Verification of Assayed Blood Gas Quality Control Ranges

Jeffrey M Haynes and Robert G Fishwick

BACKGROUND: Blood gas quality control (QC) is an essential and mandatory part of a laboratory's quality plan. The acceptable QC range should be 2 SD from the mean value. The use of assayed QC material does not negate the responsibility of the laboratory to calculate the mean and 2 SD ranges of QC measurements for verification. Verifying assayed QC ranges is a Clinical Laboratory Improvement Amendment (CLIA) requirement. This study shows the results of assayed QC mean and 2 SD range verification from a blood gas analyzer. METHODS: QC data from a blood gas analyzer were compared to manufacturer-provided mean and ranges. The percent difference between the measured mean and the manufacturer-provided mean was calculated to assess agreement. The measured SD was used to determine how many SD the manufacturer-provided ranges were from the measured mean. RESULTS: The largest difference in mean values was 2.27% > the manufacturer-provided mean. Forty-eight percent of all mean value comparisons showed a difference of 0%, and 71% were < 1%. The manufacturer-provided ranges were considerably wider than the measured 2 SD range, ranging from 2.4–75 SD. None of the manufacturer-provided ranges were deemed acceptable for clinical use. CONCLUSIONS: Our analysis validates the CLIA mandate and American Association for Respiratory Care Clinical Practice Guideline recommendation that laboratories must verify manufacturer-provided QC means and ranges and adjust QC means and ranges to match the performance of their blood gas analyzer. *Key words: quality control; laboratory errors; quality events; blood gas analysis; patient safety; quality assurance.*
[Respir Care 2022;67(4):428–432. © 2022 Daedalus Enterprises]

Introduction

Quality control (QC) is an essential and mandatory part of a laboratory's quality plan. However, there is significant variability in QC practices even among some of the most prestigious hospitals.[1] Many laboratories that perform blood gas analysis use assayed materials to perform QC testing. Assayed QC materials have manufacturer-provided ranges, whereas unassayed QC materials require the user to calculate the mean and ranges from a series of measurements. However, the use of assayed QC materials does not negate the responsibility of the laboratory to calculate the mean and ranges of QC measurements. The laboratory must make these calculations to verify that the manufacturer-provided QC mean and ranges are appropriate for their analyzer. Verifying assayed QC ranges is a Clinical Laboratory Improvement Amendment (CLIA) requirement (§493.1256 standard: control procedures [d][10][ii]): "the laboratory may use the stated value of a commercially assayed control material provided the stated value is for the methodology and instrumentation employed by the laboratory and is verified by the laboratory" (https://ecfr.federalregister.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-K/subject-group-ECFRc96daead380f6ed/section-493.1256#p-493.1256(d)(10)(ii). *Accessed August 2, 2021*). In addition, the American Association for Respiratory Care (AARC) Clinical Practice Guideline for blood gas analysis and hemoximetry states that "the laboratory director or designee should determine the acceptable range for QC results, based on statistically relevant or medical-needs criteria."[2]

If the manufacturer-provided QC mean values and ranges are not representative of the true performance of the

Mr Haynes is affiliated with Pulmonary Function Laboratory, Elliot Health System, Manchester, New Hampshire. Mr Fishwick is affiliated with Respiratory Care Department, Elliot Health System, Manchester, New Hampshire.

Mr Haynes discloses a relationship with Morgan Scientific. Mr Fishwick has disclosed no conflicts of interest.

Correspondence: Jeffrey M Haynes RRT RPFT FAARC, Pulmonary Function Laboratory, Elliot Health System, 185 Queen City Avenue, Manchester, New Hampshire 03101. E-mail: jhaynes@elliot-hs.org.

DOI: 10.4187/respcare.09342

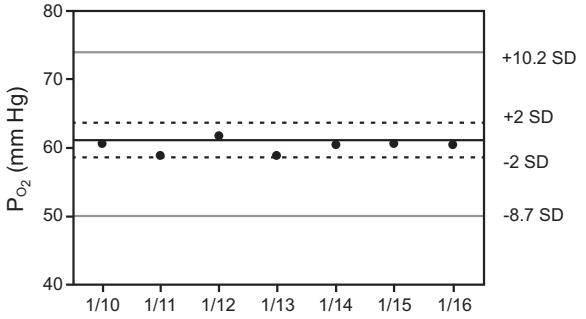

Fig. 1. Levey-Jennings plot of level 1 P_{O_2} quality control data. The center line represents the measured mean value. The 2 SD dashed lines were calculated from serial testing, and the −8.7 and +10.2 SD (gray lines) are based on manufacturer-provided ranges.

analyzer, QC testing may not alert the laboratory's clinicians that a problem exists. Such a situation may affect the accuracy of patient data and potentially affect patient care. Unfortunately, some laboratories that perform blood gases analysis use manufacturer-provided QC mean values and ranges without performing verification. Whereas there is a paucity of data regarding the frequency of blood gas laboratories using unverified QC ranges, based on informal conversations we are confident that this is not an uncommon practice. This study shows the results of assayed QC mean and range verification from a blood gas analyzer and underscores the importance of verifying and adjusting QC mean values and ranges to match the performance of the analyzer.

Methods

QC data from a blood gas analyzer (Roche cobas b 221, Roche Diagnostics, Mannheim, Germany) were compared to manufacturer-provided mean values and ranges. Data points identified as random errors according to Westgard's rules (http://www.westgard.com. *Accessed August 2, 2021*) were excluded from analysis. The manufacturer-provided QC material (AUTO-TROL PLUS B, Roche Diagnostics) lot numbers were as follows: level 1, 21480768; level 2, 21480863; level 3, 21480962. Measured QC mean values and SD were calculated by the Roche cobas b 221 software. Figures 1 and 3 were generated using statistical software (Prism 4, GraphPad Software, San Diego, California). Figure 2 was generated using Microsoft Excel (Microsoft, Redmond, Washington).

The percent difference between the measured mean and the manufacturer-provided mean was calculated to assess agreement. The measured SD was used to determine how many SD the manufacturer-provided ranges were from the measured mean.

Results

The agreement between the measured and manufacturer-provided QC mean values and ranges are listed in Table 1.

QUICK LOOK

Current knowledge

Some laboratories that perform blood gases analysis use assayed materials to perform quality control (QC) testing. Assayed QC materials have manufacturer-provided mean values and ranges, whereas unassayed QC materials require the user to calculate the mean values and ranges from a series of measurements. Clinical Laboratory Improvement Amendment (CLIA) and the American Association for Respiratory Care (AARC) recommend that blood gas laboratories verify that manufacturer-provided QC mean values and ranges are appropriate for their blood gas analyzer.

What this paper contributes to our knowledge

Our analysis validates the CLIA mandate and AARC Clinical Practice Guideline recommendation that laboratories must verify manufacturer-provided QC means and ranges and adjust QC means and ranges to match the performance of their blood gas analyzer. Verification of assayed QC mean values and ranges is necessary to ensure quality testing and patient safety.

The largest difference in mean values was observed in the level 2 P_{CO_2}, where the measured mean was 2.27% (1 mm Hg) > the manufacturer-provided mean. Forty-eight percent of all mean value comparisons showed a difference of 0%, and 71% were < 1%. The manufacturer-provided ranges were considerably wider than the measured 2 SD range, ranging from 2.4 SD (low range of level 1 P_{CO_2}) to 75 SD (levels 2 and 3 methemoglobin). An example of the assayed ranges in comparison to the measured 2 SD range is shown in a Levey-Jennings plot for level 1 P_{O_2} in Figure 1, where the low range is 8.7 and the high range is 10.2 SD from the measured mean. None of the manufacturer-provided ranges was deemed acceptable for clinical use.

Discussion

An analysis of an electronic error reporting system from 30 health care organizations showed that laboratory errors accounted for 14.1% of all quality events.[3] Laboratory errors may occur during the pre-analytical, analytical, and post-analytical phases. Whereas pre-analytical errors account for most laboratory events,[3] a malfunctioning analyzer is an obvious source of error and is the primary focus of a QC program.

QC is an essential and mandatory part of a laboratory's quality plan. The acceptable QC range should be 2 SD from the mean value. As mentioned earlier, manufacturers may provide "assayed" expected mean values and ranges, but

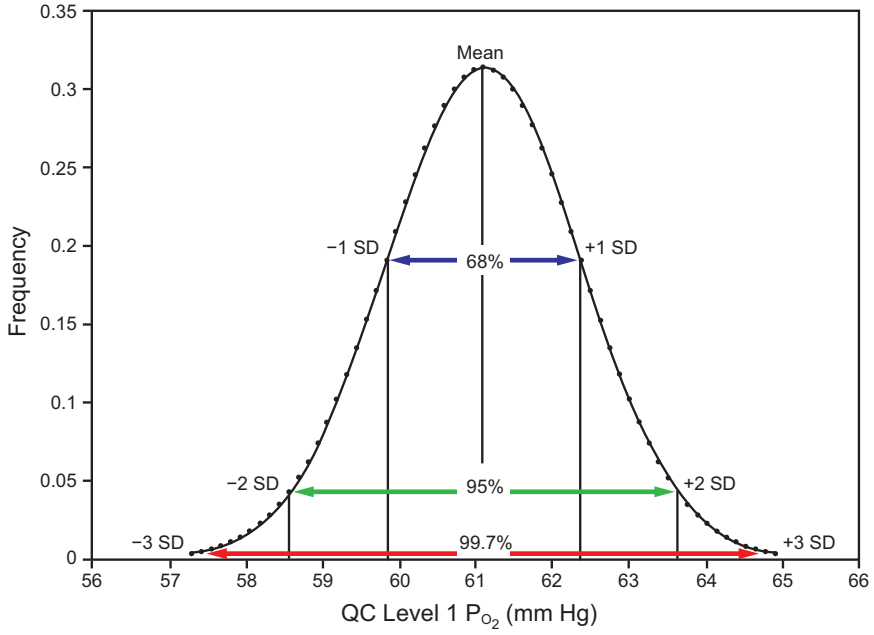

Fig. 2. Distribution or bell curve of level 1 P_{O_2} quality control (QC) data. The mean value is at the center of the curve. According to the empirical rule, if the analyzer is operating normally and the QC material is unadulterated, 68%, 95%, and 99.7% of QC measurements should fall between 1 SD (blue arrow), 2 SD (green arrow), and 3 SD (red arrow), respectively.

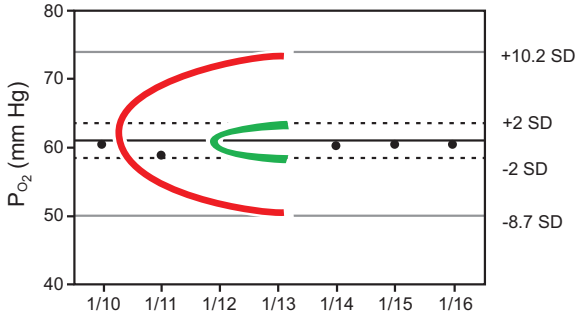

Fig. 3. Levey-Jennings plot of level 1 P_{O_2} quality control data with cartoon bell curves inserted to depict the differences in the bell curve size of manufacturer-provided (red curve) and measured quality control ranges (green curve). The solid line represents the measured mean value.

these must be verified as accurate reflections of the analyzer's performance. To verify assayed QC mean values and ranges, a minimum of 20 QC measurements (same level from the same lot) should be collected from multiple successive days. Any obvious and infrequent outliers or "random errors" (eg, expected P_{O_2} value $\sim$ 50 mm Hg, measured value 100 mm Hg) should be excluded from analysis. If random errors are occurring frequently, the analyzer should be inspected for malfunctions. Using $<$ 20 samples to calculate the mean and SD may not provide an accurate assessment of the analyzer's performance, and using a very large sample (eg, $>$ 100 values) unnecessarily adds time to the process. Computer software can quickly and accurately calculate mean and SD values; however, laboratory personnel should still understand how these values are derived and what they represent.

The mean or average (denoted by the symbol $\bar{x}$ or μ) is derived by adding the sum of all recorded values and dividing that value by the number of measurements as shown in the equation below:

$$\bar{x} = \Sigma^x / n$$

Where
$\bar{x}$ = mean
Σ^x = sum of recorded values
n = number of values.

SD is a measure of the variability around the mean. Two blood gas analyzers could produce identical mean values for a QC measure (eg, level 1 P_{O_2}); however, the analyzer with the smaller SD has less variability and more precision. Less variability and more precision impact the confidence one should have in patient data because repeating analysis on a blood specimen will not yield exactly the same results. It stands to reason that the data variability present in QC testing may also apply to results reported from patient sampling. For example, consider two identical blood gas analyzers, both report similar mean values for level 1 P_{O_2}, 60 mm Hg; however, the observed range of measured values for analyzer #1 is 55–65 mm Hg, and the observed range of measured values is 45–75 mm Hg for analyzer #2. Clinicians should have less confidence that a patient's reported P_{O_2} of 65 mm Hg is truly $>$ 60 mm Hg if analyzer

Table 1. Agreement Between Measured and Manufacturer-Provided QC Mean and Ranges

	Measured				Assayed Means and Ranges				Assayed SD Offsets*	
	Samples	Mean	SD	Range	Mean	Δ	% Δ	Range	SD < Mean	SD > Mean
Level 1										
pH	62	7.21	0.005	7.20–7.22	7.21	0	0	7.18–7.24	6	6
P_{CO_2}	62	65.7	1.12	63.5–67.9	67	−1.3	1.98	63.0–71.0	2.4	4.7
P_{O_2}	62	61.1	1.27	58.6–63.6	62	−0.9	1.47	50.0–74.0	8.7	10.2
THb, g/dL	62	7.1	0.05	7.0–7.2	7.1	0	0	6.40–7.80	14	14
O_2Hb, %	62	48.9	0.25	48.4–49.4	49	−0.1	0.20	45.0–53.0	15.6	16.4
COHb, %	62	22.1	0.11	21.9–22.3	22.1	0	0	19.6–24.6	22.7	22.7
MetHb, %	62	11.6	0.06	11.5–11.7	11.6	0	0	10.1–13.1	25	25
Level 2										
pH	60	7.44	0.005	7.43–7.45	7.43	0.01	−0.13	7.400–7.460	8	4
P_{CO_2}	60	44.0	0.74	42.5–45.5	43	1	−2.27	40.0–46.0	5.4	2.7
P_{O_2}	60	102.2	1.42	99.4–105.0	101	1.2	−1.17	89.0–113.0	9.3	7.6
THb, g/dL	60	11.9	0.08	11.7–12.1	11.9	0	0	10.9–12.9	12.5	12.5
O_2Hb, %	60	75.6	0.11	75.4–75.8	75.5	0.1	−0.13	71.5–79.5	37.3	35.5
COHb, %	60	10.5	0.05	10.4–10.6	10.6	–0.1	0.95	8.1–13.1	48	52
MetHb, %	60	5.7	0.02	5.66–5.74	5.7	0	0	4.2–7.2	75	75
Level 3										
pH	61	7.58	0.004	7.57–7.59	7.58	0	0	7.550–7.610	7.5	7.5
P_{CO_2}	61	24.6	0.39	23.8–25.4	25	−0.40	1.63	22.0–28.0	6.7	8.7
P_{O_2}	61	153.7	1.35	151.0–156.4	152	1.70	−1.11	140.0–164.0	10.1	7.6
THb, g/dL	61	17.9	0.1	17.7–18.1	17.8	0.10	−0.56	16.40–19.20	15	13
O_2Hb, %	61	93.8	0.11	93.6–94.0	93.8	0	0	89.8–97.8	36.4	36.4
COHb, %	61	2.7	0.05	2.6–2.8	2.7	0	0	0.2–5.2	50	50
MetHb, %	61	1.6	0.02	1.56–1.64	1.6	0	0	0.1–3.1	75	75

* The number of SD the manufacturer-provided ranges is offset from the measured mean.
THb = hemoglobin
O_2Hb = oxyhemoglobin
COHb = carboxyhemoglobin
MetHb = methemoglobin

#2 was used as opposed to analyzer #1. The precision of a blood gas analyzer can impact clinical decision-making especially when small differences in blood gas values change clinical classifications such as qualifying for long-term O_2 therapy[4] or determining ARDS severity.[5]

The lower-case Greek letter σ is used to denote SD. SD calculation starts by summing the difference between each measured value and the mean. For example, if the QC mean is 60, a value of 55 has a difference score of −5, and a value of 65 has a difference score of +5. Because the mean is by definition at the center of all values, the sum of all difference scores will always be zero. To calculate SD, the difference scores will need to be summed; and because some are negative and some are positive, they are squared to transform all difference scores to positive numbers. Now the SD can be calculated by the following equation:

$$\sigma = \sqrt{\Sigma(x_i - \overline{x})^2/n - 1}$$

where
σ = SD
Σ = sum
x_i = individual values
$\overline{x}$ = mean of all values
n = number of values.

The SD equation subtracts 1 from the number of measurements (one degree of freedom) to calculate a slightly larger SD. Using a larger SD makes the sample variance more generalizable to the entire population. For example, if you were calculating the mean and SD of resting S_{pO_2} from a sample of 1,000 subjects in a city with a population 100,000, using the number of samples without subtracting 1 might indicate less variance around the mean (ie, smaller SD) than if you sampled all 100,000 residents. This process (n-1) is intended to produce a less biased estimate of variance. Because the SD of blood gas QC is calculated from a sample (eg, 20 QC measurements) and not the total number of measurements projected for the QC lot (eg, 365 measurements for a lot that expires in 1 year), a slightly higher SD is desirable to avoid underestimating the true variance that might be higher if calculated from all 365 measurements in a 1-y period. Using a SD smaller than

the true variance for blood gas QC will result in more false "out of control" conditions.

A 2 SD range is used for blood gas QC because the empirical rule states that if the blood gas analyzer is operating normally, and the QC material is unadulterated, 95% of recorded QC values should be within 2 SD of the mean.[6] Therefore, a QC value 2 SD from the mean should only occur once in 20 measurements (5%). If a recorded QC measurement is between 2–3 SD from the mean, a "warning" condition exists because this is an unusual value (1/20) based on past performance. The empirical rule also states that 99.7% of QC results should be within 3 SD of the mean, so a QC value > 3 SD from the mean should never occur if the analyzer is operating normally and the QC material is unadulterated. Figure 2 shows the distribution or "bell" curve for level 1 P_{O_2} data. The mean value is at the center of the distribution curve; 68%, 95%, and 99.7% of QC measurements should fall between 1 SD, 2 SD, and 3 SD, respectively. A Levey-Jennings plot charts QC data against the mean and SD as a function of time; whereas the distribution curve is not displayed, it is essentially flipped 90°. Figure 3 shows a cartoon drawing of the differences in the size of the distribution curves for measured versus assayed ranges on a Levey-Jennings plot for level 1 P_{O_2}.

The software of modern blood gas analyzers will alert the laboratory personnel if a QC measurement is out of range, and many devices will lock out the analyzer from performing clinical testing until the QC passes. It is, therefore, important that QC ranges are an accurate reflection of the blood gas analyzer's true performance.

Our analysis shows the importance of verifying manufacturer-provided QC ranges. Whereas the manufacturer-provided means were essentially identical to our measured mean values, with the exception of level 1 low P_{CO_2} and level 2 high P_{CO_2}, the manufacturer-provided ranges were markedly outside of 2 SD. This level of discordance could result in undetected analyzer malfunction and potentially impact clinical decision-making.

Conclusions

Our analysis validates the CLIA mandate and AARC Clinical Practice Guideline recommendation[2] that laboratories must verify manufacturer-provided QC mean values and ranges and adjust the QC mean value and ranges to match the performance of their blood gas analyzer.

REFERENCES

1. Rosenbaum MW, Flood JG, Melanson SE, Baumann NA, Marzinke MA, Rai AJ, et al. Quality control practices for chemistry and immunochemistry in a cohort of 21 large academic medical centers. Am J Clin Pathol 2018;150(2):96-104.
2. Davis MD, Walsh BK, Sittig SE, Restrepo RD. AARC clinical practice guideline: blood gas analysis and hemoximetry: 2013. Respir Care 2013;58(10):1694-1703.
3. Snydman LK, Harubin B, Kumar S, Chen J, Lopez RE, Salem DN. Voluntary electronic reporting of laboratory errors: an analysis of 37,532 laboratory event reports from 30 health care organizations. Am J Med Qual 2012;27(2):147-153.
4. Branson RD. Oxygen therapy in COPD. Respir Care 2018;63(6):734-748.
5. The ARDS Definition Task Force. Acute respiratory distress syndrome: the Berlin definition. JAMA 2012;307(23):2526-2533.
6. Pilbeam SP. Fundamentals of biostatistics. Respir Care 1980;25(1 Pt 2):154-172.