Automated Selection of Statistical Quality-Control Procedures to Assure Meeting Clinical or Analytical Quality Requirements, James O. Westgard1,2 and Bernard Stein2

1 Dept. of Pathol. and Lab. Med., Univ. of Wisconsin Med. School, Madison, WI 53792 (address for correspondence; fax 608-263-0910), and 2 Westgard Quality Corp., Ogunquit, ME 03907; e-mail www.westgard.com

Efforts to totally automate laboratory testing processes must address the issue of how to assure the quality of the final test result. A recent survey by Tetrault and Steindel[1] reported that laboratories are using the same quality-control (QC) procedures today that they used 10 years ago. Actually, more than half of the laboratories indicated they are still using control limits set as the mean ± 2s (the 1s rule), a practice that dates to the 1950s and ‘60s[2,3], when statistical QC was first used with manual methods and the first generation of automated Technicon AutoAnalyzer systems. Other laboratories indicated they are using variations of the multirule type of QC procedure introduced in the early 1980s[4].

Tetrault and Steindel[1] recommend that “the best set of control rules will vary from method to method and cannot be determined through simple algorithms or formulas. The laboratorian has to balance true error-detection capabilities against the probabilities of falsely rejecting a good run.” The information needed about the error-detection and false-rejection characteristics of stable QC procedures became available in the clinical chemistry literature almost 20 years ago[5–8] and was extended by Cembrowski et al.[9–12] to consider patient data algorithms. This information was incorporated into laboratory QC texts[13,14] and clinical chemistry texts[15,16] in the 1990s and continues to be expanded and improved by Parvin[17–20] and by Smith and Kroft[21].

Guidelines and strategies for selecting QC procedures were described in 1986[13], and some detailed applications have been published to demonstrate the selection and design of QC procedures[22–24]. QC selection grids were introduced in 1990[25] to provide a simple table “look up” approach for selecting QC procedures on the basis of the size of systematic error that must be detected and on an assay’s expected stability or frequency of errors. More quantitative quality-planning models were introduced in 1991[26,27], and a new planning tool—the OPSpecs chart[3]—was developed[28,29] to show the relation between allowable precision and accuracy as well as what QC is necessary to assure detection of critical-sized errors that would otherwise cause method performance to exceed a defined analytical or clinical quality requirement. For applications involving analytical quality requirements and commonly used QC procedures, a Windows3-based PC program (QC Validator4, Version 1.1; Westgard Quality Corp., Ogunquit, ME) was developed to prepare power function graphs, critical error graphs, and OPSpecs charts[31].

Quality-planning models describe the mathematical relationships between a quality requirement and the factors that can cause variation in the final test result. Some of these factors are analytical, such as the stable imprecision (smean%), and inaccuracy of the method and the sensitivity of the QC procedure to detect unstable random and systematic errors (∆Sεcrit). Others are pre-analytical, such as the within-subject biological variation (swsub), which describes the changes in concentration about the subject’s true homeostatic set point. We have now developed an automated process to select statistical control rules and numbers of control measurements (N) that will assure the quality required by clinical decision intervals (Dint) criteria or analytical total error (TEa) criteria. The quality-planning models used in the automatic process are essentially the same as those described earlier[26,27], except that the effect of replicate measurements on method performance and QC design are more completely and directly considered by entering the number of replicate samples analyzed. Here, we describe the automatic process and demonstrate that it provides results comparable with those in earlier studies documented in this journal.

The computer program used, QC Validator Version 2.0, runs under Microsoft Windows 3.1 or Windows 95. It requires an IBM or IBM-compatible computer with a 386 or higher processor, 1 MB hard disk space, 2 MB RAM, a VGA interface and monitor, a printer supported by Windows, and TrueType fonts. The program includes an installation utility, a detailed HELP facility, and a manual that provides tutorials, a reference guide to program function and operation, and a review of the technical background for the program. The computer program allows the user to enter information about the characteristics of a method’s performance (e.g., smean%, inaccuracy, expected frequency of errors, swsub), and the quality required (clinically important change or Dint, TEa). The user then initiates automatic selection on the basis of the number of control materials to be analyzed (i.e., 1, 2, or 3), and the program constructs a chart of operating specifications (OPSpecs chart) that displays the selected control rules and N. The automatic QC selection process is based on user-editable criteria for the types of control rules that can be implemented by the laboratory, the total numbers of control measurements that are practical, the maximum percentage of false rejections that can be tolerated, and the minimum percentage of error detection that is acceptable for detection of medically important ∆Sεcrit or random errors.

The table of candidate QC procedures contains the power curves and OPSpecs calculation parameters for 94
different QC procedures, including: constant-limit single rules $1_{2s}$, $1_{2.5s}$, $1_{3s}$, and $1_{3.5s}$, with N values from 1 to 8; multirules such as $1_{3s}/2_{2s}/R_{4s}/4_{ns}$ with N values of 2 and 4 applied over runs from 1 to 5; multirules such as $1_{3s}/2_{3s}/3_{3s}/12_{ns}$, with N values of 3 and 6 applied over runs from 1 to 4; multirules such as $1_{3s}/2_{2s}/R_{4s}/4_{1s}$, with N values of 2 to 8; mean and range rules such as $x_{0.05}/R_{0.05}$, $x_{0.01}/R_{0.01}$, and $x_{0.002}/R_{0.002}$, with N values from 2 to 8; and extended limit rules such as $1_{4s}, 1_{5s}$, and $1_{6s}$, with N values from 1 to 8. The program includes an editor utility that allows users to add power curves and OPSpecs calculation parameters for other rules and N values.

OPSpecs charts were prepared for the examples illustrated earlier by spreadsheet calculations of allowable imprecision and allowable inaccuracy for both the clinical model [26] and the analytical model [27]. Fig. 1 (top) shows the OPSpecs chart for a cholesterol example, for which $D_{int} = 20\%$ and $s_{wsub} = 6.5\%$. The x-intercepts for maximum allowable imprecision range from 2.4\% to 3.5\%, which agrees with the range of 2.4\% to 3.5\% observed previously (Fig. 2a in [26]). For individual QC procedures, the largest difference was observed for the $1_{3s}$ rule with N = 4, for which the current intercept is 2.8\% (vs the previous estimate of 2.95\%). Performing duplicate cholesterol tests and using the average of the two measured concentrations for diagnostic classification provides a maximum allowable imprecision of 3.3\% to 4.7\%, compared with the earlier estimate of 3.2\% to 4.9\% (Fig. 2b in [26]).

For a cholesterol example in which TE$_a$ is 10\%, Fig. 1 (bottom) shows the estimated allowable imprecision (x-intercepts) to be 1.9\% to 2.6\%, the same as the 1.9\% to 2.6\% documented earlier (Fig. 4 in [27]). In the same example, when error detection is optimized for random error rather than systematic error, the allowable imprecision is estimated as 1.2\% to 2.3\%, very similar to earlier estimates of 1.2\% to 2.4\% (Fig. 6 in [27]).

The differences between the program output and the earlier spreadsheet calculations are the result of the different calculation parameters for the sizes of systematic error that can be detected with the specified probability. These parameters were estimated as the x-values that correspond to the 0.90 intersection with the power curves for the different control rules and N values, which in the spreadsheet calculation were constructed on a point-to-point basis and then later fitted with smoothed curves to improve the estimates.

The automatic QC selection process was tested by comparing the QC procedures selected by the program that uses the OPSpecs methodology with those selected in an earlier study in which critical error graphs were used manually to select control rules and N values for 18 different analytes on a multistest chemistry analyzer [23]. In the earlier study, the design strategy fixed N at 2 per run, considered only single-rule QC procedures, and varied the control rules to match the error detection capability to the test method performance. Table 1 shows the test, total error requirement, observed imprecision, calculated $\triangle SE_{crit}$, and the QC procedure selected manually from critical error graphs or selected automatically from OPSpecs charts. For the automatic selection process, the frequency of errors parameter was set to $<2\%$, to permit use of 50\% Analytical Quality Assurance (AQA) charts in those situations where 90\% AQA could not be achieved.

For 14 tests, both the manual and automatic approaches selected the $1_{3.5s}$ rule with N = 2; for another test (albumin), both approaches selected the $1_{2.5s}$ rule with N = 2. For two tests (chloride and total CO$_2$), the $1_{2.5s}$ rule with N = 2 had been selected manually, whereas the

![Fig. 1. OPSpecs charts for (top) a clinical decision interval quality requirement of 20% and (bottom) a TE$_a$ quality requirement of 10%.
Both panels: The operating limits (top to bottom) correspond to the following QC procedures: $1_{3s}/2_{2s}/R_{4s}/4_{ns}$ with N = 4; $1_{3s}/R_{4s}$ with N = 4; $1_{2s}$ with N = 4; $1_{2s}$ with N = 2; $1_{3s}/2_{2s}/R_{4s}$ with N = 2; and $1_{2s}$ with N = 2. The operating point (bottom panel) represents the National Cholesterol Education Program’s recommended 3\% specifications for imprecision and inaccuracy.](https://academic.oup.com/clinchem/article-abstract/43/2/400/5640716)
The automatic selection process identified a multirule procedure with $N = 4$. In these latter two cases, if the automatic QC selection criteria were changed to restrict $N$ to 2, then the automated process would select a multirule procedure with $N = 2$; if the selection logic for 50% AQA charts was changed to prefer single rules over multirules, then a $1_{2.5s}$ rule with $N = 2$ would be selected. That is, the automatic QC process can be modified to implement the same preferences used earlier in the process of manual selection from critical error graphs. For one test (calcium), method performance relative to the quality desired was so marginal that the assay required a special effort to make duplicate measurements, which was necessary to improve method performance and process control. The effect of making duplicate measurements could be considered directly with the automatic selection process, which then identified a $1_{2.5s}$ rule with $N = 4$ as necessary to provide 90% detection of medically important errors.

Therefore, although the QC designs and outcomes of the automatic and manual selections were similar, one thing was not: The time required to assess the QC designs for these 18 tests and to document the selection with printouts of OPSpecs charts was ~30 min with the automatic QC selection process, whereas the earlier study took several months.

In the future, one can envision an automatic QC selection process integrated into laboratory instrument software, on-line QC monitors, data management workstations, laboratory information systems, and automated laboratory process control software, thereby permitting laboratory scientists to focus on defining the quality needed for a test, rather than worrying about the control rules and numbers of control measurements that should be used. When coupled with on-line data acquisition and appropriate software to estimate method performance characteristics, the information needed for QC design will itself be automatically available. When linked to software for on-line QC monitoring, the selected rules can be implemented automatically. Thus, an integrated system for analytical quality management is possible when an automatic QC selection process is linked with method performance data and on-line QC monitoring, allowing the management of analytical quality with an appropriate QC design based on current information on method performance.

Such an integrated quality-management system provides the opportunity for a dynamic QC system that can automatically adjust to changes in method performance, e.g., changing to more-sensitive control rules and higher $N$ values if performance deteriorates and relaxing the rules and reducing $N$ if performance improves. Changes in method stability can be factored into the dynamic QC process and can also be used to adjust run length, as suggested by recent design work on the application of “average of normals” patient data algorithms to measure process stability and determine when to requalify the testing process [32]. Thus, technology for the next-generation system for analytical quality management can be envisioned now that automatic QC selection is a reality.

Westgard Quality Corp. supported the development of this software. Robert Kennedy performed the coding, and Sten Westgard developed the HELP facility and program.

### Table 1. QC procedures selected by automated QC selection process using OPSpecs charts compared with those selected manually by using critical error graphs.

<table>
<thead>
<tr>
<th>Test</th>
<th>$TE_{%}$</th>
<th>$S_{max, %}$</th>
<th>$\Delta SE_{crit}$</th>
<th>Manually</th>
<th>Automatically</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium</td>
<td>3.08</td>
<td>0.52</td>
<td>4.27</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>Potassium</td>
<td>10.0</td>
<td>1.17</td>
<td>6.90</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>Chloride</td>
<td>4.0</td>
<td>1.04</td>
<td>2.20</td>
<td>$1_{2.5s}, N = 2$</td>
<td>MR, $N = 4$</td>
</tr>
<tr>
<td>Total CO₂</td>
<td>10.0</td>
<td>2.50</td>
<td>2.35</td>
<td>$1_{2.5s}, N = 2$</td>
<td>MR, $N = 4$</td>
</tr>
<tr>
<td>Glucose</td>
<td>8.0</td>
<td>1.20</td>
<td>5.02</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>Urea N</td>
<td>10.0</td>
<td>1.33</td>
<td>5.87</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>Creatinine</td>
<td>30.0</td>
<td>3.00</td>
<td>8.35</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>Calcium</td>
<td>5.0</td>
<td>1.68</td>
<td>1.33</td>
<td>Special case</td>
<td>MR, $N = 4$</td>
</tr>
<tr>
<td>Phosphorus</td>
<td>10.0</td>
<td>1.28</td>
<td>6.16</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>Uric acid</td>
<td>10.0</td>
<td>1.10</td>
<td>7.44</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>10.0</td>
<td>1.35</td>
<td>5.76</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>Total protein</td>
<td>12.0</td>
<td>1.84</td>
<td>4.87</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>Albumin</td>
<td>10.0</td>
<td>2.13</td>
<td>3.04</td>
<td>$1_{2.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>Total bilirubin</td>
<td>20.0</td>
<td>2.20</td>
<td>7.44</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>GGT</td>
<td>10.0</td>
<td>1.17</td>
<td>6.90</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>ALP</td>
<td>10.0</td>
<td>1.17</td>
<td>6.90</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>AST</td>
<td>20.0</td>
<td>3.0</td>
<td>5.02</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
<tr>
<td>LD</td>
<td>20.0</td>
<td>3.0</td>
<td>5.02</td>
<td>$1_{3.5s}, N = 2$</td>
<td>Same</td>
</tr>
</tbody>
</table>

MR, multirule; GGT, γ-glutamyltransferase; ALP, alkaline phosphatase; AST, aspartate aminotransferase; LD, lactate dehydrogenase.


**References**


**Rapid Screening for α1-Antitrypsin Z and S Mutations, Christopher W.K. Lam, Chi-Pui Pang,* Priscilla M.K. Poon, Chang-Hong Yin, and Geetha Bharathi (Dept. of Chem. Pathol., Chinese Univ. of Hong Kong, Prince of Wales Hosp., Shatin, N.T., Hong Kong; *author for correspondence: fax 852 26365090, e-mail cppang@cuhk.hk)**

α1-Antitrypsin (A1AT) is a serine protease inhibitor required for the prevention of proteolytic tissue damage, principally in the lung, by neutrophil elastase released by inflammatory cells [1]. While severe A1AT deficiency is the major factor leading to emphysema and related pulmonary diseases, it is also associated with neonatal hepatitis and cirrhosis [1, 2]. A1AT deficiency is an autosomal codominant disorder with a prevalence of about 1:3000 in Caucasians [3]. The A1AT gene has 7 exons spanning ~12 kb. The most common gene defect resulting in A1AT deficiency is that of a protease inhibitor (PI)-system Z mutation Glu342 to Lys, which is a single base substitution [1, 2]. A1AT deficiency is an autosomal codominant disorder with a prevalence of about 1:3000 in Caucasians [3]. The A1AT gene has 7 exons spanning ~12 kb. The most common gene defect resulting in A1AT deficiency is that of a protease inhibitor (PI)-system Z mutation Glu342 to Lys, which is a single base substitution of G to A in exon 5 [4, 5]. The S mutation, a Glu264 to Val change, is caused by an A to T substitution in exon 3 [6]. Individuals with SS are unaffected, SZ may be symptomatic, and ZZ results in the most severe clinical symptoms. In Caucasians the prevalence of the S allele ranges from 5% to 10% and Z allele 2% to 5% depending on geographical location [7, 8]. Although the frequencies are unknown in the Chinese, geographical variability of the A1AT alleles is evident by phenotypic analysis of the PI variants [9]. We have established a rapid screening procedure involving multiplex PCR to detect the Z and S mutations in Chinese in Hong Kong who came from southern China.

EDTA-whole-blood specimens were obtained from local Chinese in Hong Kong who attended the Prince of Wales Hospital for routine checkup or for treatment of diabetes mellitus. Genomic DNA was extracted from the blood specimens by the salting-out method [10]. Our procedure for mutation analysis was modified from the PCR-mediated site-directed mutagenesis method of Tazelaar et al. [11] with their primers for the Z and the S mutations, labeled as primers ZF and ZR and primers SF and SR respectively. Each PCR mixture, in a final volume of 25 µL, contained 0.2 mmol/L deoxynucleoside triphosphates (Boehringer Mannheim, Mannheim, Germany), 1.5 mmol/L magnesium chloride, 0.1 g/L gelatin, 2.5 pmol each of primers ZF, ZR, SF, and SR (synthesized by Gibco BRL, Gaithersburg, MD), 0.2 µg of DNA, 0.5 U of Taq polymerase (Gibco BRL), and 1× PCR buffer from Gibco.