
CLIN. CHEM. 39/2, 297–304 (1993)

Total Error Assessment of Five Methods for Cholesterol Screening

W. Greg Miller,1 James M. McKenney,2 Margaret R. Conner,2,4 and Vernon M. Chinchilli2

We report the accuracy, imprecision, total analytical errors, and patient misclassification errors for cholesterol measured from capillary whole blood, venous whole blood, and venous plasma samples by five devices used in public cholesterol screening environments: Reflotron, Vision, Ektachem DT-60, QuickRead, and Liposcan. None of the methods met the National Cholesterol Education Program (NCEP) performance recommendations of 3% CV with 3% bias. The Vision and Reflotron methods used with venous samples gave individual results with total errors consistent with a combined CV and bias in the 4–5% range; capillary blood samples had total errors >5% (combined CV and bias criteria). The DT-60 performance was near the 5% total error criterion for capillary samples and was >5% for venous samples. Misclassification of individuals into desirable or referral groups for venous samples was as great as 5.1% for the DT-60, 5.7% for the Vision, and 7.1% for the Reflotron. Misclassifications for capillary blood samples were as great as 6.7%, 18.3%, and 14.1% for DT-60, Vision, and Reflotron, respectively. The QuickRead and Liposcan results were substantially poorer than those obtained by the other methods.

Indexing Terms: variation, source of · intermethod comparison · fingerstick blood collection

Heart disease is a major health problem in the United States. One of the modifiable risk factors associated with heart disease is high blood concentrations of cholesterol. An estimated 36% of the population has hypercholesterolemia, as defined by the National Cholesterol Education Program (NCEP) (1). Clinical trials demonstrate that the incidence of heart disease is reduced when patients with high blood cholesterol are treated with diet and drugs (2). The NCEP has recommended that all patients older than 20 years have their cholesterol concentration measured at least every five years (3). This recommendation has created a need for accurate and widely available cholesterol testing in our society.

Several compact devices that measure total cholesterol from capillary blood samples are now available to make cholesterol testing more accessible to the public.

1 Section of Clinical Chemistry, Department of Pathology, and Departments of 2 Pharmacy and Pharmaceutics, and 3 Biostatistics, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA 23298-0597.
2 Present address: Lloyd M. Parks College of Pharmacy, Ohio State University, Columbus, OH 43210.

Received February 3, 1992; accepted December 4, 1992.

CLINICAL CHEMISTRY, Vol. 38, No. 2, 1993
The Laboratory Standardization Panel of the NCEP recommends that the accuracy of cholesterol testing be within 3% of the result by the Reference Method at the Centers for Disease Control and that imprecision (CV) be within ±3% (4). These guidelines apply to all settings in which cholesterol is measured, including public screenings and physicians' offices, where desk-top devices are likely to be used, and to all sample types, including capillary blood. We report here the accuracy, imprecision, total analytical errors, and misclassification rates of cholesterol measured in capillary whole blood, venous whole blood, and venous plasma by five devices used in public cholesterol screening environments.

**Materials and Methods**

The cholesterol-measuring devices tested were Reflotron (Boehringer Mannheim Diagnostics, Indianapolis, IN), Vision (Abbott Labs., Irving, TX), Ektachem DT-60 (Eastman Kodak, Rochester, NY), QuickRead (Photest Diagnostics, Little Falls, NJ), and Liposcan (Home Diagnostics, Eatontown, NJ). Each manufacturer provided equipment, reagents, controls, calibrators, and instruction. All devices were operated exactly according to the manufacturer's instructions by two pharmacists, each of whom operated each system on various days. Each operator had limited previous laboratory testing experience. During an initial familiarization period, each investigator performed ~50 tests with each device. The Liposcan required visual interpretation of a color change on a reagent strip and gave semiquantitative results. One of the investigators, who had passed a standard test for color blindness and was unaware of the results being generated by the other systems, consistently interpreted the Liposcan results.

A single lot of reagent, which had been validated by the Virginia State Health Department Cholesterol Screening Standardization Program, was used for the Reflotron. Single lots of reagents were also used for the DT-60 and Liposcan assays. Two lots of reagents were used for the Vision and QuickRead systems, the lots being changed after the analysis of samples from the first 40 subjects. Each lot of Vision reagents gave comparable results and the data were combined for analysis; the first lot of QuickRead serum/plasma reagents gave higher results than the second lot, and only data for the second lot were used. In all cases, the reagents, calibrators, and controls were lots released by the manufacturers for customer use.

The comparison method for determining accuracy was an enzymatic cholesterol procedure in which Boehringer Mannheim Diagnostics reagents (no. 692905) and calibrators (no. 125512) were used with a Cobas-Bio analyzer (Roche Diagnostics, Nutley, NJ). The enzymatic procedure was standardized by participation in the Centers for Disease Control Lipid Standardization Program. All comparison assays were performed in duplicate.

We recruited 100 asymptomatic, ambulatory adult subjects (>18 years) for the study. No subject was evaluated by all five devices because not all systems were used on every test day. The number of test days that each method used is indicated in the Figures. For each instrument tested, we analyzed duplicate capillary blood samples collected by fingerstick from each subject with an automatic lancet and standard aseptic technique. Subjects had been sitting for ≥5 min and were oriented to the testing procedure. The order in which capillary samples were collected was randomized. Blood cholesterol was quantified from capillary samples typically within 10 min and always within 1 h after collection.

Subjects then walked to an outpatient blood-collection station, where venous blood samples were collected into both heparin and EDTA-containing evacuated tubes within 2 h of collection of the capillary samples. The heparinized venous whole blood was mixed, sampled with the appropriate capillary pipette, and assayed in duplicate with each device. The heparinized whole blood was then centrifuged and the samples assayed in duplicate with each device. Because the Ektachem DT-60 can assay only serum or plasma, we did not assay venous whole blood with this instrument. Because the Liposcan system can assay only whole blood, plasma samples were not measured with this device. The heparinized venous plasma samples were stored refrigerated and analyzed within 24 h of collection by the accuracy comparison method for cholesterol and for triglycerides; they were visually inspected for turbidity and bilirubinemia. The EDTA venous blood samples were assayed for hemocrit with an ELT-800 device (Ortho Diagnostic Systems, Braintree, MA).

Three subjects' results, for which the heparinized venous plasma sample was hemolyzed or turbid, were excluded from the data. All other venous samples were clear: nonhemolyzed and nonicteric. All samples had triglyceride concentrations <3 g/L except four, which contained 3.29, 3.38, 3.69, and 6.90 g/L. Capillary blood samples could not be evaluated directly for hemolysis because the whole blood was analyzed without delay. Data for individual samples were included in the analysis only if results were obtained for both duplicate samples; this exclusion led us to omit 6 capillary and 1 venous plasma results for Reflotron, 27 capillary results for Vision, 5 capillary and 1 venous plasma results for DT-60, 11 capillary and 2 venous blood results for QuickRead, and 10 capillary and 8 venous blood results for Liposcan. The lack of results for one (or both) of the paired samples was occasioned either by technical limitations, such as a clotted capillary tube, or by a hemolysis message with no result (Vision). For all data analyses, we used the mean of the duplicate assays unless stated otherwise.

Three frozen pooled human serum samples were assayed in duplicate at the beginning and end of each of 10 days (n = 40, Reflotron), 9 days (n = 36, Vision), or 6 days (n = 24, DT-60, and QuickRead). Pools were prepared from freshly collected human serum that had been frozen and thawed before pooling. Aliquots were stored at -70 °C. A new aliquot was thawed each day of
analysis and stored at room temperature during the day.

Interference from triglycerides and hemoglobin was evaluated in both whole blood and plasma samples. Variable volumes of erythrocyte hemolysate and isotonic saline were added to heparinized blood or plasma to produce a constant 3.2% dilution of the samples, to give final hemoglobin concentrations of 0, 0.5, 1.0, and 3.3 g/L. An analogous technique with Intralipid (Cutter Biologics, Berkeley, CA) produced a constant 1.6% dilution of the samples, to give final triglyceride concentrations of 0, 2.6, 4.6, and 10.5 g/L. Each specimen was assayed in triplicate.

Results

During the time of this evaluation, 30 frozen samples from individual patients [cholesterol 4.27–9.02 mmol/L (1.65–3.49 g/L)] were assayed with both the standardized enzymatic method used for accuracy evaluation and the Centers for Disease Control Abell–Kendall Reference Method. The standardized enzymatic method used in this study had an average −0.8% bias when compared with the Reference Method, with 29 of 30 samples being within ±3% and one sample having a −4.0% bias. The imprecision (CV) of the comparison method was 2.5%.

The subjects sampled in this evaluation had cholesterol values of 3.05–8.33 mmol/L (1.18–3.22 g/L). Plots of the percent bias compared with the venous plasma result assayed with the standardized method are shown in Figures 1–4. Substantial scatter in the data is immediately apparent and some results for individual patients are significantly biased from the correct cholesterol value. Because the Liposcan method was semiquantitative, it was evaluated only for the misclassification rate of individual results.

The average accuracy of the cholesterol screening instruments is shown in Table 1. The DT-60 results for capillary and venous samples and the Reflotron results for venous samples were within the NCEP 3% recommendations. Both the Reflotron and Vision had higher results for capillary samples than for venous whole-blood samples. The QuickRead results were clearly unacceptable for capillary or venous whole-blood samples. The QuickRead product insert states that the method gives erroneous results for blood samples with hematocrit <39%; however, the performance remained unacceptable even when only results for samples with hematocrit ≥39% were considered. Thirty-three percent of the apparently healthy ambulatory population sampled in this study had hematocrits <39% (overall range 27–49%), and all QuickRead data were used for subsequent evaluation of that method.

Linear-regression analysis of percent bias vs hematocrit had correlation coefficients <0.16 for all sample types for the Vision and DT-60 methods and for plasma samples for the QuickRead method. The QuickRead method had correlation coefficients of 0.58 for capillary and 0.52 for venous whole-blood samples, with both slopes significantly different from zero (P = 0.0001), which supports the manufacturer’s statement of sensitivity to hematocrit. The Reflotron had a correlation coefficient of −0.3 with all sample types, including plasma, and slopes significantly different from zero (P = 0.011 to 0.005), which may represent a small concentration-dependent calibration effect rather than sensitivity to hematocrit. Linear-regression analysis of percent bias vs comparison method cholesterol values had correlation coefficients and slopes similar to those for the hematocrit plots for all methods, which suggests that calibration effects are probably more important than hematocrit in the performance of these methods.

The total imprecision of the screening methods measured by replicate analysis of three frozen human serum pools is shown in Table 2. Table 3 presents the within-run imprecision estimated from the duplicate assays of each donor’s samples. A random effects analysis of variance (ANOVA) (5) was used to determine the within-subject variance, and the CV was calculated for the mean cholesterol of the population sampled. The total and within-run imprecision estimates were generally similar. None of the methods consistently met the NCEP 3% CV recommendations.

The total analytical errors of these methods is shown in Table 4 as the 95% tolerance intervals for the ratio of intra-subject mean results of test to comparison methods. A statistical description of this tolerance interval determination is presented in the Appendix. The tolerance interval describes the performance range within which individual patients’ values fall 95% of the time. The acceptable limits for the tolerance interval include contributions from the accuracy and imprecision of the

<table>
<thead>
<tr>
<th>Table 1. Average Percent Bias for Cholesterol Determined in Individual Subjects’ Samples*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias, % (and n)</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Reflotron</td>
</tr>
<tr>
<td>Vision</td>
</tr>
<tr>
<td>DT-60</td>
</tr>
<tr>
<td>QuickRead</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

n/a, not applicable; hct, hematocrit.
* No. of days on which fresh samples were collected and tested are given in Figs. 1–4.

<table>
<thead>
<tr>
<th>Table 2. Total Imprecision (CV, %) Determined with Frozen Serum Pools*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pool 1 (5.04 mmol/L, 1.95 g/L)</td>
</tr>
<tr>
<td>-------------------------------</td>
</tr>
<tr>
<td>Reflotron</td>
</tr>
<tr>
<td>Vision</td>
</tr>
<tr>
<td>DT-60</td>
</tr>
<tr>
<td>QuickRead</td>
</tr>
</tbody>
</table>

* Pools were assayed in duplicate at the beginning and end of each of 10 (n = 40, Reflotron), 9 (n = 36, Vision), or 6 days (n = 24 each, DT-60 and QuickRead).
* Cholesterol concentration in pool is listed in parentheses.
Table 3. Within-Run Imprecision from Assay of Subjects' Samples in Duplicate

<table>
<thead>
<tr>
<th></th>
<th>Capillary blood</th>
<th>Venous blood</th>
<th>Venous plasma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflotron</td>
<td>5.1</td>
<td>5.9</td>
<td>5.8</td>
</tr>
<tr>
<td>Vision</td>
<td>7.3</td>
<td>2.5</td>
<td>1.9</td>
</tr>
<tr>
<td>DT-60</td>
<td>4.0</td>
<td>n/a</td>
<td>4.1</td>
</tr>
<tr>
<td>QuickRead</td>
<td>9.9</td>
<td>6.2</td>
<td>8.5</td>
</tr>
</tbody>
</table>

n as in Table 1.

Table 4. Total Error for Individual Subjects' Measurements

<table>
<thead>
<tr>
<th></th>
<th>Capillary blood</th>
<th>Venous blood</th>
<th>Venous plasma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflotron</td>
<td>0.86–1.23</td>
<td>0.81–1.18</td>
<td>0.85–1.16</td>
</tr>
<tr>
<td>Vision</td>
<td>0.85–1.34</td>
<td>0.88–1.21</td>
<td>0.92–1.16</td>
</tr>
<tr>
<td>DT-60</td>
<td>0.85–1.22</td>
<td>n/a</td>
<td>0.78–1.26</td>
</tr>
<tr>
<td>QuickRead</td>
<td>0.91–1.47</td>
<td>0.92–1.42</td>
<td>0.74–1.33</td>
</tr>
</tbody>
</table>

Number of observations as in Table 1. Performance that met combined 3% precision (CV) and accuracy performance would have had tolerance intervals between 0.87 and 1.13 (see text for details).

method being evaluated and from the imprecision of the comparison method. The ratio of test to comparison methods results will have a value of 1.00 if the two results agree perfectly. The range of acceptable tolerance intervals consistent with specified performance guidelines is given by equations 2 and 3 in the Appendix. For total cholesterol with specifications of 3% bias with 3% CV, a comparison method CV of 2.5%, and K factor of 2.349, the acceptable tolerance interval range is 0.87–1.13. Similarly, for specifications of 4% and 5% combined bias and CV, the acceptable tolerance intervals would be 0.83–1.17 and 0.79–1.21, respectively. The K factor is dependent on the number of observations and varied from 2.236 (n = 98) to 2.349 (n = 56). However, this range of K factors has minimal impact on the tolerance interval evaluation criteria; e.g., at K = 2.236, the acceptable tolerance interval is 0.88–1.12 for combined 3% bias and CV performance criteria. Thus, none of the methods had a total error consistent with the NCEP 3% performance guidelines. The Reflotron and Vision had total errors for venous samples consistent with a 4–5% combined bias and CV performance criterion but had total errors for capillary samples greater than the 5% criterion. The DT-60 performance was near the 5% total error criterion for capillary samples and exceeded the 5% criterion for venous samples. The QuickRead total errors substantially exceeded the 5% combined bias and CV criterion for all sample types.

The definition of tolerance interval (Appendix equation 6) includes the term K, which is a constant whose value depends on the number of observations and on the underlying statistical distribution of the test/reference ratio values (6). The numerical value of K is determined here for a gaussian distribution. However, the assumption of normality has minimal impact on the evaluation of the estimated tolerance intervals because the same K factor is used in Appendix equations 10 and 11, which set the lower and upper bounds for acceptable tolerance interval on the basis of specific performance criteria. Because the same assumption regarding ratio distribution is used for both calculation of the test method tolerance interval and for determination of the acceptability criteria for that tolerance interval, a valid judgment of performance can be made even if the distribution of ratios is non gaussian. However, if the distribution is non gaussian, the numerical values of the tolerance intervals may not be appropriate to describe the actual range of test method values expected for a population.

The ability of cholesterol screening methods to categorize individuals into desirable or referral groups is shown in Table 5. Categorizations were based on the NCEP decision point of 5.17 mmol/L (2.00 g/L), with approximately one-half of the results being above and one-half below this value. The data are presented with allowance for imprecision in measurement for values close to the decision point. That is, results that would produce misclassification vs the comparison method were not counted when they were within 5.4% of the comparison method value. The allowance of 5.4% is derived from the 3% precision recommendation of the NCEP and the 2.5% CV of the comparison method as follows: 1.96 [(CV^2 + CV_r^2)/2]^{1/2}, where 1.96 gives the one-tail 95% confidence limit; CV and CV_r are the CVs for screening and comparison methods, respectively; and the denominator 2 is used because all measurements were in duplicate. Thus, misclassifications due to imprecision of the assay systems are excluded. The Reflotron and Vision results for capillary blood samples were on average 4–5% greater than for the corresponding venous blood samples and had a substantially larger false-positive rate than for whole-blood or plasma samples. The patient categorization performance of the QuickRead and Liposcan methods was unacceptable.

Table 6 shows that triglycerides did not interfere with any of the quantitative methods. The Reflotron was unaffected by moderate concentrations of hemoglobin but showed a negative interference at very high concen-
trations of hemoglobin. The Vision was sensitive to low concentrations of hemoglobin. The Vision system detected and flagged samples with above-normal hemoglobin concentrations but failed to flag results that were biased by low hemoglobin concentrations.

**Discussion**

By all evaluation criteria, the QuickRead and Liposcan methods produced unreliable cholesterol results and are not considered further.

The commonly reported measures of analytical performance, mean bias vs a comparison method and imprecision from replicate assays of pooled material, have limitations when extended to the performance of a method for an individual subject's specimen. Mean bias can be near zero when individual results have substantial positive and negative bias. Imprecision as SD or CV based on a relatively small number of replicates is adversely affected by a large confidence interval for the statistic and by variability in the frozen or freeze-dried material assayed. Because of these limitations, we have developed an evaluation protocol based on total error of individual results. We believe that total error gives a realistic assessment of the performance to be expected for individual subjects' measurements.

We have used the 95% tolerance interval for the ratio of test to comparison methods results to express total error. This tolerance interval determination is based on the bias of a result vs a comparison method and the within-subject variance based on replicate (in this case duplicate) measurements of each subject's cholesterol. The tolerance interval describes a range of values within which an individual subject's results are expected to fall 95% of the time. Thus, this statistic describes the combined effects of measurement bias and imprecision and gives a realistic parameter to evaluate a method's performance for individual results.

The importance of evaluating total error is illustrated graphically in the bias plots (Figures 1–4), which show that a large proportion of subjects' results differ substantially from the correct cholesterol values. As an example, let us consider venous plasma measurements with the Reflotron analyzer: the average bias was –0.3% and yet 16% of individual results were >8.4% from the correct value because of the imprecision of the assays. The value 8.4% is the uncertainty range predicted by a gaussian distribution for 3% bias with 3% CV for the test method and 2.5% CV for the comparison method with all measurements in duplicate (3% + 1.96 [(3² + 2.5²)/2]¹⁄²). The other methods showed generally similar performance characteristics for individual results. Although some of the separate measures of average bias did meet the NCEP 3% guidelines, e.g., DT-60 with capillary and venous samples and Reflotron with venous samples, none of the methods had adequate combined bias and precision for the total error to satisfy the NCEP recommendations.

The performance of these three methods in a screening environment can also be evaluated by their ability to correctly categorize an individual into a low-risk or a referral group. Methods with a greater accuracy bias and (or) poorer imprecision had more misclassifications. Thus the Reflotron did a good job with venous plasma samples and a poorer job with venous whole blood, but

<table>
<thead>
<tr>
<th>Table 6. Interferences with Cholesterol Screening Instruments*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Reflotron</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Vision</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>DT-60</td>
</tr>
<tr>
<td>QuickRead</td>
</tr>
</tbody>
</table>

* Plasma cholesterol before supplementation was 5.17 mmol/L (2.00 g/L). Hemoglobin was tested at 0, 0.5, 1.0, and 3.3 g/L, triglycerides at 0, 2.6, 4.8, and 10.5 g/L.
reported many false-positive results with capillary samples because of its positive bias with this sample type. The Vision had no false-negative results but did have the largest false-positive rate, particularly with capillary samples, because of its positive average bias. The DT-60 had the smallest overall misclassification rate but still had an excessive error rate (>5%) with capillary samples.

Capillary samples gave unacceptable total error or misclassification rates with all methods. Capillary samples averaged 4–5% higher results than for venous samples with the Vision and Reflotron analyzers. In contrast, the DT-60 method showed no difference between these samples. Each of these methods showed the greatest variability in results for capillary samples, which may have been affected by collection technique. Although every effort was made to collect free-flowing blood, the collection of six to eight capillary samples from each subject may have resulted in some compromised specimens. Other investigators have reported differences in cholesterol measurements between capillary and venous samples. Kupke et al. (7) found capillary results 8.7% lower than venous results in an electrophoretic procedure. Boerma et al. (8) found capillary results 2.1% lower than venous results with a Reflotron. Ishikawa et al. (9), using a gas-chromatographic procedure, found capillary results 0.5% different from venous results. Greenland et al. (10) found capillary results 3.6% higher than venous results with an Ektachem 700. Bachorik et al. (11) found capillary results 7.0% higher than venous results with a Reflotron. These contradictory results indicate a lack of consensus in the literature regarding the average bias in cholesterol values between capillary and venous blood samples. Ammirati et al. (12) recently reported an optimized fingerstick collection technique that gave cholesterol results equivalent to venous blood samples with the AccuMeter (ChemTrak, Sunnyvale, CA) and Spectrum (Abbott) methods. Our data, along with those of others, suggest that the capillary collection technique is critical and must be standardized to obtain reliable cholesterol results.

In summary, none of the methods met the NCEP performance recommendations of 3% CV with 3% bias for cholesterol measurement. The Reflotron, Vision, and DT-60 methods had generally better performance with venous samples than with capillary blood samples. Capillary samples were subject to more variability in all methods, probably because of variability in sample collection.

References
Appendix

Suppose there are \( m \) measurements from the test product (T) and \( m \) measurements from the reference product (R) for each of \( n \) subjects. For the \( i \)th subject, \( i = 1, 2, \ldots, n \), these are denoted as

\[
T_{i1}, T_{i2}, \ldots, T_{in} \quad \text{and} \quad R_{i1}, R_{i2}, \ldots, R_{in},
\]

respectively. The statistical model is assumed to be

\[
T_{ij} = \mu_T + \delta_i + \epsilon_{TY} \quad \text{and} \quad R_{ij} = \mu_R + \delta_i + \epsilon_{RY},
\]

where \( \mu_T \) and \( \mu_R \) denote the true population means for the test and reference product, respectively; \( \delta_i \) is the random effect for the \( i \)th subject; and \( \epsilon_{TY} \) and \( \epsilon_{RY} \) are random error terms for the test and reference products, \( i = 1, 2, \ldots, n \) and \( j = 1, 2, \ldots, m \). It is assumed that the \( \delta_i \) and the \( \epsilon_{TY} \) and \( \epsilon_{RY} \) are all independent with null means and variances \( \sigma^2_\delta \), \( \sigma^2_{\epsilon_{TY}} \), and \( \sigma^2_{\epsilon_{RY}} \) which denote the intersubject variance, the test intrasubject variance, and the reference intrasubject variance, respectively.

Let

\[
T_i = \frac{1}{m} (T_{i1} + T_{i2} + \ldots + T_{in}) \quad \text{and} \quad R_i = \frac{1}{m} (R_{i1} + R_{i2} + \ldots + R_{in})
\]

and let

\[
\bar{A} = \frac{1}{n} (A_1 + A_2 + \ldots + A_n)
\]

and

\[
S^2_A = \frac{1}{n-1} ((A_1 - \bar{A})^2 + (A_2 - \bar{A})^2 + \ldots + (A_n - \bar{A})^2)
\]

denote the sample mean and sample variance, respectively, based on the \( n \) individual ratios.

A tolerance interval for the individual ratios is of interest because it is desirable to determine the effectiveness of the test product with respect to the reference product for each individual rather than for the average individual in the population. If the \( A_i \) values are normally distributed, then a 95% tolerance interval for 95% of the population is given by
\[ A \pm K \cdot S_A, \]  

where \( K = 2.305 \) when \( n = 68. \)

If the test product compares well with the reference product, then the tolerance interval will be narrowly centered about 1. It is possible to establish lower and upper endpoints, based on proposed bounds for the absolute relative bias and the coefficients of variation, which should enclose the calculated tolerance interval if the test product performs well. Then the (within-subject) absolute relative bias is

\[ \beta = |\mu_T - \mu_R|/\mu_R \]  

and the (within-subject) coefficients of variation are

\[ \xi_T = \sigma_T/\mu_T \] and \( \xi_R = \sigma_R/\mu_R. \)  

For example, with serum cholesterol, it is desirable for these values to have \( \beta \leq 0.03, \xi_T \leq 0.03, \) and \( \xi_R \leq 0.015. \)

Using a Taylor series expansion, we find that the true variance of any subject’s ratio is approximately equal to

\[
\left( \frac{\mu_T}{\mu_R} \right)^2 \left[ \frac{1}{m} \left( \xi_T^2 + \xi_R^2 \right) + \sigma_3^2 \left( \frac{1}{\mu_R} - \frac{1}{\mu_T} \right)^2 \right] = \\
\left( 1 + \frac{\mu_T - \mu_R}{\mu_R} \right)^2 \left[ \frac{1}{m} \left( \xi_T^2 + \xi_R^2 \right) + \sigma_3^2 \left( \frac{\mu_T - \mu_R}{\mu_R \mu_T} \right)^2 \right],
\]

and, assuming that \( \sigma_3^2/\mu_T^2 \) is no larger than 1, this is bounded above by

\[ (1 + \beta)^2 \left[ \frac{1}{m} \left( \xi_T^2 + \xi_R^2 \right) + \beta^2 \right]. \]

If the test product is performing well, then the calculated tolerance interval should be enclosed by \((K_L, K_U),\) where

\[ K_L = 1 - \left\{ \beta + K(1 + \beta) \left[ \frac{1}{m} \left( \xi_T^2 + \xi_R^2 \right) + \beta^2 \right] \right\}^{1/2}, \]

\[ K_U = 1 + \left\{ \beta + K(1 + \beta) \left[ \frac{1}{m} \left( \xi_T^2 + \xi_R^2 \right) + \beta^2 \right] \right\}^{1/2}. \]

**Clinical and Analytical Evaluation of a Continuous Enzymatic Method for Measuring Pancreatic Lipase Activity**

Mauro Panteghini, Franca Pagani, and Roberto Bonora

We report the evaluation of a new commercial kit for the determination of pancreatic lipase activity. The kit is based on the use of a 1,2-diglyceride as substrate and a specific monoglyceride lipase. The detection step is the continuous colorimetric measurement of hydrogen peroxide produced from glycerol by glycerol kinase, glycerol-3-phosphate oxidase, and peroxidase reactions. The procedure appears to be precise (between-day CV <9%) and the results show good correlation with those obtained by alternative procedures (vs turbidimetry, \( r = 0.965; \) vs ultraviolet absorbance–enzymatic method, \( r = 0.995; \) vs Ektachem, \( r = 0.976; \) vs immunonometry, \( r = 0.970). \) However, the method is susceptible to interference by increased concentrations (>4.5 mmol/L) of serum triglycerides. We estimated the reference interval for healthy adults to be 8–44 U/L. When we evaluated clinical efficacy by using receiver-operating characteristic curves and the overlap index, no significant differences were found between the commercial kit and a common turbidimetric assay for diagnosing patients with acute pancreatitis; both methods performed satisfactorily.

Enzymatic testing is often used to evaluate patients with acute abdominal distress and possible pancreatitis (1). Although amylase (EC 3.2.1.1) has been historically used for this purpose, the clinical nonspecificity of total amylase concentrations has often been emphasized (2, 3) and other assays have been developed to improve on the specificity of the diagnosis (3–5). Recently, we were able to confirm the diagnosis of acute pancreatitis by serum pancreatic lipase (triaclyglycerol acylhydrolase, EC 3.1.1.3) by using a turbidimetric determination (6–9). However, some analytical shortcomings (narrow measuring range, low sensitivity in the subnormal range, and negative activities in rare specimens containing low lipase activities) limit the reliability of turbidimetric measurement of lipase activity and discourage the use of this assay as a first diagnostic test for pancreatic disease (9). When we simplified the lipase

---

1° Laboratorio Analisi Chimico-Cliniche, Spedali Civili, 25123 Brescia, Italy.  
Received February 11, 1992; accepted August 24, 1992.