A Self-Consistent Set of Reference Values for 23 Clinical Chemical Analytes

Ursula E. Spichiger and Dieter J. Vonderschmitt

Heparinized plasma of 528 blood donors was subjected to the 23 most frequently ordered chemical and enzymatic tests. The central fraction of the distribution of all results for each test was estimated. Out of the 528 donors a reference population has been selected. Because of the lack of other criteria, the result for any test of a blood donor was selected as a value belonging to the reference population if the results for the other 22 analytes of this particular donor lay within their own central fraction. On this basis an iterative procedure for the selection was programmed, considering the interaction between tests. The procedure was stopped when the reference limits for all 23 tests were converging. Fractions from 0.90 to 0.98 were applied to results for men and women donors separately. The elimination procedure and the criteria to select the best fitted fraction are discussed. The derived reference intervals are designated a "self-consistent set of reference values."

Additional Keyphrases: blood donors · normal range · reference interval · multiple screening

From 1979 to 1983, IFCC section no. 2 published a series of recommendations on the theory of reference values (1). By convention the reference interval for an analyte was defined as the central 0.95 fraction of all values from samples collected from a reference population. If such samples are subjected to 23 different laboratory tests, the probability of finding individuals with all 23 results within the 0.95 central fraction should exceed 0.3 (0.9523) and equal 0.3 under the condition of total independence of all 23 variables.

In reality, and applying the currently used reference limits, we found that only the results for 0.06 of all blood donors were within the reference intervals for all 23 variables. We concluded that the reference intervals for our laboratory were not properly estimated and (or) the population of blood donors does not constitute an optimal reference population. Also, it is questionable whether or not the central 0.95 fraction is generally suitable as a decision limit for reference intervals of the different analytes in multiple screening. In this study a system to develop reference intervals by multiple screening of blood donors is described, which does not a priori use the central 0.95 fraction to eliminate values from the tailing ends of the frequency distribution of an analyte. The best fitted central fraction and the criteria to define a set of results as reference values is developed from the data set of 23 different analytes itself.

The "self-consistent set of reference values" is adapted to multiple screening. It is derived from blood donors resembling a population of ambulant patients. The procedure takes into account the interdependence of analytes without the mathematical effort essential for multivariate methods.

Materials and Methods

Sampling group. Specimens were collected from 331 male and 197 female unselected blood donors under the following conditions. Donors presented themselves between 0800 and 1000 h after a light breakfast; they had abstained from alcohol and medication (anticonceptives and antihypertensives could not be excluded), and felt subjectively well. Blood was drawn from the first 15 donors in the morning after blood pressure and hemoglobin concentration were checked. Body weight could not be evaluated. The age distribution of the men was normal (Kolmogorov-Smirnov goodness-of-fit test, 5% level of significance) and in the range of 20 to 83 years. The age distribution of the women was platted (curtosis = -1.1, skewness = -0.05) and in the range of 20 to 66 years. The two groups were treated separately.

Samples. Specimens were collected in heparinized Vacutainer Tubes (Becton-Dickinson, Rutherford, NJ 07070) after a collecting period of 5 to 10 min with slightly opened tourniquet. For the analysis of glucose, a fluoride complex-tube was used. Specimens were promptly centrifuged. Plasma was separated approximately 1 h later and analyzed within 2 h of centrifugation. An aliquot was frozen and stored at less than -20 °C and used to re-examine a donor's plasma in case of missing results or of extreme values needing confirmation.

Apparatus. The following 12 tests were performed on an SMA II continuous-flow system (Technicon Instruments Inc., Tarrytown, NY 10591): Na⁺, K⁺, total calcium, inorganic phosphate, Cl⁻, total CO₂, urea, creatinine, total protein, albumin, alkaline phosphatase, and total bilirubin. Eleven tests were performed in a Hitachi 705 discrete microchemistry analyzer (Boehringer Mannheim Diagnostics, Inc., Houston, TX): glucose, uric acid, creatine kinase, CK-MB, total lactate dehydrogenase, ASAT, ALAT, GGT, triglyceride, cholesterol, and α-amylase.1

Chemical methods. We used the test methods listed in Table 1.

Calibrating materials. The SMA II was calibrated by one-point calibration with serum reference standard (ECS, Anaheim, CA; no. C-310-50).

Quality control. Accuracy was assured by "Kontrollogen-L" (Behringwerke AG, Marburg, F.R.G.; lot 623123 E) and "Lytool P" (api bio-Mérieux SA, CH-1202 Geneva, Switzerland; lot 01502/09/86) at the beginning and at the end of a series of 20 to 25 samples. Values for the quality-control samples, compared daily with their assigned values, lay within the tolerance interval. A possible drift was detected and precision assured through use of an evaluated pooled plasma. Residues of fresh plasma of patients were pooled, filtered, and enriched with bilirubin (Bilirubin purity: Fluka, Buchs, Switzerland; no. 14369) (6). The pooled plasma was transferred to centrifugal cups in aliquots of 1.5 to 2 mL and stored below -20 °C. Every cup was centrifuged before introducing the sample as day-to-day control. The

1 Nonstandard abbreviations: CK, creatine kinase; ASAT, aspartate aminotransferase; ALAT, alanine aminotransferase; and GGT, γ-glutamyltransferase.
Table 1. Methods of Chemical Analysis Used in the SMA II and Hitachi 705

<table>
<thead>
<tr>
<th>Test</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium</td>
<td>Flame emission spectroscopy, 589 nm</td>
</tr>
<tr>
<td>Potassium</td>
<td>Flame emission spectroscopy with Li-reflection, 786 nm</td>
</tr>
<tr>
<td>Total calcium</td>
<td>o-Cresolphthalein, 570 nm (8-hydroxyquinoline-KOH)</td>
</tr>
<tr>
<td>Inorganic phosphate</td>
<td>Phosphomolybdate, reduction with stannous chloride (Bi-catalysis), 660 nm</td>
</tr>
<tr>
<td>Chloride</td>
<td>According to Schales, colorimetric</td>
</tr>
<tr>
<td>Total carbon dioxide</td>
<td>Cresol red, colorimetric</td>
</tr>
<tr>
<td>Urea</td>
<td>DAM, *90 °C, end point, 520 nm</td>
</tr>
<tr>
<td>Creatinine</td>
<td>Jaffé, final point, 505 nm</td>
</tr>
<tr>
<td>Total protein</td>
<td>Biuret, end point, 546 nm</td>
</tr>
<tr>
<td>Albumin</td>
<td>Bromcresol green, end point, 635 nm</td>
</tr>
<tr>
<td>Alkaline phosphatase</td>
<td>p-Nitrophenol, AMP, AMP, AMP, end point, 410 nm</td>
</tr>
<tr>
<td>Total bilirubin</td>
<td>Jendrassik and Gröf (4), end point, 591 nm</td>
</tr>
<tr>
<td>Glucose</td>
<td>Enzymatic at 340 nm, GlcDH</td>
</tr>
<tr>
<td>Glucose</td>
<td>Peridochrom (Boehringer) GOD-PAP according to Trinder (5), enzymatic at 510 nm, 37 °C</td>
</tr>
<tr>
<td>Uric acid</td>
<td>Uricase-POD, enzymatic at 570 nm, 37 °C</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>Cholesteroxidase-PAP according to Trinder (5), 500 nm, end point</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>GPO-PAP, enzymatic at 505 nm, [Boehringer kit, according to Wahlenburg/Bergmeyer (6, 7), end point]</td>
</tr>
<tr>
<td>Creatine kinase</td>
<td>Optimized standard method, kinetic at 340 nm, NAC*, activated, 37 °C</td>
</tr>
<tr>
<td>CK isoenzyme MB</td>
<td>Immunologic (Boehringer)</td>
</tr>
<tr>
<td>Lactate dehydrogenase</td>
<td>Standard method SFBC, kinetic at 340 nm, 37 °C</td>
</tr>
<tr>
<td>Aspartate aminotransferase</td>
<td>Standard method IFCC, kinetic at 340 nm, 37 °C</td>
</tr>
<tr>
<td>Alanine aminotransferase</td>
<td>Standard method</td>
</tr>
<tr>
<td>Gamma-glutamyltransferase</td>
<td>Kinetic at 415 nm, as glutation transferase, 37 °C</td>
</tr>
<tr>
<td>a-Amylase</td>
<td>Kinetic as a-glutamyltransferase, as glucose oxidase as 415 nm as PNP, and maltototiolase, kinetic unit factor = 0.54</td>
</tr>
</tbody>
</table>


Pooled plasma was analyzed three times in every series of 20 to 25 samples during 12 weeks as a blind control. Earlier investigations showed that it was stable for more than one year. The analytical variation sα from day to day was estimated by using one to two values obtained for every quality-control sample in a day during 22 days. Except for Na⁺, the analytical variation sα was less than 1% of the reference interval used in our laboratory (9). Variations in total calcium, chloride, total protein, glucose, lactate dehydrogenase, and creatinine as established with the unknown pooled sample came close to this limit.

The analytical coefficient of variation (CV) for enzymes in the reference interval did not exceed 0.08, except for CK-MB. The analytical CVs for metabolites in the reference interval ranged from 0.008 (Na⁺) to 0.038 (creatinine), 0.052 (total bilirubin), and 0.062 (triglycerides), as established with the blind controls. As a routine procedure of the laboratory the quality assessment plan of the Swiss Center for Quality Assessment, CH-2301 La-Chaux-de-Fonds, was followed. All the results reported to the Center were approved with respect to accuracy and precision.

**Computing system.** The procedure for evaluating reference intervals was programmed in the SAS (Statistical Analysis System; SAS Institute Inc., Box 8000, Cary, NC) run on an IBM 3033 under VM/CMS of the computing center of the University of Zurich. The same system was used for storage of data and statistical analysis.

**Procedure, Program**

1) In accordance with the recommendations given in reference 1, the values for all results of each test were sorted and the limits of the central fraction F of the frequency distribution per test were calculated and stored (0.90 < F < 0.98, see 8).

2) If all the values obtained for tests 2 to 23 for the first individual were within the stored limits of the central fraction F, the value of test number 1 of the first observation was stored as a value of the reference population. If only one value for tests 2 to 23 failed to meet this condition, the whole set of values for this blood donor was ignored.

3) The process was repeated for all observations. The new limits of the central fraction F were calculated based on the remaining values (values of the reference population).

4) This process was repeated for the second analyte. If the values for test 1 were within the new limits of the central fraction F and the values of tests 2 to 23 were within the original limits of the fraction F, the values of analyte number 2 were stored for all observations. At the end the new limits of the central fraction F for analyte number 2 were fixed and stored.

5) This process was repeated for all analytes.

6) When all observations and all analytes were treated, the result was a set of changed limits of the fraction F for all 23 analytes and a smaller number of observations, corresponding to a new (smaller) reference population.

7) The whole procedure was then repeated iteratively with the same central fraction F and the new (smaller) reference population until the upper and lower decision limits and the number of observations stored remained constant. The limits of the central fraction F of this last frequency distribution of every analyte were taken as reference values and are called "self-consistent set of reference values."

8) The iterative procedure was repeated with all central fractions between 0.98 and 0.90 in increments of 0.01. The particular fraction used for the iterations is called Fint.

**Mechanism of Elimination of Values and Observations through Interdependence**

Analytical variables often correlate to some extent with one another. This correlation may be strong or nearly nonexistent. We now describe the pattern of elimination of observations corresponding to individual donors from the total number of observations in the iterative process for the two extreme situations where there is total dependence (strong correlation) or total independence (no correlation) between parameters (Figure 1).

**Total dependence of analyte A1 with another laboratory parameter A2 (linear or inverse).** If values that do not belong to the central fraction F are eliminated from the tailing ends of the distribution of A2, the matching values in the distribution of A1 are also eliminated from the tailing ends. This is especially true for the first iteration. With every step
of the iterative process, however, the proportion of the results that are removed from the tailing ends of \( A_1 \) decreases and a higher proportion of values from the zone of high density of the distribution is eliminated. The number of observations tends very slowly to zero. The following frequency function \( f^p \) describes this type of eliminating process by iteration, regardless of the strength of dependence.

\[
f^p(x_0) = x_0 - [x_0(1 - F_{i0}) - x_0(1 - F_{i1}) - \ldots - x_0(1 - F_{in})] 
\]

where \( x_0 \) = total number of observations at the beginning, and \( x_i \) = number of observations after \( i \) iterations, \( i = 1 \) to \( n \).

Total independence of analyte \( A_1 \) with all other analyte \( A_k \):
If values that do not belong to the central fraction \( F \) of a particular distribution of \( A_k \) are eliminated from the tailing ends, each result in \( A_1 \) has the same chance to be removed or added again in every step of the iteration. Because of the random elimination of results according to a fixed percentage \((1 - F_{i0})\), convergence is reached after the first step. The number of stored observations depends on the central fraction \( F_{it} \) only. There are \( x_0 \cdot (F_{it})^p \) observations to be stored. The frequency function can be formulated as

\[
f^p(x_0) = x_0 \cdot (F_{it})^p
\]

where \( p \) = number of analytes and \( n \) = number of iterations, \( n = 1 \).

These theoretical considerations do not prove that the iterations must ultimately lead to convergence. With total dependence of parameters, the frequency function converges slowly to zero; with total independence 70% of the original observations are eliminated if the central 0.95 fraction of each parameter is retained. Because neither of the two extreme situations is likely to be found in reality, we hoped to be able to deduce reference values by the proposed procedure.

Results and Discussion

Elimination Procedure

1) The limits of the intervals of all tests converge after a number of iterations for all central fractions in the range \( 0.90 \leq F_{it} \leq 0.98 \).

\[
\begin{array}{cccc}
A1 & A2 & A1 & A2 \\
\end{array}
\]

Fig. 1. Mechanism of elimination of test results by interaction
A1, A2, analyte 1, analyte 2; \( F_{it} \) central fraction used to eliminate extreme values during the iterative process. Left: Total dependence of two analytes. Limits move very slowly to the middle of the range. If the frequency of values is high, values stay constant and limits do not move any longer. Right: Total independence of two analytes. Limits do not move. The number of results that are eliminated from the tailing ends correspond to \( F_{it} \).

2) The number of individuals and values representing the reference population depends on the width of the central fraction \( F_{it} \) chosen for the iterating process (Figure 2). It is not clear from the data that we obtained whether the number of iterations indicates the strength of interdependence of the results of the tests. Test results for male and female blood donors, however, do not show the same interdependence.

3) The number and position of the test results eliminated from the tailing ends is different for different analytes (Figure 3). If, after convergence, the values that have been eliminated during the whole procedure from within the final reference interval were added again, a theoretical central fraction \( F_{ec} \) would result. The theoretical fraction is calculated as

\[
F_{ec} = (n - n_{el})/n
\]

where \( n \) = total number of male or female reference individuals and \( n_{el} \) = number of eliminated values outside the reference intervals. This number is different for each analyte.

The resulting central fraction \( F_{ec} \) is different for different analytes. This becomes obvious by iteration with smaller central fractions \( F_{it} \). The fraction \( F_{ec} \) depends on the concerted action of three factors that influence the elimination: the density of the frequency distribution, the interdependence of analytes, and the selected fraction \( F_{it} \). \( F_{ec} \) has the largest value if results for \( A_1 \) (Figure 1) corresponding to values in the tailing ends of \( A_2 \) are mainly eliminated from the center of the frequency distribution of \( A_2 \). The number of eliminated values is then proportional to the frequency of its occurrence and, therefore, a smaller percentage is removed from the tailing ends with every step of the iterative process. Because only integral numbers or none of the observations can be eliminated, the elimination of each value corresponds to at least 1% or 0.5%, respectively, for the remaining
number of 197 male or 96 female donors after several iteration steps. Further studies with a larger sample group will therefore be necessary to differentiate between these influences.

Selection of the Best Fitted $F_R$

**Probabilistic criterion.** Assuming that there exists a certain dependence between analytes, the remaining fraction of blood donors describing the self-consistent set of reference values cannot be lower than 0.3, supposing that $F_{sc}$ of each analyte should average 0.95. Only the fraction $F_{hr}$ of 0.96 and 0.97 for men, and 0.95 and 0.96 for women fulfill this criterion. (For $F_{hr}$ 0.97 for men the product of all $F_{hr}$ is 0.43, the average of all $F_{hr}$ is 0.964.)

**Criterion of the density of the frequency distributions.** The kurtosis of the frequency distributions of stored values diminishes from original data to the reduced number of observations by iteration with $F_{hr}$ from 0.96 to 0.95. The male reference population determined from iteration with $F_{hr} = 0.97$ shows values of kurtosis that are no longer significantly different from the normal distribution for all analytes except ASAT, ALAT, triglyceride, CK, and CK-MB. With $F_{hr} = 0.96$ the exceptions are total bilirubin, potassium, and CK-MB. The kurtosis of the eliminated values for both fractions $F_{hr}$ and most analytes is different from the normal distribution, with high significance. Skewness shows similar effects. With $F_{hr} = 0.95$ most test results for men are normally distributed (exceptions: sodium, total bilirubin, ALAT, GGT, cholesterol). The 0.97 fraction for men and the 0.96 fraction for women is preferred. Further elimination with $F_{hr} 0.96 for men and 0.96 for women has nearly the same effects on density and normality of distributions but strongly reduces the number of observations. The selection of $F_{hr}$ by the criterion of the density of the frequency distribution will be analyzed by cluster analysis (10, 11). This will be the aim of further studies and will be published separately.

**Diagnostic Criterion**

Specificity and sensitivity of tests change with changing reference intervals. The diagnostic usefulness of the central 0.95 fraction is generally recognized (1). If reference values are adapted to the analytical performance, it affords a diagnostic specificity of 97.5% with respect to a particular disease or for multiple screening, if only higher or lower values are pathognomonic. If both high and low values are pathognomonic, the specificity for multiple testing is still 95.0%.
The self-consistent set of reference intervals is adapted to blood donors and is characteristic for this group. The high values for GGT, triglyceride, and glucose were expected and suitable for study of the eliminating mechanism. In some cases the described procedure eliminates—by correlation between parameters—more test results from the diagnostically important upper end of the distribution than does the central 0.95 interval. This is true for ASAT, ALAT, alkaline phosphatase, GGT, triglyceride, and glucose. The opposite holds true for cholesterol and α-amylase. For CK and total bilirubin in plasma only two values for men show very independent behaviors. They may clearly be eliminated by inspection of the frequency distribution (see Figure 3). The upper limit is then 25 μmol/L for total bilirubin and 299 U/L (37 °C) for CK. By iteration with the 0.96 fraction for women, CK values of 514, 445, 312, 211, and 179 U/L (37 °C) are eliminated, as are total bilirubin concentrations of 40 and 33 μmol/L. The upper CK limit for women is 189 U/L (37 °C), the second value is 166 U/L. The upper limit for total bilirubin is 22 μmol/L. For clarity, test results for the women are only reported when their difference from results for men exceeds the analytical variation.

Compared with our current laboratory reference values, the intervals of the "self-consistent set of reference values" are larger, the corresponding specificity higher, and the sensitivity lower. Some intervals are dislocated (total calcium, total lactate dehydrogenase, uric acid), some diagnostically important upper limits also are lower (ASAT, ALAT, total protein, alkaline phosphatase). Generally, the elimination of values from the center of the distribution by interdependence decreases the number of values that are eliminated from the tailing end. Because of the unknown strength of interdependence and redundancy between analytes, the central 0.95 fraction of test results in multiple screening cannot be estimated correctly. Observations are eliminated from the center and diminish the number of removed observations in the tails. These results in general explain the small number of "healthy" blood donors. Also for multivariate methods, high specificity and low sensitivity were observed (12). A comparison with the "self-consistent set of reference values" would be interesting.

Conclusion

The elimination of values from the tailing ends of the frequency distributions of analytes by the central fraction \( F_{0.05} \) is combined with the elimination of values by interaction between analytes. The theoretically resulting fraction \( F_{0.05} \) of all results for one analyte after subtraction of the number of values outside the reference interval differs from one analyte to another. This fraction \( F_{0.05} \) is developed from the distribution of values itself. Because of the elimination of values by interaction, the fraction \( F_{0.05} \) is, on the average, larger than 0.95. The fraction \( F_{0.05} \) is not a central fraction.

The programming of the procedure is simple and requires no specialized mathematical "know how" (standardization, transformation, handling of covariance, and correlation matrix).

The described method does not require any mathematical or statistical assumptions and can be performed without involved mathematical procedures.

Observations with missing values may be included without disturbing the procedure. It may be applied to every set of laboratory tests with a continuous scaling. Large sets of analytes can be treated simultaneously. The result is a "self-consistent set of reference values." The values from iteration with the 0.97 fraction are close to the conventional 0.95 fraction limits, but extreme values for dependent analytes are cut more strictly. Extreme values for independent analytes may in some cases be retained, which may mainly be due to the limited number of original observations, but are easily detectable. The results presented here are characteristic for blood donors. They are biased by food intake ("light breakfast") and intake of anticonceptives by the female donors, by the subject's position during blood collection, and by venous constriction. They may be corrected, but assays with other sample groups and comparisons with multivariate methods will be necessary.

References