Design of Quality-Control Specimens for Use with a Small Multi-Channel Analyzer

Timothy E. Caragher and George F. Grannis

Specimens having linear relationships between concentration and instrument response for seven analytes were prepared for use with a small multi-channel analyzer. The specimens had inter-specimen and inter-constituent relationships that facilitated the performance evaluation of all channels simultaneously with each individual specimen. Techniques are described for use of the specimens in various aspects of quality control.

Additional Keyphrases: analytical error • calibration materials • performance evaluation • precision and accuracy • continuous-flow analysis • proficiency • use and limitations of the correlation coefficient • linear regression analysis

Automated instruments that are used to analyze specimens for several analytes simultaneously ("multi-channel" analyzers) are now commonplace in clinical chemistry laboratories. Most of these instruments are standardized by the technique commonly referred to as "single-point calibration." The technique usually provides reasonably accurate results near the calibration concentration, but for a variety of reasons significant errors can result at other concentrations (1–4). Even when the instrument is calibrated correctly initially, the baseline or calibration set-points, or both, may change during the analytical run owing to changes in the reagents or the instrumental system. In monitoring and controlling this drift it is customary to analyze calibration "check" specimens at frequent intervals and to readjust the initial set-points appropriately (5). However, repeated readjustments of either the baseline or the calibration set-point during a series of analyses may result in progressively inaccurate values, even at the calibration concentration (6, 7). For these and other reasons it is also customary to analyze control specimens in each analytical run to verify that the analyses were within acceptable limits.

Sets of serum-based specimens that have linear inter-relationships have been described (8–12). With such specimens it is possible to evaluate method linearity, accuracy, and precision over a wide range of analyte concentration, and to use a variety of statistical and graphical techniques to condense and summarize the control data. Linearity related specimens should be particularly useful with multi-channel analyzers because these instruments quickly generate large numbers of results and require intensive bench-level quality control, owing in part to the uncertainties inherent to the technique of single-point calibration.

The purpose of this report is to describe the design and use of inter-related control specimens in the performance evaluation of a small multi-channel analyzer.1

Materials and Methods
Preparation of the Specimens

Preparation of the base serum pool: Various commercial lyophilized serum preparations were reconstituted with distilled water to 80% of their normal reconstitution volumes, and combined. This concentrated pool was analyzed for sodium, potassium, chloride, total carbon dioxide, urea nitrogen, creatinine, and glucose, and was dialyzed (0.0010-inch thickness dialysis tubing; A. H. Thomas Co., Philadelphia, Pa. 19104) at 4°C to approximate equilibrium, against two volumes of an aqueous solution of these analytes. The analyte concentrations of the aqueous solution were those calculated to yield, at equilibrium, the desired concentrations in the dialyzed pool. The equilibrium dialysis was done to decrease the concentrations of all dialyzable solutes except those of interest in the present study, while

---

1 Detailed information concerning the design, preparation, and use of inter-related specimens may be found in the dissertation "The Design of Serum-based Quality Control Material Yielding Enhanced Interpretative Information for Multi-Channel Analyzers," available from University Microfilms International, Ann Arbor, Mich. 48106.
concomitantly adjusting the analytes of interest to approximately the desired concentrations. The dialyzed pool had the following concentrations of analytes of interest: sodium, 149.9 mmol/liter; potassium, 3.5 mmol/liter; chloride, 99.4 mmol/liter; total carbon dioxide, 11.0 mmol/liter; urea nitrogen, 129 mg/liter; creatinine, 9.4 mg/liter; and glucose, 976 mg/liter.

Preparation of specimens: We prepared the following six aqueous solutions, using analytical-grade chemicals in the case of sodium chloride, potassium nitrate, and ammonium carbonate or National Bureau of Standards Standard Reference Materials (SRM) in the case of urea, creatinine, and glucose: sodium chloride, 802.5 mmol/liter; potassium nitrate, 90.50 mmol/liter; ammonium carbonate, 619.2 mmol/liter; urea, 8312.4 mg of urea nitrogen/liter; creatinine, 1121.4 mg/liter; and glucose, 42697 mg/liter. Six portions of the dialyzed serum pool then were diluted, respectively, with exactly 1/4 volume of each of the aqueous solutions. These six serum pools, each of which had been supplemented with one of the analyte solutions, were combined in various proportions as shown in Table 1, to yield six admixtures. Two-milliliter aliquots were dispensed, frozen, and stored at −20 °C.

An additional set of dialyzed specimens was prepared by diluting portions of the six admixtures with exactly 1/4 volume of distilled water. Two-milliliter aliquots were dispensed, frozen, and stored at −20 °C.

Analytical Methods

Continuous-flow analyses were performed with three SMA 6/60 instruments (Technicon Instruments Corporation, Tarrytown, N.Y. 10591) by the following methods: sodium and potassium, flame photometry (13); chloride, mercuric thiocyanate (13); total carbon dioxide, cresol red (13); urea nitrogen, diacetyl monoxime (13); creatinine, alkaline picrate (13); and glucose, glucose oxidase (14). Manual analyses for sodium, chloride, and total carbon dioxide were done, respectively, with the IL (Instrumentation Laboratories) Model 343 flame photometer, the Corning Model 920 Chloridometer, and the Kopp–Natelson microgasometer technique (15).

Assignment of Expected Analytical Values

Expected values for the specimens were calculated from analysis of the base pool and knowledge of the amount of each analyte weighed into the diluents used in the preparation of the supplemented pools, and of the proportions of the supplemented pools used in the preparation of the admixtures from which the specimens were dispensed. Except for total carbon dioxide, the analyte concentrations of the base pool were determined by concomitant analysis with three specimens of Survey Validated Reference Material [SVRM; College of American Pathologists’ Survey Serum (16, 17)]. The reference values for the latter specimens were the mean values established in the CAP Survey Program by hundreds of laboratories (18). Because the SVRM did not have an assigned value for carbon dioxide, the carbon dioxide content of the base pool was determined by the Kopp–Natelson microgasometric technique (15).

Data Collection

Two of the three SMA 6/60 instruments (no. 1 and no. 2) were used in analyses for sodium, potassium, chloride, total carbon dioxide, urea nitrogen, and glucose. The third instrument (no. 3) was used in analyses for sodium, potassium, urea nitrogen, creatinine, and glucose. All specimens of the sets were initially analyzed in single analytical runs, to evaluate the specimens. Thereafter, the specimens were analyzed in various combinations to evaluate instrument performance and to provide the ancillary data for this study.

Results and Discussion

Method and Rationale of Data Analysis

The analytical results for each analyte were subjected to linear regression analysis (ordinate, reported values; abscissa, % of supplemented pool) and were summarized as the correlation coefficient, the X- and Y-intercepts, slope, and apparent percent recovery of added analyte. The reason for using these parameters to summarize results is based on (a) the fundamental proposition that the results produced by any quantitative analytical procedure should be linearly correlated with the known results (1, 2, 9) and (b) the fact that the specimens were designed to have known quantitative linear inter-relationships, which a properly calibrated analytical system should reflect. The failure to find the expected values should be indicative of various kinds of problems, either with the system itself or with the specimens.

This basic concept is illustrated in Figure 1, in which the reported values for some analyte are graphed against the expected values. Ideally, the results should fall along

<table>
<thead>
<tr>
<th>Specimen no:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urea pool</td>
<td>0.00</td>
<td>33.33</td>
<td>26.67</td>
<td>20.00</td>
<td>13.33</td>
<td>6.67</td>
</tr>
<tr>
<td>NaCl pool</td>
<td>6.67</td>
<td>0.00</td>
<td>33.33</td>
<td>26.67</td>
<td>20.00</td>
<td>13.33</td>
</tr>
<tr>
<td>K⁺ pool</td>
<td>13.33</td>
<td>6.67</td>
<td>0.00</td>
<td>33.33</td>
<td>26.67</td>
<td>20.00</td>
</tr>
<tr>
<td>CO₂ pool</td>
<td>20.00</td>
<td>13.33</td>
<td>6.67</td>
<td>0.00</td>
<td>33.33</td>
<td>26.67</td>
</tr>
<tr>
<td>Creatinine pool</td>
<td>26.67</td>
<td>20.00</td>
<td>13.33</td>
<td>6.67</td>
<td>0.00</td>
<td>33.33</td>
</tr>
<tr>
<td>Glucose pool</td>
<td>33.33</td>
<td>26.67</td>
<td>20.00</td>
<td>13.33</td>
<td>6.67</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Proportion by volume
line 1, representing a perfect correlation of determined and expected values. However, quantitative analytical procedures are always subject to proportionate or constant biases, or both, that may cause results to deviate from the ideal relationship \((1, 2, 19)\). In Figure 1, lines 2 and 3 represent the effects of proportionate and constant biases, respectively, and line 4 represents the effects of a combination of such biases.

The lines shown in Figure 1 have been called "operational" lines because every clinical laboratory, in every analytical procedure, produces results that fall along some line when graphed as in Figure 1 (2). One of the basic objectives in quality control is to determine the laboratory's actual operational lines and, so far as possible, attain and maintain the ideal relationship, a line of slope 1.0 (line 1). The laboratory's operational line for each analyte can be determined most easily and reliably by using specimens known to have quantitative linear inter-relationships.

The specimens were prepared by making quantitative admixtures of portions of a serum pool that had been supplemented with the analytes of interest. When the determined concentrations of an analyte in a set of such specimens are graphed against the known amount (i.e., percentage) of supplemented pool in the specimens, as in Figure 2, the results should fall along a straight line that can be characterized by linear regression analysis. The *correlation coefficient*, \(r\), provides an estimate of the correlation of the analytical values with the known specimen composition, and may be considered a measure of the linearity of the analyses. In general, a value for \(r\) of less than 0.98 indicates that one or more of the analyses was noticeably removed from the best straight line fit of the data (see Appendix). Such a result may be due to (a) a single "outlier"; (b) generally disperse, or imprecise, data; or (c) a curvilinear relationship. The Y-intercept of the best straight line is an estimate, based on the analyses of the full set of specimens, of the concentration of analyte in the base pool (i.e., the specimen with 0% of supplemented pool, see Figure 2). The expected analyte concentration of this pool was determined by analysis of the pool in parallel with the SVRM. That is, the SVRM was used as a standard for the analysis of the base pool. Thus, a comparison of the expected and determined values of the Y-intercept indicates how well the analytical system is calibrated with respect to the hundreds of laboratories who collectively established the target value of the SVRM.

The X-intercept and slope have particular relevance for evaluation of instrumental systems that use the technique of single-point calibration. Implicit in this technique are the assumptions that (a) an appropriate blank, or baseline, setting representing zero concentration of analyte can be made; (b) that the instrumental readout is directly proportional to analyte concentration; and (c) that the proportionality constant established when the instrument is calibrated is appropriate for the analysis of serum specimens. The X-intercept is a value extrapolated from the analyses of specimens having a range of analyte concentrations, and it corresponds to the composition (as per cent of supplemented pool, see Figure 2) of a hypothetical specimen having zero concentration of analyte. Consequently, the X-intercept is most appropriate for evaluating the baseline setting of an instrument. Deviations of the X-intercept from the expected value should be indicative of analytical bias similar to that represented by line 3 or 4, Figure 1.

The *slope* is a measure of the concentration gradient that was established during the preparation of the specimens, and the expected slope is calculated from the amount of weighed analyte known to have been added to the specimens. Consequently, the slope determined by regression analysis of the analytical results provides an indication of how well the system is calibrated to account for added analyte. The apparent analytical recovery of added analyte is conveniently expressed as a percent, which is calculated by multiplying the ratio...
of the determined slope to the expected slope by 100.² The percent recovery is a measure of the deviation of the slope of the operational line from the ideal, and such deviations are indicative of biases illustrated by line 2 or 4, or both, in Figure 1.

Evaluation of Analytical Biases

A complete set of specimens was analyzed with the three SMA 6/60 instruments on several occasions, and the linear regression parameters were averaged to provide representative performance parameters for each instrument at the time of the study. The average correlation coefficients, after exclusion of occasional outliers, were greater than 0.99, which indicates that the data were linearly distributed. The X- and Y- intercepts and slopes are compared with expected values in Table 2, and the results for the instruments showing the worst performance (i.e., the largest deviations from expected values) are illustrated in Figure 3 A–G. These results will be discussed by analyte:

² In the equation of a straight line, \( Y = mx + b \), \( m \) is the slope and \( b \) is the Y-intercept. Rearranging, \( Y - b = mx \). If \( Y \) is the concentration of analyte in a supplemented pool and \( b \) is the concentration of analyte in the base pool, then \( Y - b \) is the concentration of analyte due to weighed-in additions to the supplemented pool. When \( X \) has a value of 100% (i.e., 100% supplemented pool), \( Y - b = 100m \). Thus, the slope multiplied by 100 is an estimate of the analyte added in the preparation of the specimen.

Potassium: The data in Table 2 show that instrument no. 3 yielded values for the X- and Y- intercepts and slopes that were virtually identical with the expected values, while instruments 1 and 2 demonstrated small deviations of the X-intercept and slope (see Figure 3A). The fact that one instrument agreed well with expected values suggested that a problem, albeit minor, existed with the other instruments. However, after routine maintenance (replacement of worn lines, dirty coils, etc.), analyses of the sets of specimens on these instruments yielded values of the X-intercepts and slopes that also agreed well with the expected values.

Sodium: Figure 3B and the data in Table 2 show that all three instruments failed to yield the expected X- and Y- intercepts and slopes. However, all three instruments were similarly biased and the recovery of added analyte was quite reproducible among the instruments.

The failure of sodium analyses to account for added analyte was reported earlier (10) and remains a small but vexing problem. We have prepared specimens with weighed-in amounts of sodium chloride on numerous occasions, and recovery of sodium has usually been about 95% with the SMA 6/60, but about 100% by manual analysis with the IL flame photometer. The fact that similarly biased results are obtained with three instruments suggests that the bias resides with either the method or the instruments.
Table 2. Comparison of the Determined Linear-Regression Parameters with Expected Values

<table>
<thead>
<tr>
<th></th>
<th>Undiluted specimens</th>
<th>Diluted specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Y-Intercept</td>
<td>X-Intercept</td>
</tr>
<tr>
<td><strong>Potassium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected value</td>
<td>2.80</td>
<td>-15.5</td>
</tr>
<tr>
<td>Instrument 1</td>
<td>2.80</td>
<td>-16.2</td>
</tr>
<tr>
<td>Instrument 2</td>
<td>2.86</td>
<td>-16.9</td>
</tr>
<tr>
<td>Instrument 3</td>
<td>2.80</td>
<td>-15.5</td>
</tr>
<tr>
<td><strong>Sodium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected value</td>
<td>119.9</td>
<td>-74.7</td>
</tr>
<tr>
<td>Instrument 1</td>
<td>122.6</td>
<td>-79.3</td>
</tr>
<tr>
<td>Instrument 2</td>
<td>121.8</td>
<td>-78.5</td>
</tr>
<tr>
<td>Instrument 3</td>
<td>121.5</td>
<td>-80.2</td>
</tr>
<tr>
<td><strong>Chloride</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected value</td>
<td>79.5</td>
<td>-49.5</td>
</tr>
<tr>
<td>Instrument 1</td>
<td>84.2</td>
<td>-63.8</td>
</tr>
<tr>
<td>Instrument 2</td>
<td>83.1</td>
<td>-61.4</td>
</tr>
<tr>
<td><strong>Carbonate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected value</td>
<td>9.0</td>
<td>-7.1</td>
</tr>
<tr>
<td>Instrument 1</td>
<td>9.8</td>
<td>-7.3</td>
</tr>
<tr>
<td>Instrument 2</td>
<td>10.3</td>
<td>-7.8</td>
</tr>
<tr>
<td><strong>Urea nitrogen</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected values</td>
<td>103</td>
<td>-6.2</td>
</tr>
<tr>
<td>Instrument 1</td>
<td>111</td>
<td>-6.6</td>
</tr>
<tr>
<td>Instrument 2</td>
<td>102</td>
<td>-6.0</td>
</tr>
<tr>
<td>Instrument 3</td>
<td>112</td>
<td>-7.1</td>
</tr>
<tr>
<td><strong>Creatinine</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected value</td>
<td>08</td>
<td>-3.3</td>
</tr>
<tr>
<td>Instrument 3</td>
<td>09</td>
<td>-3.9</td>
</tr>
<tr>
<td><strong>Glucose</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected value</td>
<td>779</td>
<td>-9.1</td>
</tr>
<tr>
<td>Instrument 1</td>
<td>772</td>
<td>-9.3</td>
</tr>
<tr>
<td>Instrument 2</td>
<td>774</td>
<td>-9.8</td>
</tr>
<tr>
<td>Instrument 3</td>
<td>767</td>
<td>-9.5</td>
</tr>
</tbody>
</table>

The data of Table 2 show that the undiluted specimens, ranging in sodium concentration from 120 to 173 mmol/liter, yielded essentially the same X-intercept and apparent recovery as did the diluted specimens, in which sodium concentration ranged from 96 to 140 mmol/liter. These results, obtained from the upper and lower halves of the analytical range, indicate that the instrument read-out was linear over this entire range of concentrations. However, the deviations of the X-intercept and slope from expected values indicate a systematic bias of the operational line. In Figure 3B it is seen that the determined and expected lines intersect near the calibration set-point (~145 mmol/liter), but that there is a rather large deviation of the X-intercept. This bias is consistent with an inappropriate setting of the baseline, which could possibly be due to a slightly nonlinear photometer response at low sodium concentrations. Figure 3B illustrates how the technique of single-point calibration may result in a clearly mis-calibrated analytical system and yet yield results that are reasonably accurate over the clinically significant range of concentrations.

Chloride: Figure 3C and the data in Table 2 show that neither instrument yielded expected values for both the X- and Y-intercepts and slopes. The expected slopes for chloride are identical to those for sodium, because both of these analytes were added to the specimens in equimolar amounts, as sodium chloride. The specimens were apparently satisfactory, because analysis for sodium and chloride by the manual procedures yielded the expected values, with the slopes for sodium and chloride in close agreement with one another.

In Table 2 the X-intercepts and apparent recoveries of the undiluted and diluted specimens are distinctively different. The undiluted specimens encompassed a range of chloride concentration from 80 to 133 mmol/liter and yielded an X-intercept that was more negative than expected and a slope that was only 82% of the expected value. Chloride concentration in the diluted specimens ranged from 64 to 106 mmol/liter and yielded an X-intercept less negative than expected, and a slope equal to or greater than expected. Thus, the specimen sets covering the lower and upper portions of the analytical range yielded different lines, indicating that the instrument read-out was nonlinear.

The instrument is calibrated by (a) setting a baseline
corresponding to zero concentration of chloride; (b) standardizing or “sloping” the instrument with two aqueous standards having chloride concentrations of 70 and 130 mmol/liter, and using an electronically expanded scale; and (c) “calibrating” the instrument with a serum-based reference specimen having a chloride concentration of 112 mmol/liter. Figure 3C illustrates how this procedure may lead to accurate results near the calibration set point but to clearly biased results at other concentrations.

As the final calibration at 112 mmol/liter always involved a downward adjustment of the calibration setpoint, there is a clear difference between the aqueous standards and the calibration serum, which is probably due to different rates of dialysis of chloride ion. When we have used serum-based materials, the chloride concentrations of which had been established by coulometric analysis, to standardize the instrument at 70 and 130 mmol/liter (in place of the recommended aqueous standards), analysis of our inter-related specimens then yield determined values that correspond well with the expected values (Figure 4). Thus the aqueous standards may be inappropriate for the initial “sloping” of the instrument for analysis of serum specimens.

Carbon dioxide: Figure 3D and the data in Table 2 show that the expected X-intercepts were obtained with both instruments. However, both the Y-intercepts and slopes were greater than expected. This failure of CO₂ analyses to yield expected values was found to be due in part to the use of a commercial calibration serum that had an erroneous assigned value, and in part to the loss of CO₂ from the calibration material during the course of several hours of its use. Repeated analyses of the calibrator by the Kopp–Natelson procedure yielded an average value of 23.5 mmol/liter rather than the assigned value of 25.0 mmol/liter. Because samples from other lots of the same material were found to have the expected values, the calibrator in use apparently had an erroneous assigned value. In addition, we have confirmed the well-known fact that when the freshly reconstituted calibration material is stored in capped vials, which are opened periodically to remove samples, CO₂ is gradually lost during 4 h. This loss is presumably due to repeated equilibration of the CO₂ in the solution with that in the air above the solution, CO₂ being lost each time the vial is opened for sampling. To correct these problems, we assigned the calibration material the assay value of 23.5 mmol/liter, and the calibration material was kept in a syringe with no air space above the sample.

Interestingly, in the total-CO₂ procedure the measured absorbance is not linearly related to analyte concentration over the full analytical range (20); the instrument read-out is linearized by the manufacturer by the use of a non-uniform scale on the recorder chart paper. In Figure 5, the concentration stated on the scale is graphed against the scale length in millimeters, and an unnatural discontinuity is noted at a concentration of 10 mmol/liter. Since most clinical specimens have a total CO₂ concentration greater than 10 mmol/liter, this method nonlinearity is of little practical consequence.

Urea nitrogen and creatinine: Figure 3E and F and Table 2 show that our urea nitrogen and creatinine analyses yielded values close to the expected values. Instruments 1 and 2 had slopes for urea nitrogen that slightly exceeded the expected value; values from instrument 3 were low. These differences are not considered consequential.

Glucose: The results for glucose (Figure 3G and Table 2) are of interest in that the determined X-intercepts agreed well with the expected values, while the slopes were variably less than expected. The decreased slopes were probably due to loss of glucose from the specimens, because no bacteriostatic agent had been added, and the earliest evaluations indicated acceptable method cali-
bration. These results provide an example of a proportionate bias that is due to gradual specimen deterioration.

Some Control Charts for Use with Linearly Related Specimens

The specimens (diluted and undiluted) were also analyzed randomly with one instrument during several weeks, to provide representative control data. Figure 6 illustrates one kind of chart that was developed to provide a simultaneous evaluation of all channels with each individual specimen. Although Figure 6 shows the chart for specimen no 1, similar charts may be prepared for the other specimens in a set. On each chart the scale for each analyte is positioned on the abscissa according to the percentage composition of that analyte, which was established in the preparation of the specimens. Although the analytical data are plotted in the usual concentration units of each analyte, the ordinate is also a relative scale ranging from 0 to 100, in which a value of 0 represents the analyte concentration in the serum base that was used in the preparation of the specimens and a value of 100 represents the highest concentration in a supplemented specimen. On each chart the analytes are ranked on the abscissa in order of their percent composition, and consequently, the expected values are ranked in the same order on the ordinate. Because both the ordinate and abscissa have the same relative range of units (0 to 100) the analytical results should correlate exactly with the expected values and fall along the dashed diagonal line of slope 1.0. In use, each analytical result from a single specimen may be graphed on the appropriate specimen chart; the result for each analyte should not only fall within specified ranges, but all results should fall along the diagonal line. Systematic deviations of all analyses from the dashed line should be indicative of problems that affect all analytes (e.g., inadequate specimen sampling) while deviations of single analytes indicate problems only with that channel. Thus these charts facilitate monitoring all analytical procedures in each analytical run.

When the specimens are used sequentially in successive analytical runs, a second series of charts may be developed for each analyte. Figure 7 shows such a chart for potassium; similar charts may be prepared for each of the other analytes, in each of which the specimens are ranked on the abscissa in order of their analyte composition and the data are again plotted on concentration scales that are convertible to the same relative units. The analytical data, obtained in successive runs, should fall within the indicated limits, and along the dashed line. Thus, these charts provide a convenient means of monitoring each analytical procedure in successive runs.

In routine use, the data acquired as in Figures 6 and 7 may be summarized by linear regression analysis, with calculation of the X-intercept and slope for comparison with expected values.

An alternative way of monitoring analytical performance by using interrelated specimens is illustrated in Figure 8. Because of the way in which the specimens were prepared it is possible to calculate, from each result, either the analyte concentration of the serum base pool, or the concentration due to added analyte in the supplemented pool.

For example, for each analyte the specimens are related according to the equation:

$$ Y_i = m(X_i + X_o) $$

(1)

in which $Y_i$ is a determined value of a specimen, $X_i$ is the percent of supplemented pool in the specimen, $m$ is the expected slope, and $X_o$ is the absolute value of the X-intercept. Similarly, the Y-intercept, $Y_o$ is given by the equation:

$$ Y_o = m X_o $$

(2)

and the concentration of analyte in the most concen-
treated specimen, $Y_{100}$, is given by the equation:

$$Y_{100} = m(33.33 + X_o)$$  \hspace{1cm} (3)

Equations 1 and 2 may be combined as in equation 4 and equations 1 and 3 may be combined as in equation 5:

$$Y_o = Y_i \frac{X_o}{(X_i + X_o)} = Y_i A_i$$  \hspace{1cm} (4)

$$Y_{100} = Y_i \frac{(33.33 + X_o)}{(X_i + X_o)} = Y_i B_i$$  \hspace{1cm} (5)

In these equations the factors $A_i$ and $B_i$ are constants for each analyte in each specimen. The constants are based on the initial calculation of the $X$-intercept and the known proportion of supplemented pool added in the preparation of the specimens. Equations 4 and 5 state that any analytical value, $Y_i$, may be multiplied by the appropriate factors, $A_i$ and $B_i$, to yield estimates of $Y_o$ and $Y_{100}$. $Y_o$ is the concentration of analyte in the serum base pool used in the preparation of the specimens, and the target value was determined by analyses provided with the CAP Survey Validated Reference Serum. $Y_{100}$ is the concentration of analyte in the most concentrated specimen, and this target value was determined by the addition of a known (weighed) amount of analyte to the base pool. Thus, $Y_o$ and $Y_{100}$ represent two reliable fixed target values to which all results may be related. Figure 8 shows a series of estimates of $Y_o$ and $Y_{100}$ based on random analyses, on one instrument over a three-week period, of the 12 inter-related specimens used in the present study. The charts show that for each analyte all 12 specimens yielded essentially the same
mean value, within reasonable limits of error, and that the data reflect the various analytical problems discussed earlier.

Figure 8 also illustrates a phenomenon that might be called "cycling" and is indicative of a mis-calibrated instrument. Figure 3C shows that the chloride channel was calibrated to yield inordinately high values at low concentrations of analyte, correct values at mid-range concentration of analyte, and inordinately low values at high concentrations of analyte. Consequently, when the set of control specimens were analyzed serially, the estimates of $Y_0$ and $Y_{100}$ "cycled" from high to low values. A period of cycling in chloride analyses is indicated in Figure 8 and the phenomenon is also evident in the data of some of the other analytes. Apparently, this type of mis-calibration is common when the technique of single-point calibration is used.

Discussion

From these experiments, we can draw certain conclusions:

Design and Use of Inter-related Specimens

The specimens used in this study were prepared to have linear interconstituent and inter-specimen relationships, to have analytical values in agreement with those for a nationally available, well-assayed reference serum, and to contain known amounts of added analytes. Collectively, these quantitative specimen characteristics help to maximize the useful information available from each analysis of a control specimen and help to minimize the effort required at the bench level to utilize quality-control data effectively.

The specimens were prepared by multiple admixture of several supplemented portions of a prepared serum pool (Table 1). Analyses of the specimens for the seven analytes of interest yielded the expected linear relationships reasonably well, or showed systematic biases (Figure 3, A-G, and Table 2). These biases (sodium, chloride, and carbon dioxide analyses) were shown to be due to systematic factors rather than to factors related to the design of the specimens. The use of multiple supplemented pools to prepare the specimen offers flexibility in the compositional design of the specimens, and consequently enhances the versatility of use of the specimens. The inter-constituent relationships make it possible to graph all results from a single specimen as in Figure 6, and to evaluate the result for each analyte in relation to those of all other analytes. The interspecimen relationships make it possible to graph results from successive specimens as in Figure 7, and to evaluate current performance for each analyte with respect to prior performance. Thus, when used regularly, these specimens should facilitate the comprehensive evaluation of instrument performance from analyses of each specimen.

The specimens were also designed to facilitate accuracy in analyses. Except for carbon dioxide, the analyte concentrations of the base serum from which the specimens were prepared were established by direct comparison with the College of American Pathologists' Survey Serum, whose assigned values for the analytes included in the present study are considered to be the most reliable of any serum-based calibration (reference) material available (16, 17). The assigned values for potassium, sodium, and chloride in this material agree closely with values obtained by the National Bureau of Standards with definitive analytical procedures (21). The analyte concentrations of the supplemented pools were established by the addition of weighed amounts of chemicals that were of analytical-reagent grade, or better. Because of the design of the specimens, the analyses of each specimen, for each analyte, may be compared directly to the value established with the SVRM ($Y_0$, Figure 8) and to the amount of weighed in analyte ($Y_{100}$, Figure 8). Thus, it is possible to relate each analysis to two fixed and reliable reference values and thereby closely monitor the relative accuracy of each analytical method.

Consideration of Analytical Problems

The results of our studies demonstrate that linearly related specimens provide a sensitive means for detecting a variety of analytical problems, some of which are correctable by the laboratory and others that seem to be inherent to the technique of single-point calibration as applied to continuous-flow analysis with the SMA 6/60 and are not easily corrected. The successful use of the technique of single-point calibration depends on the final instrumental read-out being directly proportional to analyte concentration. An instrument may be designed to yield a proportionate read-out, even when the reaction used in the analysis is nonlinear, by a variety of means such as mechanical or electronic linearization of the detector output (22). Another simple design ploy that was frequently used in the development of the multi-channel continuous-flow analyzer was to construct a nonlinear scale for the recorder chart paper (see Figure 5), which thereby provided an apparently linear readout of concentration even though the reaction and detector response remained distinctly nonlinear (20). To be effective in widespread use, this ploy required that the analytical method as performed in each laboratory must be exactly as nonlinear (non-ideal?) as it was in the manufacturer's laboratory at the time the printed scale was constructed. This technique was used with early models of the multi-channel continuous-flow analyzer for sodium and potassium analyses because the flame photometer clearly did not yield a linear response to these analytes (20). The flame photometer used in the newer SMA 6/60 instruments, such as those we used in the present study, are believed by the manufacturer's representatives to have an "adequately linear response," and consequently a linear scale is now used on the recorder-chart paper.3 Our studies confirm that the response to sodium and potassium is linear in the range

---

3 Personal communications, various technical service representatives of Technicon Instruments Corp., Tarrytown, N. Y.
of clinical interest, but that the response to sodium may be biased at low concentrations. This bias would result in the setting of an inappropriate baseline, which in turn would cause an inappropriate slope, as our data (Figure 3B, and Table 2) illustrate.

Results by the mercuric thiocyanate procedure for chloride analyses are known not to obey Beer's law, and this procedure also was formerly linearized by using a nonlinear printed scale on the recorder paper (20). Newer instruments, however, are supplied with a linear scale on the chart paper. The reason for this change apparently is that the manufacturer believes that modifications incorporated with the newer instruments result in "an adequately linear response within the range of 70–130 mmol/L chloride". Our results show that neither of the instruments used in our laboratory yielded a linear read-out in this concentration range. The principal difference between our method and the manufacturer's recommended method is our use of the original aqueous mercuric thiocyanate reagent, rather than its methanolic counterpart. While this difference might account for our results, we believe the problem is more fundamental because (a) our specimens have been analyzed with SMA 6/60 instruments in several other laboratories and reveal the same kind of bias that we observe and (b) the problem was alleviated by using well-assayed protein-based calibration materials to standardize the instrument at 70 and 130 mmol/liter, rather than the recommended aqueous standards (Figure 4).

Thus, the ability of a laboratory to attain accurate analyses is limited in part by the design of the instrumental system, important details of which are not always readily available to the laboratory. The linearly related specimens that we have described provide one means by which a laboratory may evaluate instrumental performance and identify specific problem areas.

We thank Dr. John A. Lott for valuable advice during the course of these studies, and Joan Mercier, MT (ASCP) and Mary Kyler, MT (ASCP) for their patient understanding and expert technical assistance. This work was supported by NIH Training Grant No. GM01805. Portions of this work were submitted by T. E. Caragher to the Graduate School of Ohio State University in partial fulfillment of the requirements for the Ph.D. degree in pathology (major: clinical chemistry).

Appendix: Use of the Correlation Coefficient and Linear Regression Parameters to Evaluate the Linearity of Analyses

One of the reviewers questioned our use of the correlation coefficient, r, to evaluate the linearity of data because, as he stated, "the value of r is strictly dependent on n (the number of observations)," and "with only six data points, the value of r is probably not significant." Although these particular statements are not true (r can have many values, even when n is held constant, and laboratory analysts daily obtain highly significant correlations—their standard curves—with even fewer than six data points), it is true that considerable confusion exists concerning the proper use of the correlation coefficient. Interested readers may find scholarly discussion of the use (and misuse) of the correlation coefficient in references 23 and 24. Our use of the correlation coefficient is based on practical, rather than theoretical, considerations: the correlation coefficient is readily obtained with linear regression statistics and, when used as we have described, has proven to be a useful indicator of the linearity of data. In this appendix we present several examples that illustrate the uses and limitations of the correlation coefficient.

Linear regression analysis is commonly used to determine the degree of correlation of two parameters. The analysis results in the calculation of the Y intercept and slope of the best line fit to the data, and of the correlation coefficient, r. The correlation coefficient is calculated from the equation:

\[ r = \frac{n \Sigma XY - \Sigma X \Sigma Y}{\sqrt{(n \Sigma X^2 - (\Sigma X)^2)(n \Sigma Y^2 - (\Sigma Y)^2)}} \]  

in which n is the number of paired values of X and Y. The value of r may range from -1.0 to +1.0. A value of 0.0 indicates that no linear correlation exists, and a value of ±1.0 indicates a perfect linear correlation, i.e., that all values fall exactly along a single straight line (23–25).

There are two major classes of problems for which the calculation of the correlation coefficient may be useful. In one application, the relationship of X and Y may be
unknown or problematical, and the correlation coefficient is used to estimate the likelihood that the two parameters are related. The example given in Figures 9A, B, and C show how, in this kind of application, the correlation coefficient is influenced by the number and distribution of the data. In each figure, the solid line is the line of direct proportionality, and the dashed line is the best line calculated by linear regression. The calculated value of \( r \) is given in each figure. Figure 9A shows a cluster of 25 data points, and the value of \( r \) is low. In Figures 9B and C, two and four data points, respectively, were removed from the central cluster and were reassigned values near the line of proportionality. In spite of the visual impression of a clear correlation in Figure 9C, the value of \( r \) is only 0.704, and the baseline fit does not correspond to the line of direct proportionality. These figures illustrate how the weight of a large central cluster of disperse data may result in a low value of \( r \) and invalid estimates of the Y intercept and slope, even though the data as a whole are linearly distributed along the line of perfect correlation. Thus, in this kind of application the correlation coefficient is strongly influenced by the number and distribution of data points.

One might suppose, from the foregoing example, that a high value of \( r \) of, say, 0.98, would indicate that the data were linearly distributed. Figures 10A, B, and C show that this is not necessarily so. Figure 10A shows two clusters of data, and \( r \) has a very high value. In Figure 10B and C, one and two points, respectively, have been removed from each cluster and reassigned values that clearly suggest a curvilinear relationship between X and Y. However, the value of \( r \) is only slightly decreased by this redistribution of the data. The high value of \( r \) in Figure 10A is due to the fact that two points (represented by the two clusters) determine a single straight line and the correlation of the data is consequently high. In Figures 10B and C it is the weight of the two clusters that prevent large changes in \( r \) as the data on the whole deviate from a linear relationship. Figures 9 and 10 show that the correlation coefficient can be a poor indicator of the linearity of data, and illustrate why, in this kind of application, the value of \( r \) is used only to suggest the possibility that a correlation may exist. In these examples \( r \) must be interpreted cautiously and with some appreciation of how the distribution of data influences \( r \).

In another kind of application, it may be known that \( X \) and \( Y \) are ordinarily linearly related and very highly correlated. This is the kind of application described in
Fig. 12. Illustration of how the dispersion of data by ±5% (Figure 12A), ±7.5% (Figure 12B), or ±10% (Figure 12C) affects the value of r. Note that the best line fit to the data (dashed line) differs little from the line of direct proportionality.

Fig. 13. Illustration of how curvilinear data affect the value of r. Note that curvilinear data cause the best-line fit of the data to deviate noticeably from the line of direct proportionality.

this and other publications on the use of interrelated specimens in clinical chemistry (10–12). Suppose that six specimens are known to be linearly related in terms of composition, and to have analyte concentrations distributed over the major portion of the analytical range. We wish to know whether a laboratory’s analyses of these specimens are also linearly correlated with specimen composition. There are only three ways in which such analyses can deviate from linearity: (a) the analytical data may include a single “outlier”, (b) the analytical data may be generally disperse (i.e., more than one “outlier” may be present), or (c) the analytical data may be curvilinear. These three kinds of deviations from linearity are illustrated in Figures 11, 12, and 13, respectively. Figure 11A shows the correlation coefficients that result as a single high analytical value is removed progressively from the line of perfect correlation. Figures 11B and C show the changes that occur in r as intermediate and low values, respectively, are removed progressively from the ideal line. Figures 12A, B, and C show the correlation coefficients that result when the analytical data are dispersed by ±5, 7.5, and 10%, respectively, about the line of proportionality. Figures 13A, B, and C show the values of r that result from various kinds of curvilinear data.

From inspection of Figures 11, 12, and 13 we conclude that in this particular kind of application, in which n is constant, the data are distributed over the range of interest, and a value of r approaching 1.000 is expected, a correlation coefficient of <0.98 should be clearly indicative that the analytical data deviate significantly from linearity, due either to an isolated outlier (Figure 11), generally disperse data (Figure 12), or curvilinear data (Figure 13).

Our choice of 0.98 as the discriminatory value of r is obviously arbitrary; others might select a lower or higher value. In any case, the value of r must be interpreted judiciously, with due regard for deviations of the intercept and slope of the best-line fit of the data from the expected or theoretical values.

References
