Reprinted from POINT OF CARE
Vol. 2 No. 1 March 2003

Validation of iQM Active Process Control Technology

*James O. Westgard, PhD, †Kevin D. Fallon, PhD, and †Sohrab Mansouri, PhD

A validation study has been performed to determine the error detection capabilities of a new quality control (QC) technology called "intelligent Quality Management (iQM)." iQM is a completely automated statistical QC process that uses frequent measurements of internal process control solutions to monitor measurement variation and signal abnormal drifts, then applies pattern recognition algorithms to identify the type of error and trigger appropriate corrective actions. The validation methodology follows National Committee for Clinical Laboratory Standards (NCCLS) and International Standards Organization (ISO) guidelines and involves characterizing method performance on the sigma-scale, characterizing instrument drift limits as statistical control rules, then assessing the probabilities of rejecting runs whose errors would exceed defined quality requirements, such as the Clinical Laboratory Improvement Amendment (CLIA) proficiency criteria for acceptable performance. Practical measures of performance are obtained by determining the average run lengths and converting them to the average times for detecting errors. With iQM, medically important errors will be detected within 0.05 to 0.5 hours for pH, PCO2, PO2, potassium, calcium, lactate, and hematocrit; 0.12 to 1.2 hours for glucose; and 0.17 to 1.7 hours for sodium. Compared to current QC practices where controls are analyzed every 8 to 24 hours, the iQM technology is expected to provide faster detection of errors.
Key Words: Quality control—Error detection—False rejection—Average run length—Average time for error detection.

Introduction

Quality control continues to be an issue in blood gas measurements and near-patient testing applications. Matrix effects and specimen handling problems have been identified as particularly critical for PO_2 measurements, requiring more complicated statistical QC procedures and higher numbers of control measurements to provide the desired error detection.[1–2] Electronic QC has been used as an alternate approach, but the efficacy of electronic QC can be challenged because it tests only the readout device and does not monitor the performance of the sensors.[3]

Recognizing the difficulties with QC procedures for new measurement technologies, NCCLS has provided guidelines for developing a quality management system that is aligned with the possible failure modes of the analytic system.[4] The NCCLS guidelines recommend that a manufacturer identify potential

From the *Department of Pathology and Laboratory Medicine, University of Wisconsin Medical School, Madison, Wisconsin; and †Instrumentation Laboratory, Lexington, Massachusetts.

Address correspondence and reprint requests to James O. Westgard, PhD, Department of Pathology and Laboratory Medicine, University of Wisconsin Medical School, Room 6169 MSC, 1300 University Avenue, Madison, WI 53706
e-mail: jowestgard@med.wisc.edu).

This work was supported by Instrumentation Laboratory, Lexington, MA 02421.

sources of errors, then develop methods of control for each of these error sources or failure modes. The resulting methods of control for all the frequent failure modes make up the quality management system. These methods of control may involve specific checks on instrument parameters, such as the open or closed position of a sample control valve, the zero or baseline reading of a sensor, or the time needed for a sensor to reach a steady state response. Other parameters, such as sensor variability, may be monitored directly by statistical QC.

There is a tendency by manufacturers and laboratories to avoid statistical QC because of the difficulties in training operators. Statistical QC is one of the most powerful techniques for monitoring analytical performance if done right. The NCCLS also provides guidelines for doing statistical QC correctly.[5] One important aspect is to select the right QC procedure (i.e., the control rules and number of control measurements). The NCCLS provides a general QC planning methodology that starts with a definition of the quality requirement for the test, then accounts for the observed performance of the method (precision and bias) and the expected rejection characteristics of the control rules. One source of quality requirements is the CLIA list of criteria for acceptable performance in proficiency testing events.[6]

This NCCLS document also provides guidelines for implementing QC procedures correctly. One important aspect is how often controls should be analyzed, which depends on the definition of the analytical run. The NCCLS defines an analytical run as "an interval (i.e., a period of time or series of measurements) within which the accuracy and precision of the measuring system is expected to be stable." [5] The length of the run depends on the stability of the analytical system, which is a function of the manufacturer's design. The NCCLS notes that, "between quality control observations, events may occur causing the measurement process to be susceptible to variations that are important to detect." Runs are inherently short in analytical systems when reagents and components change frequently and independently, thereby causing the measurement system to be susceptible to errors. Runs may be long when there are few changes to the system and the stability of reagents and components can be assured.

Guidelines for validation of manufacturer's recommended procedures for user quality control are being developed by the ISO.[7] These guidelines prescribe information that should be included in recommendations to users such as: (a) the type of error that the quality control procedure is intended to detect; (b) control materials that may be used; (c) control materials that may not be used; (d) recommended analyte concentrations; (e) guidelines for determining acceptability criteria (control limits); and (f) the probabilities of detecting and not detecting an inaccurate result.

The purpose of this paper is to apply current QC practice guidelines[5–7] to assess the performance capabilities of the iQM quality control technology and describe those capabilities in a manner that can be compared with the performance available from current QC practices. To accomplish this, we make use of a practical measure of performance: the average time needed to detect medically significant analytical errors. This measure is easily interpreted, the shorter the detection time, the better the QC procedure.

Methods and Materials

The GEM Premier 3000 analyzer (Instrumentation Laboratory, Lexington, MA) performs tests on whole blood samples. Different test cartridges provide different combinations of tests and accommodate different volumes of testing. Cartridges for blood gas tests typically include sodium, potassium, and ionized calcium and may also include glucose, lactate, and hematocrit. The cartridge provides all the reagent solutions and sensors needed for the selected group of tests. Installation of a new cartridge involves user verification of results on external reference materials. After verification, performance is monitored by the use of three internal or on-board reference solutions, called process control solutions (PCS). Process control solution A is analyzed every 1 to 4 hours, depending on the life cycle of the cartridge. Solution B is analyzed every 3 to 30 minutes, depending on the patient workload. Solution C is analyzed every 24 hours.

Validation methodology

Fig 1 illustrates the validation approach. The quality requirement is defined for each test and the precision

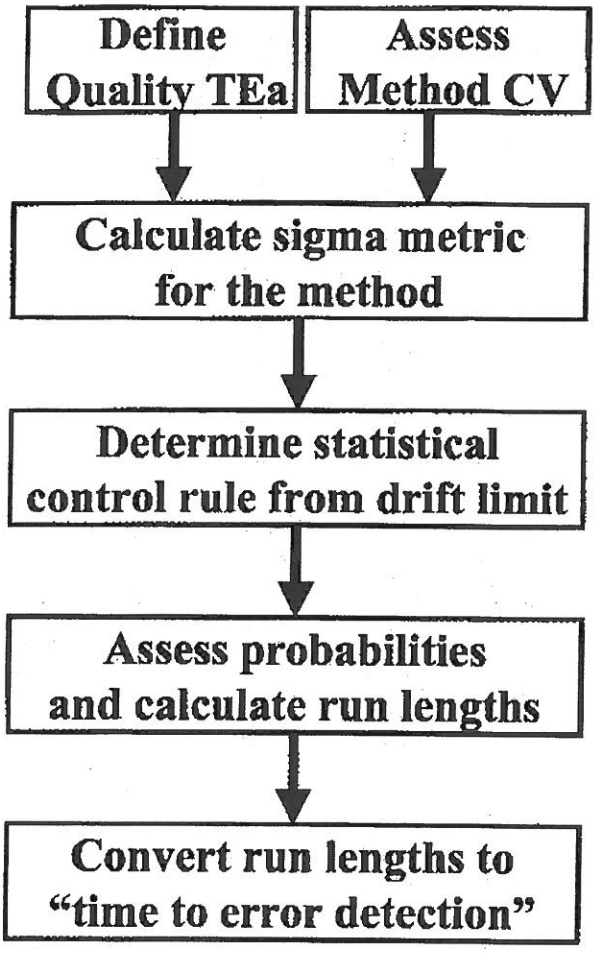

Fig 1 Quality control assessment approach.

performance of the method is determined. The sigma metric for method performance is calculated by dividing the quality requirement by the precision observed for the method.[8] The statistical control limit is determined by dividing the instrument's specified drift limit by the precision observed for the method. The difference between these two metrics, the sigma of the process and the number of standard deviations (SDs) of the control limit, is used to assess the probabilities of error detection, which are then used to calculate the average number of runs (or average number of measurements) needed to detect a problem. The average run length can be converted to units of time by multiplying by the time interval for analysis of the different reference materials.

Quality requirements

In the United States, the CLIA acceptability criteria define the minimum quality that must be achieved in laboratory testing. They include allowable total errors (TE_a) of 0.04 pH units, 5 mm Hg for PCO2, 4.0 mmol/L for sodium, 0.5 mmol/L for potassium, and 0.25 mmol/L for ionized calcium. For hematocrit, the CLIA criterion of 6% was applied to calculate an allowable total error of 1.8% packed red cell volume at a level of 30% packed red cell volume. For PO2, CLIA defines the acceptability criterion as 3 SD of the group variation observed in a proficiency testing event or external quality assessment survey, which we earlier determined to be approximately 10% based on the group variation observed in Wisconsin proficiency testing (PT) surveys.[2] For glucose in whole blood, the Clark error grid analysis for clinical accuracy employs a 20% total error criterion.[9] For lactate, an allowable total error of 0.45 mmol/L was used.

Precision performance

Data from 24 different cartridges were analyzed to provide estimates of precision that would be representative of method performance. These cartridges represented a wide variety of uses, many of which were returned from customers in the field because of QC problems. The average number of measurements per cartridge was approximately 85 for process control solution A and 725 for process control solution B.

Results

Table 1 summarizes the GEM method and QC performance characteristics. The top half of the table is for process control solution A and the bottom half for solution B. The rows of the table list the data parameters. The columns represent different tests. To illustrate the data analysis, consider the PO2 test and the performance observed for solution A. At a mean of 118.88 mm Hg, the average method SD observed on 24 cartridges was 1.647 mm Hg, which gives a coefficient of variation (CV) of 1.39% [(1.647/118.9)*100]. The CLIA allowable error of 10% translates to a quality requirement of 11.9 mm Hg (118.88*0.10), which divided by the method SD of 1.647 gives a sigma metric of 7.22 for the method (11.9/1.647). The meaning of the sigma metric is that plus or minus 7.22 SDs will fit into the quality requirement, as shown in Fig 2. When method performance is stable, the observed error distribution will fall well within the allowable limits of error.

Table 1 Method and quality control performance characteristics

	PH	PCO2	P02	Na+	K+	CA++	Glucose	Lactate	Hct
Material A	pH units	mmHg	mmHg	mmol/L	mmol/L	mmol/L	mmol/L	mmol/L	%
Mean	6.90	64.09	118.88	102.4	6.73	2.63	7.83	3.13	
SD	0.0063	1.249	1.647	0.725	0.0434	0.0340	0.206	0.0898	
CV (%)		1.95%	1.39%	0.71%	0.64%	1.29%	2.63%	2.87%	
Quality TEa	0.04	5.0	11.89	4.0	0.50	0.25	1.565	0.45	
Method sigma	6.23s	4.00s	7.22s	5.52s	11.52s	7.35s	7.61s	5.01s	
Drift limit	0.02	3.0	6.0	3.0	0.20	0.12	0.777	0.30	
Control rule	3.18s	2.40s	3.64s	4.14s	4.61s	3.53s	3.78s	3.34s	
P_{fr}	0.001	0.0164	0.001	0.001	0.011	0.001	0.001	0.001	
ARL_{fr}	1000	62.5	1000	1000	1000	1000	1000	1000	
Z for P_{ed}	1.53	–0.05	1.92	–0.27	5.26	2.17	2.19	0.02	
P_{ed}	0.937	0.480	0.973	0.394	1.000	0.985	0.986	0.508	
ARL_{ed}	1.07	2.08	1.03	2.54	1.00	1.02	1.01	1.97	
Average detection time 1.0 hour sampling	1.07 hr	2.08 hr	1.03 hr	2.54 hr	1.00 hr	1.02 hr	1.01 hr	1.97 hr	
Average detection time 4.0 hour sampling	4.27 hr	8.33 hr	4.11 hr	10.2 hr	4.00 hr	4.06 hr	4.06 hr	7.87 hr	
Material B									
Mean	7.40	32.48	176.16	144.3	3.60	1.16	(3.33)	0.0	11.0
SD	0.0020	0.408	1.368	0.895	0.0109	0.00888	0.076	0.0466	0.0865
CV		1.26%	0.78%	0.62%	0.30%	0.77%	2.28%		0.79
Quality TEa	0.04	5.0	17.62	4.0	0.50	0.10	0.667	0.45	1.80
Method sigma	20.0s	12.2s	12.9s	4.47s	45.9s	11.3s	8.78s	9.66s	20.8s
Drift limit	0.03	3.0	10.0	3.0	0.30	0.06	0.555	0.3	1.0
Control rule	15.0s	7.35s	7.31s	3.35s	27.5s	6.76s	7.32s	6.44s	11.6
P_{fr}	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
ARL_{fr}	1000	1000	1000	1000	1000	1000	1000	1000	1000
Z for P_{ed}	3.35	3.25	3.92	–0.53	16.7	2.89	–0.19	1.57	7.60
P_{ed}	1.000	1.000	1.000	0.298	1.000	1.000	0.425	0.942	1.000
ARL_{ed}	1.00	1.00	1.00	3.36	1.00	1.00	2.35	1.06	1.00
Average detection time 0.05 hour sampling	0.05 hr	0.05 hr	0.05 hr	0.17 hr	0.05 hr	0.05 hr	0.12 hr	0.05 hr	0.05 hr
Average detection time 0.5 hour sampling	0.50 hr	0.50 hr	0.50 hr	1.68 hr	0.50 hr	0.50 hr	1.18 hr	0.53 hr	0.5 hr

The QC characteristic of interest during stable performance is the probability for false rejection (i.e., the chance that a rejection signal occurs when there is no error except for the inherent imprecision of the method.) For the PO2 example, the statistical control limit is 3.64s as determined by dividing the method drift specification by the observed method SD (6/1.647). Once the control limit is known, the probability of false rejection (P_{fr}) can be determined, as shown in Fig 3. The control limits cut the distribution at plus and minus 3.64s. The areas in the tails of the distribution (i.e., the areas exceeding 3.64s) can be looked up in a table of areas under a normal or gaussian curve (available in any statistics textbook), which gives a probability of false rejection less than 0.001.

If method performance becomes unstable, then the QC procedure is supposed to detect the problem and

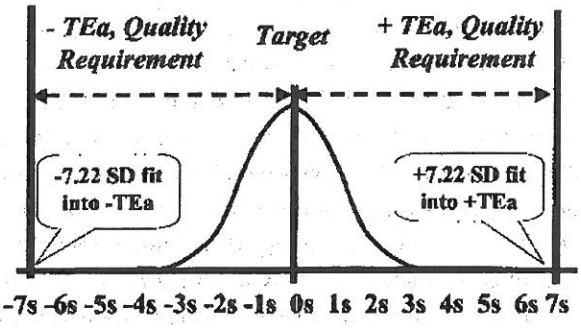

Fig 2 Illustration of the sigma metric for process performance.

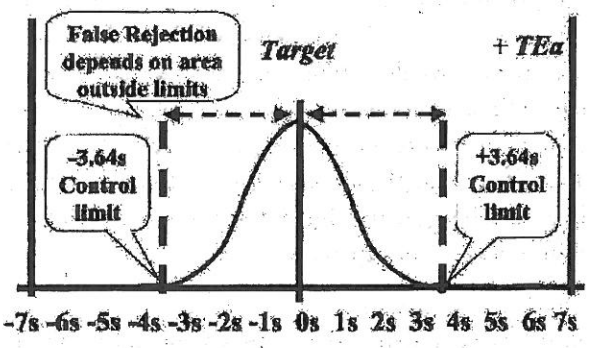

Fig 3 Interpretation of a drift limit as a statistical control limit.

alert the operator. Fig 4 illustrates the situation where the QC procedure needs to sound an alarm. It shows an error distribution that has shifted to the right until there is a 5% risk of a bad result, i.e., 5% of the area of the measurement distribution is in the tail exceeding the quality requirement, as shown by the portion to the right of the solid line. The probability of detecting a shift of this magnitude can be assessed from the area of the distribution that exceeds the control limit shown by the dashed line at 3.64s, again using a table of areas under a normal curve.

The actual numbers for this assessment are shown in Fig 5. The difference between the method sigma and the control limit is 3.58s (7.22 to 3.64). The method sigma cuts the tail of the error distribution at 1.65s above the mean (to limit the risk to 5% or 0.05 in the area in the tail). The control limit cuts the error distribution at 1.93s below the mean (3.58 to 1.65). From a table of areas under a normal curve, the area in the lower tail is found to be 0.0260, leaving 0.974 of the area above the control limit. The probability of detecting a medically important systematic error is 0.974, i.e., there is a 97.4% chance of detecting this error.

Probabilities for rejection

Ideally, the probability for error detection (P_{ed}) should be 1.00 and the probability for false rejection (P_{fr}) should be 0.00. In practice, values of 0.90 and 0.05 are often used for QC design. For process control solution A, P_{ed} is more than 0.90 for pH, PO2, potassium, calcium, and glucose (Clark whole blood criterion). P_{ed} is 0.480 for PCO2, 0.394 for sodium, and 0.508 for lactate. P_{fr} is generally 0.001 or less, except for PCO2 where it is expected to be 0.0164. For process control solution B, P_{ed} is more than 0.90 for pH, PCO2, PO2, potassium, calcium, lactate, and

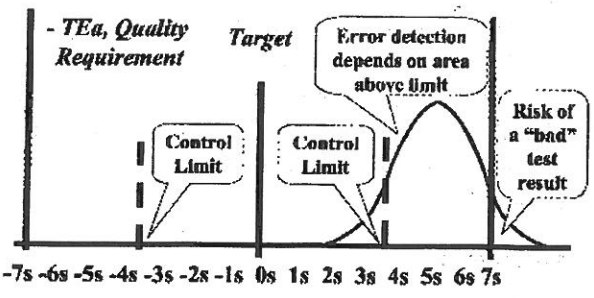

Fig 4 Determination of the probability of error detection.

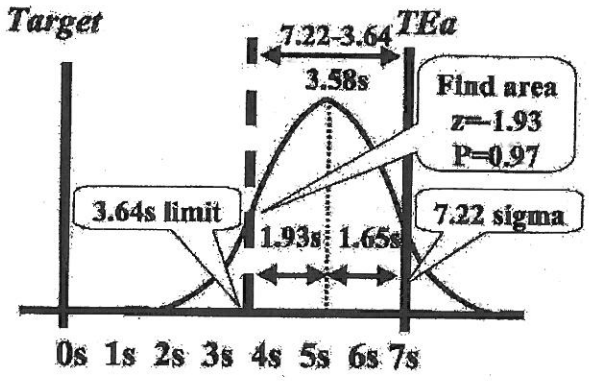

Fig 5 Illustration of the calculation of the probability of error detection.

hematocrit. For sodium, P_{ed} is 0.298 and 0.425 for glucose (Clark whole blood criterion). P_{fr} is less than 0.001 for all tests.

Average run lengths

These probabilities for rejection can be converted to average run lengths (ARL) by taking their reciprocals (ARL = 1/P).[10] An ARL is the average number of runs before a rejection occurs. When method performance is stable, the value of P_{fr} can be used to calculate the average run length for false rejection (ARL_{fr}), which should be long, e.g., if P_{fr} is 0.001 then ARL_{fr} is 1000, which means a false rejection should be observed only once in every 1000 measurements of reference material A. When method performance is unstable, the value of P_{ed} can be used to calculate the average run length for error detection (ARL_{ed}), which should be short, e.g., if P_{ed} is 0.97 (as for PO2 on process control solution A), the problem will usually be detected by a single measurement (ARL_{ed} is 1/0.97 or 1.03).

Average time to detection

In Table 1, the bottom lines for solutions A and B show the average time to detection in hours. The ARLs can be converted to units of time by multiplying by the sampling interval of the process control solutions. Given that solution A is analyzed every 1 to 4 hours (depending on life cycle of the cartridge), a problem with pH, PO2, potassium, calcium, and glucose (Clark whole blood criterion) would be detected within 1 to 4 hours. For PCO2, the average time to detection would be 2.08 to 8.3 hours; for sodium, 2.54 to 10.16 hours; for lactate 1.97 to 7.88 hours.

Given that solution B is analyzed every 0.05 hours (3 minutes) during heavy patient workload and every 0.5 hours (30 minutes) during minimum workload (or standby), problems will be detected much faster, usually within 0.05 to 0.5 hours (3 to 30 minutes) for pH, pCO2, pO2, potassium, calcium, lactate, and hematocrit. For sodium, the average time to detection will be 0.17 hours (10.1 minutes) during heavy patient workload and 1.7 hours (101 minutes) during minimal workload (or standby); for glucose (Clark whole blood criterion), 0.12 to 1.2 hours (7.1 to 71 minutes).

Discussion

The evaluation of QC performance in terms of the average time to detection is consistent with Parvin's recommendations to use measures of QC performance that are appropriate for system operation.[11] In the GEM application, process control solutions are analyzed frequently, whether patient specimens are being analyzed or not, therefore it is difficult to define an analytical run in traditional terms related to the size of a batch or numbers of patient specimens. Time is the dimension that is relevant and ARLs are the appropriate performance characteristics for interpreting QC performance in practical terms, the time it takes to detect analytical problems. The advantage of describing QC performance in units of time is that it is easy to compare performance to what can be expected with current or conventional QC practices.

For blood gas measurements, compliance with the minimum QC requirements can be equated to the analysis of one control every 8 hours. For many other chemistry measurements, compliance requires analyzing two controls every 24 hours. This means that it currently takes laboratories a minimum of 8 hours to detect analytical errors in blood gas analysis and up to 24 hours for many chemistry tests, assuming the QC procedures achieve ideal error detection (which is highly unlikely[12]). In comparison, the iQM technology can be expected to provide much earlier detection of analytical errors, within 3 to 30 minutes for pH, PCO2, PO2, potassium, calcium, lactate, and hematocrit and within 2 hours for sodium and glucose. These shorter times for detection of errors are the result of the frequent analysis of process control solution B and the reduced variability in the gas parameters, owing to the elimination of operator effects on sample handling.[12]

The iQM performance characteristics have been assessed relative to the CLIA criteria for acceptability and the Clark error grid criterion for whole blood glucose, but similar performance would be expected for the quality criteria established in other countries. For example, the new German quality regulations[13] for near-patient testing applications specify allowable total errors of 0.06 for pH, 12% for PCO2, 12% for PO2 more than 100 mm Hg or 12 mm Hg when less than 100 mm Hg, 5% for sodium, 9.1% for potassium, 15% for ionized calcium less than 1 mmol/L and 0.15 mmol/L when more than 1 mmol/L, 15% for glucose more than 60 mg/dL or 9 mg/dL when less than 60 mg/dL, and 9% for hematocrit. These requirements are nearly the same, or slightly less demanding, except for glucose, which is somewhat more demanding than the clinical requirement from the Clark error grid analysis (15% versus 20%). The average times to detection will be at least as good, if not better, for all tests except glucose, which would also have a significantly longer detection time if the CLIA criterion for serum glucose were used.

Another important advancement with the iQM technology is the failure pattern recognition software[14] that can identify certain causes and initiate appropriate corrective actions, such as flushing the system with process control solution C to remove micro-clots that may have formed on the electrode sensors, then automatically recalibrating and revalidating sensor performance. Additional details of the failure mode effect analysis and pattern recognition algorithms will be presented in another report.

Conclusion

Better error detection together with error pattern recognition yields corrective actions that optimize test performance for near-patient testing and decentralized laboratory operations. These improvements in QC technology simplify operations for analysts both trained and untrained, and provide greater assurance that quality test results will be produced in any setting.

References

1. Burnett RW. Matrix effects in blood gas proficiency testing. *Arch Pathol Lab Med* 1993;117:365–368.

2. Olafsdotter E, Westgard JO, Ehrmeyer SS, Fallon KD. Matrix effects and the performance and selection of quality-control procedures to monitor PO2 measurements. *Clin Chem* 1996;42:392–396.
3. Westgard JO. Electronic quality control, the total testing process, and the total quality control system. *Clin Chim Acta* 2001;307:45–48.
4. National Committee for Clinical Laboratory Standards. Quality management for unit-use testing. Document EP-18P. Wayne, PA: NCCLS; 1999.
5. National Committee for Clinical Laboratory Standards. Statistical quality control for quantitative measurements: principles and definitions. Document C24-A2. Wayne, PA: NCCLS; 1999
6. United States Department of Health and Human Services. Medicare, Medicaid, and CLIA programs: regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA). Final rule. *Fed Regist* 1992;57:7002–7186.
7. International Organization for Standardization. In vitro diagnostic medical devices – Manufacturer validation of recommended procedures for user quality control. ISO/DIS 15198. Geneva, Switzerland: ISO; 2002.
8. Westgard JO. Six Sigma Quality Design and Control: Desirable precision and requisite QC for laboratory measurement processes. Madison WI: Westgard QC, Inc; 2001:35–43.
9. Clark WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy on systems for self-monitoring of blood glucose. *Diabetes Care* 1987;10:622–628.
10. Duncan AJ. Quality control and industrial statistics, 4th ed. Homewood, IL: Richard D. Irwin, Inc.; 1974:374.
11. Parvin CA, Gronowski AM. Effect of analytical run length on quality control (QC) performance and the QC planning process. *Clin Chem* 1997;43:2149–2154.
12. Fallon KD, Mansouri S. IQM-The next generation for instrument quality control. [Abstract]. *Clin Chem* 2002;48:A154.
13. Richtlinie der Bundesarztekammer zur qualitatssicherung quantitativer laboratoriumsmedizinischer Untersuchungen: Beschluss des Vorstandes der Bundesarztekammer vom 24 August 2001. *Deutsches Arzteblatt* 2001;98:A2747-A2759.
14. Mansouri S, Fallon KD, Eames P. Intelligent quality management in the GEM Premier 3000 critical care analyzer. [Abstract]. *Clin Chem* 2002;48:A6-A7.

Instrumentation
Laboratory

FOR MORE INFORMATION, VISIT
www.ilus.com OR CALL **800-955-9525**

WESTiQM.CS 3/30