

**Subject:** some new terms

**From:** "Case, James T." <jtcase@ucdavis.edu>

**Date:** Sun, 20 Nov 2005 15:24:37 -0800

**To:** "Kathy Mercer" <kmercer@regenstrief.org>, "Clem McDonald" <cmcdonald@regenstrief.org>

These terms were requested by the HL7 LabSIG to support the interpretation criterion for a particular analytical observation. I have attached a reference that describes the reference change value. While these proposed terms seems very generic, they are only used in the context of an actual test value. If we do not use these generic terms then we would need to create an RCV term for every analyte that we want to use with this value. I don't think that is what we want to do, is it?

James T. Case MS, DVM, PhD  
Professor, Clinical Diagnostic Informatics  
California Animal Health and Food Safety Laboratory  
School of Veterinary Medicine  
University of California  
P.O. Box 1770  
Davis, CA 95617-1770  
530-752-4408  
jtcase@ucdavis.edu

<b>submit20051118.mdb</b>	<b>Content-Description:</b> submit20051118.mdb
	<b>Content-Type:</b> application/msaccess
	<b>Content-Encoding:</b> base64

<b>submit20051118.mdb</b>	<b>Content-Description:</b> submit20051118.mdb
	<b>Content-Type:</b> application/msaccess
	<b>Content-Encoding:</b> base64

<b>25.pdf</b>	<b>Content-Description:</b> 25.pdf
	<b>Content-Type:</b> application/octet-stream
	<b>Content-Encoding:</b> base64

S_ROW	S_LOCAL_CD	S_COMPO	S_PROP	S_TIME	S_SYS	S_SCALE	S METH	S_REL_NAM
1	RELMA23AB	REFERENCE	NUM	XXX	^POPULATION QN		CALCULATED	RCV 0.99
2	RELMAADF4	REFERENCE	PCT	XXX	^POPULATION QN		CALCULATED	RCV 0.99
3	RELMACF0A	REFERENCE	PCT	XXX	^POPULATION QN		CALCULATED	RCV 0.95
4	RELMA5CE2	REFERENCE	NUM	XXX	^POPULATION QN		CALCULATED	RCV 0.95

## On the Calculation of a "Reference Change" for Comparing Two Consecutive Measurements

Eugene K. Harris<sup>1</sup> and Toshio Yasaka<sup>2</sup>

We describe a statistical method for calculating a "reference change," defined as that difference between two consecutive test results in an individual that is statistically significant in a given proportion of all similar persons. By allowing for variation in within-person variances, this procedure computes a reference change that is more specific (i.e., less prone to false positives) than that obtained directly from the distribution of observed differences between measurements. Moreover, the method may easily be extended to a test for trend in three successive measurements. The method has been applied to semi-annual measurements of serum calcium and alkaline phosphatase in 698 men and women enrolled in a large health-maintenance program. We believe that these ideas may also be usefully applied to successive laboratory tests in carefully defined patient populations—but this introduces special problems, which are discussed briefly.

**Additional Keyphrases:** *statistics • reference interval • meaningful differences in serial data on an individual • sex-related differences • analytical variation • normal vs patient populations*

How significant is an observed change between two successive measurements? This must be a common problem in the clinical interpretation of laboratory data, as attested by surveys (e.g., refs. 1, 2) of physicians' opinions on the smallest changes in selected constituents that would indicate the need for follow-up activity in different types of patients. However, statistical studies of observed changes between test results have generally been concerned with establishing Delta-checks for improved control of laboratory procedures (3-7) rather than developing reference values for clinical use. One exception has been an analysis of weekly series of blood constituents in 37 healthy male volunteers (8). In this study, a method of calculating "reference changes" was proposed based on the distribution of within-person variances. In the present note, we apply this method to a much larger group of apparently healthy men and women for whom data were collected semi-annually for many years. The results should be more realistic for health-maintenance programs; for example, long-term analytical variation will show its effect. On the other hand, application of this method to hospital patients presents special problems, which we discuss briefly at the end.

Our data base, described in a recent paper (9) on subject-specific reference ranges, consists of results from 698 individuals: 412 men and 286 women, enrolled in the Perfect Liberty (PL) Health Control System in Japan. Under this program, screening, referral, and limited clinical services are currently provided in Osaka and Tokyo to over 16 000 persons,

ages 20-70 years. Those contributing to the present study have undergone 15-18 semi-annual screening examinations. All blood samples were drawn between 0900 and 1100, after overnight fasting, and analyzed by continuous flow (Technicon SMA 12/60). Because our primary purpose here is to illustrate a method, we will present results for only two analytes: calcium and alkaline phosphatase (EC 3.1.3.1).

At present, the PL Health Control System is applying the concept of a reference change to successive differences observed among the first few serial measurements of a constituent, before the individual's own normal pattern over time has clearly emerged. The reference change is derived from the population of differences between consecutive results that have been obtained from all persons in the health-maintenance program (10, 11). Such populations have been found to be approximately gaussian with mean zero and variance, say  $\sigma_{\Delta j}^2$ , for the  $j$ -th constituent. The attending physician is alerted if the current result for that analyte differs by more than  $2\sigma_{\Delta j}$  from the previous result. Despite this assumption that the universe of differences is normally distributed, the method is basically nonparametric, and could be made completely so by selecting upper and lower percentiles instead of  $\pm 2\sigma_{\Delta j}$ . We say this because the distribution of observed differences represents the combined effects of many parameters which are not estimated under the current procedure. They include, for example, within-person variances each of which depends, in turn, on the serial correlation between observations of the analyte in that individual. We propose to deal with these underlying parameters explicitly, believing for reasons stated below that this will lead to a better choice of reference change values.

### Statistical Methods

As stated in ref. 8, the essential statistical idea is to select as a reference change that difference between consecutive values that would be statistically significant (e.g.,  $p \leq 0.05$ ) in a large majority of the individuals concerned. This requires taking into account within-person variance under steady-state conditions and the distribution of such variances across the selected population. In addition, some reasonable time-series model should be postulated to describe the expected statistical characteristics of repeated measurements within an individual. For example, such a model would allow for serial correlation, the size of which is likely to depend on the time interval between measurements. Assuming a simple autoregressive model, the variance of a difference between two successive observations in the  $i$ -th individual is given by the formula

$$\begin{aligned} \text{Var}(x_{it} - x_{(t-1)i}) &= (\text{let us say}) \sigma_{\Delta i}^2 \\ &= 2\sigma_i^2\{(1 - \rho_i) + \rho_i\sigma_a^2/\sigma_i^2\} \quad (1) \end{aligned}$$

where  $\sigma_i^2$  denotes the variance of repeated measurements in that individual;  $\rho_i$ , the serial correlation between two consecutive measurements; and  $\sigma_a^2$ , the analytical variance included in  $\sigma_i^2$ . For the  $i$ -th individual, a suitable reference change might be a difference  $D_{0i}$ , such that  $|D_{0i}|/\sigma_{\Delta i}$  equals 1.96, assuming that we are concerned with changes either up or down and that differences between measurements within

<sup>1</sup> Laboratory of Applied Studies, Division of Computer Research and Technology, National Institutes of Health, Bethesda, MD 20205.

<sup>2</sup> Perfect Liberty Medical Data Center, Tondabayashi, Osaka, Japan.

Received July 21, 1982; accepted Sept. 7, 1982.

an individual are normally distributed. For most constituents,  $\sigma_i^2$  is not a constant, but varies from one person to another depending on individual behavior and physiological control mechanisms. Therefore, to arrive at a single reference change applicable to all individuals in the group, one must take this variation into account, and select a difference  $D_0$  such that  $|D_0|/1.96$  will exceed some specified proportion  $p$  of the  $\sigma_{di}$  values in the population (e.g., 95%). Expressing this condition more formally and writing Pr for "the probability that," we desire a value  $D_0$  such that

$$\Pr \{ \sigma_{di}^2 \leq (D_0/1.96)^2 \} = p$$

Substituting for  $\sigma_{di}^2$  the formula given in equation 1, this equation may be written

$$\Pr \{ (2\sigma_i^2[(1 - \rho_i) + \rho_i\sigma_a^2/\sigma_i^2]) \leq (D_0/1.96)^2 \} = p$$

or

$$\Pr \{ \sigma_i^2 \leq [(D_0/1.96)^2 - 2\rho_i\sigma_a^2]/2(1 - \rho_i) \} = p \quad (2)$$

Undoubtedly,  $\rho_i$  also varies from person to person, but to achieve any practical result, we shall have to substitute an average value for  $\rho_i$ , say  $\rho$ . In studies of healthy individuals or outpatients on periodic recall, the interval between measurements is generally sufficiently long to reduce serial correlation to zero, at least on the average. This simplifies the preceding equation to the form,

$$\Pr \{ \sigma_i^2 \leq \frac{1}{2}(D_0/1.96)^2 \} = p \quad (3)$$

In this case, knowledge of the analytical variance  $\sigma_a^2$  is not required. In the general case, when  $\rho \neq 0$ , an appropriate value for  $\sigma_a^2$  is necessary; for example, it should include long-term variance if the time between measurements covers several weeks or months.

A further difficulty lies in the fact that the individual's true variance  $\sigma_i^2$  must be estimated from a limited, often quite small, number of serial observations. Thus, even if  $\sigma_i^2$  were the same for all individuals in the group, the observed variances  $s_i^2$  would still exhibit statistical sampling variation. To resolve this problem, we use an equation previously derived (12), which allows estimating the variance of  $\sigma_i^2$  (Var  $\sigma_i^2$ ) from the mean and variance of  $s_i^2$ . For ease of reference, this equation is restated in the appendix as equation A1. If Var  $\sigma_i^2 = 0$  (i.e., all  $\sigma_i^2$  are equal), then the variation in  $s_i^2$  is due entirely to random sampling around a constant true variance. In this case, the estimate of Var  $\sigma_i^2$  will be approximately zero, or at least very small relative to Var  $s_i^2$ .

### Normality of Differences between Consecutive Measurements

In an individual series, this assumption may be tested empirically by plotting the differences on normal probability paper to see if an approximately straight line results. This is not feasible when hundreds of individuals and multiple analytes are involved. Therefore, as in earlier work (13), we used the Shapiro-Wilk test for this purpose. This powerful small-sample test produces a statistic  $W$ , small values of which indicate nonnormality.<sup>3</sup> Non-overlapping differences (e.g.,  $x_2 - x_1, x_4 - x_3$ , etc.) were tested to avoid introducing extraneous correlation. Of course, this reduced the number of differences to only half the number of actual measurements. With so many

<sup>3</sup> A theoretically more elegant test would treat each individual's vector of differences  $(x_2 - x_1), (x_3 - x_2), \dots, (x_t - x_{t-1})$  as a variable. Then, assuming that as far as differences between observations are concerned the individuals form a homogeneous group, these vectors could be tested for multivariate normality in a single  $F$ -test. However, the procedure described above offers more useful information because the individual  $W$ -statistics can be understood directly by referral to the original data.

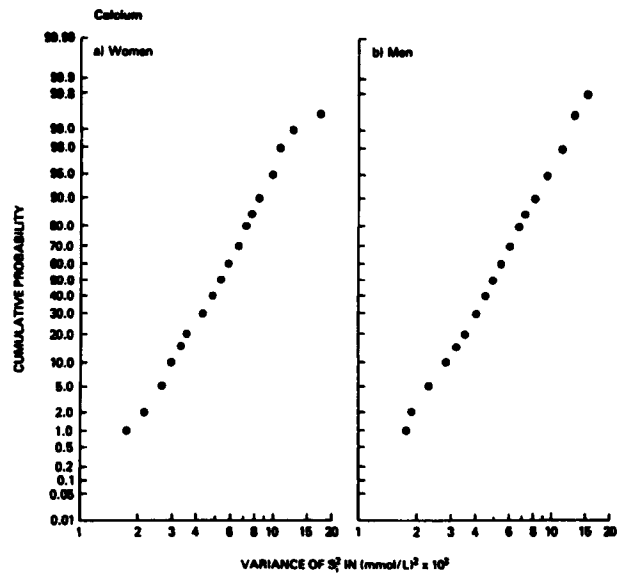


Fig. 1. Cumulative probability distributions of within-person variances in semi-annual measurements of calcium, by sex

individuals, it was clearly not appropriate to compare each value of  $W$  against, say, the 0.05 probability point of the theoretical distribution under normality. Instead, we compared observed and expected proportions of  $W$ -values equal to or below the published fifth and tenth percentile points at sample size eight. In all cases, observed proportions were slightly higher, indicating an overage of smaller  $W$ -values. For calcium, the proportions at or below the theoretical 5 and 10% points were 6.7 and 12.6%, respectively. For alkaline phosphatase, these respective proportions were 8.5 and 17%, respectively. Thus, for calcium, less than 3% of the subjects, and for alkaline phosphatase, about 7%, showed significantly nongaussian distributions of differences. Since the trimming procedure described below protects against extreme observations, we decided to continue the analysis, using the data in original form.

### Observed Distributions of Variance

Figures 1 and 2 show the cumulative distributions of within-person variances in semi-annual measurements of calcium and alkaline phosphatase among 698 men and women with long-term membership in the PL health maintenance program. The distribution of calcium variances in each sex appears to conform fairly well to a log-normal form, but the alkaline phosphatase distributions include some relatively large variances which distort the upper ends of the graphs. To test for outliers and, at the same time, obtain robust estimates of the true mean and standard deviation of nonoutlying  $s_i^2$ -values, we used Healy's trimming procedure (15) in each case, deleting about 2% of the observed variances in each tail (eight values per side in men, six per side in women). Healy's estimating formulas were applied to the remaining  $\log_e s_i^2$ , because in this method it is assumed that the underlying (homogeneous) distribution of the variable is normal. The number of outliers (exceeding  $\pm 3.4$  standard deviations) turned out to be remarkably few: none in calcium, and four variances in alkaline phosphatase among men and two among women. In each case, the estimated mean and variance of the underlying distribution of  $\log_e s_i^2$  values were converted to original scale (formulas for this are given as equations A2 and A3 in the appendix), and equation A1 was applied to obtain

**Table 1. Means and Standard Deviations of Observed Within-Person Variances ( $s_i^2$ ), Adjusted for Outliers, and Estimated Means and Standard Deviations of True Within-Person Variances ( $\sigma_i^2$ ) for Two Analytes, by Sex**

Analyte	$s_i^2$		$\sigma_i^2$	
	Mean	SD	Mean	SD
<b>Calcium (mmol/L)<sup>2</sup></b>				
Men	$5.41 \times 10^{-3}$	$2.35 \times 10^{-3}$	$5.41 \times 10^{-3}$	$1.21 \times 10^{-3}$
Women	$5.58 \times 10^{-3}$	$2.35 \times 10^{-3}$	$5.58 \times 10^{-3}$	$1.11 \times 10^{-3}$
<b>Alk. phos. (U/L)<sup>2</sup></b>				
Men	101	78.0	101	64.6
Women	141	132	141	114

estimates of the standard deviation of the true within-person variances  $\sigma_i^2$ . Results are given in Table 1. This table invites comparison between the estimated standard deviation of  $\sigma_i^2$  and the observed standard deviation of  $s_i^2$ . However, the proper comparison is between the respective squares of these numbers (i.e., the estimated variance of  $\sigma_i^2$  compared with the observed variance of  $s_i^2$ ). Thus, in calcium, the variance in  $\sigma_i^2$  accounts for about 25% of the variance in  $s_i^2$   $[(1.2/2.4)^2]$ . The remaining 75% is due to statistical sampling variability, which would exist even if  $\sigma_i^2$  were constant for all individuals. In alkaline phosphatase, on the other hand, over 70% of the observed variance in  $s_i^2$  is attributable to variation in  $\sigma_i^2$ , less than 30% to sampling variability.

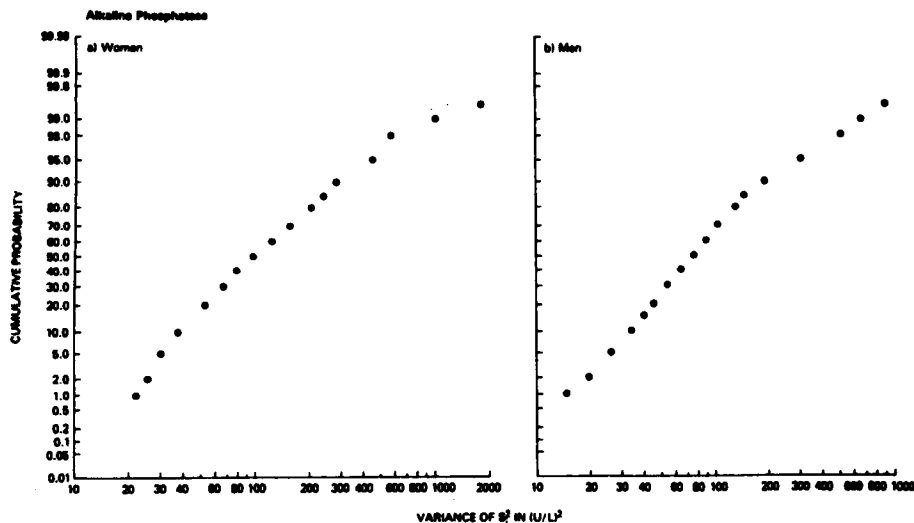
To this point, we have estimated the mean and standard deviation of  $\sigma_i^2$  for each analyte, by sex, and have ascertained that  $\sigma_i^2$  follows an approximately log-normal distribution in each of the four groups. This last conclusion is implied by the approximately log-normal distribution of  $s_i^2$  after adjusting for outliers (see ref. 12 for fuller discussion). The one remaining step before applying equation 2 to determine a reference change  $D_0$  in each analyte/sex group is to examine the distribution of observed serial correlation values, each based on 15-18 semi-annual measurements. The means (medians) and 2.5-97.5 percentile ranges of these correlations are listed in Table 2.

Assuming  $\rho = 0$ , observed correlation values should be normally distributed with variance approximately equal to  $1/n$ , where  $n$  is the number of repeated measurements (16). Using the average value  $n = 16$ , the standard deviation of these correlations should then be 0.25, leading to a middle 95% range

**Table 2. Mean (Median) and 2.5-97.5 Percentile Range of Correlations between Successive Observations, by Analyte and Sex**

Analyte	Mean (median)	2.5-97.5 percentile range
<b>Calcium</b>		
Men	-0.0032 (-0.0023)	-0.43 to +0.42
Women	0.024 (0.017)	-0.40 to +0.52
<b>Alk. phos.</b>		
Men	0.0085 (0.0075)	-0.51 to +0.54
Women	0.12 (0.14)	-0.48 to +0.71

of -0.50 to +0.50. The standard error of the mean correlation would be 0.012 for men and 0.015 for women. Results in Table 2 support the hypothesis  $\rho = 0$  for calcium in both sexes and alkaline phosphatase in men. In all these cases, observed correlation coefficients appeared normally distributed when plotted on probability paper. However, for alkaline phosphatase in women, the mean correlation between successive measurements was significantly greater than zero and the 97.5th percentile higher than expected. In this case, determination of a reference change requires equation 2 and, therefore, a value for  $\sigma_0^2$  including both long-term and within-day analytical variance. A recent study (17) covering the first six years of operation of the PL Tokyo Health Control Center's laboratory estimated the within-day coefficient of analytical variation (CV) for alkaline phosphatase to be 5%



**Fig. 2. Cumulative probability distributions of within-person variances in semi-annual measurements of alkaline phosphatase, by sex**

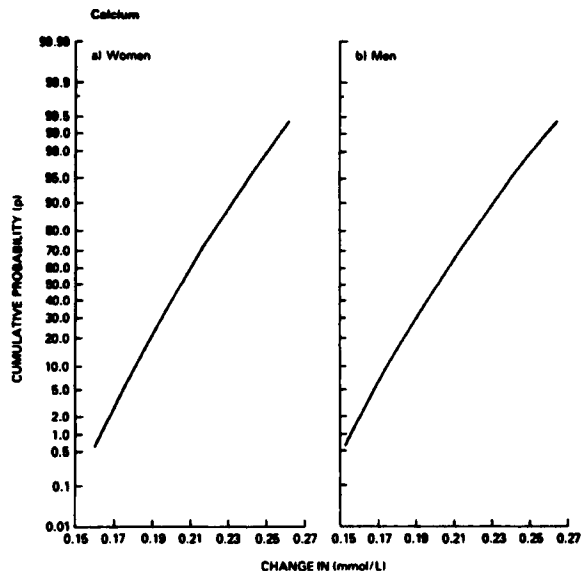


Fig. 3. Graph of the expected proportion of individuals in whom a specified change between two calcium measurements six months apart would be statistically significant at the 5% probability level, by sex

and the long-term CV to be 13%, for a combined CV of 14%. Therefore, given the mean value of 67.6 U/L for alkaline phosphatase in the women studied here,  $\sigma_a^2 = 90$  (U/L)<sup>2</sup>. The Technicon SMA 12/60 has been in operation throughout the time these data were acquired.

#### Calculation of Reference Changes

For calcium in both sexes and alkaline phosphatase in men, equation 3 was used; for alkaline phosphatase in women, we relied on equation 2, setting  $\rho_1 = \rho = 0.12$  and  $\sigma_a^2 = 90$ , but we also used equation 3 for comparison. To obtain the probabilities specified by the left-hand sides of these equations for selected values of  $D_0$ , we assumed in each case a log-normal distribution of  $\sigma_1^2$  with mean and standard deviation given by the figures in the right-hand columns of Table 1. Actual calculations were performed through a computer program in BASIC (available from E.K.H. on request). Results are shown in Figures 3 and 4. In these graphs, the abscissa covers a range of possible reference changes ( $D_0$  values), while the ordinate indicates the cumulative probability  $p$  (right-hand side of equation 2) associated with a specific reference change. For convenience, a normal probability scale has been used. In the case of calcium (Figure 3), the curves are similar for both men and women. For example, a change of 0.20 mmol/L between two consecutive observations would be statistically significant

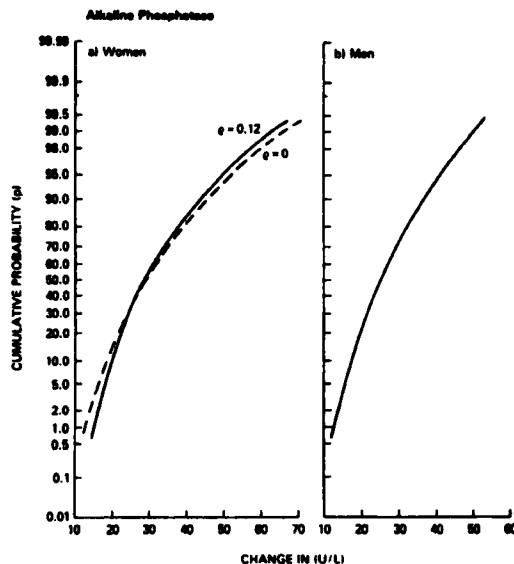


Fig. 4. Graph of the expected proportion of individuals in whom a specified change between two alkaline phosphatase measurements six months apart would be statistically significant at the 5% probability level, by sex

(i.e., expected by chance no more than 5% of the time) in only 40–50% of the individuals in this study. However, a change of 0.25 mmol/L would be significant in about 95% of these apparently healthy persons. This latter difference would then represent a more conservative reference change.

In the case of alkaline phosphatase (Figure 4), the results for men and women differ considerably. For example, a change of 25 U/L would be statistically significant in about half the men in this population but in only about 32% of the women. To be significant in 95% of healthy individuals, reference changes of 41 U/L for men and 50 U/L for women would be required. The higher figure for women reflects a 40% greater within-person variance, on the average, in women and a wider distribution of such variances than among men. This may be due, in part, to the inclusion of women who crossed the menopause during the 7.5–9 year span of this study. Post-menopausal women may undergo marked increases in alkaline phosphatase (e.g., 18–20), so that a set of measurements of this enzyme at pre- and post-menopausal times could show relatively large variation. An examination of the effects of age decade (by sex) on within-person variances has not been included in the present analysis, but we hope to undertake such a study soon for all blood constituents measured in the PL Health Control System.

Note that in Figure 4a, the curve for  $\rho = 0.12$  is somewhat steeper than that for  $\rho = 0$ . It is true in general that as  $\rho$  increases through its range from  $-1$  to  $+1$ , the curve of  $D_0$  vs  $p$  becomes increasingly steeper. Because the selected reference change would probably be that value of  $D_0$  for which  $p = 0.90$  or  $0.95$  (certainly  $>0.5$ ), the higher the serial correlation between successive values, the smaller the reference change. This is as expected in view of the fact that the variance of differences between adjacent observations in an autoregressive model decreases as  $\rho$  gets larger. For example, a negative value of  $\rho$  implies an oscillatory series, which would produce a relatively high variance in the difference between consecutive values.

#### Comparison with Current Procedure

In a previous section, we discussed the analysis of observed differences to test if they were normally distributed. As part

Table 3. Standard Deviations of Observed Differences by Analyte and Sex ( $s_\Delta$ ) and Reference Changes Computed According to Current Practice

Analyte	Reference changes	
	$s_\Delta$	$\pm 2s_\Delta$
Calcium, mmol/L		
Men	0.10	0.20
Women	0.10	0.20
Alkaline phosphatase, U/L		
Men	13.6	27.2
Women	14.5	29.0

of that analysis, the standard deviation of these differences was computed for each analyte/sex group ( $s_{\Delta}$ ), allowing us to calculate reference changes according to the procedure now being used, simply  $\pm 2s_{\Delta}$ . The figures are given in Table 3.

These reference changes correspond to p-values around 50%, as indicated in Figures 3 and 4; that is, they represent statistically significant changes in only about half the healthy individuals in this study. This is because the standard deviation of the distribution of differences is not affected by variation in within-person variances, but only by the mean within-person variance (and also by person-to-person variability in mean differences, but this will be relatively small). On the other hand, kurtosis of the distribution is increased by variation in within-person variances. We might expect, therefore, that use of the completely nonparametric procedure that selected the 2.5th and 97.5th percentiles of the distribution of observed differences would produce more conservative reference changes.

For calcium, this was not the case. The percentiles were  $\pm 0.20$  mmol/L, identical to  $\pm 2s_{\Delta}$  (Table 3). This result is not surprising given the low CV of  $\sigma_i^2$  (Table 1). However, even for alkaline phosphatase, where the CV of  $\sigma_i^2$  was much higher, the 2.5th and 97.5th percentiles were very close to  $\pm 2s_{\Delta}$ . The values were  $-26.9$  and  $28.2$  U/L, respectively, in men and  $-29.4$  and  $31.0$  U/L in women. These results indicate that reference changes derived solely from the distribution of observed differences are likely to generate many false alarms.

## Discussion

Suppose we define a reference change, at least for apparently healthy individuals, as that difference between two successive values which would be statistically significant ( $p \leq 0.05$ ) in 95% of such persons. This is perhaps too conservative a criterion, but we propose it as a basis for discussion. In this study, as noted above, the values  $D_0 = 0.25$  mmol of calcium per liter in both sexes, 41 U/L for alkaline phosphatase in men, and 50 U/L in women satisfy this definition of a reference change. Of course, such values may depend not only on the population of individuals observed and the analytical procedures used, but also on the time interval between successive measurements. For example, in the earlier study by Harris and Brown (8), the comparable  $D_0$ -value for calcium was 0.13 mmol/L, about half the value found here. In that study, the subjects (healthy men) were observed weekly for about five months. In the present study the time interval was six months and the span of data, 7.5–9 years. The average within-person standard deviation in calcium was roughly twice as high, presumably owing to the much greater time span of data. Not only biologically induced variation, but analytical variance as well would probably increase with a longer time span. For purposes of a health-maintenance program, a reference change based on measurements made six months apart seems more suitable than weekly or monthly observations. In such programs, reference changes should be used only to compare early measurements in a continuing series. After a record of four or more serial observations has been established for a given subject, tests of current deviations should be based on time-series models applied to the entire sequence of previous data (e.g., ref. 21).

Thus, reference changes represent a bridge between conventional (cross-sectional) reference values, which are entirely population-based, and subject-specific reference values, which are derived from past results for a single individual. The reference change utilizes within-person variances but considers their distribution across a population and is based, in the end, on selection of a cumulative probability ( $p$ ) characterizing that population. The reference change concept may also be extended to the difference between the first and third observa-

tions in a series in order to identify a significant trend within the first three observations. In this case, the appropriate formula analogous to equation 1, based on the autoregressive model, is

$$\text{Var}(x_{ti} - x_{(t-2)i}) = 2\sigma_i^2[(1 - \rho^2) + \rho^2\sigma_i^2/\sigma_i^2] \quad (4)$$

replacing  $\rho_i$  by its average value  $\rho$ .

This equation is identical to equation 1 except for  $\rho^2$  in place of  $\rho$ . Therefore, for  $\rho > 0$ , and the same value of  $p$ , we would obtain a larger value of the reference change  $D_0$  since  $\rho^2 < \rho$ . For example, consider alkaline phosphatase in women ( $\rho = 0.12$ ), and assume we were interested only in positive differences (increases). In equation 2, we would substitute 1.65 for 1.96 to account for the switch from a two-tailed to a one-sided test of significance at the 0.05 level. Then, the reference change ( $D_0$  value at  $p = 0.95$ ) between the first and second test result becomes 42 U/L. Between the first and third observation, this reference change would increase to 44 U/L. Thus an increase of, say, 25 U/L at the second observation would be too small to be alarming in itself, but if that same increase were repeated over the next time interval, it would be clear that a statistically significant trend had occurred.

The problem of reference changes in hospital patients is more complicated than in healthy subjects, both conceptually and statistically. One aspect of this problem was discussed in the most recent report on Delta-checks (7), namely, the unpredictable and often substantial changes between two consecutive test results caused by disease processes or intervening treatment. Undoubtedly, there are kinds of patients in whom this difficulty is not likely to arise, at least in the short run, and for whom the development of reference changes would be useful. However, these patient groups must first be carefully defined and might well differ in type from one hospital to another. Further, for many such patients, no more than two or three successive results may exist under steady-state conditions. These results are likely to have been obtained at relatively short intervals—e.g., only a day or two apart—introducing perhaps a much higher serial correlation than that found in semi-annual examinations of healthy persons or outpatients. Also, the range of results for a given constituent may be much wider than in healthy subjects, and different reference changes might be appropriate for different sub-ranges of the constituent (6). None of these statistical problems is insurmountable; there are valid ways of adapting the procedures described in this paper to meet all these situations. The result would be a set of parametric reference changes which might be compared, for example, with the nonparametric Delta-checks proposed by Wheeler and Sheiner (6) based on the frequency distribution of observed differences between consecutive test values. The critical problem remains that of defining suitable patient categories in which to collect data.

We are indebted to Mr. George Shakarji and Mr. F. David Van Sant, whose considerable skills in programming and data processing were required for the completion of this work. We are also grateful to Dr. Adelin Albert for his careful review and comments on an earlier draft.

## Appendix

Formulas useful in calculating reference changes include:

1. Estimating the variance of true within-person variances ( $\sigma_i^2$ ) from the mean and variance of observed variances ( $s_i^2$ ):

$$\begin{aligned} \text{(Estimated) Var } \sigma_i^2 \\ = [\text{Var } s_i^2 - (2/n - 1)(\text{mean } s_i^2)^2] (n - 1)/(n + 1) \quad (A1) \end{aligned}$$

where  $n$  is the average number of within-person observations.

2. Converting the mean and variance of  $\log_e s_1^2$  to the mean and variance of  $s_1^2$ , when  $\log_e s_1^2$  is normally distributed: Let  $M$  = mean  $\log_e s_1^2$ , and  $V$  = variance  $\log_e s_1^2$ . Then,

$$\text{mean } s_1^2 = \exp[M + (V/2)] \quad (\text{A2})$$

$$\text{var } s_1^2 = \exp(2M + V)[\exp(V) - 1] \quad (\text{A3})$$

where  $\exp(\quad)$  denotes the exponential function of the quantity in parentheses.

## References

1. Skendzel, L. P., How physicians use laboratory tests. *J. Am. Med. Assoc.* **239**, 1077-1080 (1978).
2. Elion-Gerritsen, W. E., Analytic precision in clinical chemistry and medical decisions. *Am. J. Clin. Pathol.* **73**, 183-195 (1980).
3. Nosanchuk, J. S., and Gottmann, A. W., CUMS and Delta checks. *Am. J. Clin. Pathol.* **62**, 707-712 (1974).
4. Whitehurst, P., DiSilvio, T. V., and Boyadjian, G., Evaluation of discrepancies in patients' results—an aspect of computer-assisted quality control. *Clin. Chem.* **21**, 87-92 (1975).
5. Ladenson, J., Patients as their own controls: Use of the computer to identify "laboratory error." *Clin. Chem.* **21**, 1648-1653 (1975).
6. Wheeler, L. A., and Sheiner, L. B., Delta check tables for the Technicon SMA 6 continuous-flow analyzer. *Clin. Chem.* **23**, 216-219 (1977).
7. Wheeler, L. A., and Sheiner, L. B., A clinical evaluation of various Delta check methods. *Clin. Chem.* **27**, 5-9 (1981).
8. Harris, E. K., and Brown, S. S., Temporal changes in the concentrations of serum constituents in healthy men. *Ann. Clin. Biochem.* **16**, 169-176 (1979).
9. Harris, E. K., Yasaka, T., Horton, M., and Shakarji, G., Comparing multivariate and univariate subject-specific reference regions for blood constituents in healthy persons. *Clin. Chem.* **28**, 422-426 (1982).
10. Yasaka, T., Health-control data-base system and subject-specific normal range. *Med. Informatics* **1**, 105-132 (1976).
11. Yasaka, T., Nakano, K., Nakano, M., and Suganuma, G., Normalized Q-sum method for clinical data evaluation (homeostasis dynamics). In *Medical Computing: Proc. of an Int. Symposium*, M. Laudet, J. Anderson, and F. Begon, Eds., Taylor & Francis Ltd., London, 1977, pp 311-328.
12. Harris, E. K., Distinguishing physiologic variation from analytic variation. *J. Chron. Dis.* **23**, 469-490 (1970).
13. Harris, E. K., Kanofsky, P., Shakarji, G., and Cotlove, E., Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. II. Estimating biological components of variation. *Clin. Chem.* **16**, 1022-1027 (1970).
14. Shapiro, S. S., and Wilk, M. B., An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591-611 (1965).
15. Healy, M. J. R., Outliers in clinical chemistry quality-control schemes. *Clin. Chem.* **25**, 675-677 (1979).
16. Box, G. E. P., and Jenkins, G. M., *Time Series Analysis, Forecasting and Control*, 2nd ed., Holden-Day, San Francisco, CA, 1976, pp 35, 178.
17. Yasaka, T., Nakano, K., Nakano, M., and Komatsu, T., The long-term quality control and clinical data evaluation of individuals in AMHTS. *Med. Informatics*, **4**, 173-185 (1979).
18. Werner, M., Tolla, R. E., Hultin, J. V., and Mallecker, J., Influence of sex and age on the normal range of eleven serum constituents. *J. Clin. Chem. Clin. Biochem.* **8**, 105-115 (1970).
19. Wilding, P., Rollason, J. G., and Robinson, D., Patterns of change for various biochemical constituents detected in well population screening. *Clin. Chim. Acta* **41**, 375-387 (1972).
20. Goldberg, D. M., and Winfield, D., Relationship of serum enzyme activities to demographic variables in a healthy population. *Clin. Chim. Acta* **54**, 357-368 (1974).
21. Harris, E. K., Cooll, B. K., Shakarji, G., and Williams, G. Z., On the use of statistical models of within-person variation in long-term studies of healthy individuals. *Clin. Chem.* **26**, 383-391 (1980).