Comparison of Estimates of Long-Term Analytical Variation Derived from Subject Samples and Control Serum

George Z. Williams,1 Eugene K. Harris,2 and Graham M. Widdowson1

Variation in the assays of uniform control serum commonly are assumed to represent day-to-day analytical variation. To test this assumption, we compared the differences between results of serum aliquots assayed immediately for 12 constituents and frozen aliquots accumulated and assayed on a single day with the results of control serum variation from the same period. One aliquot of each weekly sample was stored frozen. Eleven subjects were sampled for 12 weeks. Storage at −20 °C for 15 weeks had a mild destructive effect on two enzymes in serum. The control serum data revealed significant linear trends in magnesium (upwards) and alkaline phosphatase (downwards) that substantially increased the respective variances. In the other 10 constituents tested, comparison of variances indicated that long-term (weeks) variation in control serum assays is similar to the difference of variation between aliquots assayed immediately and those frozen and assayed at the same time. For these constituents, this finding justifies the use of control serum to estimate long-term analytical variation.

Additional Keyphrases: analytical variation resulting from storage, for control serum and samples • variation, source of

As pointed out by Williams et al. (1) and Bokelund et al. (2), sources of variation in serum constituents in a time-series of measurements include:
• conditions of subject preparation (efficacy of diet and activity control, fasting, posture, emotional state, etc.)
• effects of specimen collection (use of tourniquet, specimen collection in an evacuated tube, temperature and duration of clotting, serum separation, etc.)
• imprecision of assay procedures
• random unidentified errors
• the subject’s physiological variation, data on which is usually the objective of the analysis.

Because it is not practical or possible to eliminate the sources of all procedural biases (the first three listed above) and random "noise" cannot be identified, the best estimate of "physiological" variation results from rigid standardization of patient-preparative and specimen-collection procedures and measurement of and correction for variability originating from analytical or laboratory factors. Past studies (e.g., 1, 3–5) have assumed that uniform serum pool samples, assayed along with subject specimens, provide a valid estimate of long-term (day-to-day) analytical variation. That is, we have accepted the premise that because daily samples from a uniform serum pool are not subject to varying collection procedures or physiological variation, any variation measured over time is caused by long-term analytical and environmental factors. By subtracting the serum pool variance from total variance of a series of measurements of each variable (after removing "within-batch" variance) a difference is obtained that is assumed to represent the subject’s physiological variation, although it may still include variation from specimen collection procedures.

Here, we describe experiments designed to test the validity of this assumption.

Materials and Methods
Protocol

A group of 11 subjects was selected for availability, a state of apparent good health, and motivation to cooperate in this study. They were members of the staff of our research institutes (Institutes of Medical Sciences) with similar occupations: scientists, technologists, and research administrators. Two women and seven men ranged in age from 25 to 35, one woman was 63, one man 64. These persons were examined to determine their state of health and rule out medication or diagnosable disease. They were instructed to eat a light evening meal the night before sampling, not later than 2000 hours, and to avoid alcoholic beverages. After an overnight fast, they presented for blood collection between 0830 and 0900 hours the following morning.
Specimens were obtained for 12 consecutive weeks. Blood samples were collected from the median cubital vein by gravity with large-bore needles, without a tourniquet; the serum was separated 1 h after collection. Each weekly serum sample of each subject was separated into six aliquots; one duplicate pair of aliquots, subsample Type A, was stored at 4 °C overnight and assayed the next morning; another pair, Type B, was stored at -20 °C for one week, thawed, thoroughly mixed, and assayed in the same run with the next Type A specimen from the same person; the remaining pair of aliquots, Type C, were stored at -20 °C until the 15th–17th week, when each constituent of the C-samples from the same subject were assayed together in one day. The purpose of the Type C-samples was to eliminate long-term analytical variance while retaining physiological variance. The use of these samples in conjunction with Types A and B is explained in more detail in the next section. Uniform control serum samples were assayed in the same runs with the subject samples, but the duplicates of both subject and control sera were placed in random positions in the sample trays.

Chemical analyses were performed with mechanized and automated instruments to avoid manual pipetting and other error-prone manipulations.

Calcium and magnesium were determined by an automated method (6) with the Varian AA120 Atomic Absorption Spectrophotometer (Varian Techtron, Walnut Creek, Calif. 94598); urea nitrogen, glucose, total protein, and albumin were determined with the Beckman DSA-560 discrete sample analyzer (Beckman Instruments, Inc., Clinical Instruments Operations, Fullerton, Calif. 92634) by the urease/phenol/hypochloride (7), hexokinase (8, 9), biuret (10), and modified bromcresol green (11) methods, respectively. The AutoAnalyzer II (Technicon Instruments Corp., Tarrytown, N. Y. 10591) was used for the continuous-flow analysis of uric acid by the phosphotungstic acid/hydroxylamine (AAI-13) method (12) and for inorganic phosphate by the ammonium molybdate/stannous chloride/hydrazine (AAI-4) method (13, 14). Aspartate aminotransferase (EC 2.6.1.1) (15), lactate dehydrogenase (EC 1.1.1.27) (16), and alkaline phosphatase (EC 3.1.3.1) (17), were determined by rate methods with the Beckman Kintrac analyzer. Cholesterol was determined manually by a method in which ferric chloride/glacial acetic acid is the reagent (18).

Statistical Procedures

The chief purpose of the statistical analysis was to compare the estimated long-term analytical variance in sample Types A and B with the estimate obtained from a parallel series of daily analyses of samples of a control serum pool. In addition to these analytical variances, the analysis provides estimates of the interindividual variances among the 11 subjects, but such variances are of no particular interest in this study. The primary statistical method was that of nested analysis of variance [described by Harris et al. (4)], applied separately to sample Types A, B, and C. Additional methods were applied to compare the means of the three sample types and to test for the existence of storage-induced trends in certain constituents. The mathematical notation we used to define the different variance components is consistent with that followed by Harris et al. (4).

In Type A samples (4 °C overnight storage) the intra-individual variance components, \( s^2_i \), includes average physiological variance plus long-term analytical variance, but not duplicate variance. The latter was determined separately in each set of samples. Since Type B samples should reflect the same sources of variation affecting Type A samples plus whatever changes are caused by frozen storage during one week, the difference \( s^2_i (B) - s^2_i (A) \) represents an estimate of average variance due to such storage. The statistical test of the hypothesis that such additional variance did not exist in this study was a paired \( t \)-test for each constituent, based on the 11 individual values \( (s^2_i (B)/s^2_i (A)) \), with use of logarithms to remove the skewness in the distribution of variance components. In none of the constituents did this test approach statistical significance (the largest \(|t|\)-value was 0.88). Therefore, the values of \( s^2_i (A) \) and \( s^2_i (B) \) were averaged (see Table 2, col. 4). We denote this average by \( s^2_i (AB) \).

The variance component \( s^2_i (C) \) contains the same physiological variance as \( s^2_i (AB) \) but does not include random long-term analytical variance. However, it may contain added variance arising from systematic trends caused by frozen storage for one to 17 weeks. Thus, the difference \( s^2_i (AB) - s^2_i (C) \) eliminates physiological variance while retaining random long-term analytical variance [contained in \( s^2_i (AB) \)], but may also be reduced in size by variance from storage-induced trends. The variance of concurrent measurements of vials from a uniform control serum \( s^2_i (B) \) (adjusted for within-day duplicate variance) contains random long-term variance and, possibly, the variance of trends. When storage-induced trends are absent or eliminated, comparison of the difference, \( s^2_i (AB) - s^2_i (C) \), with the control serum variance provides a test of the assumption that daily data from control serum yield unbiased estimates of the long-term analytical variation contained in a series of weekly subject samples.

We also compared the three sample types with respect to differences of mean values. Since the variances of results from Type C samples may be expected to be significantly less than from Type A or B samples because long-term analytical variation is eliminated in Type C, a standard analysis of variance is not entirely valid. Instead, a nonparametric ranking procedure was applied whereby the smallest of the three sample means for each individual and test was assigned rank 3, the next rank 2, and the largest rank 1. Ranks were then arranged by sample type over all subjects and the averages compared using Friedman's \( x^2 \) test for ranks (19).

---

3 Reconstituted lyophilized control serum; Hyland Division, Travenol Laboratories, Inc., Costa Mesa, Calif. 92626.
Table 1. Mean Values for Various Serum Constituents In the Variously Treated Samples and Control Serum

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Sample type</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Calcium, mmol/liter</td>
<td>2.36</td>
<td>2.36</td>
<td>2.40</td>
<td>1.88</td>
<td></td>
</tr>
<tr>
<td>Magnesium, mmol/liter</td>
<td>0.79</td>
<td>0.79</td>
<td>0.84</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>Urea nitrogen, mmol/liter</td>
<td>5.75</td>
<td>5.78</td>
<td>5.78</td>
<td>4.82</td>
<td></td>
</tr>
<tr>
<td>Glucose, mmol/liter</td>
<td>4.94</td>
<td>4.95</td>
<td>4.90</td>
<td>3.90</td>
<td></td>
</tr>
<tr>
<td>Total protein, g/liter</td>
<td>71.6</td>
<td>71.4</td>
<td>71.2</td>
<td>79.0</td>
<td></td>
</tr>
<tr>
<td>Albumin, g/liter</td>
<td>40.4</td>
<td>40.3</td>
<td>40.5</td>
<td>45.1</td>
<td></td>
</tr>
<tr>
<td>Cholesterol, mmol/liter</td>
<td>5.54</td>
<td>5.54</td>
<td>5.59</td>
<td>3.87</td>
<td></td>
</tr>
<tr>
<td>Phosphate, mmol/liter</td>
<td>1.09</td>
<td>1.09</td>
<td>1.10</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Uric acid, mmol/liter</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>Aspartate aminotransferase, U/liter</td>
<td>12.7</td>
<td>12.4</td>
<td>11.3</td>
<td>6.52</td>
<td></td>
</tr>
<tr>
<td>Lactate dehydrogenase, U/liter</td>
<td>47.7</td>
<td>47.9</td>
<td>49.9</td>
<td>27.7</td>
<td></td>
</tr>
<tr>
<td>Alkaline phosphatase, U/liter</td>
<td>45.0</td>
<td>44.0</td>
<td>39.9</td>
<td>30.8</td>
<td></td>
</tr>
</tbody>
</table>

On the null hypothesis of no difference among average ranks the statistic

$$\chi^2 = \left[ 12 \sum_{i=1}^{p} S_i^2/cp(p+1) \right] - 3c(p+1),$$

is distributed approximately in a chi-square distribution with \( p - 1 \) degrees of freedom, where \( S_i \) is the sum of ranks for the \( i \)-th sample type, \( p \) is the number of such types (3), and \( c \) is the number of subjects (11).

Results

Comparison of Means

The chi-square test of ranks showed significant differences (\( P < .05 \)) among the means of the different types of samples with respect to calcium, magnesium, phosphate, and the three enzymes. These results were confirmed by inspection of the mean values for each individual and sample types. Average differences (Table 1) were small in the case of calcium and phosphate, but larger for the enzymes and magnesium. There was a consistent loss of activity of aspartate aminotransferase and alkaline phosphatase in all Type C samples, averaging 11.3% and 11.5%, respectively. For lactate dehydrogenase the average difference was in the opposite direction. This was a net effect, because of an increase of activity in eight of the 11 subjects, averaging 7.6%, and a decrease averaging 1.8% in samples from the other three individuals. In nine of 11 individuals, the calcium increased slightly (1.9%) in sample C; magnesium increased consistently in all individuals by 3 to 10% (average 6%); phosphate increased an average of 1.4% in 10 of the 11 subjects. For the other six constituents, there were very small (average, <1%) negative and positive changes.

Thus, in half the blood constituents measured, long-term storage (about eight weeks, on the average) generally produced a significant change in concentration (Table 1). On the other hand, frozen storage for one week only (Type B samples) produced little or no effect, a slight decrease in aspartate aminotransferase and alkaline phosphatase, averaging about 2%, being the only statistically significant finding. This would suggest that the changes in the enzymes and magnesium on long-term storage represented a continuing trend rather than a step change that occurred soon after storage began.

However, individual graphs revealed that in almost all cases where a trend was observed in data from Type C samples, it was accompanied by similar trends in the corresponding results for Types A or B samples, or both. Moreover, the increases or decreases (enzymes and magnesium) in Type C samples as compared with Types A and B samples in these individual graphs appeared to occur in randomly varying amounts independent of length of storage. Thus, there was no need to adjust Type C variances for storage-induced trends.

Table 1 also lists the mean values of the control serum measurements. In all constituents except total protein and albumin, the serum control means were less than the subject sample means. The percentage differences were comparatively minor in most cases but were larger for cholesterol, uric acid, and the enzymes.

Regression analysis of the data for the control serum, which included about three times as many daily means as the Type C subject samples, showed increasing trends in magnesium and lactate dehydrogenase, and a decreasing trend in alkaline phosphatase. The trend in lactate dehydrogenase, although statistically significant, introduced only a small increase in variance (12%). However, the trends in magnesium and alkaline phosphatase increased the variance of the control serum data by at least 50%.

Comparison of Variances

For each constituent we can now compare the average intra-individual variances \( s^2_{\text{AB}} \) and \( s^2_{\text{C}} \) and see whether their difference may be accounted for by long-term analytical variance estimated from control serum data \( (s^2_\text{e}) \). Table 2 presents these statistics as well as the separate variance components \( s^2_{\text{A}} \) and \( s^2_{\text{B}} \).
In magnesium and alkaline phosphatase, the values of $s_i^2$, after elimination of the linear trends are also given.

The agreement between the differences $s_{AB}^2(AB) - s_{AB}^2(C)$, (which we will call, $s_i^2$), in column 6 and the control-serum variances $s_i^2$ (column 7) is very close in most cases. Except for magnesium (unadjusted $s_i^2$ used), none of the discrepancies exceed 1.5 times the standard error of the corresponding $s_i^2$ values. A gaussian-distribution test is reasonable here, because the $s_i^2$ values are the means of 11 individual differences that may be either negative or positive.

**Discussion**

Because aspartate aminotransferase and alkaline phosphatase show consistent decreases in Type C samples of all subjects (unidirectional differences, although small), we conclude that frozen storage for 15 weeks is mildly destructive. Differences in lactate dehydrogenase of Type C samples were smaller, inconsistent, and probably chance events. Average increases in calcium and magnesium in frozen samples can be explained by slight sample evaporation. But it also is possible that incomplete solution and non-uniform dispersion of calcium and magnesium complexes in thawed samples in the analyzer cups (despite thorough mixing) permitted pipetting of slightly different concentrations.

The highly significant trends in measurements of magnesium and alkaline phosphatase in the control serum indicate that such results should not be used to estimate the long-term analytical variance of these constituents in serial patients' samples, which usually are analyzed immediately after collection.

Although the difference between $s_i^2$ and $s_i^2$, for cholesterol was not statistically significant, the control serum variance in this case possibly is an underestimate of the long-term analytical variance in the subject samples. As shown in Table 1, the mean cholesterol in the control serum was noticeably lower than the subject means, and variance may be directly correlated with mean value. On the other hand, in lactate dehydrogenase and uric acid $s_i^2$ was very close to or slightly higher than $s_i^2$, although the control serum means for these constituents were also less than the corresponding subject means.

Finally, we note again, as we did at the beginning, that part of the variation observed in measurements of subject samples arises from sources not associated with the analytical procedures [called “preinstrumental” sources by Bokelund et al. (2)], and therefore not contributing to the variation seen in control serum data. We have attempted to minimize the effects of pre-instrumental sources by rigid adherence to a defined protocol for the collection of fasting samples. In any case, despite potential differences in sources of variation, our results indicate close similarity between control serum variance $s_i^2$ and the comparable variance $s_i^2$ obtained from Types A, B, and C samples. In all constituents tested, except magnesium and alkaline phosphatase, these data support the practice of using analyses of uniform batch control sera to estimate long-term analytical variance in patients' samples. However, further studies of this difficult question would be welcome.

This study was supported by research grant R01-GM17360 from the NIGMS, NIH, Bethesda, Md.

**References**

1. Williams, G. Z., Young, D. S., Stein, M. R., and Cotlove, E., Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. I. Objectives, subject selection,