Aids to the Evaluation of Diagnostic Performance, as Applied to Immunological Creatine-Kinase MB Assay in Myocardial Infarction

By *M. Werner* and *R. J. Mohrbacher*

Division of Laboratory Medicine, The George Washington University Medical Center, Washington, DC

Summary: Immunoassay of creatine kinase-MB provides numerical information, which makes it possible to estimate quantitatively the diagnostic performance following myocardial infarction. Two graphical methods for such an evaluation are presented. The relationship between technical sensitivity and specificity was analyzed using a continuous function, the receiver-operator characteristic curve. Using this function, a diagnostic threshold ("upper normal limit") was chosen, which balances both technical specificity and sensitivity on the first day following infarction. On the second and third days this threshold caused a progressive loss of technical sensitivity. The relationship between effectiveness and the prevalence of myocardial infarction in the tested population was evaluated with a nomogram correlating these two quantities. With the chosen diagnostic threshold, effectiveness is independent of prevalence on the first day, but the loss in technical sensitivity on subsequent days causes effectiveness to decay when the prevalence is high.

Bewertungsmaße der diagnostischen Leistungsfähigkeit, angewendet auf die immunologische Kreatinkinase-MB-Bestimmung bei Herzinfarkt

Zusammenfassung: Die immunologische Bestimmung von Kreatinkinase-MB verschafft zahlenmäßige Ergebnisse, welche die Anwendung quantitativer Bewertungsmaße der diagnostischen Leistungsfähigkeit nach Herzinfarkt gestattet. Zwei Mittel für eine solche Wertbestimmung werden dargestellt. Die Wechselbeziehungen zwischen technischer Empfindlichkeit und Spezifität wurden mittels einer kontinuierlichen Funktion, der „Receiver-Operator Characteristic Curve" analysiert. Ein diagnostischer Schwellenwert ("obere Normalgrenze") wurde gewählt, bei welchem sowohl die technische Empfindlichkeit wie die technische Spezifität am ersten Tag nach Infarkt im Gleichgewicht sind. Am zweiten und dritten Tag verursacht dieser Schwellenwert einen zunehmenden Verlust an technischer Empfindlichkeit. Die Beziehung der Effektivität (effectiveness) der Untersuchung zur Häufigkeit von Herzinfarkt in der untersuchten Population wurde mit Hilfe eines Nomogramms bewertet, welches diese zwei Kenngrößen miteinander korreliert. Mit dem gewählten Schwellenwert ist die Effektivität der Untersuchung am ersten Tag unabhängig von der Krankheitshäufigkeit, aber der Verlust an technischer Empfindlichkeit verursacht bei hoher Krankheitshäufigkeit in den folgenden Tagen eine Verminderung der Effektivität.

Introduction

Creatine kinase (creatine kinase EC 2.7.3.2) and more specifically its "heart isoenzyme" creatine kinase-MB are generally considered the best laboratory indicators of myocardial infarction. In the past, assay of creatine kinase-MB in the clinical laboratory relied on either chromatographic or electrophoretic isolation, although other techniques have been proposed (1). Technical limitations of these methods (1–3) have compromised the quantitative assessment of diagnostic information provided by different values of creatine kinase-MB activity.

Recently, a creatine kinase-MB assay has been developed which relies on the immunological blocking of enzymatic activity due to the M-subunit of creatine kinase (4,5). The numerical information generated by this test is of a quality comparable to that of standard enzyme assays. Thus, it becomes possible to analyze the performance of creatine kinase-MB assays for the diagnosis of myocardial infarction in quantitative terms. This paper introduces two graphical representations to aid such an evaluation, the receiver-operator characteristic (ROC) function and a nomogram relating test effectiveness to disease prevalence.

Experimental Procedure

Subjects

Patients with myocardial infarction (N=42) and patients with suspected but unconfirmed myocardial infarction (N=98) were investigated. These diagnostic classifications were established by a review of clinical history, laboratory data and electrocardiographic data following the patient's discharge.

Samples were obtained on three days following infarction or following admission for suspected infarction (days 1, 2 and 3). Venous blood was collected in vacuum tubes (CORVAC[R]; Corning Medical, 11 North Street, Medfield, MA 02052), analyzed immediately or stored for no more than 18 hours at 4°C until assayed.

Immunological assay of creatine kinase-MB

A spectrophotometer equipped with a cuvette holder thermostatted at 37°C was used (Model 25; Beckman Instruments Inc., Fullerton, CA 92634). The assay mixture consisted of 10 μl reagent containing: 100 mmol/l triethanolamine buffer pH 7.0; 35 mmol/l creatine phosphate; 20 mmol/l glucose; 9 mmol/l glutathione; 10 mmol/l AMP; > 1.2 kU/l glucose 6-phosphate dehydrogenase; 11.2 kU/l hexokinase, and 800 U/l creatine kinase-MM inhibiting antibodies from the goat (E. Merck, Darmstadt, West-Germany). Diadenosine pentaphosphate was added to this assay mixture in a final concentration of 10 μmol/l. Following addition of the sample, the mixture was preincubated for 7 minutes and the absorbance at 340 nm was then recorded for 3 minutes to calculate enzyme activity using a value of 6,300 l · mol^{-1} · cm^{-1} as the molar absorbance coefficient of NADPH. Results were expressed in U/l.

Electrophoretic assay of creatine kinase-MB

Creatine kinase isoenzymes were separated on 10g/l agarose gel and visualized by overlay with the creatine kinase assay mixture of Oliver (7). Relative activities of isoenzyme fractions were estimated by densitometric scanning at 340 nm. Absolute activity of creatine kinase-MB was calculated by multiplying the relative activity of this fraction (percent) with the total creatine kinase activity.

Assays of total creatine kinase, aspartate aminotransferase (EC 2.6.1.1), lactate dehydrogenase (EC 1.1.1.27), and 2-hydroxybutyrate dehydrogenase

All four enzymes were assayed at 37°C on a mechanized, kinetic enzyme analyzer (System TR; Beckman Instrument Inc.). The creatine kinase assay of Oliver (7), the aspartate aminotransferase assay of Amador & Wacker (8), the lactate dehydrogenase assay of Gay, McComb & Bowers (9) and the aspartate aminotransferase assay of Rosalki et al (10) were used.

Statistical analysis

The terms technical sensitivity, technical specificity and effectiveness were used as previously defined (11). Technical sensitivity is the probability of a positive test among patients with an infarct, and technical specificity is the probability of a negative test among patients without an infarct:

Technical Sensitivity =

$$= \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} = \frac{DT}{DT + \overline{D}T}$$

Technical Specificity =

$$= \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} = \frac{\overline{DT}}{\overline{DT} + D\overline{T}}$$

Effectiveness (E) is the sum of the population proportions of true positives and true negatives:

$$E = P(DT) + P(\overline{DT})$$

Results

To establish the relationships of the immunological creatine kinase-MB assay to other laboratory indicators of myocardial infarction, the activities measured by this assay were correlated with the creatine kinase-MB activities derived from electrophoretic separation, as well as with the total creatine kinase, aspartate aminotransferase, 2-hydroxybutyrate dehydrogenase and lactate dehydrogenase activities (tab. 1). Correlations with creatine kinase-MB by electrophoresis, total creatine kinase and aspartate aminotransferase (three parameters known to be early indicators of myocardial infarction), were essentially equivalent, ranging from 0.62 to 0.68. Conversely, correlations both with 2-hydroxybutyrate dehydrogenase and with lactate dehydrogenase, which are known to rise later following infarction, were lower.

Tab. 1. Correlation coefficients, r, for creatine kinase-MB immunological assays and other enzyme assays in patients with myocardial infarction. N = 114.

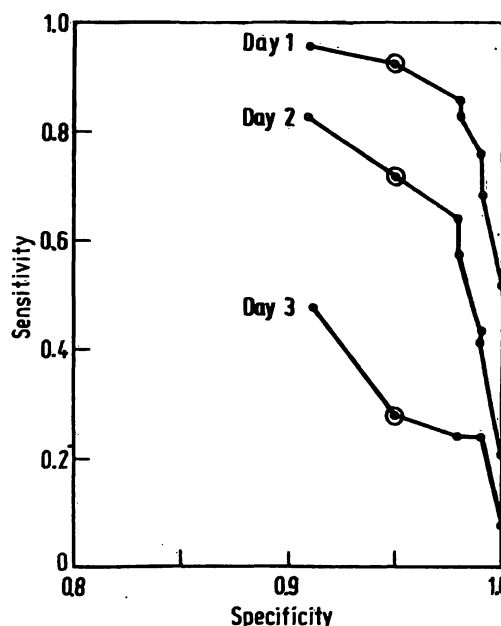| Enzyme	r	95% Confidence interval for r	Significance
Creatine kinase-MB, Electrophoresis	0.68	0.56–0.77	P < 0.001
Creatine kinase	0.62	0.48–0.74	P < 0.001
Aspartate aminotransferase	0.64	0.51–0.74	P < 0.001
2-Hydroxybutyrate dehydrogenase	0.41	0.23–0.56	P < 0.001
Lactate dehydrogenase	0.20	0.01–0.38	P < 0.01

Fig. 1. Receiver-Operator Characteristic (ROC) curves for immunological creatine kinase-MB assays on the three days following infarction. The ideal test which is characterized by a 1.0 sensitivity and a 1.0 specificity would plot on the upper right-hand corner. The operating points corresponding to a decision threshold of 15 U/l are circled. (Note different scales used for specificity and sensitivity).

The interdependence of technical sensitivity and specificity of immunological creatine kinase-MB assays was analyzed by the ROC function (fig. 1). Findings from the first, second and third days (24 hour periods) following infarction were separated. Using different creatine kinase-MB values as decision thresholds to distinguish normal from elevated findings, results from infarcted patients were classified into true positives and false negatives to calculate technical sensitivity for each day. All results from uninfarcted patients were classified into true negatives and false positives to calculate technical specificity. Thus, as the decision threshold was raised stepwise from zero to 60 U/l in analyzing data from day one, technical sensitivity decreased while technical specificity increased, since a decreasing number of patients were classified as positives.

When an activity of 15 U/l was used to separate negatives from positives, the two technical properties were equivalently balanced with a sensitivity of 93% and a specificity of 95%. We have subsequently used this value as our criterion to distinguish normal from elevated creatine kinase-MB values. On the second and third days, the overall discriminatory power diminished increasingly as the ROC curves evidenced a less favorable relationship between sensitivity and specificity. Using again 15 U/l as the discriminatory threshold, sensitivity on the second day was 72%, and on the third day 28%.

Figure 2 illustrates the dependence of the effectiveness of immunological creatine kinase-MB assays upon the
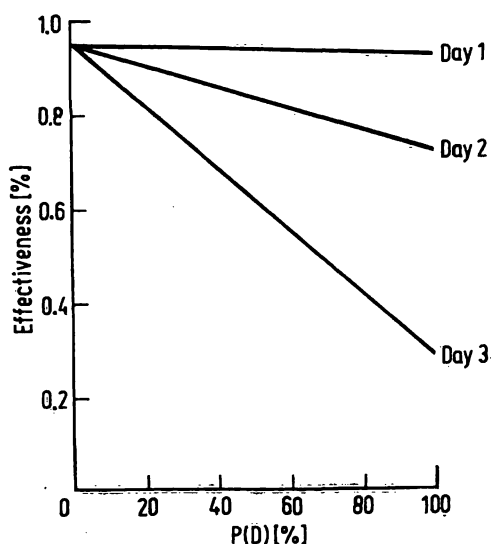

Fig. 2. Effectiveness of immunological creatine kinase-MB assays on different days following infarction as a function of the prevalence of patients with myocardial infarction in the tested population. If all patients are infarcted (P (D) = 100%), effectiveness is equal to technical sensitivity; if no patients are infarcted (O(D) = 0%), effectiveness is equal to technical specificity. Effectiveness for other prevalences is obtained by linear interpolation between these limiting values.

prevalence of patients with myocardial infarction in the tested population. With a discriminatory threshold of 15 U/l effectiveness was practically insensitive to disease prevalence on the first day. Using the same decision threshold on subsequent days produced a dramatic decay of effectiveness at high disease prevalence, whereas effectiveness at low prevalences was not similarly affected.

Discussion

Traditionally, the evaluation of new laboratory assays focuses on technical and biochemical properties. Much less attention is paid to the diagnostic performance of the assays, even though this is the most crucial determinant of practical utility. Opinions about the immunological creatine-kinase-MB assay illustrate these views. From a purely biochemical viewpoint, the inability of this procedure to distinguish activity due to creatine kinase-BB from that due to creatine kinase-MB has caused concern. From a diagnostic viewpoint, a lack of chemical specificity in principle need not invalidate a procedure (12). This is particularly true if interferences, such as creatine kinase-BB in blood, are rarely encountered (6).

Concern with technical aspects of creatine kinase-MB assay has led to modifications of reaction temperature (4, 5, 13, 14), sulfhydryl activators (17, 18), inhibitors of adenylate kinase (EC 2.7.4.3) (l.c. (16)) and blanking procedures. These variables are of secondary relevance for the present purpose. The important technical feature of immunological creatine kinase-MB assay is the quantitative nature of results which facilitates the statistical evaluation of diagnostic performance.

A useful assessment of diagnostic performance rests on two foundations: Scrupulous retrospective case review and analytical summation which facilitates cogent conclusions. Many previous evaluations of laboratory indicators of myocardial infarction fall short of dealing adequately with the following issues: The definition of the normal range used to classify the results, attention to the elapsed time since the pathogenic event, and inclusion of the prevalence of infarction in the tested population. Experimental observations can only become relevant to clinical practice if these points are properly considered.

First, a reference range derived previously from observations in subjects without myocardial infarction is generally used to separate physiological from pathological findings (14, 19, 20). However, from a diagnostic viewpoint such a threshold usually is not the most appropriate. Joint consideration of findings typical for diseased subjects as well as nondiseased subjects is preferable (21). The ROC function provides a convenient way to consider these groups simultaneously (22). Technical sensitivity can be related to technical specificity in a continuous function. Thus, different decision thresholds can be selected to tailor optimal outcome in different

circumstances. Further, the ROC function permits comparison of the same test applied in different conditions such as different stages in the evolution of myocardial infarction (fig. 1). Similarly, the overall discriminatory power of different tests can be contrasted with each other to select the best one.

Second, the diagnostic performance of enzyme assays is typically evaluated on the basis of serial measurements following myocardial infarction. If the calculated diagnostic properties, such as sensitivity and specificity do not relate the statistical odds to sampling at a specific time, biased conclusions result.

Frequently, the rules for the serial sampling or for the choice of selected values on which the calculation of technical sensitivity is based are not detailed (14, 19). In other instances only peak enzyme activities in the series are utilized for analysis even though these occur at different times following infarction and can only be identified retrospectively (20). However, the test's varying capabilities at different points in time is relevant, as patients present themselves for diagnosis at different intervals following infarction. For this reason, we have analyzed all observations from all patients after classifying them according to time since the infarct. The diagnostic power of creatine kinase and creatine kinase-MB is greatest within the first 24 hours after the onset of retrosternal pain (19, 23). The later deterioration is due to a loss of technical sensitivity, while technical specificity remains unaffected (fig. 1).

Third, any evaluation of diagnostic utility should include the prevalence of the diagnosed condition in the tested population. Two parameters of test performance which take prevalence into account have been proposed. These quantities have been called alternatively "positive and negative predictive value" (24) and "diagnostic sensitivity and specificity" (11). The experience contained in these values can be condensed into a *single* number which has been termed the effectiveness of a test (11). This quantity considers the odds of a true positive and of a true negative jointly. Construction of a simple nomogram to survey effectiveness as a continuous function at different prevalences of disease requires only the values of technical sensitivity and specificity. Figure 2 shows that the diagnostic threshold chosen for creatine kinase-MB renders effectiveness independent of disease prevalence on the day of infarction when the assay has its greatest utility. On subsequent days that same threshold provides less technical sensitivity than a lower one might have, and consequently would compromise effectiveness most were there a high prevalence of infarcted patients in the tested population.

The inferences from these propositions are:

1. The relationship between technical sensitivity and specificity which should determine normal limits is conveniently visualized in the ROC curve.
2. The elapsed time since infarction should be explicitly considered when test performance is evaluated.
3. A nomogram of effectiveness provides a simple means to include and evaluate the influence of prevalences on test performance.

References

1. Morin, L. G. (1977), Clin. Chem. 23, 205–210.
2. Griffiths, J. & Handschuh, G. (1977), Clin. Chem. 23, 567–570.
3. Lum, C. & Levy, A. L. (1975), Clin. Chem. 21, 1601–1604.
4. Würzburg, U., Hennrich, N., Orth, H.-D., Lang, H., Prellwitz, W., Neumeier, D., Knedel, M. & Rick, W. (1977), J. Clin. Chem. Clin. Biochem. 15, 131–137.
5. Gerhardt, W., Ljungdahl, L., Börjesson, J., Hofvendahl, S. & Hedenäs, B. (1977), Clin. Chim. Acta 78, 29–41.
6. Lang, H., Würzburg, U., Neumeier, D., Knedel, M., Prellwitz, W., Kattermann, R., Schlebusch, H. & Schürmann, J. (1978), Klin. Wochenschr. 56, 641–646.
7. Oliver, J. T. (1955), Biochem. J. 61, 116–122.
8. Amador, E. & Wacker, W. E. C. (1962), Clin. Chem. 8, 343–350.
9. Gay, R. J., McComb, R. B. & Bowers, G. N. (1968), Clin. Chem. 14, 740–753.
10. Rosalki, S. B. (1960), Nature 188, 1110–1111.
11. Werner, M., Brooks, S. H. & Wette, R. (1973), Hum. Pathol. 4, 17–30.
12. Werner, M., Brooks, S. H. & Cohnen, G. (1972), Clin. Chem. 18, 116–123.
13. Neumeier, D., Prellwitz, W., Würzburg, U., Brundobler, M., Olberman, M., Just, H.-J., Knedel, M. & Lang, H. (1976), Clin. Chim. Acta 73, 445–451.
14. Pfisterer, J., Ritz, R., Scholer, A. & Vonderschmitt, D. (1978), Schw. Med. Wochenschr. 108, 24–27.
15. Frithz, G., Ericsson, P. & Ronquist, G. (1976), Upsala J. Med. Sci. 81, 155–158.
16. Szasz, G., Gerhardt, W., Gruber, W. & Bernt, E. (1976), Clin. Chem. 22, 1806–1811.
17. Klotzsch, S. G. (1976), Clin. Chem. 22, 1236–1237.
18. Weidmann, V. G. (1973), Z. Klin. Chem. Klin. Biochem. 11, 134–135.
19. Ljungdahl, L., Hofvendahl, S., Gerhardt, W. & Borjesson, J. (1977), Clin. Chim. Acta 78, 43–53.
20. Roark, S. F., Wagner, G. S., Izlar, Jr., H. L. & Roe, C. R. (1976), Circulation 53, 965–969.
21. Galen, R. S., Reiffel, J. A. & Gambino, S. R. (1975), J. Am. Med. Assoc. 232, 145–147.
22. Werner, M. (1977), Med. Welt. 28, 1254–1261.
23. Neumeier, D., Prellwitz, W., Sandel, P. & Knedel, M. (1978), Klin. Wochenschr. 56, 449–456.
24. Vecchio, T. J. (1966), N. Engl. J. Med. 274, 1171–1173.

Prof. Dr. M. Werner
Division of Laboratory Medicine
The George Washington University
Medical Center
910 Twenty-third Street, N.W.
Washington, D. C. 20037
USA