not significantly improve performance. In the report of Van Lente and Kazmierczak (6), no information about how the tests were combined is given. In the report of Viel et al. (7), logistic regression was used to combine the tests; however, amylase determination was explicitly involved in the assessment of acute pancreatitis. Recommendations for ordering both tests together in previous work (2, 8, 9) were based on general surveys and reviews of the issue (8, 9) and on actual results from the application of the "AND" and "OR" rules, using the optimal univariate cutoffs when combining tests (2).

Our results confirm that the best combinations used in most previous studies offer no advantages over using lipase results alone. However, use of a logistic regression discriminant function does result in statistically significant superior performance over lipase alone. Inspection of areas under the ROC curves suggests that this improvement may be clinically insignificant, at least with regard to the global comparison over all cutoffs. This contrasts with our finding for a particular operating point defined by simultaneously optimized sensitivity and specificity that the resulting discriminant function appears to have clinically improved performance. Thus, a judicious combination of serum amylase and lipase determinations may give better performance than lipase alone in the assessment of acute pancreatitis. However, we believe that successful clinical application of such an approach will demand the formulation of a discriminant rule specific to local populations of patients and analytical techniques. Further, logistical difficulties associated with the implementation of clinical on-line decision-making algorithms will need to be simplified and overcome. One approach is to have the discriminant score calculated and reported by the laboratory information system along with appropriate cutoffs and (or) interpretative comments derived by computer algorithm.

Fig. 3. ROC curves for univariate serum lipase (thick line) and for logistic regression discriminant function [lipase + 0.2539 amylase] > 2.847] (thin line) in the assessment of acute pancreatitis.

References