Comparison of Quality-Control Rules Used in Clinical Chemistry Laboratories

J. Bishop and A. B. J. Nix

Numerous papers have been written to show which combinations of Shewhart-type quality-control charts are optimal for detecting systematic shifts in the mean response of a process, increases in the random error of a process, and linear drift effects in the mean response across the assay batch. One paper by Westgard et al. (Clin Chem 1977;23:1857–67) especially seems to have attracted the attention of users. Here we derive detailed results that enable the characteristics of the various Shewhart-type control schemes, including the multirule scheme (Clin Chem 1981;27:493–501), to be calculated and show that a fundamental formula proposed by Westgard et al. in the earlier paper is in error, although their derived results are not seriously wrong. We also show that, from a practical point of view, a suitably chosen Cusum scheme is near optimal for all the types and combinations of errors discussed, thereby removing the selection problem for the user.

Indexing Terms: statistics · data handling

Quality-control samples are widely used in clinical chemistry laboratories to assess the quality and stability of routine analytical methods. Different laboratory managers have different views as to how many quality-control samples should be inserted into a given analytical run and where these samples should be placed. Whatever the regime in operation, it is standard practice to plot the quality control values on Shewhart-type (1) [or Levey–Jennings (2)] control charts. These charts are well understood and they give rise to a wide range of possible control charts that could be used. For example, one rule might be to declare an out-of-control situation as soon as one quality-control value is more than two standard deviations (>2σ) away from the target value. A second example might be to conclude that the process is out of control on the first occasion of seven consecutive observations falling on the same side of the target value. A third scheme might be to run both of the schemes simultaneously and conclude that the system has gone out of control when at least one of the component schemes has gone out of control.

These examples are meant to illustrate the vast potential for defining control rules of the Shewhart type. The papers by Westgard and others (3–12) consider several such schemes. The problem the user faces is, which of all of these potential rules is best and for what type of analytical error? Most authors seem to have tried to solve this problem by simulation studies, although Parvin (11) made some theoretical progress by also considering the distinction between the modeling of “intermittent” and “persistent” error. Unfortunately, most of these authors, apart from Parvin (12) in his recent paper, have not defined a clear criterion for deciding which rule is best; consequently, the user still faces the problem of selecting an efficient control scheme.

Here, we show that all of the popular Shewhart-type control rules considered in the literature can be put into a Markov Chain (13) framework, thereby avoiding simulation and its associated sampling error; with a clearly defined assessment criterion, the schemes can be ranked in order of efficiency. We also show that a suitably chosen Cusum scheme can perform at least as well as the best Shewhart-type rule considered so far. A more general conclusion about the efficiency of Cusum schemes has already been reached (14) for continuous industrial quality-control schemes.

Principles and Methods

Performance Characteristics of Control Rules

Historically, for industrial applications, the properties of quality-control rules, when monitoring the performance of continuous production processes, have been assessed by the determination of their average run length (ARL) profiles (15–21). The ARL in these applications is defined to be the average number of quality-control samples inspected before an out-of-control message is indicated. To emphasize that this approach depends on the sequential use of quality-control samples, we use the modified notation ARLQC. For these industrial applications, assessment based on the ARLQC profile makes sense because the process is continuous. The run length arising from a particular control scheme can take any value between 1 and infinity; the average is the “typical value” associated with that scheme.

Westgard et al. (4), however, point out that for clinical applications the situation is different. The clinical biochemist usually decides how many quality-control samples (N) are to be placed in a given assay batch and then decides the quality-control status of that batch after all N quality-control values have been obtained. The results from different batches may not be combined on the same control chart; instead, the control rule is reinitialized for each assay batch. For this application, the ARLQC for a given assay batch is inappropriate because it will always have a value of N. The batch characteristics that are appropriate are (a) the probability of false rejection (pR), which is the probability that an out-of-control signal will be issued for a given batch.
when the process is in control, and (b) the probability of error detection ($p_{ed}$), which is the probability that an out-of-control signal will be issued for a given batch when the process is out-of-control. In this approach, the average number of assay batches processed before an out-of-control message is given is $1/p_r$ for the in-control situation and $1/p_{ed}$ for the out-of-control situation. Westgard refers to these expressions as the ARL of the scheme for the in-control and out-of-control situations, respectively. The unit being assessed here is the batch and not an individual quality-control sample. To avoid confusion with ARL_{QC}, we refer to Westgard's ARL as ARL_{batch}.

A modification to the above protocol has been considered by Westgard (4) and crystallized by Parvin (11). Parvin defines the above procedure, which relies only on information from within a run, as being analysis of "intermittent" error. The situation where results from different assay runs are combined so that quality-control rules can apply across runs is referred to as "persistent" error. In this latter case, Parvin points out that the parameters $p_r$ and $p_{ed}$ are ill-defined and should be replaced by either the ARL concept or by plots of their cumulative probabilities of error detection, i.e., $\Sigma p(R)$, where $p(R)$ is the probability that the process first issues an out-of-control signal on the $i$th run (batch). Both of these approaches have been reviewed by Nix et al. (10). Whatever approach is adopted, it is clear that an assessment criterion is required to rank the vast array of control schemes in order of efficiency and that the schemes should be compared under the same environmental conditions (persistent or intermittent errors).

An assessment criterion for quality-control procedures in industrial applications has been put forward by Nix et al. (10) and recently applied in clinical chemistry by Parvin (12). The assessment criterion, based on ARL profiles, concludes that of two schemes, A and B, A is better than B if they both have the same out-of-control ARL and A has the greater in-control ARL. In terms of $p_{ed}$ and $p_r$, it seems reasonable to assess quality-control schemes for clinical use, which are designed for intermittent error, by arranging for them to have the same $p_r$ value and then choosing the scheme that has maximum $p_{ed}$. Although earlier papers adopted the spirit of this assessment criterion, they did not apply it rigorously and rather vague conclusions were drawn.

Control Rules Studied

It is not the purpose of this paper to do an exhaustive assessment of all Shewhart-type rules but merely to show how their properties can be easily assessed by a Markov Chain approach, thereby removing the need for the simulation exercise used by Westgard and others and consequently the effects of sampling error on the results. As a result, we consider only those Shewhart-type schemes proposed by Westgard et al. (4):

1_{ARL}: One control observation exceeds $\bar{x} \pm ks$ (the usual action limits on Shewhart control charts).

2_{ARL}: Two consecutive control observations exceed the same limit, which is either $\bar{x} + ks$ or $\bar{x} - ks$.

3_{ARL}: Three consecutive control observations exceed the same limit, which is either $\bar{x} + ks$ or $\bar{x} - ks$.

4_{ARL}: Four consecutive control observations exceed the same limit, which is either $\bar{x} + ks$ or $\bar{x} - ks$.

5_{ARL}: Seven consecutive observations fall on the same side of $\bar{x}$.

10_{ARL}: Ten consecutive observations fall on the same side of $\bar{x}$.

$R_{ARL}$: At least one observation is greater than the upper limit of $\bar{x} + ks$ and at least one observation is less than the lower limit of $\bar{x} - ks$.

Calculation of $p_r$ and $p_{ed}$ for Shewhart-Type Schemes (Intermittent Error)

We begin by highlighting an error in Westgard's formula for calculating $p_r$ for two-sided schemes. The method used by Westgard is as follows:

For a one-sided rule,

$$p_r = 1 - p_N.$$  

where $p_N$ is the probability of no rejection for a sequence of $M$ consecutive control observations in a run containing $N$ control observations. $p_N$ is calculated from the formula

$$p_N = (1 - p')^{(P - 1)} + p'^{(P - 2)} + \ldots + (p')^{M - 1}p_{N - M}$$

where $p'$ is the probability that a single control observation exceeds the limit set for the control rule and where $p_0 = p_1 = p_{M-1} = 1$. For a two-sided rule,

$$p_r = 2(1 - p_N) - (1 - p_N)^2.$$

To illustrate the error in the last formula, we consider the $2_{ARL}$ scheme. Suppose the probability that a control observation exceeds the upper limit is $\alpha$. Then, the probability that both observations exceed the upper limit is $\alpha^2$. Using Westgard's formula for the one-sided rule ($\alpha = p'$)

$$p_2 = (1 - \alpha)p_1 + \alpha(1 - \alpha)p_0 = (1 - \alpha) + \alpha(1 - \alpha) = 1 - \alpha^2$$

Hence $p_r = 1 - p_2 = \alpha^2$, in agreement with our earlier result. But consider now the two-sided scheme. Because
an out-of-control signal can be given only if both control observations lie above the upper limit or below the lower limit \( p_{tr} = 2 \alpha^2 \). Westgard’s formula gives

\[
p_{tr} = 2(1 - p_2) - (1 - p_2)^2 = 2\alpha^2 - \alpha^4
\]

Although this extra term \((-\alpha^4)\) is small in this example, its presence does illustrate that the two-sided formula proposed by Westgard is in error.

We now show how this particular Shewhart scheme can be cast in a Markov chain framework, which will apply equally well for in-and-out-of-control situations, unlike the approach of Westgard. For the \( 2_{as} \) rule we define the following four states:

- **S(0,0):** No contribution to an out-of-control signal from above the upper or below the lower limit.
- **S(1,0):** One control observation contributes to the out-of-control signal from above the upper limit.
- **S(0,1):** One control observation contributes to the out-of-control signal from below the lower limit.
- **A:** The signal is declared out of control (an absorbing state).

An example of how the system moves between these states is shown in Figure 1.

The first observation lies between the control limits and so cannot contribute to an out-of-control signal. The system is still in state S(0,0).

The second observation lies above the upper limit and so could be the first one of two that contribute to the out-of-control signal. The state of the system is thus S(1,0).

The third observation lies within the control limits, so the second observation can no longer contribute to an out-of-control signal. The system returns to state S(0,0).

The fourth observation lies below the lower limit and could be the first of two observations leading to an out-of-control signal. The system is now in state S(0,1).

The fifth observation lies above the upper limit. For this observation to be part of an out-of-control signal, it would be the first observation of two above the upper limit, and so the state is now S(1,0).

The sixth observation is also above the upper limit and so combines with the fifth observation to give an out-of-control signal. The system is now in state A.

With these states defined, the probabilities of moving between the various states in one step is summarized by the one-step transition matrix shown below:

<table>
<thead>
<tr>
<th>Initial state</th>
<th>S(0,0)</th>
<th>S(1,0)</th>
<th>S(0,1)</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>S(0,0)</td>
<td>(p_0)</td>
<td>0</td>
<td>(p_1)</td>
<td>0</td>
</tr>
<tr>
<td>S(1,0)</td>
<td>(p_0)</td>
<td>0</td>
<td>(p_1)</td>
<td>(p_1)</td>
</tr>
<tr>
<td>S(0,1)</td>
<td>(p_0)</td>
<td>(p_1)</td>
<td>0</td>
<td>(p_1)</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

where \( p_0 \) = probability that the observation falls between the control limits
\( p_1 \) = probability that the observation falls above upper control limit
\( p_1 \) = probability that the observation falls below lower control limit

\[
[P_1P_2\ldots P_N]_{14},
\]

where \( P_i \) is the one-step transition matrix when the \( i \)th quality-control observation is made.

For all other Shewhart-type control rules listed, earlier states can be defined and the transition matrices formed. [In Appendix 1 are given the one-step transition matrices for the schemes \( 1_2, 2_2, R_{2s}, 4_{2s} \), and the combined rule \( 1_3/2_2/R_{2s} \); the transition matrices for the multirules \( 1_3/2_2/R_{2s}/4_{1s} \), and \( 1_3/2_2/R_{2s}/4_{1s}/10_2/64 \times 64 \) and \( 1198 \times 1198 \), respectively, and are not shown.]

Thus the proposed method is quite general and applies equally well to one-sided schemes, two-sided schemes, multirule schemes, in-and-out-of-control situations, and situations in which the errors in the system are related to imprecision, shift, or drift.

**Calculation of Control Limits for Fixed \( P_s \) As \( N \) Varies (Intermittent Error)**

In application it is quite common for a laboratory manager to specify \( N \) and \( p_{tr} \) and then seek the positions of the control limits to meet the specification. Westgard et al. (4) present such a table (Table 1 in their paper), which is derived under various approximations. Because we are able to calculate the exact \( p_{tr} \) value for the various Shewhart schemes, it is quite easy to find the exact positions of the control limits for specified \( p_{tr} \).

In most applications, the two-sided scheme can be assumed to be symmetrical, e.g., \( p_1 = p_{tr} = p \). In this situation and for the \( 2_{as} \) scheme we need to find the value of \( p \) that satisfies the equation

\[
[P^N]_{14} = p_{tr},
\]

for specified \( N \) and \( p_{tr} \). Having found the value of \( p \), it is a simple task to invert the probability function to obtain
the values of the control limits. The results of this analysis are presented later for a variety of Shewhart control schemes.

Control Rule Characteristics for Persistent Error

As pointed out by Parvin (11), the concept of \( p_r \) and \( p_{sd} \) is ill-defined when control rules are used across runs (batches). Parvin describes several possible alternatives. For completeness we show how the Markov Chain method can be used to obtain Parvin's results. The unconditional probability that an error has persisted and gone undetected for \((i - 1)\) runs \((i - 1)\) runs have been accepted) and is detected on the \( i \)th run \((i \)th run is rejected) is denoted by \( p(R_i) \). The conditional probability that the error is detected on the \( i \)th run, given that the \((i - 1)\) previous runs with error have been accepted, is denoted by \( p(R_i|A_{i-1}) \). The cumulative probability that the error has been detected on or before the \( i \)th run is denoted by \( p(\leq R_i) \). For convenience, suppose there are \( m \) quality-control results per run; then, by the end of the \( i \)th run, \( mi \) quality-control results have been obtained. The cumulative probability \( p(\leq R_i) \) is the probability that the system has reached the absorbing state (out of control) by the \( mi \)th quality-control result. For the \( 2s_m \) rule, this is just \([P^{mi}]_{i,4}\); that is,

\[
p(\leq R_i) = [P^{mi}]_{i,4}. \]

Therefore, the probability that the system first goes out of control on the \( i \)th run is

\[
p(R_i) = [P^{mi}]_{i,4} - [P^{mi-1}]_{i,4}. \]

Finally, the probability that the system has not reached an out-of-control state by the \( i \)th run is

\[
p(A_i) = 1 - [P^{mi}]_{i,4}. \]

The above results are all that are required to obtain Parvin's formulae. However, the Markov Chain approach we present here is more general than the approach outlined by Parvin and much easier to use. Parvin's approach cannot be applied to multirules (he recommends simulation). Parvin also restricts the control rules in such a way that they count consecutive values above, or below, a control limit; that is, the \( R_{4s} \) rule that requires one observation above the \( \bar{x} + 2s \) limit and one below the \( \bar{x} - 2s \) limit cannot be modeled. Finally, rules such as the \( 2s_m \) rule cannot be properly used across runs because the approach disallows the possibility of the last result of one run combining with the first result of the next run to produce an out-of-control signal. One of our referees has indicated that when the \( 2s_m \) rule is used for the detection of persistent error, in the case where there are \((e.g.)\) two quality-control samples per run, it is customary to use the rule only within a run and not across runs, so that Parvin's results apply. Also the referee indicated that a \( 4s_m \) rule would be used only across two consecutive batches and would not use the quality-control samples from one batch with the last and first quality-control samples of the previous and subsequent batch, respectively. We will return to this point in the Results, when we show that the most efficient way to use rules across batches is to combine all of the quality-control results and ignore the batch classification. Apart from being a more efficient way to proceed, it makes the implementation of the control rules more straightforward. The Markov Chain approach presented here does not place the restrictions on the process that Parvin's approach does. Not withstanding the constraints just mentioned, Parvin has extended the discussion to the use of quality-control rules across runs. His preferred method of assessment is the cumulative probability \( p(\leq R_i) \), and this is what we now concentrate on. Within this framework, Parvin considers two further aspects of persistent error, the "start-up" and "ongoing" cases. "Start-up" refers to the system being in the out-of-control situation from the first run, whereas "ongoing" refers to the system starting at the in-control situation and then moving to the out-of-control situation during the operation of the quality-control scheme. Both cases can be modeled by our methods but in the interest of space we analyze only the start-up situation, just described.

Calculation of \( P_{fr} \), \( P_{sd} \), and Cumulative Probabilities for Cusum Schemes

The statistical properties of the Cusum schemes referred to in this text have been obtained by a simulation exercise, with 50 000 simulations to evaluate \( p_r \) values and 30 000 simulations to evaluate \( p_{sd} \) values. Determination of the cumulative distribution profiles involve distinguishing probabilities that are very similar in value; for these cases, \( 10^6 \) simulations were carried out.

Results

No Analytical Errors (Intermittent Case)

Table 1 shows the exact values for \( p_{fr} \), calculated by using Markov Chain theory, for the group of Shewhart-type control rules considered in this paper that were also considered by Westgard et al. (4). The figures in parentheses are those quoted by Westgard et al., calculated with their formula. As can be seen, the errors in their figures are slight, although they do seem more pronounced for the \( 3s_m \) rule and certainly increase in magnitude with larger \( N \). It is clear from both sets of figures that \( p_{fr} \) increases as \( N \) increases, unacceptably so for the scheme \( 1_{1.96} \).

The correct way of comparing these schemes is considered later, but part of the comparison involves determining the parameters of the control scheme that give rise to specified \( p_r \) values for each value of \( N \). A table of such parameter values is given by Westgard et al. for the control rules \( 1_{0.05}, 3_{0.05}, 1_{0.01}, 3_{0.01}, 1_{0.025}, \) and \( 3_{0.025} \). The exact parameter values are shown in Table 2 together with the Westgard et al. figures in parentheses. (A referee has also pointed out that, for all \( 1_r \)-type rules, the Westgard et al. figures are not consistent with the method of calculation suggested at the foot of their table.)

Tables 1 and 2 have been included for completeness so
Table 1. Probability of False Rejection, $p_{tr}$, Obtained from Markov Chain Theory [and Values Quoted Incorrectly by Westgard et al. (4)]

<table>
<thead>
<tr>
<th>N</th>
<th>$1_{.000}$</th>
<th>$1_{.000}$</th>
<th>$2_{.000}$</th>
<th>$3_{.000}$</th>
<th>$4_{.000}$</th>
<th>$7\bar{x}$</th>
<th>$10\bar{x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0020</td>
<td>0.0500</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.0020)</td>
<td>(0.0494)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0040</td>
<td>0.0975</td>
<td>0.0013</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.0040)</td>
<td>(0.0963)</td>
<td>(0.0012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0080</td>
<td>0.1426</td>
<td>0.0025</td>
<td>0.0080</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.0080)</td>
<td>(0.1409)</td>
<td>(0.0025)</td>
<td>(0.0079)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0080</td>
<td>0.1855</td>
<td>0.0037</td>
<td>0.0147</td>
<td>0.0013</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.0080)</td>
<td>(0.1833)</td>
<td>(0.0037)</td>
<td>(0.0146)</td>
<td>(0.0013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0100</td>
<td>0.2262</td>
<td>0.0049</td>
<td>0.0214</td>
<td>0.0023</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.0100)</td>
<td>(0.2237)</td>
<td>(0.0049)</td>
<td>(0.0213)</td>
<td>(0.0023)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0120</td>
<td>0.2649</td>
<td>0.0061</td>
<td>0.0281</td>
<td>0.0034</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.0119)</td>
<td>(0.2620)</td>
<td>(0.0061)</td>
<td>(0.0279)</td>
<td>(0.0034)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.0139</td>
<td>0.3016</td>
<td>0.0073</td>
<td>0.0348</td>
<td>0.0045</td>
<td>0.0156</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.0139)</td>
<td>(0.2985)</td>
<td>(0.0073)</td>
<td>(0.0345)</td>
<td>(0.0045)</td>
<td>(0.0156)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.0159</td>
<td>0.3386</td>
<td>0.0065</td>
<td>0.0414</td>
<td>0.0055</td>
<td>0.0234</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.0159)</td>
<td>(0.3331)</td>
<td>(0.0065)</td>
<td>(0.0411)</td>
<td>(0.0055)</td>
<td>(0.0233)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0198</td>
<td>0.4012</td>
<td>0.0110</td>
<td>0.0544</td>
<td>0.0077</td>
<td>0.0391</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>(0.0198)</td>
<td>(0.3973)</td>
<td>(0.0110)</td>
<td>(0.0540)</td>
<td>(0.0076)</td>
<td>(0.0387)</td>
<td>(0.0020)</td>
</tr>
<tr>
<td>12</td>
<td>0.0238</td>
<td>0.4596</td>
<td>0.0134</td>
<td>0.0673</td>
<td>0.0098</td>
<td>0.0547</td>
<td>0.0039</td>
</tr>
<tr>
<td></td>
<td>(0.0238)</td>
<td>(0.4553)</td>
<td>(0.0134)</td>
<td>(0.0668)</td>
<td>(0.0098)</td>
<td>(0.0539)</td>
<td>(0.0039)</td>
</tr>
<tr>
<td>16</td>
<td>0.0316</td>
<td>0.5596</td>
<td>0.0182</td>
<td>0.0925</td>
<td>0.0140</td>
<td>0.0854</td>
<td>0.0078</td>
</tr>
<tr>
<td></td>
<td>(0.0316)</td>
<td>(0.5552)</td>
<td>(0.0182)</td>
<td>(0.0919)</td>
<td>(0.0140)</td>
<td>(0.0839)</td>
<td>(0.0078)</td>
</tr>
<tr>
<td>20</td>
<td>0.0383</td>
<td>0.6415</td>
<td>0.0230</td>
<td>0.1171</td>
<td>0.0182</td>
<td>0.1151</td>
<td>0.0117</td>
</tr>
<tr>
<td></td>
<td>(0.0383)</td>
<td>(0.6387)</td>
<td>(0.0234)</td>
<td>(0.1182)</td>
<td>(0.0182)</td>
<td>(0.1130)</td>
<td>(0.0117)</td>
</tr>
<tr>
<td>28</td>
<td>0.0546</td>
<td>0.7822</td>
<td>0.0325</td>
<td>0.1642</td>
<td>0.0268</td>
<td>0.1715</td>
<td>0.0185</td>
</tr>
<tr>
<td></td>
<td>(0.0546)</td>
<td>(0.7757)</td>
<td>(0.0324)</td>
<td>(0.1630)</td>
<td>(0.0268)</td>
<td>(0.1683)</td>
<td>(0.0184)</td>
</tr>
<tr>
<td>40</td>
<td>0.0770</td>
<td>0.8715</td>
<td>0.0486</td>
<td>0.2302</td>
<td>0.0390</td>
<td>0.2495</td>
<td>0.0310</td>
</tr>
<tr>
<td></td>
<td>(0.0770)</td>
<td>(0.8680)</td>
<td>(0.0485)</td>
<td>(0.2266)</td>
<td>(0.0390)</td>
<td>(0.2449)</td>
<td>(0.0309)</td>
</tr>
<tr>
<td>100</td>
<td>0.1816</td>
<td>0.9841</td>
<td>0.1139</td>
<td>0.4886</td>
<td>0.0898</td>
<td>0.5423</td>
<td>0.0867</td>
</tr>
<tr>
<td></td>
<td>(0.1816)</td>
<td>(0.9637)</td>
<td>(0.1138)</td>
<td>(0.4870)</td>
<td>(0.0867)</td>
<td>(0.5342)</td>
<td>(0.0863)</td>
</tr>
</tbody>
</table>

that direct comparisons with the Westgard et al. figures can be made. The important issue those authors raised was, which of all the schemes considered is the best, and for what type of analytical error? This question is easily answered by the method outlined earlier in this paper; i.e., arrange for all control schemes to have the same $p_{tr}$ value and then choose that control scheme that has maximum $p_{sd}$. We consider three types of analytical error: systematic shift, systematic drift, and an increase in measurement error (imprecision). For brevity we only consider the cases $N = 5$ and $N = 15$.

**Systematic Shift (Single Schemes, Intermittent Error)**

A systematic shift occurs when there is a sudden increase in the mean response of the process but the imprecision remains constant throughout. Figure 2 (left) shows a plot of $p_{sd}$, when $p_{tr} = 0.05$, as a function of the change in the mean, for various Shewhart and Cusum schemes when $N = 5$. The Shewhart schemes $7\bar{x}$ and $10\bar{x}$ cannot be included in this plot, because these schemes require a minimum of 7 and 10 observations, respectively. Figure 2 (left) shows that the Cusum scheme should be the preferred control scheme, whatever the magnitude of change in the mean, because the Cusum profile is uniformly greater. Figure 2 (right) shows a similar plot for the case when $N = 15$. The larger the N value, the better the relative performance of the Cusum scheme—a result consistent with the observations by Rowland et al. (14). The Shewhart schemes $7\bar{x}$ and $10\bar{x}$ are omitted because neither scheme can have a $p_{tr}$ value of 0.05. To assess the relative performance of the $7\bar{x}$ and $10\bar{x}$ schemes, we must consider $p_{tr}$ values of 0.03906 and 0.00684 when $N = 10$ and $N = 15$, respectively. Figure 3 shows plots similar to Figure 2 but with the above-specified $p_{tr}$ and N values. Because the $7\bar{x}$ and $10\bar{x}$ schemes clearly are not as efficient as the Cusum, in the interests of space we will not consider these schemes further.

**Systematic Drift (Single Schemes, Intermittent Error)**

This type of analytical error occurs when there is a gradual change in the mean response of the process across the run, rather than a sudden change (as in a systematic shift). Which statistical model to adopt depends on the assay protocol being followed. For example, if there are N quality-control samples in the assay run, do you start with a quality-control sample before processing the first set of unknowns or do you process the first set of unknowns before any quality-control samples? Both situations are illustrated in Figure 4.

Suppose the change in the mean across the run is q. In the first design, the first quality-control sample
would be on target, with the remaining quality controls having mean values incremented by \(qs/(N - 1)\). For the second design, the first quality-control sample would have a mean value of \(qs/N\), and this would be the increment in the mean for subsequent quality-control samples. Westgard et al. do not make it clear which design they considered so, for the purposes of this paper, we consider the first design.

Plots of \(p_{ad}\) for various magnitudes of drift are shown in Figure 5, the two panels corresponding to a \(p_r\) value of 0.05 and analytical run lengths of \(N = 5\) and \(N = 15\), respectively. As illustrated, the Cusum scheme has properties far superior to those of the Shewhart schemes considered.

Random Error (Single Schemes, Intermittent Error)

This type of error occurs when the imprecision of the process suddenly deteriorates while the mean response of the process remains on target. Plots of \(p_{ad}\) for various increases in analytical error are shown in Figure 6, the two panels corresponding to a \(p_r\) value of 0.05 and analytical run lengths of \(N = 5\) and \(N = 15\), respectively. Again, the Cusum schemes perform very favorably, compared with most Shewhart-type schemes. The exception is the \(I_{0.05}\) rule, which seems to perform as well as the Cusum schemes. This might be expected when random error is increasing, because random error, by its very nature, is likely to produce large erratic swings in individual results rather than a consistent trend. Consequently, a \(I_{0.05}\) rule would respond quickly to these individual distant responses. Having said this, we show in Figure 6 that it is still possible to find an equally efficient Cusum scheme. For random error, the Cusum scheme that provides the better properties has a reference value of \(k = 1\), whereas for systematic drift and shift the Cusum scheme with reference value \(k = 0.5\) is best. This observation is consistent with the literature (22), where it is known that the optimal reference value for a Cusum scheme to detect a systematic shift of \(qs = q_{2/6}\), i.e., midway between the in- and out-of-control situations. There is no equivalent result available for random error, so that, from Figure 6, the midpoint rule apparently does not hold. As a consequence, it would be in the user’s interest to experiment a little to find the optimal reference value for the particular analytical error to be detected if the error is not of a drift or shift nature.

Multirule (Intermittent Error)

Because the multirule \(1_3/2_2/4_1/10_6\) proposed by Westgard et al. involves the use of the \(10_6\) rule, we

<table>
<thead>
<tr>
<th>(N)</th>
<th>(I_{0.05})</th>
<th>(I_{0.05})</th>
<th>(I_{0.01})</th>
<th>(I_{0.01})</th>
<th>(I_{0.005})</th>
<th>(I_{0.005})</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.236</td>
<td>1.002</td>
<td>2.806</td>
<td>1.471</td>
<td>3.290</td>
<td>1.857</td>
</tr>
<tr>
<td></td>
<td>(2.24)</td>
<td>(1.01)</td>
<td>(2.81)</td>
<td>(1.47)</td>
<td>(3.27)</td>
<td>(1.86)</td>
</tr>
<tr>
<td>3</td>
<td>2.388</td>
<td>1.199</td>
<td>2.934</td>
<td>1.639</td>
<td>3.403</td>
<td>2.005</td>
</tr>
<tr>
<td></td>
<td>(2.39)</td>
<td>(1.22)</td>
<td>(2.93)</td>
<td>(1.64)</td>
<td>(3.36)</td>
<td>(2.01)</td>
</tr>
<tr>
<td>4</td>
<td>2.491</td>
<td>1.314</td>
<td>3.022</td>
<td>1.735</td>
<td>3.481</td>
<td>2.089</td>
</tr>
<tr>
<td></td>
<td>(2.50)</td>
<td>(1.33)</td>
<td>(3.01)</td>
<td>(1.74)</td>
<td>(3.44)</td>
<td>(2.09)</td>
</tr>
<tr>
<td>5</td>
<td>2.589</td>
<td>1.393</td>
<td>3.089</td>
<td>1.801</td>
<td>3.540</td>
<td>2.147</td>
</tr>
<tr>
<td></td>
<td>(2.58)</td>
<td>(1.41)</td>
<td>(3.06)</td>
<td>(1.81)</td>
<td>(3.49)</td>
<td>(2.15)</td>
</tr>
<tr>
<td>6</td>
<td>2.631</td>
<td>1.452</td>
<td>3.143</td>
<td>1.851</td>
<td>3.588</td>
<td>2.191</td>
</tr>
<tr>
<td></td>
<td>(2.64)</td>
<td>(1.47)</td>
<td>(3.13)</td>
<td>(1.86)</td>
<td>(3.52)</td>
<td>(2.19)</td>
</tr>
<tr>
<td>8</td>
<td>2.727</td>
<td>1.540</td>
<td>3.226</td>
<td>1.926</td>
<td>3.662</td>
<td>2.257</td>
</tr>
<tr>
<td></td>
<td>(2.74)</td>
<td>(1.56)</td>
<td>(3.21)</td>
<td>(1.93)</td>
<td>(3.59)</td>
<td>(2.26)</td>
</tr>
<tr>
<td>10</td>
<td>2.800</td>
<td>1.603</td>
<td>3.289</td>
<td>1.990</td>
<td>3.719</td>
<td>2.305</td>
</tr>
<tr>
<td></td>
<td>(2.81)</td>
<td>(1.62)</td>
<td>(3.27)</td>
<td>(2.00)</td>
<td>(3.64)</td>
<td>(2.30)</td>
</tr>
<tr>
<td>12</td>
<td>2.858</td>
<td>1.653</td>
<td>3.340</td>
<td>2.022</td>
<td>3.765</td>
<td>2.342</td>
</tr>
<tr>
<td></td>
<td>(2.86)</td>
<td>(1.67)</td>
<td>(3.31)</td>
<td>(2.03)</td>
<td>(3.66)</td>
<td>(2.34)</td>
</tr>
<tr>
<td>16</td>
<td>2.948</td>
<td>1.727</td>
<td>3.419</td>
<td>2.087</td>
<td>3.836</td>
<td>2.400</td>
</tr>
<tr>
<td></td>
<td>(2.94)</td>
<td>(1.74)</td>
<td>(3.38)</td>
<td>(2.09)</td>
<td>(3.72)</td>
<td>(2.40)</td>
</tr>
<tr>
<td>20</td>
<td>3.016</td>
<td>1.783</td>
<td>3.480</td>
<td>2.135</td>
<td>3.891</td>
<td>2.443</td>
</tr>
<tr>
<td></td>
<td>(3.02)</td>
<td>(1.80)</td>
<td>(3.43)</td>
<td>(2.14)</td>
<td>(3.75)</td>
<td>(2.44)</td>
</tr>
<tr>
<td>28</td>
<td>3.118</td>
<td>1.863</td>
<td>3.569</td>
<td>2.205</td>
<td>3.972</td>
<td>2.506</td>
</tr>
<tr>
<td></td>
<td>(3.11)</td>
<td>(1.88)</td>
<td>(3.51)</td>
<td>(2.21)</td>
<td>(3.78)</td>
<td>(2.51)</td>
</tr>
<tr>
<td>40</td>
<td>3.220</td>
<td>1.944</td>
<td>3.661</td>
<td>2.276</td>
<td>4.056</td>
<td>2.570</td>
</tr>
<tr>
<td></td>
<td>(3.21)</td>
<td>(1.96)</td>
<td>(3.60)</td>
<td>(2.28)</td>
<td>(3.85)</td>
<td>(2.57)</td>
</tr>
</tbody>
</table>

Fig. 2. Plots of \(p_{ad}\) for various Cusum (CS) and Shewhart-type rules as a function of the magnitude of systematic shift

The notation for the various rules is defined in text.
Persistent Error

We first begin by making the case that the most efficient way of operating Shewhart-type control schemes, including the multirules, across runs is to combine the quality-control results and ignore the batch classification. To be specific, and to reduce space, we consider the $2_{2x}$ and $4_{1x}$ rules in detail. For ease of explanation we make the following definitions.

- Pooled approach: Combine all quality-control samples and disregard batch classification.
- Batch approach: Use control schemes in such a way that, if a batch is involved in operating a control scheme, then all of the quality-control samples within that batch are involved.

Consider a $2_{2x}$ rule for the case of persistent error with two quality-control observations per batch. The cumulative probabilities for both the in-control and out-of-control situations, calculated by using the Markov Chain approach, are shown in Table 3, where the
out-of-control condition is a shift in the mean value of 1σ. As can be seen, the out-of-control profile of the pooled approach is far superior to that of the batch approach, although we note that the in-control profile is not so good. If we then place the decision lines of the 2σ̅ rule at ±2.14σ̅, we obtain the results in Table 4. Table 4 shows that, to the accuracy given, the pooled approach has an in-control cumulative probability profile that is the same as that for the batch approach and has an out-of-control profile that is almost uniformly greater. We thus conclude that, in the situations considered, the pooled approach is the more efficient method. Tables 5 and 6 are equivalent to Tables 3 and 4, but for the 4σ̅ rule. Again, the pooled approach is the better one, although the difference between the two approaches is less pronounced than for the 2σ̅ rule. We would expect this pattern of behavior because the difference between the two approaches relates to the “end effect” problem, when not all quality-control samples within a batch are included in the decision rule; this effect is minimized for control rules that make use of many more quality-control samples than are within a given batch. Because we wish to compare the Cusum with the best possible Shewhart type of approach, we will assume, in what follows, that the pooled approach has been adopted.

Multi-rule (Persistent Error, Startup Case)

In the interests of space we consider only the multi-rule mentioned earlier, which was also considered by Parvin. In Parvin's comparison the cumulative probability of an out-of-control message being issued was plotted up to run 15. Figure 8 shows, for the start-up case, the cumulative probability of error detection for the multi-rule when the system is in control and when the system is assumed to be out of control from run 1 with a systematic shift in the mean response of the process equal to 1σ. The slight "kinks" in the plot occur at points where the component schemes start to operate. Also plotted are the corresponding quantities for the Cusum scheme with decision interval h = 3.385 and
reference value \( k = 0.5 \). As Figure 8 shows, the profile for the Cusum scheme at in-control conditions is almost uniformly below that of the multirule; when the system is out of control, the Cusum profile is almost uniformly greater than the multirule profile. The cumulative probability profile of the Cusum scheme at in-control can be lowered either by increasing the decision interval \( h \) or reference value \( k \) or both. If the changes in \( h \) and \( k \) are not too large, the out-of-control profile will not be affected much. These results mean that for the first 15 runs there is a Cusum scheme that will perform much better than the multirule \( \frac{1}{2}2 \alpha R_{L}/R_{H}/10\sigma \). The ARLs (remember the pooled approach has been used, so we are referring to ARL\(_{50}\)) of the multirule and Cusum schemes when the system is in control are 55 and 74, respectively, with the corresponding results for the out-of-control situation being 12.8 and 7.2, respectively. The ARLs of the two schemes are not equal at the in- or
out-of-control situations and so cannot be properly compared by using this criterion. However, the nomogram provided by Kemp et al. (8) shows that the Cusum scheme \( C_{.06} \) has an in-control ARL of 55 and an out-of-control ARL of 6.8. This Cusum has the same in-control ARL as the multirule but has a much lower out-of-control ARL. Again, it would seem that a Cusum scheme can be found with superior properties. Clearly, the optimal values for \( h \) and \( k \) depend on the assessment criterion used.

Discussion

We have outlined a procedure, based on simple Markov Chain theory, that enables the user to calculate all of the salient properties of single and multirule Shewhart-type quality-control schemes for both intermittent error and persistent error. This procedure avoids the simulation process adopted by many authors and consequently removes the sampling error in estimating \( p_h \) and \( p_{ed} \) for intermittent error and the cumulative probability profile for persistent error. As an example of the effects of sampling error, Westgard et al. used 1600 simulations to estimate \( p_h \) and 400 simulations to estimate \( p_{ed} \). For a quoted value of 0.05 for \( p_h \) and 0.3 for \( p_{ed} \), the corresponding widths of the 95% confidence intervals are 0.02 and 0.09, respectively. These errors are not insignificant and should be borne in mind when comparing different quality-control schemes. In this paper, only the Cusum scheme had to be simulated. We used 50 000 for \( p_h \) and 30 000 for \( p_{ed} \); for \( p_h = 0.05 \) and \( p_{ed} = 0.3 \), these give 95% confidence intervals with widths of 0.004 and 0.01, respectively.

We have also shown that a fundamental formula used by Westgard et al. to calculate \( p_h \) is in error. The Markov Chain approach discussed earlier shows that, although the formula is wrong, the results are not grossly in error. For our readers’ information, we have provided tables equivalent to those of Westgard et al., showing the correct figures.

Westgard’s main efforts, in several of his papers (3-6), are directed at comparing the various quality-control procedures to see which have the “best” performance characteristics in the presence of different types of analytical error. Most of his comparisons involve plotting \( p_{ed} \) as a function of the magnitude of error in the system or as a function of the size of the analytical run for a given magnitude of error. These plots have been derived by simulation procedures in the main and some of the \( p_{ed} \) values have been obtained by using an incorrect formula, as outlined earlier. His multiple plots have schemes with totally different \( p_h \) and \( p_{ed} \) values and consequently it is extremely difficult, if not impossible, to make a proper judgment about the optimal characteristics of each scheme. We have defined a strict comparison criterion for dealing with intermittent error that makes comparisons objective and shows that, whatever analytical error is present, a Cusum rule can be found that has better properties than the single or multirule Shewhart schemes considered. We have conjectured (10) that a Cusum scheme is best of all. We have also offered evidence to suggest that when control rules are used across batches, to detect persistent error, the most efficient way of utilizing the quality-control information is to pool all quality-control sample values and disregard the batch classification. For persistent error, as described by Parvin, there seems to be two approaches, comparing schemes by their ARLQC profiles (10) or by plotting the cumulative distribution of run length. We have considered both approaches for the multirule \( 1_{.06}2_{.06}R_{.06}/4_{.06} \) and again conclude that a Cusum scheme can be found with superior properties. We clearly cannot claim that a Cusum scheme is better than all Shewhart schemes because an infinite number of such schemes is possible; however, we do feel that if the Cusum scheme can be bettered, it will only be very marginally. This is, in our view, the most important result of the paper because it removes the need for the user to select which quality-control procedure to use. Also, it is our experience that, although academic papers might further refine the Shewhart multirules, the general user would much prefer a simple-to-use scheme that is easy to plot and interpret. One referee indicated that to encourage the use of Cusum schemes software should be made available. As a result we have included in the Appendix 2 a program listing in Microsoft Quickbasic that will simulate a two-sided Cusum decision interval scheme with decision interval \( h \) and reference value \( k \).

Finally, Westgard et al. consider the effects of the various types of analytical error on each of the various schemes when many of the schemes are primarily designed to detect just one component of error. Nix et al. report (10) that the most efficient way to detect and discriminate between the different types of error is to design Cusum schemes that are individually efficient at detecting the different error components. We conjecture, therefore, that just three Cusum schemes would provide a very efficient way of detecting systematic shift, drift,
and imprecision, while at the same time providing the user with a simple-to-use procedure.

We express our thanks to the referees, who made a number of useful comments that have improved the text.

Appendix 1

Listed here are the one-step transition matrices for a selection of single and multirule Shewhart schemes.

1. Here there are only two states.

- ISO point within three standard deviations (3s) limits (probability \( p_0 \)).
- A, point outside of 3s limits (probability 1 - \( p_0 \)).

\[
\begin{align*}
S(0,0) & \rightarrow S(1,0) \\
S(1,0) & \rightarrow S(0,0) \\
S(0,0) & \rightarrow S(0,0) \\
A & \rightarrow A
\end{align*}
\]

where \( p_0 \) = probability that the point falls between \( \bar{x} + 2s \) and \( \bar{x} - 2s \).

- \( p_u \) = probability that the point falls above the \( \bar{x} + 2s \) limit,
- \( p_l \) = probability that the point falls below the \( \bar{x} - 2s \) limit.

2. Given in text.

R\(_{4s}\): Here the four states—S(0,0), S(1,0), S(0,1), A—are the same as those for the 2s rule (see text).

\[
\begin{align*}
S(0,0) & \rightarrow S(1,0) \\
S(1,0) & \rightarrow S(0,0) \\
S(0,0) & \rightarrow S(0,0) \\
A & \rightarrow A
\end{align*}
\]

where \( p_0 \) = probability that the point falls between \( \bar{x} + 1s \) and \( \bar{x} - 1s \),

- \( p_u \) = probability that the point falls above the \( \bar{x} + 1s \) limit,
- \( p_l \) = probability that the point falls below the \( \bar{x} - 1s \) limit.

The following program, written in Microsoft Quickbasic 4.5, simulates the use of a two-sided Cusum rule with decision interval \( h \) and reference value \( k \).

1. **ISO point within three standard deviations (3s) limits (probability \( p_0 \)).**

\[
\begin{align*}
S(0,0) & \rightarrow S(1,0) \\
S(1,0) & \rightarrow S(0,0) \\
S(0,0) & \rightarrow S(0,0) \\
A & \rightarrow A
\end{align*}
\]

where \( a \) = probability that the point falls between \( \bar{x} + 1s \) and \( \bar{x} - 1s \),

- \( b \) = probability that the point falls between \( \bar{x} + 1s \) and \( \bar{x} + 2s \),
- \( c \) = probability that the point falls between \( \bar{x} - 1s \) and \( \bar{x} - 2s \),
- \( d \) = probability that the point falls between \( \bar{x} + 2s \) and \( \bar{x} + 3s \),
- \( e \) = probability that the point falls between \( \bar{x} - 2s \) and \( \bar{x} - 3s \),
- \( f \) = probability that the point is \( > \bar{x} + 3s \),
- \( g \) = probability that the point is \( < \bar{x} - 3s \).

Note that the states ISO-2S(1,0)-RS(0,0), ISO-2S(1,0)-RS(1,0), ISO-2S(0,1)-RS(0,0), and ISO-2S(0,1)-RS(1,0) are not accessible and so have a row of zeroes. These states should be removed but have been left in to maintain the block pattern of the matrix. These rows will not affect probability calculations.

Appendix 2

The following program, written in Microsoft Quickbasic 4.5, simulates the use of a two-sided Cusum rule with decision interval \( h \) and reference value \( k \).

\[
\begin{align*}
S(0,0) & \rightarrow S(1,0) \\
S(1,0) & \rightarrow S(0,0) \\
S(0,0) & \rightarrow S(0,0) \\
A & \rightarrow A
\end{align*}
\]

where \( \sigma \) = probability that the point falls between \( \bar{x} + 1s \) and \( \bar{x} - 1s \),

- \( \nu \) = probability that the point falls above the \( \bar{x} + 1s \) limit,
- \( \mu \) = probability that the point falls below the \( \bar{x} - 1s \) limit.

The program uses a set of matrices for the multirule Shewhart schemes. In an obvious notation the one-step transition matrix is:

\[
\begin{align*}
S(0,0) & \rightarrow S(1,0) \\
S(1,0) & \rightarrow S(0,0) \\
S(0,0) & \rightarrow S(0,0) \\
A & \rightarrow A
\end{align*}
\]

where \( a \) = probability that the point falls between \( \bar{x} + 1s \) and \( \bar{x} - 1s \),

- \( b \) = probability that the point falls between \( \bar{x} + 1s \) and \( \bar{x} + 2s \),
- \( c \) = probability that the point falls between \( \bar{x} - 1s \) and \( \bar{x} - 2s \),
- \( d \) = probability that the point falls between \( \bar{x} + 2s \) and \( \bar{x} + 3s \),
- \( e \) = probability that the point falls between \( \bar{x} - 2s \) and \( \bar{x} - 3s \),
- \( f \) = probability that the point is \( > \bar{x} + 3s \),
- \( g \) = probability that the point is \( < \bar{x} - 3s \).

Note that the states ISO-2S(1,0)-RS(0,0), ISO-2S(1,0)-RS(1,0), ISO-2S(0,1)-RS(0,0), and ISO-2S(0,1)-RS(1,0) are not accessible and so have a row of zeroes. These states should be removed but have been left in to maintain the block pattern of the matrix. These rows will not affect probability calculations.

The quality-control observations have been placed in the data array \( X(i) \) \( i = 1, \ldots, N \), where \( X(i) \) is the ith quality-control observation arranged in time order. The variable RUNNO corresponds to the first point at which the Cusum rule declares an out-of-
control message. If the out-of-control state is not reached by the end of the data array, RUNNO takes the value 0. The array FREQ(i) is the frequency distribution of run lengths.

References