Are Reference Limits for Serum Creatine Kinase Valid in Detection of the Carrier State for Duchenne Muscular Dystrophy?

Hanna-Dieter Gruemer,¹ W. Greg Miller,¹ Vernon M. Chinchilli,² Robert T. Lesher,³ Carol R. Hassler,⁴ Peter A. Blasco,¹ Walter E. Nance,⁵ and Barbara M. Goldsmith⁶

We evaluated serum creatine kinase (CK) as an index to heterozygosity in Duchenne muscular dystrophy. When the 97.5th percentile of the CK normal reference interval was selected as the cutoff point, only 31% of 28 obligate carrier mothers and 24% of 43 mothers of simplex cases (those with only one occurrence of dystrophy in the kindred) exceeded this cutoff value. The outcome depended to some degree on the method used for establishing the reference limit for 379 presumably non-carrier, ambulatory women. The considerable overlap of CK activities between the control and carrier population as well as the non-gaussian distribution of the data permitted no satisfactory approach for differentiating these two populations. Neither the application of likelihood ratios, which evaluates a continuum of results without the dichotomy of a cutoff point, nor the application of predictive value based on sensitivity and specificity, which involves use of a cutoff value for decision making, provided a reliable estimate of carrier status. There was no significant difference \((2\alpha = .19)\) between the median CK activity of obligate carrier mothers and mothers of simplex cases. The serum CK test does not provide data that either support or reject the Haldane hypothesis.

Additional Keyphrases: reference interval • carrier detection • likelihood ratios • predictive value • statistical treatment of population differences in cases of overlap

Determination of creatine kinase (CK, EC 2.7.3.2) activity in serum is currently the preferred laboratory method for detecting and confirming carrier status in Duchenne muscular dystrophy (DMD). Methodological problems such as standardization, long-term accuracy, and interlaboratory compatibility in the interpretation of CK results have received little attention. Grannis et al. (1), in their 1978 College of American Pathologists survey, commented on the large biases in determination of serum CK activity among the 450 participating laboratories. They ascribed these biases to differences in reagent composition and instrumentation used. No primary standards are available for assuring long-term consistency of analytical performance. Manufacturers change reagent composition to comply with recommendations of the International Federation of Clinical Chemistry and, as a result, "normal values" for serum CK activity have increased over the years.

Geneticists have developed reasonable criteria for defining carrier status in X-linked disease, but no such criteria have been developed for interpreting serum CK activity results in the evaluation of carrier status. This was shown by Bullock et al. (2), who mailed serum CK samples to 166 laboratories for evaluation of carrier status for DMD. Not only did values for serum CK activities disagree from one laboratory to another, there was also considerable disagreement among laboratories in the assigned probability value for carrier status. Several other studies, attempts to define criteria for a positive CK result, have been hampered by small numbers of samples (usually fewer than 100), which fail adequately to define the markedly skewed upper reference limit. Cutoff values between healthy and carrier populations have been assigned by applying gaussian statistical analysis, either directly to the data or to a logarithmic transformation of it. Logarithmic and other transformations will not "normalize" distributions that are as markedly skewed towards higher values as are results for serum CK activity in a healthy population, where data from approximately 400 observations are necessary to define the skewed tail region adequately (3). One large series in the literature (4), with which many smaller studies are in harmony, confirms the markedly skewed distribution of CK results in control populations. The distribution of CK results for carriers of DMD in this and many previous studies overlap markedly with that for controls, making CK assay rather unsuitable as a laboratory test for carrier status.

Because the use of CK for carrier detection in DMD is clouded by inadequate reference populations and cutoff values, we undertook this prospective study of the relationship between the distribution of CK values in control individuals and carriers of DMD. More broadly, our objective was to apply quantitative statistical methods to understand the problems involved when a medical diagnosis relies on a test for which the results are non-gaussian and distributions overlap between non-diseased and diseased populations.

Materials and Methods

Serum was sampled from the ambulatory control group after a medical history was taken and a physical examination performed on each individual to assure that she was in good health. The samples were collected during a four-year period from 580 females, ages one to 77 years, none of them known to be pregnant. When an individual was tested on more than one occasion, we used only the first recorded observation in this study.

Serum specimens were also obtained from 28 obligate carriers and 43 mothers of simplex cases of DMD during this same four-year period.

Specimens were allowed to clot, centrifuged within a few hours of sampling, and the supernatant serum was separated and stored refrigerated. Control specimens were never stored longer than about 48 h; each patient's specimen was usually analyzed the same day it was collected. Hemolized serum specimens were not used in this study, to avoid the
problem of contamination with adenylate kinase (EC 2.7.4.3) from erythrocytes.

We measured creatine kinase activity by use of a modified coupled Rosalki reaction (5) in a Model KA-150 kinetic enzyme analyzer (Perkin-Elmer Corp., Norwalk, CT), with reagents from Perkin-Elmer. This instrument provides a 6-min preincubation of serum with buffer, co-factors, and li-thiothreitol, followed by addition of substrate and measurement of the rate of reaction at 37 °C. During the early part of this study—i.e., prior to March 1979—we measured CK activity in a continuous-flow analysis system (SMAC; Technicon Instruments Corp., Tarrytown, NY) by a modification of the method of Siegel and Cohen (6), using secondary standards for calibration to produce a result equivalent to that produced by the kinetic analyzer (7).

After the data collection for this study was terminated, we replaced the Perkin-Elmer instrument with a Cobas-Bio centrifugal analyzer (Roche Analytical Instruments, Nutley, NJ), updated the reagent formulation to conform to the recommendations of the Scandinavian Committee on Enzymes, and re-analyzed some of the specimens (8). Regression analysis by use of the Deming approach (9) between the Cobas method (y) and the Perkin-Elmer method (x) for 81 patients' specimens (activity range, 6–1075 U/L) gave \( y = 1.30x + 4 \). We then multiplied all results in this study by the regression coefficient (1.30) before statistical analysis. The negligible intercept error was ignored. This adjustment permitted us to directly apply the reference intervals determined earlier in the study to results obtained with the optimized and widely used recommended method of the Scandinavian Committee on Enzymes.

Results

Figure 1 shows the population density and the correlation of CK activities with age in 580 females. The median for serum CK activities of various age groups decreases during the first two decades of life, with no significant changes thereafter. Because the number of individuals in each age group is small, one or two extreme values may substantially alter the estimate of the 97.5th percentile value. The change in CK activity with age combined with a skewed distribution and substantial overlap between populations make it difficult to interpret CK results for carrier detection in children. The virtually identical medians for the CK activities in the various adult age groups and the large data base (379 observations) justify the establishment of a common upper limit to the reference interval of about 240 U/L.

Frequency plots of the serum CK activities in 379 healthy women, 28 obligate or probable carrier mothers, and 43 mothers of simplex cases (Figure 2) demonstrate the markedly skewed distribution of data toward higher values. The shaded areas under the curves on the left side of Figure 2 illustrate the considerable overlap between the control group (vertical lines) and carrier group (horizontal lines). Use of the logarithmic probability scale on the ordinate causes some distortion. From these data, cumulative frequency plots may be constructed, as is shown on the right-hand side of Figure 2 for the normal women, the obligate carriers, and the mothers of simplex cases. Even though the distribution curves for the obligate carriers and for the simplex mothers are separate in the Figure, statistical analysis with the chi-square (2x = .19) and Mann-Whitney U tests (2x = .21) does not reject the null hypothesis that the CK activities of obligate carrier mothers and mothers of simplex cases have a common median. For the chi-square test, a 2 × 2 contingency table was constructed around the median of the obligate carriers with which the distribution of data for the simplex mothers was compared. The effect of the skewed distribution on interpretation of CK results is illustrated by dividing the range of results that includes 97.5% of the reference population into equal 60-U/L quarters. The first quarter contains almost half (48%) of the reference population, the second quarter 39%, the third quarter 8.5%, and the fourth quarter only 2%. Some investigators (9–12) replotted the CK enzyme activities as their logarithm (log10) in an attempt to transform a skewed distribution into a normal one. When we did so, the normalization was so incomplete (Figure 3) that we abandoned this approach. Our attempts to describe the distribution by five additional mathematical models were also unsuccessful (3).

As stated in the Appendix, likelihood ratios describe a continuum rather than a single cutoff point to assess the probability of an individual being normal or diseased. A linear relationship between CK activities and likelihood ratios (L(X)) is obtained (Figure 4) if the relative probabilities of the CK activities in the healthy-women group in relation to one of the carrier groups are calculated as:

\[
L(X) = \frac{P(X|D)}{P(X|\bar{D})}
\]

where X represents the CK activity, P(X|D) denotes the probability of X in the presence of carrier state and P(X|\bar{D}) the probability of X in the absence of carrier state. The ratios for the mothers of simplex cases, obligate carrier...
combined (posterior) probability for disease is given by:

$$P(D|X) = \frac{p L(X)}{(1 - p) + p L(X)}$$

(2)

Prevalence for this consideration is a characteristic of the mode of inheritance—in the case of DMD all daughters of obligate carriers have a 50% chance of being carriers. In our examples for CK activities of 120 and 240 U/L and a prevalence of carrier status of 0.50, the combined probabilities (and their 95% confidence limits) are 50% (38 to 61%) and 79% (67 to 86%), respectively.

A less-conservative estimate for combined probabilities may be calculated from prevalence, test sensitivities, and test specificities:

$$PV = \frac{[(\text{sensitivity}) \cdot \text{(prevalence)}]/[(\text{sensitivity}) \cdot \text{(prevalence)}] + \{1 - \text{prevalence} \cdot (1 - \text{specificity})\}}{\text{(prevalence)}}$$

(3)

The equation in this form is better known as the "predictive value of a positive test" (13, 14). For cutoff values of 120 and 240 U/L and for a prevalence of 0.5, the predictive values (and 95% confidence limits) equal 83% (74 to 92%) and 91% (80 to 99%), respectively. Likelihood ratios, sensitivities, and specificities may be calculated from the curve on the right-hand side of Figure 2 as detailed in the Appendix.

The long-term stability of the CK results for the control group is shown in Figure 5, a plot of the medians of sequential groups of 40 individual results. We consider these data to be evidence that the population and the analytical technique were stable during this four-year period. We observed an increased number of extreme values during the summer of 1979 and spring of 1981, but not the consistent seasonal changes reported by others (15, 16). The quality-control data show no methodological problems during this study, and we therefore assume that these results reflect the biological variability in the population (3).

Discussion

It is generally agreed that the distribution of serum CK activities is skewed towards high values (4, 10–13, 17–18). Such anomalies require a larger population size than is available to many laboratories before the upper limit to the normal reference interval can be estimated, and so the log transformation is commonly used in an attempt to "normalize" the data to fit a gaussian distribution. We find that normalization by this approach is incomplete (Figure 3), as did Griffiths (18). Other authors (10–13) have claimed complete normalization, but inspection of the data reported by Percy et al. (10) and Nicholson et al. (11) suggests that some extreme values were ignored. So as not to introduce a bias into our analysis, we included all results, without exception, in our statistical evaluation.

We did not attempt to limit physical exercise. Such efforts might prove futile because normal control individuals are frequently reluctant to change their life-style habits and the effect of exercise on CK activity is not predictable (19). Our populations reflect as closely as possible the real-life situation.

The preferred statistical method in this study has been the cumulative frequency plot illustrated in Figure 2 because it is relatively robust to outliers and permits easy graphic interpolation of mildly dispersed data for an optimal nonparametric estimate of the upper normal reference interval. Here the normal reference interval includes the central 95% of results and describes a statistical observation of the population distribution. It thus provides for clinical purposes a convenient reference point with which the magni
These individuals, different no already ining

...likelihood statistic serum...individuals prior (Figure 4). If one wants to test for the reliability of the ratio of 1.8 with 95% confidence, then 95 times out of 100 the ratio will be between 1.0 and 2.7. Such wide confidence intervals minimize the information value obtained from CK measurements. Even if the simplex and obligate carriers were combined into one group (N = 71), just to increase population size, the 95% confidence interval improves only slightly—1.25 to 2.6 at 180 U/L.

The advantage in the use of likelihood ratios as an estimate of carrier status in women at risk lies in the ability to relate the magnitude of an individual value to the continuum of results, both in non-carrier and carrier states. However, likelihood ratios become increasingly unreliable, for the purpose of genetic counseling, as the serum CK activities decrease into the normal range, because the large population overlap increases the possibility of misinformation on carrier status in individual cases, regardless of whether the test is used by itself or in combination with others.

The misinformation may even be magnified if the concepts of sensitivity (positivity of a test in disease) and specificity (negativity of a test in health) are used (13, 14), both of which require by definition a common cutoff value. Sensitivity and specificity have a seesaw relationship when the distribution curves of normality and abnormality overlap. For CK activities, the change in specificity above 120 U/L is relatively small as compared to the change in sensitivity (Figure 6). It is at high specificity values exceeding 120 U/L that the serum CK measurement eventually becomes useful for confirming carrier status. Because of the high incidence of the disease in the population seeking genetic counseling, the predictive value of a result exceeding 120 U/L for the mothers of simplex cases and daughters of obligate and possible carriers is about 90%. As CK cutoff values increase above 120 U/L (likelihood ratio of 1.0), the predictive value increases also. For instance, at a carrier prevalence of 0.5 (50%) and a CK cutoff value of 180 U/L, the predictive value of a positive test is 85%. Calculation of the combined probability from the likelihood ratio at 180 U/L is lower, namely 64%. The difference in value between the two methods is explicable: likelihood ratios reflect a continuum of results, but the predictive-value approach from test sensitivities and specificities dichotomizes information into values below and above the cutoff value. The latter is applicable to any result that exceeds the cutoff point. Thus, when applied to a CK result near the cutoff point, the predictive value overestates the combined probability, while the likelihood ratio yields a more nearly accurate result.

The value of either method in estimating the probability of DMD carrier status from CK results is limited by the substantial overlap between healthy and carrier populations, which greatly contributes to the broadness of the confidence interval for the likelihood ratio and the predictive value.

This analysis illustrates that the CK test is of marginal value in delineating carrier status in DMD unless the results exceed those for the healthy population interval, a

**Table 1. Calculation by Various Statistical Methods of the Upper (97.5th Percentile) Reference Limits for Serum CK Activities of 379 Healthy Women**

<table>
<thead>
<tr>
<th>Method</th>
<th>Upper limit, U/L</th>
<th>% below this limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonparametric (percentile)</td>
<td>224</td>
<td>68</td>
</tr>
<tr>
<td>Mean ± 2 SD</td>
<td>185</td>
<td>58</td>
</tr>
<tr>
<td>Mean ± 2 SD (log₁₀-transformation)</td>
<td>195</td>
<td>61</td>
</tr>
<tr>
<td>Cumulative frequency plots</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-transformed data</td>
<td>240</td>
<td>69</td>
</tr>
<tr>
<td>Linear portion of log₁₀-transformed data (Fig. 3)</td>
<td>141</td>
<td>47</td>
</tr>
</tbody>
</table>

Fig. 5. Long-term precision of CK activities: Data on 440 sera from 379 healthy women

These data were accumulated over a four-year period. Each dot represents 40 individuals. Some 61 determinations were from duplicate sets of determinations (440 - 379 = 61) not used in the establishment of reference ranges.
criterion met by only a small proportion of carriers.

Some authors (10, 11, 15) have attempted to determine the mutation rate for DMD on the basis of CK activity determinations in obligate carriers and mothers of sporadic cases. The assumption underlying this approach is that a single cutoff point can be used to differentiate carriers from non-carriers. Figure 2 clearly demonstrates that such cannot be the case, owing to the extensive overlap of the distribution curves of the normal female population and the obligate-carrier population. The lack of a significant difference between the CK activities of obligate carriers and mothers of simplex cases indicates that CK activity measurements will not distinguish these two populations. We cannot confirm or reject Haldane’s hypothesis—that a third of all afflicted offspring with a genetically lethal disease are the result of a spontaneous mutation—as applied to muscular dystrophy on the basis of total CK activity measurements.

We are indebted to Dr. Daniel B. Drachman, Johns Hopkins University School of Medicine, and Dr. Thaddeus Kelly, University of Virginia School of Medicine, for providing us with blood samples from mothers of patients with Duchenne dystrophy. This investigation was supported in part by the Muscular Dystrophy Association, Inc.

Appendix

The Likelihood Ratio Technique for Diagnosing Diseased and Non-Diseased States

Let D and \( \overline{D} \) denote the diseased and non-diseased states, respectively, and let \( X = (X_1, X_2, \ldots, X_k) \) denote a vector of \( k \) different diagnostic variables. The likelihood ratio is defined as the ratio of the conditional probability of \( X \) given D to the conditional probability of \( X \) given \( \overline{D} \), i.e.,

\[
L(X) = \frac{P(X|D)}{P(X|\overline{D})}.
\]

\( L(X) \) is important in the determination of the disease state for an individual with a particular realization of \( X \). \( L(X) \) is always a positive number and if it is much larger than 1, then the individual is more likely to be diseased. If \( L(X) \) is near to zero, then the individual is more likely to be non-diseased. If \( L(X) \) is close to 1, then the particular realization of \( X \) does not indicate any useful information about the disease state.

It has been well documented in the statistical literature, by such authors as Efron (26) and Press and Wilson (27), that

\[
L(X) = \exp(\alpha_0 + \alpha_1 X_1 + \ldots + \alpha_k X_k)
\]

is an efficient and robust model of the likelihood function. This formulation of \( L(X) \) is called the logistic likelihood because the log of the ratio of the conditional probabilities yields the linear combination \( \alpha_0 + \alpha_1 X_1 + \ldots + \alpha_k X_k \), where \( \alpha = (\alpha_0, \alpha_1, \ldots, \alpha_k) \) is the vector of unknown parameters to be estimated from the sample data.

Albert (28) provides an excellent discussion of the use of the logistic likelihood function in clinical chemistry and diagnostic situations. Further, he shows that the estimators \( \hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_k \) of the parameters can be obtained from any computer routine that conducts a logistic regression or logistic discrimination. Such programs are available in the statistical software packages SAS and BMDP. If there remains some doubt as to the appropriateness of the logistic likelihood for a particular data set, the goodness-of-fit test of Tsatis (29) can be invoked.

Assuming that the logistic function models the likelihood ratio adequately, the likelihood for an individual is estimated by

\[
\hat{L}(X) = \hat{\alpha}_0 + \hat{\alpha}_1 X_1 + \ldots + \hat{\alpha}_k X_k.
\]

For convenience, rewrite the row vector \( X \) as \( X = (X_0 = 1, X_1, X_2, \ldots, X_k) \), and let \( X' \) denote the transposed vector so that \( X' \) is a column vector. Then \( \hat{L}(X) \) can be written in vector notation as \( \hat{\alpha}X' \). In order to construct a confidence interval for \( L(X) \), it is necessary to know the estimated variances and covariances of the estimators \( \hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_k \). These can be expressed in matrix form as

\[
\begin{bmatrix}
V_{00} & V_{01} & \cdots & V_{0k} \\
V_{10} & V_{11} & \cdots & V_{1k} \\
\vdots & \vdots & \ddots & \vdots \\
V_{k0} & V_{k1} & \cdots & V_{kk}
\end{bmatrix}
= \begin{bmatrix}
\text{Var}(\hat{\alpha}_0) & \text{Cov}(\hat{\alpha}_0, \hat{\alpha}_1) & \ldots & \text{Cov}(\hat{\alpha}_0, \hat{\alpha}_k) \\
\text{Cov}(\hat{\alpha}_0, \hat{\alpha}_1) & \text{Var}(\hat{\alpha}_1) & \cdots & \text{Cov}(\hat{\alpha}_1, \hat{\alpha}_k) \\
\vdots & \vdots & \ddots & \vdots \\
\text{Cov}(\hat{\alpha}_0, \hat{\alpha}_k) & \text{Cov}(\hat{\alpha}_1, \hat{\alpha}_k) & \cdots & \text{Var}(\hat{\alpha}_k)
\end{bmatrix}
\]

and usually they are included in the output from the logistic regression/logistic discrimination computer routines. This results in a nice expression for the variance of the log of the estimated likelihood ratio as the quadratic form

\[
\text{Var}[\log \hat{L}(X)] = XVX' = \sum_{i=0}^{k} \sum_{j=0}^{k} X_iX_jV_{ij}.
\]

For the simple case of only one diagnostic variable (\( k = 1 \)), \( \text{Var}[\log \hat{L}(X)] \) reduces to \( V_{00} + 2X_1V_{01} + X_1^2V_{11} \). Regardless of the number of diagnostic variables, the 100(1 - \( \gamma \)% confidence interval for \( L(X) \), \( 0 < \gamma < 1 \), is given by

\[
\hat{L}(X) \exp\left(\pm z_1 - \gamma/2 \sqrt{\text{Var}[\log \hat{L}(X)]}\right)
\]

where \( z_1 - \gamma/2 \) is the upper 1 - \( \gamma/2 \) reference point from the standard normal distribution. Rather than relying upon \( \hat{L}(X) \) alone, the confidence interval for \( L(X) \) provides a more reliable assessment of an individual’s likelihood for disease. Also, note that the confidence interval for the likelihood ratio yields a confidence interval for the combined risk.

---

Fig. 6. Relation between sensitivity and specificity for serum CK activities between 40 and 300 U/L, plotted as a Receiver Operator Characteristic (ROC) curve.

Sensitivities were calculated from data on the population of 28 obligate carriers, the specificities from data on the 379 normal women. The predictive value (PV) was estimated from these data for daughters of obligate carriers, assuming prevalence of 5. The 95% confidence limits were calculated for predictive values [% PV (0.5)] of 83, 88, and 91%, resulting in ranges of 75 to 92%, 76 to 97%, and 75 to 99%, respectively. The dotted line demonstrates perfect ROC (24), for comparison.
(posterior probability) defined in equation 2. Let \( L(X) \) and \( L_0(X) \) denote the lower and upper endpoints of the confidence interval for \( L(X) \), respectively. If \( p \) is the prior probability of disease, then the confidence interval for the combined risk of disease, given a particular value of \( X \), is

\[
\left[ \frac{pL_0(X)}{(1 - p) + pL_0(X)}, \frac{pL(X)}{(1 - p) + pL(X)} \right].
\]

Confidence intervals are more informative than point estimators because they reflect the amount of information in the samples and the goodness-of-fit of the statistical models. Therefore, confidence intervals for the likelihood ratio and the combined risk should be examined in diagnostic situations. As a point of illustration, a particular individual may have an estimated combined risk of .75. However, suppose the 95% confidence interval for this combined risk is [.35, .87]. This interval contains .50, which indicates that there is not much predictive power with the diagnostic test for this individual. The wide confidence interval could be due to small sample sizes, a poor fit of the statistical model, or overlap of the two distribution curves.

With CK as the diagnostic variable in the sample of 379 non-carriers and 28 obligate carriers, the estimated likelihood function is

\[
\hat{L}(CK) = \exp(-1.3549 + 0.0108 \cdot CK)
\]

with

\[
\text{Var} \{\log \hat{L}(CK)\} = 0.122797 - 0.0010316 \cdot CK + 0.0000038 \cdot CK^2.
\]

For an individual with \( CK = 180 \), it turns out that \( \hat{L}(180) = 1.8024 \) and \( \text{Var} \{\log \hat{L}(180)\} = 0.0602 \). Then the 95% confidence interval for the likelihood at 180 is \([\hat{L}(180), \hat{L}_0(180)] = [1.1143, 2.9154] \). If the individual has prior probability of .5 of being a carrier, then the estimated combined risk is .6432 and the 95% confidence interval for the combined risk is \([.5270, .7446] \).

The Dichotomous Approach for Diagnosing Diseased and Non-Diseased States

The classical method for identifying diagnostic variables has been to examine the relationship between one dichotomous variable, i.e., positive (P) or negative (N) outcomes, and the disease state. The important measures of a variable’s diagnostic value are

\[
\rho_1 = \text{Sensitivity} = P(D|P)
\]

and

\[
\rho_2 = \text{Specificity} = P(N|\neg D).
\]

Sensitivity represents the proportion of diseased patients with a positive test result, and specificity represents the proportion of healthy individuals with a negative result. Sensitivity indicates how “sensitive” the diagnostic test is to the diseased state, while specificity indicates how “specific” the diagnostic test is to the disease of interest and not the healthy state or some other disease. If \( p \) is the prior probability of disease, then the combined risk (posterior probability or predictive value) is

\[
P(D|P) = \frac{\rho_1}{\rho_1 + (1 - \rho_1)(1 - \rho_2)}.
\]

Suppose that the results of a diagnostic test can be summarized by the following two-by-two table:

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>\neg D</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>n_{11}</td>
<td>n_{12}</td>
</tr>
<tr>
<td>N</td>
<td>n_{21}</td>
<td>n_{22}</td>
</tr>
</tbody>
</table>

Then the maximum likelihood estimators for sensitivity and specificity are

\[
\hat{\rho}_1 = \frac{n_{11}}{n_1}, \quad \hat{\rho}_2 = \frac{n_{22}}{n_2},
\]

respectively. For adequate sample sizes (\(n_1 \) and \(n_2\)), the 100(1 - \( \gamma \))% confidence intervals for sensitivity and specificity are given by

\[
\left[ \hat{\rho}_1 \pm z_1 - \gamma \sqrt{\frac{\hat{\rho}_1(1 - \hat{\rho}_1)}{n_1}}, \hat{\rho}_1 \right] \quad \text{and} \quad \left[ \hat{\rho}_2 \pm z_1 - \gamma \sqrt{\frac{\hat{\rho}_2(1 - \hat{\rho}_2)}{n_2}} \right],
\]

respectively. In order to find the confidence level for the combined risk, however, the joint 100(1 - \( \gamma \))% confidence intervals for \( \rho_1 \) and \( \rho_2 \) are needed. Because \( \hat{\rho}_1 \) and \( \hat{\rho}_2 \) are independent, this is given by the separate 100(1 - \( \gamma \))% confidence intervals for \( \rho_1 \) and \( \rho_2 \). In other words, the joint confidence level is

\[
[\hat{\rho}_{1\ell} \leq \rho_1 \leq \hat{\rho}_{1u}, \hat{\rho}_{2\ell} \leq \rho_2 \leq \hat{\rho}_{2u}],
\]

where \( [\hat{\rho}_{1\ell}, \hat{\rho}_{1u}] \) and \( [\hat{\rho}_{2\ell}, \hat{\rho}_{2u}] \) are the individual 100(1 - \( \gamma \))% confidence intervals for \( \rho_1 \) and \( \rho_2 \). Then the 100(1 - \( \gamma \))% confidence interval for the combined risk is

\[
\left[ \frac{\hat{\rho}_{1\ell} \hat{\rho}_{2\ell}}{\hat{\rho}_{1\ell} + (1 - \hat{\rho}_{1\ell})(1 - \hat{\rho}_{2\ell})}, \frac{\hat{\rho}_{1u} \hat{\rho}_{2u}}{\hat{\rho}_{1u} + (1 - \hat{\rho}_{1u})(1 - \hat{\rho}_{2u})} \right].
\]

Using the criteria that \( CK \geq 180 \) is a negative test result, the following two-by-two table summarizes the data presented in this manuscript:

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>\neg D</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>362</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>379</td>
</tr>
</tbody>
</table>

The estimates of sensitivity and specificity are

\[
\hat{\rho}_1 = 13/28 = 0.46 \quad \text{and} \quad \hat{\rho}_2 = 362/379 = 0.95.
\]

Notice that the cutoff value of 180 provides an insensitive, but specific, test. If a 95% confidence level for the combined risk is desired, then the individual 97.47% confidence intervals for \( \rho_1 \) and \( \rho_2 \) are needed. These are given by

\[
[\hat{\rho}_{1u}, \hat{\rho}_{1u}] = [0.2490, 0.6710] \quad \text{and} \quad [\hat{\rho}_{2u}, \hat{\rho}_{2u}] = [0.9249, 0.9751].
\]

When an individual has a prior probability of .5 of being a carrier, the estimate of the combined risk is 0.902 and its 95% confidence interval is [0.7883, 0.9642].

References
3. Miller WG, Chinchilli VM, Gruener HD, Nance WE. Sampling from a skewed population distribution as exemplified by estimation