Assessment of current National Cholesterol Education Program guidelines for total cholesterol, triglyceride, HDL-cholesterol, and LDL-cholesterol measurements

Samuel P. Caudill,* Gerald R. Cooper, S. Jay Smith, and Gary L. Myers

We examine the effect of systematic bias and random error, quality control, and intraperson biological variation on the National Cholesterol Education Program (NCEP) clinical classifications for reported lipid measurements. We consider misclassification to occur if a true lipid homeostatic set point is within a desirable range but the reported lipid value is in a high-risk range, or if a true lipid homeostatic set point is in a high-risk range but the reported lipid value is in a desirable range. To evaluate the overall adequacy of the NCEP guidelines to ensure correct patient classification, we construct operating characteristic curves for total cholesterol, triglycerides, high-density lipoprotein cholesterol, and low-density lipoprotein cholesterol. We demonstrate that if laboratories are meeting the NCEP guidelines for inherent bias and analytic precision and are using standard quality-control (QC) procedures incorporating at least two QC samples per analytical run from each of two QC pools (for a total of 4 QC samples), the current NCEP guidelines are adequate to ensure (probability >0.90) correct patient classifications regardless of the size of the systematic bias of the laboratory or increased random analytic error. Thus we suggest that at least two concentrations of QC material be included in the QC scheme to ensure that the measurement system is operating within desired specifications across the entire range of desirable and high-risk lipid concentrations and to ensure with high probability that patients are correctly classified.

The National Cholesterol Education Program (NCEP) Laboratory Panels have recommended analytical performance guidelines for accurate clinical cut point classifications (1) of patient results for total cholesterol (TC) (2,3), triglycerides (TGs) (4), HDL-cholesterol (HDLc) (5) and LDL-cholesterol (LDLc) (6) measurements. In 1990, it was observed that ~10% of physicians were not following the NCEP recommendations on classification and treatment (7). Even when the NCEP recommendations are used, there is still a small probability that misclassification will occur, leading to unnecessary expensive treatment or lack of appropriate treatment (8,9). Recent reports have suggested that the NCEP Laboratory Panel recommendations inadequately consider the quality-control (QC) requirements necessary to detect medically important errors for TC (10), TG, HDLC, and LDLC (11) measurements used to distinguish between desirable and high-risk lipid categories. Other reports have suggested that large numbers of serial specimens are necessary but not sufficient to reliably distinguish between desirable and borderline or between borderline and high-risk lipid categories. These reports usually rely on confidence intervals to indicate the probability that a calculated range of values includes the true value. Difficulties arise, however, when physicians try to distinguish to which of adjacent risk categories patients should be assigned when measurements are near a cut point. Mogadam et al. (12) reported (on the basis of four weekly lipid and lipoprotein measurements) that 75% of subjects had >20% deviation in serum TC concentrations, 95% of subjects had >20% deviation in LDLc, and 65% of subjects had >20% deviation in HDLC. On retesting, 40% of the subjects changed one risk category and 10% changed two risk categories, such as from desirable to high risk, or vice versa. Bookstein et al. (13) observed in a study of 51 volunteers that when NCEP risk classifications were used, a single measurement was reliable for assigning a subject to the desirable category if the TC value was <4.78 mmol/L (<185 mg/dL), to the
borderline risk category if the value was 5.56–5.81 mmol/L (215–225 mg/dL), and to the high-risk category if the value was >6.59 mmol/L (>255 mg/dL). Similarly, they observed that when LDLc risk classifications were used, a single measurement was reliable for assigning a subject to the desirable category if the measured value was <3.00 mmol/L (<116 mg/dL), to the borderline category only if the value was exactly 3.7 mmol/L (144 mg/dL), and to the high-risk category if the value was >4.50 mmol/L (>174 mg/dL). Roebuck et al. (14) used simulation experiments to compare expected patient misclassification rates in hypothetical populations (with rates based on TC intraindividual variability estimates from studies where measurements were taken 1 week apart) with misclassification rates in studies where measurements were taken 2 years apart. Their results demonstrate the influence of the numbers of and the time intervals between repeat measurements on the probability of patient misclassification.

Using TC measurements taken 1 year apart from 14,600 men and women, Thompson and Pocock (15) studied the implications of serum TC measurement variability on screening and monitoring lipid risk classification. They observed a within-subject total CV of 7.4% and a within-subject biological CV of 6.5%. They concluded that a single measurement of TC could be used to reliably distinguish (i.e., with >95% probability) between true values above and below the high 75th percentile of 6.9 mmol/L (265 mg/dL) only when the true TC value was >7.8 mmol/L (300 mg/dL) or <6.1 mmol/L (235 mg/dL). Gillman et al. (16) found that among 24 subjects 6.1–8.8 years of age, one TC measurement allowed reliable assignment to the acceptable category (<4.4 mmol/L, or 170 mg/dL) only if the measured value was <4.01 mmol/L (154.9 mg/dL) and to the high category (>5.17 mmol/L, or 200 mg/dL) only if the measured value was >5.56 mmol/L (215.1 mg/dL). With one TC measurement, no value allowed assignment to the borderline-high (4.40–5.17 mmol/L, or 170–199 mg/dL) category, and one reading <4.78 mmol/L (184.9 mg/dL) allowed reliable classification below the high (5.17 mmol/L, or 200 mg/dL) cut point.

These concerns about reliable lipid risk classification have led us to examine the effect of systematic bias and random error, QC, and intraperson biological variation (CVb) on the NCEP clinical classifications for reported values of lipid measurements. For the purposes of our analyses, misclassification is considered to occur if a true lipid homeostatic set point is within the range for desirable risk but the reported lipid value is in the range for high risk, or if a true lipid homeostatic set point is within the range for high risk but the reported lipid value is in the range for desirable risk. These criteria for misclassification represent a medically useful approach because they address the practical clinical situation faced by physicians who do not want to unnecessarily treat a patient whose lipid concentration is in a desirable risk category or fail to treat a patient whose lipid concentration is in a high-risk category and who want to avoid the unrealistic and practically impossible situation of trying to distinguish between desirable and borderline risk categories or between borderline and high-risk categories when lipid values are near a cut point. Misclassification as defined here is of greatest concern because of its potential to create a financial or psychological burden on the patient. Our approach examines the joint probability of the following two events that must occur simultaneously for incorrect patient classification: (a) the laboratory obtains a measured value within the range for high risk when the true homeostatic set point is in the range for desirable risk, or the laboratory obtains a measured value within the range for desirable risk when the true homeostatic set point is in the range for high risk; and (b) the QC sample(s) measured during the analytical run in which the patient specimen was analyzed are within acceptable limits (17). The conclusions we reach by this approach indicate that a laboratory satisfying the NCEP recommendations and performing adequate QC procedures can attain correct classifications for TC, TGs, HDLC, and LDLc with probability $\geq 0.97$, except for LDLc (with probability $>0.90$) when the systematic bias is between 0.5 SD$_{Analytic}$ and 2.0 SD$_{Analytic}$

Materials and Methods

To evaluate the overall adequacy of the NCEP guidelines to ensure correct patient classification, we used computer simulation and theoretic computations to construct operating characteristic (OC) curves for TC, TGs, HDLC, and LDLc. The OC curves are based on the estimated joint probability of the following two independent events that determine whether a patient will be incorrectly classified: (a) the laboratory obtains a measured value within the range for high risk when the true homeostatic set point is in the range for desirable risk, or the laboratory obtains a measured value within the range for desirable risk when the true homeostatic set point is in the range for high risk; and (b) the QC sample(s) measured during the analytical run in which the patient specimen is analyzed are within acceptable limits. Because QC information is obtained for the very purpose of preventing the laboratory from reporting a result when the instrument is not performing according to specifications, it must be incorporated into the computation of the likelihood of incorrect patient classification. If the QC samples are out of control, patient results are not used, no classification is permitted, and thus any extreme patient value will not be incorrectly classified as desirable (or vice versa).

Each OC curve corresponds to the estimated probability of correct classification and is determined by subtracting the joint probability estimated above from the number one. The probability associated with whether a laboratory will obtain a result beyond a specified decision limit is based on gaussian distribution theory, where the mean of the distribution is assumed to be determined by the true
value of the sample plus the specified laboratory bias, and the variance is assumed to be determined by the specified CVb and the specified laboratory precision. The probability associated with whether a patient result will be reported is based on a computer simulation of a QC procedure using a multirule Shewhart chart (18) with multirules 1S, 1S, 2S, 2S, R 4S, 4S, and 10S applied to two QC samples per analytical run from either one QC pool (for a total of only two QC samples) or two QC pools (for a total of four QC samples). When two QC pools are used, we apply the multirules to each pool separately and assume that patient results will only be reported if analytical runs for both QC pools corresponding to the patient specimen measurement are acceptable. Each probability estimated from a QC simulation is based on 5000 analytical runs and is computed by dividing the number of analytical runs with at least one reject signal by the total number of analytical runs. A reject signal from a multirule that spans more than one analytical run is considered to occur only when the multirule condition has been satisfied for the designated number of analytical runs required to produce a reject signal.

The joint probability of correct classification is applicable to an individual patient with specified CVb whose health status is being determined on the basis of a single randomly collected specimen and analyzed in a laboratory with a specified accuracy and precision. The determination of health status is assumed to be based on whether the measured patient lipid result is within the “desirable” or “undesirable” range. These ranges are presented in Table 1 and correspond to the NCEP Laboratory Panel’s recommended analytical performance guidelines (1). The specified CVbs are also presented in Table 1 and correspond to those estimated by Smith et al. (19). The specified accuracy and precision, also presented in Table 1, correspond to the maximum allowable values recommended in the NCEP guidelines.

The OC curves provide an overall indication of the likelihood of correct patient classification by mapping performance characteristics as functions of inherent and systematic bias and inherent and increased random analytic error. Also provided with the OC curves for correct classification probabilities are OC curves for the individual probabilities associated with the two independent events on which the joint probabilities were computed.

Results

The OC curves for TC and LDLC, are presented in Figs. 1 and 2, respectively. Similar figures for TGs and HDLC are also available on request. Figs. 1A and 2A correspond to increases in systematic bias, with the scale of the horizontal axis expressed in analytic SD units. Thus an abscissa value of 2.5 represents a systematic bias that is 2.5 times the analytic SD. An abscissa value of 0.0 represents no systematic bias, but does correspond to an inherent bias equal to the NCEP maximum allowable bias. Figs. 1B and 2B correspond to increases in random analytic error, with the scale of the horizontal axis expressed in relative analytic variation (CVa) units. Thus an abscissa value of 4.0 represents an increase in CVa threefold larger than the NCEP maximum allowable CVa. An abscissa value of one represents no increase in random analytic error above the NCEP maximum allowable CVa.

The vertical axis of each plot represents the probability associated with a labeled event. The ordinate value of each point labeled with an “L” represents the probability that a laboratory result on a specimen from a patient whose true homeostatic lipid mean is at the limit of the desirable range will exceed the limit of the undesirable range under the condition specified by the abscissa, assuming that the patient has the specified CVb and that the laboratory has the NCEP maximum allowable CVa and bias. Thus, in Fig. 1A, the probability that a laboratory result on a specimen from a patient with a true TC mean concentration of 5.17 mmol/L (200 mg/dL) and a CVb of 6.1% will exceed 6.21 mmol/L (240 mg/dL) is 0.10 in a laboratory with an inherent CVa of 3%, an inherent bias of 3%, and a systematic bias 2.5 times the inherent analytic SD.

The ordinate value of each point labeled with an “R” represents the probability that a patient result will be reported based on the single-pool (solid line) or two-pool (dashed line) multirule Shewhart QC procedure described in Materials and Methods under the condition specified by the abscissa, assuming that the laboratory has the NCEP maximum allowable CVa and the NCEP maximum allowable bias. Thus, in Fig. 1A, the probability that a patient result will be reported is 0.31 for single-pool QC or 0.10 for two-pool QC in a laboratory with an inherent CVa of 3%, an inherent bias of 3%, and a systematic bias 2.5 times the inherent analytic SD. The ordinate value of each point labeled with a “P” represents the joint probability that a patient result will be correctly classified as being in the desirable range if single-pool QC (solid line) or two-pool QC (dashed line) is used under the condition specified by the abscissa, assuming that the patient has the specified CVb and that the laboratory has the NCEP maximum allowable CVa and the NCEP maximum allowable bias. Thus, in Fig. 1A, the probability that a laboratory result (on a specimen from a patient with a true TC mean

<table>
<thead>
<tr>
<th>Table 1. Parameters used to construct OC curves.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analyte</strong></td>
</tr>
<tr>
<td>TC</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>TGs</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>HDLC</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>LDLC</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Fig. 1. Total cholesterol.
Various probabilities vs (A) laboratory systematic bias as a multiple of the analytic SD ($\Delta SE$) and (B) relative increase in total CVa ($\Delta RE$). The true average patient TC is assumed to be 5.17 mmol/L (200 mg/dL); the CVb is 6.1%; the inherent total CVa is 3%; and the inherent bias is 3%. The ordinate value of each point labeled with an L corresponds to the probability that a laboratory result on a single random patient specimen will exceed 6.21 mmol/L (240 mg/dL). The ordinate value of each point labeled with an R represents the probability that a patient result will be reported based on a single-pool (solid line) or two-pool (dashed line) multirule Shewhart QC procedure with two samples from each pool. The ordinate value of each point labeled with a P represents the joint probability that a patient result will be correctly classified as being in the desirable range if single-pool QC is used (solid line) or two-pool QC is used (dashed line).
concentration of 5.17 mmol/L (200 mg/dL) and CVb of 6.1%) will be correctly classified as having a TC value ≤5.17 mmol/L (200 mg/dL) is 0.97 for single-pool QC and 0.99 for two-pool QC in a laboratory with an inherent CVa of 3%, an inherent bias of 3%, and a systematic bias 2.5 times the inherent analytic SD.

We obtained results similar to those presented in Figs. 1 and 2 using Shewhart mean and range charts rather than...
multirule QC procedures. The OC curves in Figs. 1 and 2 (as well as similar curves for TG and HDLC) for correct classification probability based on single- or two-pool QC procedures never drop below 0.96 when the inherent bias and CVa are at the extremes of the NCEP recommendations and there is no systematic bias or increased random analytic error. In addition, the curves for two-pool QC never drop below 0.97 when there are increases in systematic bias or random analytic error, except for HDLC when the systematic bias is between 0.5 SDAnalytic and 2.0 SDAnalytic. In this range, the probability that the patient HDLC result will be reported has not declined enough to mitigate the influence of systematic bias on the probability that the laboratory result will exceed 4.14 mmol/L (160 mg/dL). The lowest probability of correct HDLC classification based on two-pool QC over the length of this interval is 0.91.

To elucidate some specific cases included in the OC charts, we demonstrate the joint probability calculations associated with correct classification of a patient, based on HDLc measured in a laboratory under “worst case" conditions (i.e., the true patient mean is at a decision limit and during characterization the laboratory is operating just within the NCEP guidelines: inherent bias is 3%, and inherent CVa is 3%). We assume that the laboratory measures a specimen from a patient who on average (throughout the course of a year) has a TC value of 5.17 mmol/L (200 mg/dL) and that the CVb for this patient is 6.1%. We then determine the probability that this patient will be misclassified (i.e., declared to have a TC value of 6.21 mmol/L [240 mg/dL] or greater) on the basis of a single patient specimen obtained at a random time period during the year. We will also assume that the laboratory that measures the patient specimen routinely measures two QC samples per analytical run from each of two QC pools (for a total of four QC samples). There are essentially four groups of scenarios under which a misclassification could occur during the analytical run in which the patient specimen is analyzed: (a) the inherent bias is 3%, there is no systematic bias, and the CVa does not exceed 3%; (b) the inherent bias is 3%, there is an additional systematic bias, and the CVa does not exceed 3%; (c) the inherent bias is 3%, there is no systematic bias, and the CVa exceeds 3%; and (d) the inherent bias is 3%, there is an additional systematic bias, and the analytic CVa exceeds 3%. The probability of correct patient classification can be computed for any specific scenario within these four groups. No probability, however, can be attached to whether one particular scenario or another will occur, because we do not know the probability that a particular systematic bias or increase in CVa will or will not occur. The joint probability calculations for the worst case scenario from each of these four groups, which are presented in detail below, produce joint probabilities of correct classification of 0.993, 0.998, 0.995, and 1.00, respectively.

(a) The measurement system is just within the NCEP accuracy and precision guidelines (i.e., as a worst case, inherent CVa is 3% and inherent bias is 3%) during the analytical run on which the patient specimen is analyzed. The joint probability of correct classification in this case is 0.993 (= 1.00 - 0.0076 × 0.98). That is, it is the complement of the product of the probabilities of two independent events: (a) the measured result (on average, M = 1.03 × 5.1720 mmol/L [1.03 × 200 mg/dL]) will be >6.2064 mmol/L (240 mg/dL), which is equivalent to a standard normal variate being >2.43 (= [6.2064 - M]/[(0.03 × M)^2 + (0.61 × M)^2]^{1/2}; this event occurs with probability = 0.0076); and (b) the QC samples measured during the analytical run in which the patient specimen was analyzed will be within acceptable limits (this event occurs with a probability approximately equal to 0.99^2 when there is no systematic bias over and above the inherent 3% bias and two QC samples per analytical run from two QC pools are used).

(b) The measurement system exceeds the NCEP accuracy guidelines during the analytical run in which the patient specimen is analyzed. For example, suppose that, in addition to its allowable inherent bias of 3%, the laboratory has the minimum systematic shift that is detectable with probability = 0.90 using the single-pool, or 0.99 using the two-pool multirule Shewhart QC procedure described in Materials and Methods. (Note: The results for other systematic shift scenarios are indicated in Fig. 1A.) For two-sample single-pool QC, the minimum detectable shift is 9.3%. (Note: 9.3% represents 3.1 SD and was determined from simulations similar to those performed by Westgard et al. (20).) Thus, in a worst case scenario (i.e., inherent CVa is 3% and inherent bias is 3%), a laboratory might have an operating bias of 12.3% (= 9.3% + 3.0%). The joint probability of correct classification in this case is 0.998 (= 1.00 - 0.1566 × 0.01). That is, it is the complement of the product of the probabilities of two independent events: (a) the measured result (on average M = 1.123 × 5.1720 mmol/L [1.123 × 200 mg/dL]) will be >6.2064 mmol/L (240 mg/dL), which is equivalent to a standard normal variate being >1.01 (= [6.2064 - M]/[(0.03 × M)^2 + (0.61 × M)^2]^{1/2}; this event occurs with probability = 0.1566); and (b) the QC samples measured during the analytical run in which the patient sample was analyzed will be within acceptable limits (this event occurs with probability approximately equal to 0.10^2 when two-pool QC is used and there is a systematic bias of 9.3% over and above the inherent 3% bias).

(c) The measurement system exceeds the NCEP precision guidelines during the analytical run on which the patient specimen is analyzed. For example, suppose that in addition to its allowable inherent CVa of 3%, the laboratory has the minimum increase in CVa that is detectable with probability = 0.80 using the single-pool, or 0.96 using the two-pool multirule Shewhart QC procedure described in Materials and Methods. (Note: Results for other increases in
CVa are included in Fig. 1B.) For two-sample single-pool QC, the minimum detectable increase in CVa is 3.5 times the characterization CVa, which corresponds to an operating CVa of 13.5% rather than 3.0%. (Note: The minimum detectable increase in CVa of 3.5 was determined from simulations similar to those performed by Westgard et al. (20).) As before, we need to compute the probability that the patient will have a measured result (on average, $M = 1.03 \times 5.1720 \text{ mmol/L} [1.03 \times 200 \text{ mg/dL}] >6.2064 \text{ mmol/L} (240 \text{ mg/dL}).$ If the total CVa deteriorates from 3% to 13.5% and if the CVb is 6.1%, then in a worst case scenario (i.e., inherent CVa is 3% and inherent bias is 3%) the laboratory would have a 3% positive bias in addition to its 13.5% CVa. The joint probability of correct classification in this case is $0.995 \quad (1.00 - 0.1326 \times 0.04).$ That is, it is the complement of the product of the probabilities of two independent events: (a) the measured result (on average, $M = 1.03 \times 5.1720 \text{ mmol/L} [1.03 \times 200 \text{ mg/dL}]$ will be $>6.2064 \text{ mmol/L} (240 \text{ mg/dL}),$ which is equivalent to a standard normal variate being $>1.115 \quad (= \left[6.2064 - M\right]/\left[(1.15 \times M^2 + (0.061 \times M)^2\right]^{1/2});$ this event occurs with probability $= 0.1326);$ and (b) the QC samples measured during the analytical run in which the patient sample was analyzed will be within acceptable limits (this event occurs with probability approximately equal to 0.20 when two-pool QC is used and there is an increase in CVa that is 3.5 times the inherent CVa of 3%).

(d) The measurement system is not meeting NCEP accuracy or precision guidelines during the analytical run on which the patient specimen is analyzed. For example, suppose that, in addition to its allowable inherent bias of 3% and its allowable total CVa of 3%, the laboratory has the minimum detectable systematic shift with power $= 0.90$ and the minimum detectable increase in total CVa with power $= 0.80,$ using the single-pool multirule Shewhart QC procedure described in Materials and Methods. I fa

In Table 2, we present the lowest expected probability of correct patient classification over the range of systematic biases or increases in total CVa considered. In each case, we assume that a single patient specimen is measured, that the CVb is as given in Table 1, that the laboratory verified that it was just meeting the NCEP guidelines for accuracy and precision during characterization of the method, and that the single- or two-pool multirule Shewhart QC procedure described in Materials and Methods is used. Column 5 of Table 2 displays the probability that a measured patient result will exceed the decision limit specified in column 3. The highest probability in column 5 is 0.1728 and corresponds to the probability that a person with an average LDLC of 4.14 mmol/L (160 mg/dL) would have a measured value $<3.36 \text{ mmol/L} (130 \text{ mg/dL})$ if the laboratory has (in addition to its inherent bias of 4%) a systematic bias equal to 1.5 analytic SD. Based on a single-pool multirule Shewhart QC procedure using two QC samples per analytical run (for a total of only two QC samples), the probability that this result would be reported is 0.74 (column 6), so that the final probability of correct classification is 0.87 (column 7). This final probability could be increased to 0.91 (column 7 in parentheses) by using a two-pool multirule Shewhart QC procedure with two QC samples per analytical run (for a total of four QC samples). These results, as well as those in Figs. 1 and 2, clearly demonstrate the important role QC plays in ensuring correct patient classification.

As mentioned in Materials and Methods, the CVbs used to generate the OC curves in Figs. 1 and 2 and Table 2 are average CVb values estimated by Smith et al. (19). If a particular individual has a larger CVb than that given in Table 1 for a particular lipid, the probability of correct classification will decrease slightly. For example, in scenario 2 above, if the CVb $= 10.0\%$ instead of 6.1%, the probability of correct classification would be 0.997 instead of 0.998. Thus, individuals with larger than average CVbs are not as likely to be correctly classified, but the decrease in likelihood is minimal.

**Discussion**

To evaluate the overall adequacy of the NCEP guidelines to ensure correct patient classification, we constructed OC curves for TC, TGs, HDLC, and LDLC. The OC curves are based on the joint probability of the following two independent events that determine whether a patient will be incorrectly classified: (a) the laboratory obtains a measured value within the range for high risk when the true homeostatic set point is in the range for desirable risk, or (the laboratory obtains a measured value within the range for desirable risk when the true homeostatic set point is in
Table 2. Lowest* probability of correct lipid classification under worst case conditions.a,b

<table>
<thead>
<tr>
<th>Lipid</th>
<th>True patient mean, mmol/L</th>
<th>Specified decision limit, mmol/L</th>
<th>Error type</th>
<th>Probability of measured result exceeding decision limit</th>
<th>Probability of QC analysis being accepted</th>
<th>Probability of correct patient classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC</td>
<td>5.17</td>
<td>6.21</td>
<td>Bias = 2.0</td>
<td>0.0688</td>
<td>0.53 (0.28)</td>
<td>0.96 (0.98)</td>
</tr>
<tr>
<td>TC</td>
<td>6.21</td>
<td>5.17</td>
<td>Bias = 2.0</td>
<td>0.1076</td>
<td>0.53 (0.28)</td>
<td>0.94 (0.97)</td>
</tr>
<tr>
<td>TC</td>
<td>5.17</td>
<td>6.21</td>
<td>CVa = 4.0</td>
<td>0.1101</td>
<td>0.25 (0.06)</td>
<td>0.97 (0.99)</td>
</tr>
<tr>
<td>TC</td>
<td>6.21</td>
<td>5.17</td>
<td>CVa = 3.0</td>
<td>0.0975</td>
<td>0.40 (0.16)</td>
<td>0.96 (0.98)</td>
</tr>
<tr>
<td>TGs</td>
<td>2.26</td>
<td>4.52</td>
<td>Bias = 2.5</td>
<td>0.0012</td>
<td>0.34 (0.12)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>TGs</td>
<td>4.52</td>
<td>2.26</td>
<td>Bias = 1.0</td>
<td>0.0274</td>
<td>0.89 (0.79)</td>
<td>0.98 (0.98)</td>
</tr>
<tr>
<td>TGs</td>
<td>2.26</td>
<td>4.52</td>
<td>CVa = 10</td>
<td>0.0496</td>
<td>0.06 (0.004)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>TGs</td>
<td>4.52</td>
<td>2.26</td>
<td>CVa = 1.1</td>
<td>0.0209</td>
<td>0.98 (0.96)</td>
<td>0.98 (0.98)</td>
</tr>
<tr>
<td>HDLC</td>
<td>0.91</td>
<td>1.55</td>
<td>Bias = 2.5</td>
<td>&lt;0.0001</td>
<td>0.32 (0.10)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>HDLC</td>
<td>1.55</td>
<td>0.91</td>
<td>Bias = 3.1</td>
<td>&lt;0.0001</td>
<td>0.10 (0.01)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>HDLC</td>
<td>0.91</td>
<td>1.55</td>
<td>CVa = 7.0</td>
<td>0.0913</td>
<td>0.11 (0.01)</td>
<td>0.99 (1.00)</td>
</tr>
<tr>
<td>HDLC</td>
<td>1.55</td>
<td>0.91</td>
<td>CVa = 9.0</td>
<td>0.0426</td>
<td>0.06 (0.004)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>LDLC</td>
<td>3.36</td>
<td>4.14</td>
<td>Bias = 1.5</td>
<td>0.1244</td>
<td>0.74 (0.55)</td>
<td>0.91 (0.93)</td>
</tr>
<tr>
<td>LDLC</td>
<td>4.14</td>
<td>3.36</td>
<td>Bias = 1.5</td>
<td>0.1728</td>
<td>0.74 (0.55)</td>
<td>0.87 (0.91)</td>
</tr>
<tr>
<td>LDLC</td>
<td>3.36</td>
<td>4.14</td>
<td>CVa = 2.0</td>
<td>0.0699</td>
<td>0.69 (0.48)</td>
<td>0.95 (0.97)</td>
</tr>
<tr>
<td>LDLC</td>
<td>4.14</td>
<td>3.36</td>
<td>CVa = 2.0</td>
<td>0.108</td>
<td>0.69 (0.48)</td>
<td>0.93 (0.95)</td>
</tr>
</tbody>
</table>

* The lowest probability of correct classification over the entire range of systematic biases or increases in total CVa considered.
\[ a \] Worst case implies that the laboratory just meets the NCEP accuracy and precision requirements during characterization and that the true patient mean is at a decision limit. Thus, decreased inherent bias, better inherent CVa, or increased difference between true patient mean and specified decision limit would increase the joint probability of correct classification.
\[ b \] Bias indicates the systematic bias as a multiple of the analytic SD. CVa indicates the relative increase in total CVa.
\[ c \] This probability is based on the assumption that a single patient specimen is measured.
\[ d \] The first probability is based on the assumption that two QC samples from a single QC pool are used. The second probability (in parentheses) is based on the assumption that two QC samples from two QC pools are used.
\[ e \] These joint probabilities represent the complement of the product of the appropriate probabilities in the two preceding columns. The first probability corresponds to single-pool QC and the second probability (in parentheses) corresponds to two-pool QC.

the range for high risk; and (b) the QC sample(s) measured during the analytical run in which the patient specimen was analyzed are within acceptable limits. Because both of these events must occur simultaneously for a patient to be incorrectly classified, the probability associated with incorrect classification is obtained by computing the joint probability of the two events. The authors of two previous articles (10, 11) and a reply to a letter to the editor (21) used only the probability associated with event 1. By not considering the probability associated with event 2, they concluded that the current NCEP recommendations (in conjunction with typical QC procedures) are not adequate to ensure a high likelihood of correct patient classifications. Because QC information is obtained for the very purpose of preventing the laboratory from reporting a result when the instrument is not performing according to specifications, it must be incorporated into the computation of the likelihood of incorrect patient classification. If the QC samples are out of control, patient results are not used, no classification is permitted, and thus any extreme (or desirable) patient value will not be incorrectly classified as desirable (or extreme). The inverse relationship between the curves labeled L and those labeled R in Figs. 1 and 2 illustrates this point. As systematic bias or random error increases, the probability that a laboratory-measured result will exceed a cut point increases; however, the probability that the laboratory will report the result decreases if one assumes that the laboratory follows acceptable QC procedures. Thus, the decreasing probability of an acceptable QC analysis with increasing systematic bias or random error tends to mitigate the influence of systematic bias or increased random error on the joint probability of patient misclassification (i.e., the probability that a laboratory result will exceed a cut point and be reported).

We have demonstrated for TC, TGs, HDLC, and LDLC that the NCEP accuracy and precision recommendations are adequate to ensure a high likelihood (>90%) of correct patient classifications. Actually, we found that a laboratory satisfying the NCEP recommendations and performing adequate QC procedures can attain correct classifications for TC, TGs, HDLC, and LDLC with probability ≥0.97, except for LDLC (with probability >0.90) when the systematic bias is between 0.5 SDAnalytic and 2.0 SDAnalytic. These analyses assume that laboratories are meeting the NCEP guidelines for inherent method bias and analytic precision and are using standard QC procedures (e.g., Shewhart mean and range chart QC or multirule Shewhart QC) that incorporate at least two QC samples from each of two QC pools (for a total of four QC samples). We suggest, therefore, that at least two concentrations of QC material be included in the QC scheme to ensure that the measurement system is operating within desired specifications across the entire range of desirable and high-risk
lipid concentrations and to ensure with high probability that patients are correctly classified.

It is important, however, to note that inherent method bias, unlike systematic bias, cannot be determined by QC procedures. Therefore, a method characterization experiment must be conducted that includes external reference materials with externally assigned target values. The estimated mean bias from such an experiment should fall within the NCEP accuracy guidelines. QC characterization analyses can be used to compute the inherent CVa, which should fall within the NCEP precision guidelines. Once a laboratory has established that the NCEP recommendations are satisfied, routine QC procedures such as described in the previous paragraph should provide adequate protection against reporting patient results when the method is no longer meeting previously verified performance standards.

Although not directly addressed in this paper, we assumed that the QC analysis length (the number of patient samples analyzed between each QC sample) has been optimized such that QC outcomes truly relate to patient samples (22). We also suggest that CVb can be reduced by obtaining two serial patient specimens at least 1 week apart (23). The relative range of the two results can be used to determine if additional patient specimens are required because of unusually high CVb. Of course, even if perfect patient lipid classification could be achieved, it would not guarantee perfect clinical diagnosis, which depends on the accuracy of the lipid screening strategy used to identify an individual at increased risk (24).

This work was performed in conjunction with the Working Group on Prerequisites for Interpretation of Serum Lipid Measurements (Chair, Gerald R. Cooper), Education and Management Division (Chair, Peter Wilding), International Federation of Clinical Chemistry.

References