Comparison of the Effectiveness of Four Clinical Chemical Assays in Classifying Patients with Chest Pain

Andre C. Van Steirteghem,1 Mark H. Zweig,2,4 E. Arthur Robertson,2 Roland M. Bernard,3 Gustaaf A. Putzeys,3 and Claude J. Bleva3

We compared the usefulness of four serum assays for classifying patients originally suspected of having an acute myocardial infarction. One of these is the long-used measurement of total creatine kinase (CK) activity. The other three are relatively new immunoassays: myoglobin by RIA, CK-BB by RIA, and CK-MB by immunoinhibition. When we evaluated test effectiveness with use of conventionally derived reference ranges, the results were misleading. However, by using receiver operating characteristic curves, we were able to effectively compare the four tests at all possible decision levels, rather than at only one. Multiple closely sequential serum specimens were obtained during the first four days after the onset of chest pain. Total CK, CK-MB, and CK-BB all behaved similarly, reaching peak diagnostic effectiveness at 18-20 h, when all three correctly classified 95% of the infarct patients, with a zero false-positive rate. However, total CK was more useful in identifying infarcts later in their courses than were the two CK isoenzymes tests. Myoglobin assay was most effective earlier in the course, at about 7 to 8 h. Our results indicate (a) that the tests for myoglobin and for CK or its isoenzymes are complementary and (b) that of the three CK tests, measurement of total CK activity provides the most information over the broadest segment of a patient's course.

Additional Keyphrases: cutoff value · heart disease · enzyme activity · relative values of diagnostic aids · CK total vs isoenzyme as marker of acute myocardial infarction · myoglobin · creatine kinase

Creatine kinase (CK),5 its isoenzymes, and myoglobin are all serum markers of myocardial insult and are useful in the diagnosis of acute myocardial infarction (AMI). Recently, immunological approaches have been used to achieve more sensitive and (or) specific assays for these markers. This study compares the performance of a traditional test, total CK activity, to three immunoassays, myoglobin by RIA, CK-BB by RIA, and CK-MB by immunoinhibition.

We obtained multiple closely sequential serum samples, timed from the onset of chest pain, from 76 patients admitted to a coronary-care unit, to characterize in detail the time course of release of these myocardial markers and to compare their relative values as indexes for distinguishing patients with an AMI from those without one. Because the effectiveness of these tests in this respect varies with time, we evaluated the performance of each test at each sampling interval. Use of reference ranges established by conventional means gave results that misrepresented the relative performance of the tests. We compared the tests, then, by using plots of true-positive rates vs false-positive rates [receiver operating characteristic (ROC) curves (1, 2)]. The use of ROC curves ensures that tests are compared with the same true- or false-positive rate for each. Furthermore, ROC curves display the relative performance of the tests over the entire spectrum of possible decision levels.

Materials and Methods

Patients studied were admitted to the Free University of Brussels Saint-Pierre Hospital Coronary Care Unit between August 1978 and June 1979 with chest pain and suspected AMI. Each patient was carefully interviewed to determine the time of onset of pain suggesting an AMI. Patients who were not admitted within 6 h of onset of chest pain were excluded. Also excluded were patients admitted to the Coronary Care Unit for reasons other than chest pain and the suspicion of an acute myocardial infarct. This included patients with arrhythmias, syncope, and patients who had just had coronary angiography. Of the 76 patients studied, 56 were ultimately diagnosed as having had an AMI based on electrocardiographic changes: 47 patients had new Q-waves while eight patients had ST and T wave changes typical of AMI. Of the latter group, four patients had "flipped" lactate dehydrogenase isoenzymes and two others had scintigraphic findings suggestive of AMI. For the 21 patients without AMI, the final diagnosis was angina pectoris in four, acute coronary insufficiency in 12, and pericarditis in five. To avoid introduction of bias, these diagnoses were made without consideration of the results of any of the four tests being evaluated.

From the time of admission, serial blood samples were collected every hour until 12 h after the onset of pain, every 2 h for the next 12 h, and every 12 h for the next three days. Serum total creatine kinase activity was determined at 25 °C by a ultraviolet kinetic method involving activation with N-acetylcysteine (Boehringer Mannheim). Serum CK-BB concentration was measured by a previously described radioimmunoassay (3), with use of anti-human CK-BB and radiolabeled human CK-BB. Whereas the original paper indicated that CK-MB cross reacted very little in this assay, we now have evidence that CK-MB cross reactivity may be great enough that the CK-MB present can contribute measurably to the CK-BB value; the amount contributed varies from patient to patient. Serum CK-MB activity was measured as residual CK-B subunit enzymatic activity after inhibition of CK-MM by sheep anti-CK-MM (E. Merck, Darmstadt, F.R.G.). This immunoinhibition assay was performed essentially as described by others (4, 5). In contrast to the RIA for CK-BB, the assay for CK-MB by immunoinhibition measures only enzymatically active CK protein. The low CK-BB enzymatic activity contributes little or nothing to the result in most cases. In the case of the RIA for CK-BB, the assay measures immunoreactive protein whether or not it is enzymatically active. Thus, both assays measure isoenzymes containing CK-B subunit, but their bases and sensitivities are

1 RIA Laboratory, Academisch Ziekenhuis, Free University of Brussels, Brussels, Belgium.
2 Clinical Pathology Department, Clinical Center, National Institutes of Health, Bethesda, MD 20205.
3 Departments of Clinical Chemistry and Immunology, and Coronary Care Unit, St. Pierre Hospital, Free University of Brussels, Brussels, Belgium.
4 Address reprint requests to this author.
5 Nonstandard abbreviations used: CK, creatine kinase; RIA, receiver operating characteristic; AMI, acute myocardial infarction; TP, true positive; FP, false positive.

Received Jan. 29, 1982; accepted Mar. 30, 1982.

CLINICAL CHEMISTRY, Vol. 28, No. 6, 1982 1319
quite different. As a result, the CK-MB assay measures primarily enzymatically active CK-MB while the CK-BB assay is directed at immunoreactive CK-BB.

Myoglobin concentration was measured with a commercially available RIA (Nuclear Medical Systems, Inc., Newport Beach, CA 92663).

Definition of terms. True-positive (TP) rate: The fraction of diseased subjects in whom the test result is “positive” (i.e., exceeds the decision level). The true-positive rate is the same as sensitivity.

False-positive (FP) rate: The fraction of unaffected subjects in whom the test result is “positive” (i.e., exceeds the decision level). This is related to specificity. Specificity refers to the fraction of unaffected subjects in whom the test result does not exceed the decision level (true negative rate). The FP rate = 1 - specificity.

ROC curve analysis. When a test is applied to a group of subjects, for any specific decision level chosen there will be a particular TP rate and a particular FP rate. An ROC curve for a test is a plot of the TP rates vs the FP rates for various decision levels. The ROC curve is actually generated by varying the decision level from the highest observed test result to zero and plotting the corresponding pairs of TP and FP rates for each decision level. We must first have test results from the diseased group of subjects to get the TP rates, as well as from the nondiseased subjects to get the FP rate. In the present case, the diseased subjects are persons with chest pain who had an AMI, while the “nondiseased” subjects are persons with chest pain who were found not to have had an AMI. A clinically effective test is characterized by a high TP rate and a low FP rate. Several tests can be compared directly by plotting their ROC curves on the same graph. The test that consistently achieves the highest TP rates at any given FP rates will have the best diagnostic efficacy.

Results

The mean results for each test for every time interval for the two groups of patients appear in Figure 1. For the patients with AMI, peaks for myoglobin, CK-BB, CK-MB, and total CK activity occurred at 8, 16, 22, and 20 h, respectively. In contrast, the curves for non-AMI patients varied over a comparatively narrow range, with no distinct peak. Although the mean values for the two groups differed markedly, the values for individual patients overlapped for all four tests at every time studied. This overlap is illustrated in Figure 2,
Concentrations. This comparison suggests that CK-BB is somewhat more sensitive and increases earlier in the course of the disease. By plotting the FP rate vs time we can also compare the specificities of these two markers (Figure 4). Note that CK-MB has a considerably lower FP rate at all times. At 16 h the FP rate is 50% for CK-BB, and only 7% for CK-MB. This suggests that while CK-BB seemed to have better sensitivity (higher TP rates), CK-MB had better specificity (lower FP rates) in the classification of this group of subjects.

At this point, a decision as to which test is better would be highly subjective. There are two problems. The first is that the two tests have different TP and FP rates. If, however, the TP rates were the same, then we could compare the two tests based on the FP rates. This requires choosing for each test a decision level ("upper limit") that gives the desired TP rate. This brings up the second problem. We are looking at test performance at a single, somewhat arbitrarily chosen, decision level. This tells us only about how the test performs at that one point. Performance can change markedly as the decision level

Fig. 2. Serum CK-BB concentrations 16 h after onset of chest pain in patients with (filled circles) and without (open circles) an AMI which contains the actual CK-BB data for the two groups of subjects at 16 h after the onset of chest pain. Note, for example, that two subjects who had an AMI had CK-BB concentrations lower than those for seven of the non-AMI subjects. Because of the overlap of the data there is no decision level above which all AMI subjects would fall (100% TP rate) and below which all non-AMI subjects would fall (0% FP rate). This kind of overlap occurred for all four tests at all 24 time points studied. Thus none of these tests perfectly distinguished our subjects as AMI or non-AMI.

Because none of the tests achieved a 100% TP rate with a 0% FP rate, we examined the rates they could achieve, to assess their relative clinical performance. Figure 3 is a plot of the true-positive rate (or sensitivity) vs time for two of the markers, CK-BB and CK-MB, based on a conventional reference range derived from testing healthy subjects. This shows the fraction of AMI subjects who exceeded the upper limit of the reference range at each time. Note that CK-BB reaches higher TP rates and remains high longer than does CK-MB. At 16 h, 100% of the AMI subjects had "elevated" CK-BB concentrations, while 97% had "elevated" CK-MB concentrations. This comparison suggests that CK-BB is somewhat more sensitive and increases earlier in the course of the disease. By plotting the FP rate vs time we can also compare the specificities of these two markers (Figure 4). Note that CK-MB has a considerably lower FP rate at all times. At 16 h the FP rate is 50% for CK-BB, and only 7% for CK-MB. This suggests that while CK-BB seemed to have better sensitivity (higher TP rates), CK-MB had better specificity (lower FP rates) in the classification of this group of subjects.

At this point, a decision as to which test is better would be highly subjective. There are two problems. The first is that the two tests have different TP and FP rates. If, however, the TP rates were the same, then we could compare the two tests based on the FP rates. This requires choosing for each test a decision level ("upper limit") that gives the desired TP rate. This brings up the second problem. We are looking at test performance at a single, somewhat arbitrarily chosen, decision level. This tells us only about how the test performs at that one point. Performance can change markedly as the decision level

Fig. 3. True-positive rate (sensitivity) for serum CK-BB and CK-MB concentrations at various times after onset of chest pain when the decision level is the conventional upper-limit-of-normal for each test

Fig. 4. False-positive rate (1 — specificity) for serum CK-BB and CK-MB concentrations at various times after onset of chest pain when the decision level is the conventional upper-limit-of-normal for each test.
Fig. 5. True- and false-positive rates for identification of AMI by CK-BB measurement at 4, 7, and 14 h after onset of chest pain

These curves were generated by varying the decision level and then plotting the corresponding true- and false-positive rates. The decision levels (CK-BB concentrations) for points A, B, and C are indicated on the 4-h curve.

changes. Note that in Figure 2, if the decision level is changed from 6.0 mg/L to 12.0 mg/L, the number of false positives in the non-AMI group falls to zero, while the TP rate falls only slightly (from 100% down to 96%). This shows how much better CK-BB can perform at another decision level and how misleading it is to have seen only the first one.

ROC curve analysis permits us to examine the entire spectrum of sensitivities and specificities that a given test can achieve by plotting the TP and FP rates for all possible decision levels, rather than just one. The relationship between the true-positive and false-positive rates is shown in the ROC curve for CK-BB at 4 h (Figure 5). The curve is generated by varying the decision level over the entire range of CK-BB results observed in our subjects. Starting with the highest CK-BB value, the true-positive and false-positive rates are both zero. As the decision level is decreased, some percentage of AMI patients (true-positive rate) and some percentage of non-AMI patients (false-positive rate) equal or exceed the limit. When the decision level reaches the lowest CK-BB result, all of the AMI and non-AMI patients equal or exceed the limit and the curve reaches 100% on both axes. In the case of a good test where there is little overlap between the two pop-

Fig. 6. True- vs false-positive rates for the identification of AMI by four myocardial markers at selected intervals after onset of pain

1322 CLINICAL CHEMISTRY, Vol. 28, No. 6, 1982
At each hour, decision levels were chosen to give a true-positive rate of 95% in the group of AMI patients. The false-positive rates were determined by using the non-AMI patients. Note that decision level and time are plotted on log scales.

In Figure 6, the four tests are compared with one another at 5, 8, 18, and 60 h. At 5 h the four tests behaved very differently. Myoglobin occupied the left-most position and achieved the best ratios of true positives to false positives. At 8 h, myoglobin exhibited its peak effectiveness, with 88% true positives and 0% false positives; it achieved 100% true positives with only about 14% false positives. After 8 h, the effectiveness of myoglobin diminished rapidly, while the three CK tests, which resemble each other in their behavior, became more effective. At 18 h, the three CK measurements were near their peak effectiveness, while the utility of myoglobin had declined markedly. At 60 h, total CK was still of appreciable value in separating AMI from non-AMI patients.

To compare these tests by using ROC curves we should have to look at 24 graphs, one for each sampling time. We can condense this comparison to a more convenient form by choosing a single TP rate and then plotting all the resulting FP rates together vs time. The choice of the target TP rate is arbitrary. (Alternatively, the converse would work equally well, that is, fixing the FP rate and plotting the corresponding TP rates.) An example of this kind of plot for each test is shown in Figure 7. We arbitrarily selected as our goal the detection of 95% of the AMI subjects (95% TP rate). The corresponding decision level and FP rate for each test at each hour is plotted. These figures are generated by using ROC curves such as Figure 5, and reading off the false-positive rate corresponding to 95% true positives. Note the dotted line in
Figure 5, which is drawn at 95% true positives. This line intersects the 4-h curve at point A, corresponding to 49% false positives. The required decision level for this point on the curve is 5.2 μg of CK-BB per liter. Both of these data are shown on the CK-BB plot of Figure 7, where the vertical dotted line at 4 h intersects the two curves. For each test an optimal period occurred during which the false-positive rate was lowest. This generally corresponded to the period when AMI patients had the highest results (see Figure 1). During this period, the decision level was high, and few non-AMI patients exceeded it. Figure 7 shows that CK-BB was most effective at 16 to 20 h, when the false-positive rate was 0%. At that time the decision level for a 95% true-positive rate was about 12 μg/L. However, CK-BB could be used to detect 95% of the AMI patients at any time (e.g., 4 h, see dotted line) by using a lower decision level, though at the cost of a higher false-positive rate.

Total CK was most effective at 16 to 22 h and CK-MB from about 18 to 22 h, both having false-positive rates of 0%. In contrast, myoglobin’s best performance for a 95% true-positive rate was at 7 to 8 h. However, 0% false-positive rate was never reached; the lowest rate of false positives was about 10% (at a 95% true positive rate). Thus myoglobin was the most useful test soon after the onset of pain, but it never matched the effectiveness of the other tests later in the patient’s course.

Discussion

Using ROC curves, we were able to compare the power of multiple tests to make a useful clinical distinction, without being dependent on any particular decision levels. When the best test for use at a particular point in the course of the disease has been selected, the choice of decision level will depend on the relative importance of the conflicting goals of maximizing the TP rate and minimizing the FP rate, both in the light of the prevalence of the condition. In this paper we do not address this complex problem of determining the “best” decision level.

In Figure 7, we did indicate the decision levels that corresponded to a TP rate of 95% for each test at each sampling time. However, this was done primarily to illustrate one way to condense and display the data. We could have just as easily chosen a TP rate of 90%—or we could have chosen a desirable FP rate and then plotted the resulting TP rates.

When a specific decision level is chosen, the predictive value of a positive and of a negative result can be calculated, using the true- and false-positive rates for the corresponding point on the ROC curve, along with information about the prevalence of the disease in the population to be tested.

In our group of subjects, the three CK tests behaved similarly. Their early time courses were about the same and with any of the three one could diagnose 95% of the true cases with 0% false positives over similar periods of several hours near the end of the first day. The plots of false-positive rates shown in Figure 7 for the three CK tests are virtually superimposable. It appears that any one of these CK tests alone provides about as much information as do all three together at the time of peak activity. However, total CK activity was useful for distinguishing AMI patients from non-AMI patients for a much longer time than were the CK isoenzymes.

Although the myoglobin assay was somewhat less effective for classifying patients, its earlier time course complements the CK tests; thus it would provide additional information when used with one of them.

It is surprising that total CK activity performed so well as compared with the CK isoenzyme tests. Total CK is usually regarded as a sensitive but nonspecific marker, especially because it is abundant in skeletal muscle. However, in this study there was no indication of any more nonspecificity for total CK activity than for the two isoenzymes containing the B subunit. Many of the patients received intramuscular injections, as is common practice. Yet, total CK was able to achieve high true-positive rates with false-positive rates just as low as those of the isoenzyme assays. Others (6–9) have found that increased total CK activity is less specific for the diagnosis of AMI than increased CK-MB activity. Our results, which appear contrary to this, may reflect our small sample or may be related to the use of an immunoinhibition test for CK-MB. If we had determined CK-MB by electrophoresis, it may well have performed better than total CK activity, as others have reported. In terms of absolute performance, our data suggest that increasing the decision level of total CK activity may diminish the number of false positives while not greatly affecting the TP rate (sensitivity).

References