Commutability Limitations Influence Quality Control Results with Different Reagent Lots

W. Greg Miller,[1*] Aybala Erek,[1,2] Tina D. Cunningham,[3] Olajumoke Oladipo,[4,5] Mitchell G. Scott,[4] and Robert E. Johnson[3]

BACKGROUND: Good laboratory practice includes verifying that each new lot of reagents is suitable for use before it is put into service. Noncommutability of quality control (QC) samples with clinical patient samples may preclude their use to verify consistency of results for patient samples between different reagent lots.

METHODS: Patient sample results and QC data were obtained from reagent lot change verification records for 18 QC materials, 661 reagent lot changes, 1483 reagent lot change–QC events, 82 analytes, and 7 instrument platforms. The significance of between-lot differences in the results for QC samples compared with those for patient samples was assessed by a modified 2-sample t test adjusted for heterogeneity of QC and patient sample measurement variances.

RESULTS: Overall, 40.9% of reagent lot change–QC events had a significant difference ($P < 0.05$) between results for QC samples compared with results for patient samples between 2 reagent lots. For QC results with differences <1.0 SD interval (83.1% of total), 37.7% were significantly different from the changes observed for patient samples. For QC results with differences ≥1.0 SD interval (16.9% of total), 57.0% were significantly different from those for patient samples.

CONCLUSIONS: Occurrence of noncommutable results for QC materials was frequent enough that the QC results could not be used to verify consistency of results for patient samples when changing lots of reagents.

© 2010 American Association for Clinical Chemistry

Good laboratory practice includes verifying that each new lot of reagent is suitable for use before it is put into service to ensure that results for patient samples are consistent when changing lots of reagents in a measurement procedure. A typical procedure is to calibrate a new lot of reagent and measure results for a group of patient samples by using both the old and new lots of reagent. Results are examined to verify consistency between results by using both lots of reagents. Ideally, a set of patient samples is used that verifies consistency over the complete measuring interval. In practice, it may be difficult to obtain patient samples that include the complete measuring interval, and a smaller interval is taken to represent the consistency of results over the complete interval. Acceptance criteria are set on the basis of what lot-to-lot reproducibility is required for the measurement to meet its intended clinical purpose and whether the observed lot-to-lot differences fit within the expectations for a measurement procedure's analytic performance capabilities.

Quality control (QC)[6] materials, as well as proficiency testing (PT) materials that are prepared similarly to QC materials, have been reported to be noncommutable with clinical patient samples when comparing results between different measurement procedures *(1–10)*. Noncommutable means the numeric relationship between results from 2 or more measurement procedures is different for QC materials than it is for clinical patient samples *(11)*. Noncommutability of QC (and PT) materials is caused by an alteration in the matrix, introduced by preparation of the QC (or PT) material, which makes it different from the matrix found in clinical patient samples *(4, 5, 11–14)*. Furthermore, the matrix of different lots of QC (or PT) materials may not be completely reproducible, which causes the magnitude of a matrix-related bias relative to clinical patient samples in a given measurement procedure to be different for different lots of QC (or PT) materials. In this report, the term "matrix-related bias" refers to an effect caused by manipulation of the sample matrix during preparation of a QC material that is different from (or in addition to) the naturally occurring differences in matrix among clinical

[1] Department of Pathology, Virginia Commonwealth University, Richmond, VA; [2] Vakif Gureba Training and Research Hospital, Istanbul, Turkey (current affiliation); [3] Department of Biostatistics, Virginia Commonwealth University, Richmond, VA; [4] Division of Laboratory and Genomic Medicine, Washington University, St. Louis, MO; [5] Department of Pathology/Laboratory Medicine, Staten Island University Hospital, Staten Island, NY (current affiliation).

* Address correspondence to this author at: PO Box 980286, Richmond, VA, 23298-0286. Fax 804-828-0353; e-mail gmiller@vcu.edu.

[6] Nonstandard abbreviations: QC, quality control; PT, proficiency testing; SDI, SD interval.

patient samples. Commutability limitations explain why QC materials have method-specific values assigned and why most PT evaluations are limited to method peer-group comparisons (15).

It would be convenient if results for QC materials were used to verify the consistency of results for patient samples when changing reagent lots. However, the same commutability limitations that affect among-method results for QC materials may also cause a different matrix-related bias to occur between a given lot of QC materials and 2 different lots of reagents (16). Reagents are engineered and manufactured to be suitable for testing patient samples. When analyzed by a particular method, QC samples, owing to their altered matrix, may show a matrix-related measurement bias, the magnitude of which varies from one lot of reagents to another. In practice, the magnitude of such matrix-related bias can be different for each individual QC material and reagent lot combination, and the magnitude of the difference is not predictable. A practical issue for laboratories is deciding whether to change QC target values due to the influence of differences in matrix-related biases when changing reagent lots. Failure to correct target values for matrix-related changes in numeric values can cause erroneous conclusions about method performance on the basis of results for QC materials (17).

We investigated the relationship between patient sample results and QC material results for 18 QC materials, 661 reagent lot changes, 1483 reagent lot change–QC events, 82 analytes, and 7 instrument platforms.

Materials and Methods

Patient sample comparison data and QC data were obtained from reagent lot change verification records from May 6, 2002, to August 30, 2005, for Ortho Vitros 950; from November 11, 2005, to February 3, 2009, for Siemens Advia 1650 and Centaur at Virginia Commonwealth University; and from March 29, 2009, to October 7, 2009, for Roche Modular, Abbott AxSym, Siemens BNII nephelometer, and Siemens Dimension RXL at Washington University. Supplemental Table 1 (in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol57/issue1) shows the 7 analyzers, 82 analytes, 1483 reagent lot change–QC events, 18 QC materials, and the number of QC concentrations and patient samples used in each reagent lot change evaluation. QC results were obtained for the 10 measurements immediately before (old lot) and after (new lot) a reagent lot change date.

The consistency of results for patient samples was typically verified by measuring an analyte in 5–10 patient samples by using both a current (old) and new

reagent lot for an analyte. In some cases, a smaller or larger number of patient samples was used, as indicated in online Supplemental Table 1. The concentrations or activities in the patient samples always spanned the interval covered by the QC concentrations and in many cases included the complete measuring interval of each assay. For a given new reagent lot verification, each patient sample was measured without replication by using the new reagent lot within 2–4 h of the old reagent lot measurement. For unstable analytes, such as CO_2 or ammonia, special precautions were used to make measurements within 15 min of reagent lot changes. Both the old and new reagent lots were calibrated according to the method manufacturer's instructions.

STATISTICAL PROCEDURES

A worked example of the statistical analysis is provided in the online Data Supplement. It was assumed the patient samples did not have matrix-related biases that were influenced by different reagent lots. The null hypothesis was the difference before and after a reagent lot change observed for a QC sample is the same as the difference observed for a set of patient samples. The difference for the patient samples between 2 reagent lots was used as the control to determine if there was a significant difference observed for the QC materials. All small differences in patient sample results seen between any given 2 reagent lots fell within each laboratory's acceptance criteria for consistency between lot changes. Consequently, any difference observed for the patient samples, for example, because of calibration effects, was included in the statistical analysis. The set of 1483 reagent lot change–QC events were evaluated as individual events, and the observations were summarized by counting how many had significant differences.

Analyte measured values were log-transformed before analysis to give an approximately constant SD over the interval of concentrations measured. We used a 2-sample t test, modified as described below to adjust for nonhomogeneity of QC and patient sample measurement variances, to test for statistical significance. The QC samples were measured independently by using the old and new reagent lots with assumed equal variance. Results for patient samples by using the old and new reagent lots were also measured independently and were paired for each patient; thus, paired differences were computed. The computed difference in means for the results from patient samples was assumed to have variances homogeneous across patients but different from the variance of the difference in means for the QC results. A Satterthwaite adjustment to the degrees of freedom, which compensates for this difference in variance, was used in the t test. This adjustment is a function of the sample sizes (always n =

Table 1. Number of significant and nonsignificant differences between results for QC samples compared with results for patient samples after changes in reagent lots for analytes grouped by category of analytical measurement procedure and instrument platform.

	All data			QC SDI <1.0[a]		QC SDI ≥1.0[b]		
	n = 1483 reagent lot change–QC sample events			n = 1232 (83.1%)		n = 251 (16.9%)		
	Reagent lot change–QC events in category	Nonsignificant[c] n = 876 (59.1%)	Significant n = 607 (40.9%)	Nonsignificant n = 768 (62.3%)	Significant n = 464 (37.7%)	Nonsignificant n = 108 (43.0%)	Significant n = 143 (57.0%)	Percent ≥1.0 SDI
General chemistry Advia	294	181[d] (61.6)	113 (38.4)	150 (63.3)	87 (36.7)	31 (54.4)	26 (45.6)	19.4
General chemistry Dimension	6	1 (16.7)	5 (83.3)	1 (16.7)	5 (83.3)	0 (0.0)	0 (0.0)	0.0
General chemistry Modular	112	87 (77.7)	25 (22.3)	83 (76.8)	25 (23.2)	4 (100.0)	0 (0.0)	3.6
General chemistry Vitros	156	81 (51.9)	75 (48.1)	61 (56.5)	47 (43.5)	20 (41.7)	28 (58.3)	30.8
Immunoassay AxSym	34	28 (82.4)	6 (17.6)	23 (82.1)	5 (17.9)	5 (83.3)	1 (16.7)	17.6
Immunoassay BNII	44	25 (56.8)	19 (43.2)	20 (52.6)	18 (47.4)	5 (83.3)	1 (16.7)	13.6
Immunoassay Centaur	514	266 (51.8)	248 (48.2)	252 (56.1)	197 (43.9)	14 (21.5)	51 (78.5)	12.6
Immunoassay Dimension	21	18 (85.7)	3 (14.3)	18 (85.7)	3 (14.3)	0 (0.0)	0 (0.0)	0.0
Immunoasssay Advia	234	150 (64.1)	84 (35.9)	120 (69.4)	53 (30.6)	28 (45.9)	33 (54.1)	26.1
ISE Advia	14	10 (71.4)	4 (28.6)	10 (71.4)	4 (28.6)	0 (0.0)	0 (0.0)	0.0
ISE Modular	28	19 (67.9)	9 (32.1)	19 (67.9)	9 (32.1)	0 (0.0)	0 (0.0)	0.0
ISE Vitros	26	12 (46.2)	14 (53.8)	11 (50.0)	11 (50.0)	1 (25.0)	3 (75.0)	15.4

[a] Includes all reagent lot changes for which the change in QC value was <1.0 SDI.
[b] Includes all reagent lot changes for which the change in QC value was ≥1.0 SDI.
[c] Nonsignificant if $P > 0.05$; significant if $P \leq 0.05$.
[d] Top number in each cell is the number of reagent lot change–QC events that met the condition for a cell; bottom number is percent of reagent lot change–QC events that met the condition for a cell.

10 for the QC sample results, but variable for the patient sample results) and estimated variances of the difference in means for the results from patients and QC measurements *(18, 19)*. P values <0.05 were considered significant.

The magnitude of a change in QC results after a change in reagent lots was evaluated for each individual reagent lot change by computing the SD interval (SDI) as follows: (mean difference in QC results between 2 reagent lots)/(SD used for QC acceptance criteria). The SD value used for QC acceptance criteria was established from the cumulative SD consistent with stable performance of a measurement procedure over an extended time interval (typically 6–12 months) and was not derived from the QC results used to evaluate a given reagent lot change.

Statistical analyses were performed by using SAS software (version 9.2; SAS Institute).

Results

Table 1 shows a summary of findings with the analytes grouped by general categories of measurement technologies. Overall, 40.9% of reagent lot change–QC events had a significant difference ($P < 0.05$) between 2 reagent lots for QC sample results compared with results for patient samples. In the various measurement technology categories, the overall number of reagent lot change–QC events with significant differences ranged from 14.3% to 83.3%. Considering all measurement technology categories, 37.7% of the QC changes that were <1.0 SDI (83.1% of total) were significantly different from those seen for patient results. For QC changes ≥1.0 SDI (16.9% of total), 57.0% were significantly different from changes seen in patient results, with immunoassay Advia, general chemistry Advia, and general chemistry Vitros having the largest per-
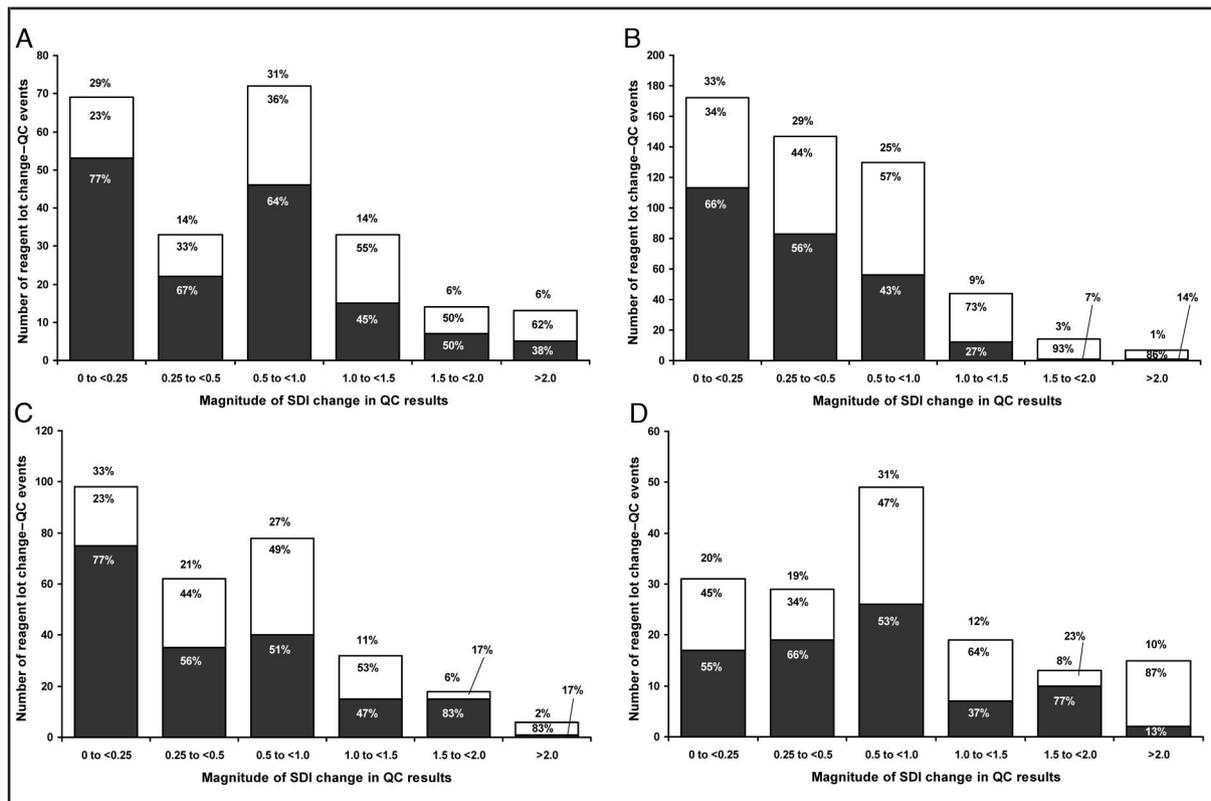
Fig. 1. Significant differences (*P* ≤ 0.05) between the results for QC samples compared with those for patient samples between 2 reagent lots subdivided by the magnitude of SDI shifts in QC results for the 4 method/instrument categories with the largest number of reagent lot change–QC events.

The nonsignificant (■) and significant (□) differences between QC and patient sample results are represented. The number within each bar segment is the percent that segment represents of the total number of reagent lot changes in a given SDI interval (*x* axis). The number at the top of the stacked segments is the percent of the total number of reagent lot changes represented by that column. (A), Immunoassay Siemens Advia; (B), immunoassay Siemens Centaur; (C), general chemistry Siemens Advia; (D), general chemistry Ortho Vitros.

centages (19% to 31%) of differences ≥1.0 SDI. Online Supplemental Table 1 shows for each analyte-instrument platform combination how many reagent lot change–QC events were statistically significant.

Fig. 1 shows differences between results for QC materials compared with those for patient samples between 2 reagent lots subdivided by the magnitude of SDI shifts in QC results for the 4 method/instrument categories with the largest number of reagent lot change–QC events examined. Note that even for relatively small changes in SDI for the QC results, there were substantial numbers of reagent lot changes for which the QC result shifts were significantly different than those observed for the patient samples. Conversely, some differences were nonsignificant at larger changes in SDI for the QC results. One reagent lot change for LDL cholesterol performed by using a Siemens Advia method is shown in the worked example in

the online Data Supplement. In this case, the QC value changed from 2.64 mmol/L (101.9 mg/dL) to 2.97 mmol/L (114.8 mg/dL), which was a change of 0.33 mmol/L (12.9 mg/dL) or 2.15 SDI. Online Supplemental Table 5 shows the QC mean values observed between 2 reagent lots, the SD used for QC acceptance criteria, and the difference as SDI for all QC differences which exceeded 1.5 SDI.

Fig. 2 shows significant differences subdivided by the magnitude of SDI shifts in QC results for 4 additional method/instrument categories. In general, these combinations of measurement procedures and instrument platforms had smaller magnitude differences in SDI for QC results with fewer significant differences between the QC results and the patient samples results. However, there were several combinations with large enough SDI changes to make the QC results unreliable to evaluate method performance.
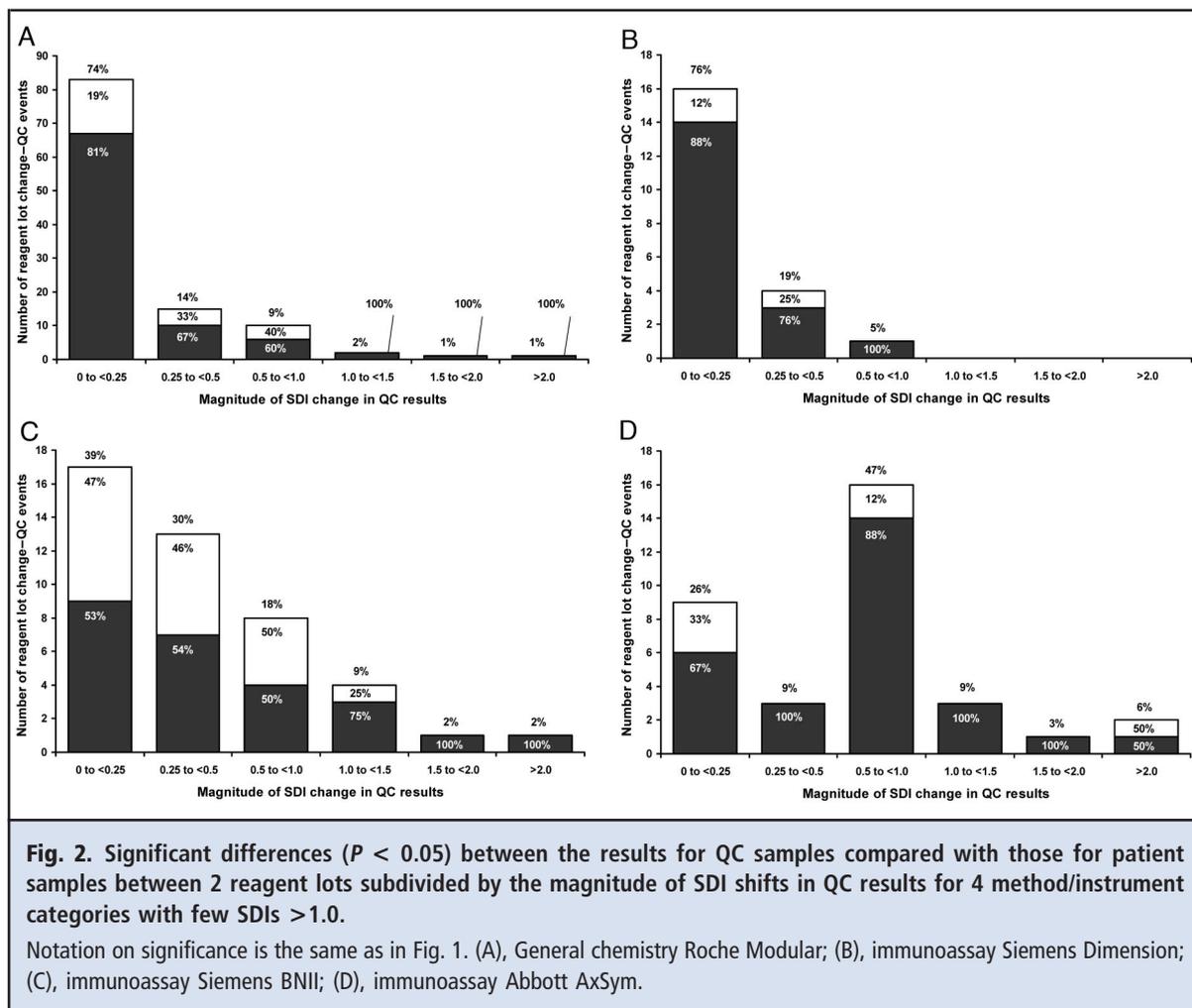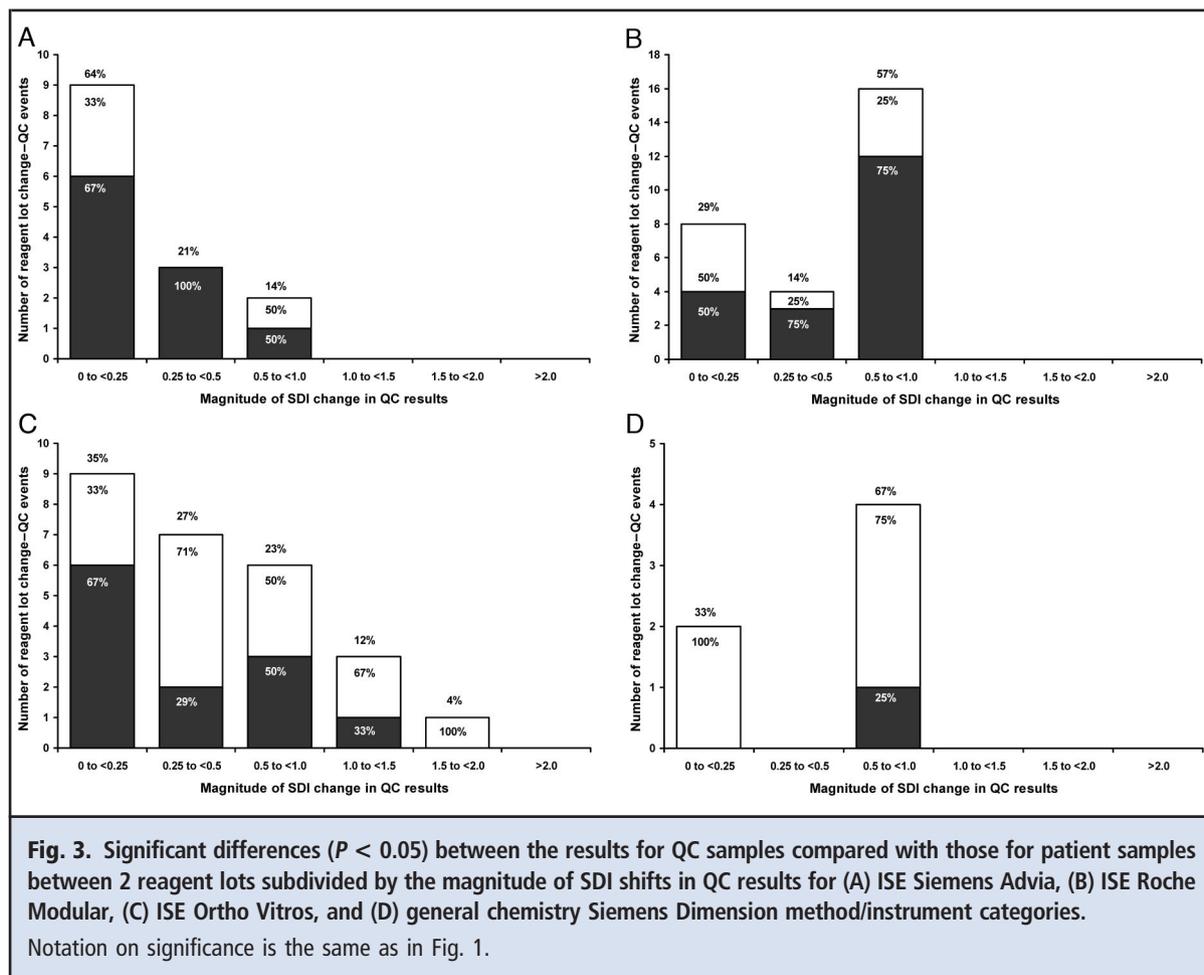
Fig. 2. Significant differences (*P* < 0.05) between the results for QC samples compared with those for patient samples between 2 reagent lots subdivided by the magnitude of SDI shifts in QC results for 4 method/instrument categories with few SDIs >1.0.

Notation on significance is the same as in Fig. 1. (A), General chemistry Roche Modular; (B), immunoassay Siemens Dimension; (C), immunoassay Siemens BNII; (D), immunoassay Abbott AxSym.

Fig. 3 shows the breakdown by magnitude of SDI change in QC results for the ion selective electrode (ISE) analytes for the Advia, Modular, and Vitros systems. Advia and Modular ISEs had no changes >1.0 SDI in QC results, whereas Vitros ISEs had changes up to 2.0 SDI magnitude, with significant differences between QC and patient sample results. Although there were only 6 reagent lot changes for the 3 general chemistry analytes examined on the Dimension analyzer, 5 of 6 instances had significant differences between QC and patient results.

Table 2 shows that nearly equivalent numbers of reagent lot changes had increases or decreases in the QC sample results compared with the changes observed for the patient samples. The percentages of increases or decreases in QC results were about the same for all reagent lot change–QC events and for events that were statistically different or not different from the changes seen in the patient samples.

Discussion

Our data demonstrate that noncommutability of QC materials with patient samples was a frequent occurrence when changing reagent lots, was observed for a wide range of different QC materials and measurement technologies, and was not predictable for a given QC material or reagent lot. Consequently, the results for QC samples should not be used to verify the consistency of results for patient samples when changing reagent lots. A laboratory must assume the possibility of noncommutability and use a set of patient samples measured by using both reagent lots to verify consistency of patient results between the reagent lots. This noncommutability limitation will likely also affect the controls used by manufacturers in reagent lot release testing and the values assigned to QC products intended for use by laboratories to validate performance of measurement procedures.

Fig. 3. Significant differences (*P* < 0.05) between the results for QC samples compared with those for patient samples between 2 reagent lots subdivided by the magnitude of SDI shifts in QC results for (A) ISE Siemens Advia, (B) ISE Roche Modular, (C) ISE Ortho Vitros, and (D) general chemistry Siemens Dimension method/instrument categories.
Notation on significance is the same as in Fig. 1.

In principle, it is possible to use a QC material to verify consistency of patient results between reagent lots as long as that QC material has been validated by either the QC or method manufacturer to be commutable with patient samples and thus suitable for that purpose. Because the commutability characteristics may be different for different lots of QC materials and/or reagents, such validation needs to be performed for each QC material and reagent lot involved. In this study, the Ortho Vitros system used QC materials provided by Ortho that had reagent lot specific values assigned. The results in Figs. 1D and 3C suggest that these

Table 2. Reagent lot changes with differences in apparent bias on the basis of QC sample results greater than (positive difference) or less than (negative difference) the actual bias observed for patient samples results.

	Overall	Nonsignificant differences between QC and patient sample results[a]	Significant differences between QC and patient sample results[a]
Positive difference	734 (49.5%)[b]	436 (49.8%)	298 (49.1%)
Negative difference	749 (50.5%)	440 (50.2%)	309 (50.9%)

[a] Nonsignificant if *P* > 0.05; significant if *P* ≤ 0.05.
[b] The number in a cell is the number of reagent lot change–QC events that satisfied the condition; the number in parentheses is the percent of the total number of reagent lot change–QC events in that column.

QC results were not adequate to verify consistency of results for patient samples between reagent lots.

Differences in matrix-related bias that may occur between a QC sample and different reagent lots do not preclude the usual practice to use QC samples to verify the consistency of results and to identify a change in a method's performance that needs to be corrected before reporting results for patient samples. As long as the same lots of QC materials and reagents are used, there will be no change in a matrix-related bias, and any changes in QC sample results should reflect changes in method performance.

The matrix-related limitation to using QC samples occurs when there is a change in reagent lots. Examination of the magnitude of SDI in Figs. 1–3 shows that matrix-related biases for QC samples could be large enough to make QC samples unreliable for indicating the performance status of a method after a reagent lot change. In these cases, the expected target values for QC samples must be adjusted to compensate for the change in matrix-related bias. Failure to adjust the expected value for a QC sample will cause a high frequency of apparent QC failures when a measurement procedure is actually performing to specifications and producing correct results for patient samples. For the worked example in the online Data Supplement, failure to adjust the target value for the change in QC mean values for LDL cholesterol from 2.64 mmol/L (101.9 mg/dL) to 2.97 mmol/L (114.8 mg/dL), which is a 2.15-SDI change, would have a large impact on the failure rate for QC results, although the results for patients were unaffected and had remained acceptable.

Both positive and negative differences in apparent bias were observed for QC results when compared with changes for patient sample results between 2 reagent lots (Table 2). There were nearly equal frequencies of positive or negative apparent biases, irrespective of whether the change in QC means was significantly or nonsignificantly different from the change observed for the patient samples. These observations indicate that there was no reliable relationship between the apparent bias for QC results between 2 reagent lots and the bias observed for patient results between the same 2 reagent lots. In all cases examined, the patient results were within each laboratory's criteria for consistency between 2 reagent lots. Consequently, a laboratory could conclude from QC results alone that there was a problem with a new reagent lot when in fact the results for patient samples were not affected or were affected to a smaller magnitude than indicated by the QC results. However, incorrect patient results could be reported if a laboratory erroneously concluded from QC results alone that there was no problem with a new reagent lot, when in fact the patient results were affected to a greater magnitude than indicated by the QC results. Because the magnitude of a matrix-related bias in QC results was different for different QC samples and reagent lot combinations, decisions about the suitability of performance of a new lot of reagents based only on QC sample results can give erroneous conclusions about performance for patient samples.

The magnitude of a matrix-related bias among reagent lots will affect interpretation of results when compared among different laboratories participating in an interlaboratory QC or PT program. Although we only investigated QC materials, the observations regarding reagent lot influence will be pertinent when evaluating results from PT materials that have been manufactured by processes similar to those used for the QC materials examined. The frequency of noncommutable QC results when changing reagent lots in this study is similar to that observed in previous investigations that examined both QC and PT materials for between-method effects (4–10). To compensate for matrix-related biases, peer-group evaluation of interlaboratory QC, or PT, results is frequently used to evaluate agreement among laboratories that use similar or identical measurement procedure technology. However, differences in magnitude of matrix-related bias caused by use of different reagent lots within a peer group will increase the SD, making QC or PT data unrepresentative of the SD expected for that method in a single laboratory. Furthermore, if a small number of laboratories were using a reagent lot different from that of the rest of the participants in a peer group, a difference in magnitude of a matrix-related bias among reagent lots could cause an apparent failure to meet the acceptance criteria for a particular comparison event. The influence of reagent lot on the matrix-related bias for QC (or PT) materials may cause estimates of analytical variability on the basis of QC (or PT) results to be larger than would be observed for patient samples.

Strengths of this study include a large number of QC materials, reagent lot changes, analytes, measurement procedures, and instrument platforms from a number of different manufacturers. Patient sample results may have differed slightly because of calibration differences between 2 reagent lots but provided a control group for any calibration difference that may have occurred. Consequently, any change in the magnitude of differences between the QC material results and the patient sample results between 2 reagent lots was caused by a change in the matrix-related bias in the QC materials.

Limitations of this study include that differences between 2 reagent lots were calculated from a relatively small number of results for patient and QC samples. This limitation was accommodated in the statistical procedure but may have produced lower statistical

power to detect important differences. In addition, methods with poorer precision are expected to be less sensitive to differences between the QC and patient sample results; consequently, small matrix-related differences in the QC results may not have been identified. It is possible that a small calibration shift may have occurred with a reagent lot change that affected both patient samples and QC samples, in which case a non-significant SDI change in the QC mean values may have been observed.

In summary, we observed significant ($P < 0.05$) differences between results for QC materials and results for patient samples for 40.9% of 1483 reagent lot change–QC events examined for 18 QC materials, 661 reagent lot changes, 82 analytes, and 7 instrument platforms. The magnitude of 16.9% of those differences exceeded a 1-SDI change in the numeric value assigned to the QC material. We concluded that occurrence of noncommutable results for QC materials was frequent enough that QC results could not be used to verify consistency of results for patient samples when changing lots of reagents.

References

1. Rej R, Jenny RW, Bretaudiere JP. Quality control in clinical chemistry: characterization of reference materials. Talanta 1984;31:851–62.

2. Franzini C, Ceriotti F. Impact of reference materials on accuracy in clinical chemistry. Clin Biochem 1998;31:449–57.

3. Cattazzo G, Franzini C, Melzi D'Eril G. Commutability of calibration and control materials for serum lipase. Clin Chem 2001;47:2108–13.

4. Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. Clin Chim Acta 2003;327:25–37.

5. Thienpont LM, Stockl D, Friedecky B, Kratochvila J, Budina M. Trueness verification in European external quality assessment schemes: time to care about the quality of the samples. Scand J Clin Lab Invest 2003;63:195–201.

6. Miller WG, Myers GL, Ashwood ER, Killeen AA, Wang E, Thienpont LM, Siekmann L. Creatinine measurement: state of the art in accuracy and inter-laboratory harmonization. Arch Pathol Lab Med 2005;129:297–304.

7. Palmer-Toy DE, Wang E, Winter WE, Soldin SJ, Klee GG, Howanitz JH, Elin RJ. Comparison of pooled fresh frozen serum to proficiency testing material in College of American Pathologists surveys: cortisol and immunoglobulin E. Arch Pathol Lab Med 2005;129:305–309.

8. Bock JL, Endres DB, Elin RJ, Wang E, Rosenzweig B, Klee GG. Comparison of fresh frozen serum to traditional proficiency testing material in a College of American Pathologists survey for ferritin, folate, and vitamin B12. Arch Pathol Lab Med 2005;129:323–7.

9. Schreiber WE, Endres DB, McDowell GA, Palomaki GE, Elin RJ, Klee GG, Wang E. Comparison of fresh frozen serum to proficiency testing material in College of American Pathologists surveys: α-fetoprotein, carcinoembryonic antigen, human chorionic gonadotropin, and prostate-specific antigen. Arch Pathol Lab Med 2005;129:331–7.

10. Miller WG, Myers GL, Ashwood ER, Killeen AK, Wang E, Ehlers GW, et al. State of the art in trueness and inter-laboratory harmonization for 10 analytes in general clinical chemistry. Arch Pathol Lab Med 2008;132:838–46.

11. Miller WG, Myers GL, Rej R. Why commutability matters [Editorial]. Clin Chem 2006;52:553–4.

12. Vesper HW, Miller WG, Myers GL. Reference materials and commutability. Clin Biochem Rev 2007;28:139–47.

13. Miller WG. Matrix effects in the measurement and standardization of lipids and lipoproteins. In: Rifai N, Warnick GR, Dominiczak MH, eds. Handbook of lipoprotein testing. 2nd ed. Washington, DC: AACC Press 2000;695–716.

14. Eckfeld JH, Copeland KR. Accuracy verification and identification of matrix effects. Arch Pathol Lab Med 1993;117:381–88.

15. Miller WG. The role of proficiency testing in achieving standardization and harmonization between laboratories. Clin Biochem 2009;42:232–5.

16. Martindale RA, Cembrowski GS, Journault LJ, Crawford JL, Tran C, Hofer TL, et al. Validating new reagents: roadmaps through the wilderness. Lab Medicine 2006;37:347–51.

17. Miller WG. Quality control. In: McPherson RA, Pincus MR, eds. Henry's clinical diagnosis and management by laboratory methods. 21st ed. St. Louis: Elsevier Saunders 2007;99–111.

18. Kutner MH, Natchsheim CJ, Neter J, Li W. Applied linear statistical models. 5th ed. Boston: McGraw-Hill Irwin; 2005.

19. Ames MH, Webster JT. On estimating approximate degrees of freedom. Am Stat 1991;45:45–50.