CLIN. CHEM. 27/3, 493-501 (1981)

proposed selected method

A Multi-Rule Shewhart Chart for Quality Control in Clinical Chemistry

Submitters: James O. Westgard, Patricia L. Barry, and Marian R. Hunt, *Depts. of Medicine, Pathology, and Clinical Laboratories, University of Wisconsin, Center for Health Sciences, Madison, WI 53792*
Torgny Groth, *Group for Biomedical Informatics, Uppsala University Data Center, S-750 02 Uppsala, Sweden*
Reviewers: Robert W. Burnett, *Clinical Chemistry Laboratory, Hartford Hospital, Hartford, CT 06115*
Adrian Hainline, Jr., *Clinical Chemistry Standardization Section, Metabolic Biochemistry Branch, Clinical Chemistry Division, Bureau of Laboratories, Center for Disease Control, Atlanta, GA 30333*
Ralph E. Thiers, *Consultant, 8710 Salty Drive, N.W., Olympia, WA 98502*
Assigned Editor: Henry Nipper, *Veterans Administration Hospital, Baltimore, MD 21218*

Introduction

A statistical control procedure is an important element in a total system of quality control. The purpose of this discussion is to outline a simple statistical procedure that is widely applicable and practical in clinical laboratories. The control procedure recommended here is for applications where stable control materials are available and are analyzed repeatedly over long periods of time. This kind of control procedure was initially described by Shewhart (*1*) and later introduced in clinical chemistry by Levey and Jennings (*2*). Control data are displayed on control charts, which are sometimes referred to as "Shewhart charts" and other times as "Levey–Jennings charts."

Control charts of this kind are now in use in most clinical laboratories. The applications from laboratory to laboratory differ primarily with (*a*) the use of single measurements or replicate measurements and (*b*) the criteria used in deciding whether the data indicate the analytical run is in or out of control. In our experience, most laboratories use control procedures based on single measurements rather than replicate measurements; thus the "single-value" type of control chart is more frequently used than charts for the mean and range of replicate measurements. Because of this, we focus on single-value charts and decision criteria appropriate for such charts. The selection of these criteria is based on some studies of their statistical properties (*3–5*), with attention being given to the interpretation of the few control observations occurring in individual analytical runs, rather than the interpretation of monthly control charts containing 20 or more observations. Thus, we focus on the immediate decisions leading to data reporting, rather than the retrospective review of large amounts of charted data.

In recommending the multi-rule Shewhart procedure, the objectives have been to provide (*a*) simple data analysis and display via control charts, such that computerized data handling is not necessary; (*b*) easy adaptation and integration into the existing control practices in clinical laboratories; (*c*) a low level of false rejections or false alarms; (*d*) an improved capability for detecting analytical errors; and (*e*) some indication of the type of analytical error occurring when a run is rejected, to aid in problem solving.

Principles

The analytical method to be controlled is first studied, to characterize its analytical performance. Measurements are made on control materials, which are assumed to be stable and to vary little in concentration from aliquot to aliquot, or vial to vial. Repeated measurements, therefore, characterize the imprecision or random errors of the analytical method. It is assumed that the distribution of these errors is gaussian and can be described by its mean ($\bar{x}$) and standard deviation (s).[1] These statistics are calculated from a replication study, generally over a 20-day period, with one measurement on each control material per analytical run and one analytical run per day.

A control chart is prepared for each control material. The chart displays concentration on the y-axis vs time on the x-axis. Horizontal lines are drawn for the mean, and for upper and lower control limits, which are calculated from the standard deviation. Several sets of control limits are included on the control chart recommended here, to permit the use of several different decision criteria or control rules.

We use the term "control rule" to indicate a criterion for judging whether the observed control measurements (or observations) represent typical or atypical (stable or unstable) performance of the analytical method. Many different control rules could be used, but they all attempt to signal when the control measurements no longer represent the expected or previously observed error distribution. Simultaneous use of several control rules—i.e., a combination of control rules—can improve the performance of a control procedure. Individual rules have different capabilities for detecting different types of analytical errors. At least two control rules need to be selected, one that detects random analytical error and another that detects systematic analytical error. When the control procedure signals that an analytical run should be rejected, the particular control rule providing the signal gives some indication of the type of analytical error that is occurring. This in turn may suggest certain sources of the error, and so aid in problem solving. In short, the rule violated gives some indi-

[1] Abbreviations used: $\bar{x}$, mean; s, standard deviation; N, number of control observations per analytical run; n, total number of control observations collected in a given time period, over many analytical runs; 1_{3s}, 2_{2s}, R_{4s}, 1_{2s}, 4_{1s}, $10_{\bar{x}}$, see "Control rules" in *Materials and Methods* section.

cation of the type of error, which in turn provides a hint about the source of error.

Note: The terminology used here (i.e., control rules, violation of control rules) is convenient for discussion purposes, both in written text and in oral presentations. However, as reviewer A.H. points out, this may cause analysts to have the feeling of doing something wrong when an out-of-control situation arises. In that sense, the choice of terminology is unfortunate. It would be more objective to use the term "decision criterion," but it is not easy to find a convenient term to use for indicating when the criterion is not met. Furthermore, the single vs plural (criterion, criteria) is difficult in oral presentations. These rules or criteria are, of course, statistical tests, but to talk of tests would be even more confusing.

Control rules should be chosen to provide a low probability for false rejection and a high probability for error detection. "False rejection" refers to the situation where the analytical process or analytical method remains stable, but a rejection signal still occurs, owing to background random error or the inherent imprecision of the analytical method. All control procedures provide some false rejections, but appropriate choice of control rules can keep the proportion low (<5%). "Error detection" refers to the situation where the analytical method has been disturbed. There are analytical errors in addition to the inherent imprecision of the analytical method. A shift or drift may have occurred, causing a systematic analytic error. The standard deviation may have increased, causing an increase in the random analytical error of the method. When such additional errors are present, the control procedure is supposed to detect them and provide a rejection signal.

In selecting control rules, it is important to first consider the probability for false rejection and to eliminate those rules where the level exceeds a probability figure of 0.05 or a percentage of 5% (*3, 5*). Then, from the remaining rules, at least one is selected that is responsive to systematic error and at least one that is responsive to increases in random error. The number of control observations per run (N) should be chosen to provide the desired probability for error detection.

In the control procedure recommended here, several control rules are used, hence the name "multi-rule" Shewhart procedure. In the daily operation of the control procedure, samples of control materials are included in each analytical run. When any one of the control rules is violated, a decision is made to reject that analytical run. A decision to accept the analytical run requires that there be no violations of any of the control rules.

Materials and Methods

Control Materials

It is not our purpose to describe the characteristics of control materials in detail. Bowers et al. (*6*) have discussed control materials in a previous volume of *Selected Methods*.

Suitable materials are generally available and in use in most laboratories, although each material may have some limitations for certain analytes. It may therefore be necessary to select control materials appropriate for different analytes, rather than use the same control materials for all methods.

In general, the most important properties are that the control materials behave like the real samples, are available in sufficient quantity for a year or so of use, are stable over the time period of use, are appropriately apportioned for convenient use, and vary little in concentration from aliquot to aliquot or vial to vial. For the control system here, two control materials having different concentrations are recommended, with one measurement being made at each concentration during each analytical run. The concentrations may be chosen to represent normal values, appropriate medical decision concentrations, or critical instrument performance limits (such as upper or lower limits of linearity).

Control Rules (Decision Criteria)

For brevity and convenience, symbols are used to represent the different control rules. The symbol has the form A_L, where A is an abbreviation for a statistic or is the number of control observations per run, and L is the control limit.

1_{2s} represents the control rule where one control observation exceeds control limits set as $\bar{x} \pm 2s$. This is the "warning" rule for a Shewhart chart and is interpreted in this discussion as a requirement for additional inspection of the control data, testing the data with the rules below to judge whether the analytical run should be accepted or rejected.

1_{3s} symbolizes the control rule where a run is rejected when one control observation exceeds control limits set as $\bar{x} \pm 3s$. These are the usual "action" or rejection limits on a Shewhart control chart.

2_{2s} is the control rule where the run is rejected when two consecutive control observations exceed the same limit, which is either $\bar{x} + 2s$ or $\bar{x} - 2s$. The rule is initially applied to the two observations within a run, one on each of two different control materials. The run is rejected when the control observations on both materials exceed their respective $+2s$ control limits or their respective $-2s$ control limits. The rule can also be applied to two consecutive observations on the same control material, one from each of two consecutive runs. When applied to consecutive observations on different materials, this will be referred to as "across" materials, to differentiate this from consecutive observations on the same material, or "within" materials.

R_{4s} is the control rule according to which the run is rejected when the range or difference between the two control observations within the run exceeds $4s$. The rule is invoked when the observation in one control material exceeds a $+2s$ limit and the observation on the other exceeds a $-2s$ limit, i.e., each observation is out by $2s$, but in opposite directions, making a total of $4s$ difference between them.

Note: Reviewer R.B. points out that this range rule could be applied when one control observation exceeds, say, $+2.5s$ and the other, say, $-1.5s$. This would be perfectly correct, though it would not be very convenient or practical without computerized data handling. There is no difficulty once an observation exceeds a $3s$ limit, because then it is out-of-control anyway. So the ambiguity in interpretation occurs when an observation is between $2s$ and $3s$. The analyst should decide how to handle this, based on what is practical in his laboratory.

4_{1s} is the control rule where the run is rejected when four consecutive control observations exceed the same limit, which is either $\bar{x} + 1s$ or $\bar{x} - 1s$. These consecutive observations can occur within one control material, which would require inspecting the observations for four consecutive runs, or across control materials, which would require inspecting only the present run and the one before it.

$10_{\bar{x}}$ is the control rule which says the run is rejected when 10 consecutive control observations fall on the same side of the mean ($\bar{x}$). These consecutive observations can occur within one control material or across control materials. This would require inspection of 10 or five consecutive runs, respectively.

A practical way of using this combination of control rules in a manual application is shown in Figure 1. The 1_{2s} rule is used as a warning rule and prompts a more detailed inspection of the data using the other control rules. If neither control observation exceeds a $2s$ limit, the analytical run is in-control and patients' data may be reported. If either observation exceeds a $2s$ limit, the control data are tested by applying the 1_{3s}, 2_{2s}, R_{4s}, 4_{1s}, and $10_{\bar{x}}$ rules. If none of these rules is violated, the run is in-control. If any one of them is violated, the

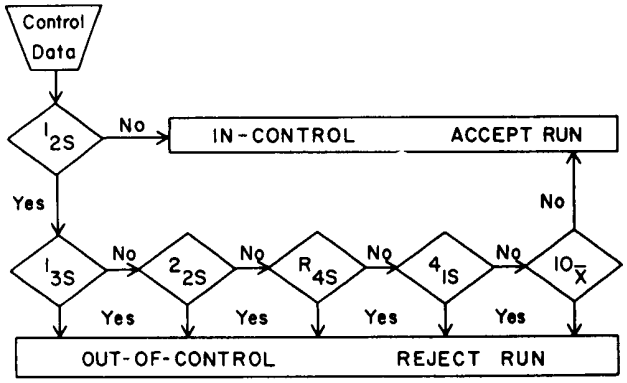

Fig. 1. Logic diagram for applying a series of decision criteria (control rules) in the multi-rule Shewhart procedure

run is out-of-control. The particular rule violated may give some indication of the type of analytical error occurring. Random error will most often be detected by the 1_{3s} and R_{4s} rules. Systematic error will usually be detected by the 2_{2s}, 4_{1s}, or $10_{\bar{x}}$ rules and, when very large, by the 1_{3s} rule.

Calculation of Control Limits

The mean ($\bar{x}$) and standard deviation (s) are calculated from the following equations:

$$\bar{x} = \Sigma x_i/n \quad (1)$$

$$s = \sqrt{\frac{n\Sigma x_i^2 - (\Sigma x_i)^2}{n(n-1)}} \quad (2)$$

where x_i is an individual control observation and n is the total number of control observations collected in the time period being analyzed. Initial estimates are often made from a data set where n is approximately 20. When n is this low, these estimates may not be reliable. They should be revised when more control observations are accumulated. This can be done by analyzing additional data sets and recording n, Σx_i, and Σx_i^2. The cumulative totals for these terms can be obtained by adding the values for the different data sets. Then these totals can be used in equations 1 and 2 to give cumulative estimates of $\bar{x}$ and s.

Control limits are calculated from $\bar{x}$ and s as follows:

$$3s \text{ control limits} = \bar{x} \pm 3s \quad (3)$$

$$2s \text{ control limits} = \bar{x} \pm 2s \quad (4)$$

$$1s \text{ control limits} = \bar{x} \pm 1s \quad (5)$$

The calculation of control limits is illustrated by the data in Tables 1–3. Table 1 shows control data collected during five months, 20 observations per month. Table 2 shows the calculated means and standard deviations, both for the individual monthly data sets and for the accumulated data. For example, for month 2, the first line gives n, Σx_i, and Σx_i^2, and the corresponding mean and standard deviation calculated from those values. The next line gives the cumulative values and the corresponding mean and standard deviation. The cumulative values for n, Σx_i, and Σx_i^2 are obtained by adding the values in the previous two lines. Observe that the standard deviation changes more from month to month for the individual monthly data sets than for the cumulative data. For these simulated data, the true mean was specified to be 100 and the true standard deviation 4.00. Note that the accuracy of the estimates improves as the cumulative number of observations increases. This shows that the control limits that are calculated from the cumulative values will be more reliable than those calculated from the individual monthly data sets. In Table 3 the calculated control limits are compared.

Preparation of Control Charts

The y-axis should be scaled to provide a concentration

Table 1. Example Control Observations for One Control Material during Five One-Month Periods

	Month				
Day	1	2	3	4	5
1	98	100	97	101	100
2	97	109	98	100	96
3	95	102	102	99	101
4	103	104	92	100	102
5	100	97	104	96	104
6	104	105	100	100	100
7	92	98	95	98	96
8	94	100	100	97	101
9	102	96	104	103	99
10	95	103	101	107	105
11	100	97	101	104	100
12	93	97	99	96	95
13	100	96	97	104	101
14	106	97	112	105	99
15	112	104	92	101	90
16	94	99	105	102	98
17	96	105	105	102	106
18	97	94	101	102	100
19	103	95	95	101	101
20	104	97	100	104	97

Table 2. Monthly and Cumulative Means and Standard Deviations Calculated from Control Data in Table 1

	Monthly (and cumulative) totals			Calculated statistics	
Month	n	x_i	x_i^2	$\bar{x}$	s
1	20	1985	197 507	99.25	5.11
2	20	1995	199 319	99.75	4.09
	(40)	(3980)	(396 825)	(99.50)	(4.46)
3	20	2000	200 434	100.00	4.78
	(60)	(5980)	(597 259)	(99.67)	(4.61)
4	20	2022	204 592	101.10	2.97
	(80)	(9002)	(801 851)	(100.00)	(4.29)
5	20	1991	198 457	99.55	3.65
	(100)	(9993)	(1 000 308)	(99.93)	(4.15)

Table 3. Control Limits Calculated for the Control Data in Table 1, with Use of the Means and Standard Deviations from Table 2

	Monthly (and cumulative) control limits		
Month	$\bar{x} \pm 1s$	$\bar{x} \pm 2s$	$\bar{x} \pm 3s$
1	94.1–104.4	89.0–109.5	83.9–114.6
2	94.7–103.8	91.6–107.9	87.5–112.0
	(95.0–104.0)	(90.6–108.4)	(86.1–112.9)
3	95.2–104.8	90.4–109.6	85.7–114.3
	(95.0–104.3)	(90.4–108.9)	(85.8–113.5)
4	98.1–104.1	95.2–107.0	92.2–110.0
	(95.7–104.3)	(91.4–108.6)	(87.1–112.9)
5	95.9–103.2	92.3–106.8	88.6–110.5
	(95.8–104.1)	(91.6–108.2)	(87.5–112.4)

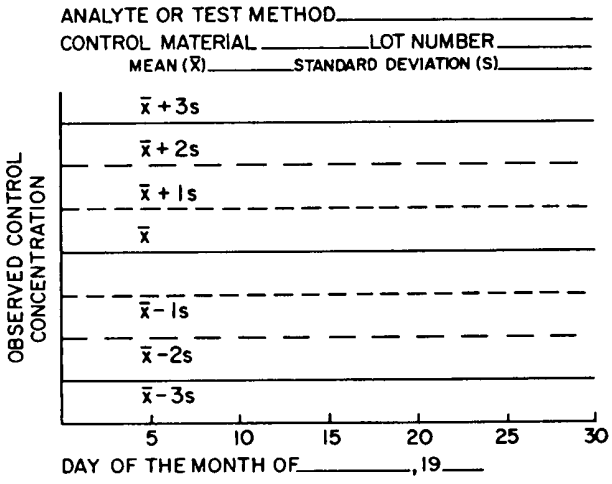

Fig. 2. Control chart for use with the multi-rule Shewhart procedure

Ordinate: concentration; abscissa: time. Control limits are drawn at $\bar{x} \pm 3s$, $\bar{x} \pm 2s$, and $\bar{x} \pm 1s$

range from $\bar{x} - 4s$ to $\bar{x} + 4s$. The x-axis should be scaled to provide the time period of interest, usually one month. Horizontal lines should be drawn corresponding to $\bar{x} + 3s$, $\bar{x} + 2s$, $\bar{x} + 1s$, $\bar{x}$, $\bar{x} - 1s$, $\bar{x} - 2s$, and $\bar{x} - 3s$. See Figure 2 for an example control chart (note that these control limits could be color coded for easier use—for example, green for $\bar{x}$, blue for $\pm 1s$, orange for $\pm 2s$, and red for $\pm 3s$). For the control procedure recommended here, it is convenient to prepare the two control charts on a single page (see Figure 2, for example).

Control Procedure

1. Analyze samples of two different control materials. Make one measurement on each control material each time when testing for statistical control.[2] Record those observations and plot them on the control charts.
2. Test the control data, using the 1_{2s} rule. *Accept* the run when both control observations are within $\bar{x} \pm 2s$ limits. Report patients' results. When at least one control observation exceeds the $\bar{x} \pm 2s$ limits, hold the patients' results and inspect the control data further, using additional control rules.
3. Inspect control data within the run.
   (*a*) Test with the 1_{3s} rule. *Reject* the run when one control observation exceeds $\bar{x} \pm 3s$ limits. Do not report patients' results.
   (*b*) Test with the 2_{2s} rule across control materials. *Reject* the run when both control observations exceed the same $\bar{x} + 2s$ or $\bar{x} - 2s$ control limit. Do not report patients' results.
   (*c*) Test with the R_{4s} rule, within the run, across control materials. *Reject* the run when one control observation exceeds a $\bar{x} + 2s$ limit and the other exceeds a $\bar{x} - 2s$ limit. Do not report patients' results.
4. Inspect control data across runs.
   (*a*) Test with the 2_{2s} rule within the control materials. *Reject* when the previous observation on the same control material exceeded the same $\bar{x} + 2s$ or $\bar{x} - 2s$ control limit. Do not report patients' results.
   (*b*) Test with the 4_{1s} rule across control materials. *Reject* when the last four consecutive control observations exceed the same $\bar{x} + 1s$ or $\bar{x} - 1s$ limit. Do not report patients' results.
   (*c*) Test with the 4_{1s} rule within control materials. *Reject* when the previous three control observations on the same control material exceeded the same $\bar{x} + 1s$ or $\bar{x} - 1s$ control limit. Do not report patients' results.
   (*d*) Test with the $10_{\bar{x}}$ rule across control materials. *Reject* when the last 10 consecutive observations fall on the same side of $\bar{x}$. Do not report patients' results.
   (*e*) Test with the $10_{\bar{x}}$ rule within control materials. *Reject* when nine previous observations on the same control material fall on the same side of $\bar{x}$. Do not report patients' results.
5. *Accept* the run when none of the rules indicates a lack of statistical control. Report patients' results.
6. When the analytical method is out-of-control:
   (*a*) Determine the type of errors occurring (random, systematic, or both) based on the control rules being violated. Note that when either the 1_{3s} or R_{4s} control rule is violated, it is more likely random error than systematic error. When systematic error is present, it is more likely to be detected by the 2_{2s}, 4_{1s}, or $10_{\bar{x}}$ rules. A review of control data on both control materials (across materials) will help detect errors that occur throughout the concentration range tested by those controls. A review of control data on a single control material (within control materials) will help detect errors in a particular concentration range.
   (*b*) Refer to a troubleshooting guide to inspect the components of the method or instrument that contribute to the type of error observed.
   (*c*) Correct the problem, then re-analyze the patients' samples and control samples, testing for statistical control by the same procedure. In assessing control of the new run, do not include the control data from the previously rejected run.
   (*d*) Consult a supervisor for any decision to report data when there is a lack of statistical control (i.e., when any of the control rules here give a rejection signal).
7. The supervisor may make a decision to report data when there is a lack of statistical control in the following situations:
   (*a*) The control problem can be shown to be due to the control materials themselves.
   (*b*) The control problem can be shown to have resulted from an isolated event that would not have affected the rest of the run (e.g., an interchange of two samples or a clerical transcription error).
   (*c*) The control problem occurs in a concentration range that is different from the concentrations of the patients' samples. The method is in-control in the range of the patients' samples.
   (*d*) The size of the analytical error is judged to be small relative to the medical usefulness requirements.[3]

Note: Reviewer R.B. noted that this list (in 7 above) should not

[2] These two samples can be analyzed each day, each shift, or each analytical run, whatever is most appropriate for the analytical method and its application. Rigorous definition of the locations, sequences, intervals, or times depends on the particular analytical methods and the laboratory application. It may sometimes be appropriate to assign the control samples randomly to positions in a run and other times to place them in specific locations that bracket the patients' samples. In some situations it may also be justifiable to analyze control samples before analyzing patients' samples, to establish that the analytical method is in a state of statistical control and can be used for patient testing.

[3] The quality of laboratory service is related not only to the size of analytical errors, but also to other factors such as the time required to obtain the result. There may be situations that require a relative judgment of the importance of the various factors involved in quality, thus it may be necessary to overrule the statistical control system. This is a professional judgment requiring knowledge of medical usefulness limits of error, an understanding of the use and interpretation of the analytical results, and experience.

be considered inclusive. There may be other situations in which it is appropriate to report the data. When encountered, these should be described and added to the list.

Discussion

Examples of Interpretations

Figure 3 shows some control data that could have been obtained when applying the control procedure recommended here. The top control chart is for a high concentration control material and the bottom chart is for a low concentration material. The observations charted are structured to illustrate how the control rules should be interpreted in many different situations.

Day 5: The control observation on the high material is within 2*s* limits. The control observation on the low concentration material exceeds the −3*s* control limit. The analytical run should be rejected. There is likely to be a random error occurring.

Day 6: The control observation on the high material exceeds the +2*s* limit, but the observation on the low material is within 2*s* limits. There is a warning of possible problems. Inspection of the control data using the 2_{2s}, 4_{1s}, R_{4s}, and $10_{\bar{x}}$ control rules does not confirm a problem. The run should be accepted.

Day 8: The control observations on both materials exceed their respective +2*s* control limits, thus the run should be rejected according to the 2_{2s} control rule (applied across control materials). There is likely to be a systematic error occurring throughout the concentration range covered by these control materials.

Day 11: The control observations on both materials exceed 2*s* control limits, but in opposite directions. The run should be rejected by the R_{4s} rule. There is likely to be random error occurring.

Day 13: The observation on the high-concentration material exceeds the −2*s* control limit. This is a warning of possible problems. Inspection of the control data with use of the 2_{2s}, 4_{1s}, R_{4s}, and $10_{\bar{x}}$ rules does not confirm a problem. The run should be accepted.

Day 14: The observation on the high-concentration material again exceeds the −2*s* control limit. The run should be rejected according to the 2_{2s} control rule (applied within one control material). There is likely to be a systematic error occurring in the high concentration range.

Day 17: The observation on the low-concentration control material exceeds its +2*s* control limit. The warning of a potential problem is confirmed by application of the 4_{1s} rule across control materials. The last two observations on each material exceed their respective +1*s* control limits, giving a total of four consecutive observations exceeding the +1*s* limit. The run should be rejected. There is likely to be a systematic error occurring throughout the concentration range covered by the controls.

Day 25: The observation on the low control material exceeds the −2*s* control limit. Inspection by the other control rules does not provide grounds for rejection. The run is accepted.

Day 27: The observation on the low control material exceeds the −2*s* control limit. Inspection reveals that the last 10 observations on that material have fallen below the mean. The run is rejected according to the $10_{\bar{x}}$ control rule. There is likely to be a systematic error occurring in the low concentration range.

Day 29: The observation on the high control material exceeds its +3*s* control limit, and the observation on the low material exceeds its +2s control limit. The run can be rejected by applying either the 1_{3s} or 2_{2s} control rule. There is likely to be a systematic error occurring throughout the concentration range covered by the control materials because both materials are exceeding their respective +2*s* control limits.

Resolving Control Problems

When the control system gives a rejection signal, a problem-solving procedure should be initiated. Often the first response by the analyst seems to be to prepare and analyze new samples of the control materials. This may not be the most productive response for the control procedure here because (*a*) the level of false rejections has been kept low by the choice of control rules, and (*b*) the difficulties with control materials themselves should have been decreased by including two different concentrations of the analyte each time the analytical method is tested for statistical control. Investigation of the analytical method itself may be a more productive response.

As a starting point when investigating the analytical method, the particular control rule violated may give an indication of the type of error that is occurring. Violation of the 2_{2s}, 4_{1s}, or $10_{\bar{x}}$ control rules suggests a systematic error, whereas violation of the 1_{3s} and R_{4s} control rules suggests a random error. Interpretation of the 1_{3s} rule can be somewhat more difficult, because it will also respond to a large systematic error. A review of the other control observation will be helpful in assessing whether the 2_{2s} rule is also being violated, in which case it is likely that there is a systematic error occurring.

The control rule violated suggests the type of analytical error that is occurring, which in turn may suggest possible cause for that problem. For example, violation of the 2_{2s} control rule, such as occurs in Figure 3 on day 8, suggests a systematic error. When the violation occurs on the two different concentrations of control material within a single run, it is unlikely to be a problem with the control materials. It is more likely to be a problem with the standards, instrument calibration, reagent blanks, or similar factors that will affect all measurements in the same direction.

When a random error occurs such as suggested by violation

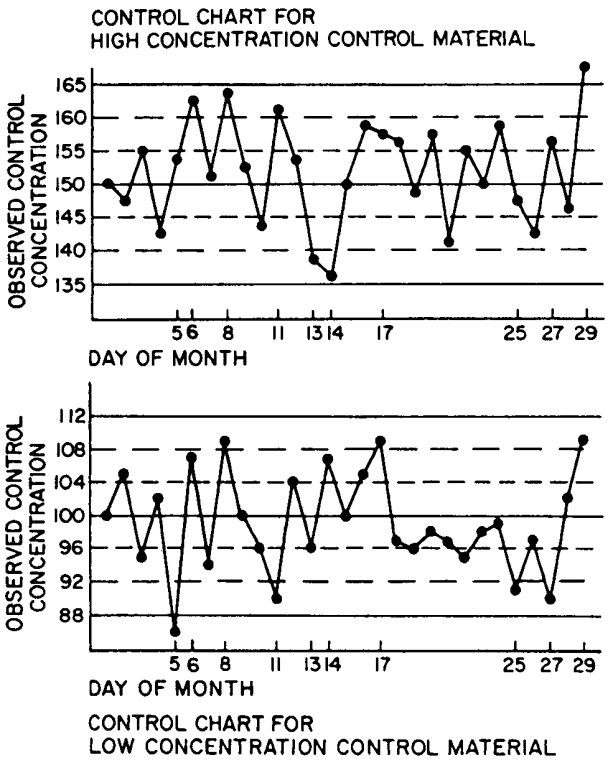

Figure 3. Example application of the multi-rule Shewhart procedure

Control charts are shown for both a high concentration control material (*top chart*) and a low concentration control material (*bottom chart*). See text for interpretation

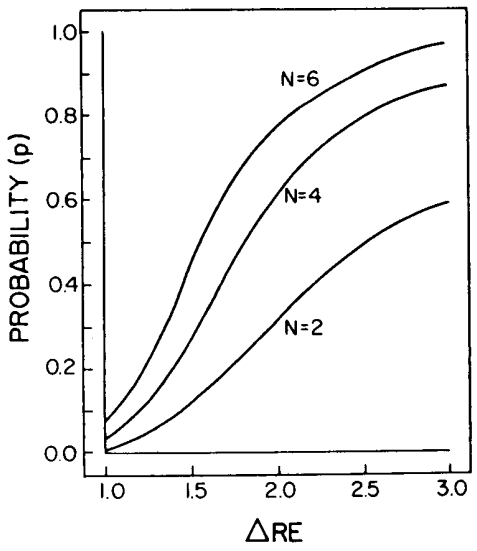

Fig. 4. Power of the $1_{3s}/2_{2s}/R_{4s}/4_{1s}/10_{\bar{x}}$ multi-rule Shewhart procedure for detecting increases in random error

The probability for rejection (p) is plotted on the ordinate vs size of the random error (ΔRE) on the abscissa. A value of 2.0 on the x-axis refers to a doubling of the standard deviation of the analytical method

of the R_{4s} rule on day 11 in Figure 3, this is usually related to different causes, such as the instability of reagents or measurement conditions, or variability in timing, pipetting, individual technique, or other similar factors. Definition of the possible sources of errors depends on the particular analytical method and the instrumentation used. The analyst will be assisted by the manufacturer's troubleshooting guidelines, documentation of reagent or instrument changes, documentation of previous problems, and experience.

When a control problem has been resolved, there remains a question of what should be done with the control data from that run—whether it should be included in further assessment of control status and in further data calculations. In assessing control after problem-solving procedures have been used, the objective should be to assess control of the newly corrected process. This is best done by increasing the number of control

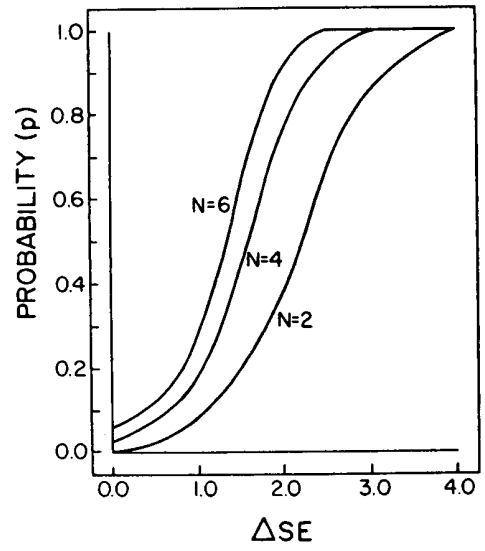

Fig. 5. Power of the $1_{3s}/2_{2s}/R_{4s}/4_{1s}/10_{\bar{x}}$ multi-rule Shewhart procedure for detecting systematic error

The probability for rejection (p) is plotted on the ordinate vs the size of the systematic error (ΔSE) on the abscissa. A value of 2.0 on the x-axis refers to a systematic shift equivalent to two times the standard deviation of the analytical method

observations in that next run, rather than utilizing any observations from a previous run. In performing calculations on control data to update the control limits, the purpose is to characterize only the *stable* performance of the analytical process. Data obtained during unstable operation should not be included.

Note: Reviewer R.B. emphasizes the importance of defining how control results from out-of-control runs should be handled. The authors' perspective is that these results do not represent the stable performance of the analytical method; therefore, if included in the summary data calculations, they would cause the standard deviation to be too large and the resulting control limits too wide. On the other hand, there is concern that elimination of these points will narrow the control limits, so that they no longer correctly characterize the tails of the error distribution. This latter problem should be minimized here because observations between 2s and 3s will be included in the final summary calculations.

Performance Characteristics

The performance characteristics of this multi-rule Shewhart procedure are summarized by the "power functions" given in Figures 4 and 5. These plots show the probability for rejecting an analytical run as a function of the size of the error occurring in the run (5). The probability for rejection (p) is plotted in the y-direction vs the size of the analytical error (ΔRE, ΔSE) in the x-direction. The point of intersection on the y-axis gives the probability for false rejection (the probability for rejecting the run when there is no analytical error except for the inherent imprecision of the analytical method). Points on the curves give the probability for error detection (the probability for rejecting the run when there is an error of the size indicated on the x-axis). In Figure 4, the size of the random error (ΔRE) is given as a factor such that a value of 2.0 means that the standard deviation of the method has doubled. In Figure 5, the size of the systematic error (ΔSE) is given in multiples of the standard deviation, such that a value of 2.0s means a systematic error equivalent to two times the standard deviation of the analytical method.

Note: Reviewer A.H. stressed the importance of understanding the concept of *inherent* random error. There is always some random error associated with a measurement process, even a stable and well-controlled process. When we discuss error detection, we actually refer to detecting error which is in addition to that inherent random error. This concept of error is important, because it helps explain why it is difficult for control procedures to detect small analytical errors. In effect, we are dealing with a signal-to-noise problem, with the inherent random error being the noise, and the additional analytical error being the signal we would like to detect.

Note: Reviewer R.T. commented that he had performed some simulations on these procedures also and that they revealed these procedures to be valid.

The different lines in Figures 4 and 5 represent the number of observations per analytical run (N). As N increases, the capability for error detection increases. However, when N = 6, the probability for false rejection exceeds 5%. This is primarily ascribable to the R_{4s} rule, and elimination of that rule will decrease the false rejections to an acceptable proportion. Based on the performance characteristics shown in Figures 4 and 5, the control procedure recommended here should be satisfactory for N from 2 to 4. If N is greater than 4, then the R_{4s} rule should be eliminated, or use of other control procedures should be considered.

Power functions should provide the basis for comparing the performance of different control procedures. However, such information is often lacking in the references and descriptions of other procedures, often because the data testing and interpretation are not sufficiently well defined. Very careful and detailed guidelines are necessary in order adequately to de-

scribe how the statistical testing is carried out. As a consequence of this careful definition, the control procedure recommended here may appear to be excessively mechanical, obviating the need for experience and skill in the interpretation of charted data. Remember that the emphasis here is on deciding the acceptability of individual analytical runs, not on the review of monthly charts of control data. Detailed guidelines for data interpretation are essential for uniform interpretation of control data by the many analysts who must make daily decisions about the acceptability of individual analytical runs. Furthermore, because that daily data analysis is carefully defined, the theoretical properties of the statistical tests (control rules) can be used to characterize the expected performance of the control procedure.

Note: Reviewer A.H. pointed out that experienced analysts are often able to make good judgments from observing the pattern of points on control charts, even though they do not use rigid rules such as those recommended here. We do not underestimate the skill in interpretation acquired through experience. The present rules are an attempt particularly to provide guidelines for data interpretation by analysts who do not have a long experience and a developed skill for making these judgments.

Comparison with Other Control Procedures

In control procedures generally recommended for application in clinical chemistry, a Shewhart type of control chart is used with either 2*s* or 3*s* control limits. Most often, 2*s* limits are recommended. Reed and Henry (*7*) illustrate control charts with both sets of limits, including labels of warning and rejection limits, respectively. However, they state that it is not necessary to have both sets of limits and that either can be used alone for rejection limits. Bermes et al. (*8*) discuss the relative merits of the 2*s* and 3*s* limits, pointing out that use of 2*s* limits frequently causes the analyst to look for problems when none exist. Although they indicate that the use of 3*s* limits will minimize this difficulty, they discourage their use because the control system will not be as sensitive for detecting analytical errors. They choose 2*s* control limits on the control charts they use to illustrate the application of statistical control in clinical chemistry.

Note: Other choices for control limits could be made. Reviewer A.H. prefers limits set to give a 1% frequency of false rejections. These limits would be the mean plus or minus 2.58 standard deviations. It is also possible to calculate limits for a selected N, such that the frequency of false rejection is fixed at 5%, 1%, or 0.2% (see reference *3*). Notice that this could also be done for rules requiring consecutive observations to exceed a specified control limit. The practical difficulty in doing so is that the control charts end up having different limits as N changes. Thus, if N = 2 for glucose and if N = 3 for serum urea nitrogen, these control charts will not have the same control limits, even though the same control rules are being used for each.

In contrast to these control procedures where one set of limits is chosen, the control procedure recommended here makes use of several sets of limits and several control rules. This use of a combination of control rules permits the response of this control system to be optimized for both a low probability for false rejection and a high probability for error detection. These improvements are achieved by a careful selection of control rules, first eliminating those rules that have too high a level of false rejections, then selecting from the remaining rules the ones most responsive for detecting different analytical errors.

Particularly critical is the use of the 1_{2s} rule as a *warning* rule, triggering the application of other rules. Although use of the 1_{2s} rule as a rejection rule is common practice, those control systems which do so will inherently have a high proportion of false rejections; for example, about 5% of the analytical runs will be rejected when N = 1, 10% when N = 2, 14% when N = 3, 18% when N = 4, and 26% when N = 6. As N increases, the level of false rejections increases. The analyst often becomes accustomed to these false alarms and usually responds by repeating the controls or the analytical run, or both, without any attempt to investigate whether any problems are occurring with the analytical method itself. The many false alarms have the effect of compromising the response to any true alarm that may occur.

Use of the 1_{2s} rule as a warning rule can decrease the false rejections, if an appropriate response to a warning signal is carefully defined. We define it here as a requirement for additional inspection of the control data, with use of additional control rules to judge whether the run is to be rejected. Patients' results should be held until this inspection is completed. When there are no additional grounds for rejecting the run, the run is judged to be in control and the patients' results are reported.

The combination of control rules recommended here is similar to that recommended by Haven (*9*), except that the R_{4s} and 4_{1s} rules have been added and the $10_{\bar{x}}$ has replaced the $7_{\bar{x}}$ rule. The R_{4s} rule is a simplified range rule adapted for the control chart recommended here and should be limited to N of 2 to 4. It would be better to determine the exact difference between the highest and lowest control values (within a run) and use control limits calculated from the within-run standard deviation instead of the total standard deviation (*4*). This becomes essential for N larger than 4. However, these complications would likely limit the use of the range rule, particularly for manual applications. The simplified range rule can be easily applied and is therefore more likely to be used.

The 4_{1s} rule has not been in common use in clinical laboratories, but has been recommended in the quality-control literature (*10*). Its probability for false rejection is low, provided the between-run standard deviation (s_b) is low. False rejections increase when the between-run standard deviation gets large, thus this rule should be limited to situations where s_b is small.

Choice of the $10_{\bar{x}}$ rule over the $7_{\bar{x}}$ rule is based on its lower probability for false rejection (*3*), but the exact number of consecutive observations is not critical as long as it is in the range 7–10. Because this type of rule will require inspection of data from two or more consecutive analytical runs, the number of observations should be chosen to be convenient: for example, for N = 2 per run, use $8_{\bar{x}}$ or $10_{\bar{x}}$ so that data from four or five runs are inspected; for N = 3, use $9_{\bar{x}}$ and three runs; for N = 4, use $8_{\bar{x}}$ and two runs, etc.

The combination of rules recommended here can also be compared to the combination of the 1_{3s} and a cumulative summation rule (*11*). The probability for detecting systematic errors is about the same, but the probability for detecting random errors may be somewhat improved owing to the addition of the R_{4s} rule. Implementation is easier because no additional data plotting is necessary.

The statistical power (probability for detecting analytical errors) can be increased by increasing the number of control observations per run (N), but consideration should also be given to the use of different control procedures as N changes. The procedure outlined here is recommended for N = 2 and could be extended for N up to 4. If N exceeded 4, the procedure should be modified by removal of the R_{4s} rule. Other control procedures would seem more appropriate for certain values of N, as summarized in Table 4. When N = 1, the only choice is between 1_{2s} and 1_{3s}. Since false rejections would be 5% for 1_{2s}, this procedure could be used, but should be restricted to only N = 1. When N = 3, a $(2 \text{ of } 3)_{2s}$ rule could be used instead of the 2_{2s} rule. When N = 4, the 4_{1s} rule will be effective. For N greater than 4, consideration should be given to mean and range procedures, such as outlined by Hainline in a forthcoming chapter of *Selected Methods* (*11*). When N

Table 4. Summary of Control Procedures Appropriate for Different Numbers of Control Observations

No. control observations	Control rules for: Individual analytical runs	Control rules for: Consecutive analytical runs
1	1_{2s}	4_{1s}
2	$1_{3s}/2_{2s}/R_{4s}$	$4_{1s}/10_{\bar{x}}$
3	$1_{3s}/(2 \text{ of } 3)_{2s}/R_{4s}$	$9_{\bar{x}}$
4	$1_{3s}/2_{2s}/R_{4s}/4_{1s}$	$8_{\bar{x}}$
4–10	Mean/range	Trend analysis (*16*)
4–20	Mean/chi-square	Trend analysis (*16*)

is very large, it would be better to substitute the chi-square test for the range test (*3*).

Note: Reviewer A.H. recommends that mean and range control procedures be considered when it is desired to obtain very tight control of an analytical method. Experiences with their use in the Lipid Reference Laboratories has been very positive. The authors' only reservation is the difficulty in applying these procedures, mainly due to the data calculations. This, of course, will not be a limitation when implemented in laboratory computer systems, microprocessor-controlled instrument systems, or micro-computers.

Mean and range, or mean and chi-square procedures, have greater statistical power when N is large and also permit the probability for false rejection to be set to a specified level. However, when N is kept low because of practical and economic reasons, the statistical power of all the procedures will be relatively low. Mandel and Nanni (*13*) prefer using the mean and range of replicates rather than treating individual observations, because the assumption of a gaussian error distribution is less tenuous. This theoretical consideration must be weighed against the practical difficulties of implementing mean and range procedures in the high-production workload of clinical laboratories. Control decisions cannot be made directly from the raw control observations, but must wait until the calculations are performed. The calculations, though not difficult, are a little more time consuming, particularly when several analytes are being measured simultaneously by multi-channel analyzers.

There also is difficulty in combining control observations when they are obtained on control materials of different concentration, such as the commonly used low-abnormal, normal, and high-abnormal materials. Unless these observations on different materials are combined, the full statistical power available from the total number of control observations will not be realized. Combining these data requires some way of normalizing the raw observations, perhaps in the manner suggested by Larsen et al. (*14*). By comparison, combining data is very simple with the control procedure recommended here. The control data are normalized by determining by how many standard deviations an observation differs from the mean for that control material. To combine results on different control materials, one needs only to count the number of observations exceeding certain limits on the control charts for the individual materials. It would, of course, be possible to combine observations that are not simply individual values or measurements, thus this approach may be useful for combining results from mean and range control procedures when two or more control materials are being analyzed.

One limitation to the application of this multi-rule Shewhart procedure may be present practices in the rounding of analytical results. Direct-readout instruments often round an analytical result to the least significant digit based on the clinical usefulness of the result. This rounding may obliterate any difference between 1*s* and 2*s* limits or 2*s* and 3*s* limits. This is sometimes the case for analytes such as albumin, urea nitrogen, CO_2, creatinine, potassium, chloride, and calcium.[4] The limitation can be overcome by obtaining an extra significant digit in the readout of results. It is important for instrument manufacturers to consider the use of the data for quality-control purposes before rounding the results for clinical significance.

Another limitation may be the application of these control rules to data from more than one analytical run (across runs). The purpose in using data from consecutive runs is to increase the number of control observations and the corresponding power for error detection. Error conditions that continue from run to run will more likely be detected by pooling the control data from the individual runs. This requires that control data from previous runs be available in a form that is convenient for inspection. If this is not practical, then some of the control rules can be eliminated. The $10_{\bar{x}}$ rule should be eliminated first, because it requires the most data and contributes the least to error detection. The 4_{1s} rule is not as difficult to apply, requiring only the previous run when applied across materials. Application within a single material is more difficult, thus the use of a rule within materials may be eliminated prior to elimination of its use across materials. In considering the use of rules across runs, it should be remembered that the R_{4s} rule is *not* intended for use across runs, only within a single run.

We think the multi-rule Shewhart control procedure is useful and practical. It has the advantages of ease of implementation and use, a low probability for false rejection, and the effective combining of results from materials of different concentrations, yielding an improved capability for error detection. Mean and range procedures, as well as cusum procedures, may be theoretically more satisfying, but they have not demonstrated their practicality in clinical laboratories. Cusum procedures, though well known, are seldom used. Mean and range procedures, as introduced in clinical laboratories by Levey and Jennings (*2*), were quickly modified by Henry and Segalove (*15*) to use with individual observations. Of the procedures that have been tried, it is the control chart for individual observations that has most influenced the practice of quality control in clinical chemistry. This must attest to the practicality of the approach. Our effort here has been to define more carefully how this approach can be successfully used.

References

1. Shewhart, W. A., *Economic Control of Quality of the Manufactured Product*, Van Nostrand, New York, NY, 1931.
2. Levey, S., and Jennings, E. R., The use of control charts in the clinical laboratories. *Am. J. Clin. Pathol.* **20,** 1059–1066 (1950).
3. Westgard, J. O., Groth, T., Aronsson, T., et al., Performance characteristics of rules for internal quality control: Probabilities for false rejection and error detection. *Clin. Chem.* **23,** 1857–1867 (1977).
4. Westgard, J. O., Falk, H., and Groth, T., Influence of a between-run component of variation, choice of control limits, and shape of error distribution on the performance characteristics of rules for internal quality control. *Clin Chem.* **25,** 394–400 (1979).
5. Westgard, J. O., and Groth, T., Power functions for statistical control rules. *Clin. Chem.* **25,** 863–869 (1979).
6. Bowers, G. N., Burnett, R. W., and McComb, R. B., Preparation and use of human serum control materials for monitoring precision in clinical chemistry. *Selected Methods Clin. Chem.* **8,** 21–27 (1977).
7. Reed, A. H., and Henry, R. J., Accuracy, precision, quality control,

[4] Cembrowski, G. S., Westgard, J. O., and Groth, T., Effects of data rounding on the performance characteristics of statistical quality control rules. *Clin. Chem.* **26,** 934 (1980). Abstract.

and miscellaneous statistics. Chap. 12 in *Clinical Chemistry, Principles and Technics*, 2nd ed., R. J. Henry, D. C. Cannon, and J. W. Wilkelman, Eds., Harper & Row, Hagerstown, MD, 1974, pp 308–316.

8. Bermes, E. W., Erviti, V., and Forman, D. T., Statistics, normal values, and quality control. In *Clinical Chemistry*, 2nd ed., N. W. Tietz, Ed., W. B. Saunders Co., Philadelphia, PA, 1976, pp 94–99.

9. Haven, G. T., Outline for quality control decisions. *Pathologist* **28,** 373–378 (1974).

10. Duncan, A. J., *Quality Control and Industrial Statistics*, 4th ed., Richard D. Irwin, Inc., Homewood, IL, 1974, p 392.

11. Hainline, Jr., A., Quality assurance. *Selected Methods Clin. Chem.* **9** (in press).

12. Westgard, J. O., Groth, T., Aronsson, T., and de Verdier, C.-H., Combined Shewhart-cusum control chart for improved quality control in clinical chemistry. *Clin. Chem.* **23,** 1881–1887 (1977).

13. Mandel, J., and Nanni, L. F., Measurement evaluation. Chap. 4 in *Quality Assurance Practices for Health Laboratories*, S. L. Inhorn, Ed., American Public Health Association, 1015 18th St., N.W., Washington, DC, p 266.

14. Larsen, L. H., Williams, R. D. B., and Nicol, G. R., Automated quality control of biomedical data, Part I: System description. *CRC Crit. Rev. Clin. Lab. Sci.* **8,** 241–257 (1977).

15. Henry, R. J., and Segalove, M., The running of standards in clinical chemistry and the use of the control chart. *J. Clin. Pathol.* **5,** 305–311 (1952).

16. Cembrowski, G. S., Westgard, J. O., Eggert, A. A., and Toren, C. E., Jr., Trend detection in control data: Optimization and interpretation of Trigg's technique for trend analysis. *Clin. Chem.* **21,** 1396–1405 (1975).