

Analytical robustness of nine common assays: frequency of outliers and extreme differences identified by a large number of duplicate measurements

Stefanie Neubig^{*1}, Anne Grotevendt¹, Anders Kallner², Matthias Nauck¹, Astrid Petersmann¹

¹Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany

²Department of Clinical Chemistry, Karolinska University Hospital, Stockholm, Sweden

*Corresponding author: stefanie_neubig@web.de

Abstract

Introduction: Duplicate measurements can be used to describe the performance and analytical robustness of assays and to identify outliers. We performed about 235,000 duplicate measurements of nine routinely measured quantities and evaluated the observed differences between the replicates to develop new markers for analytical performance and robustness.

Materials and methods: Catalytic activity concentrations of aspartate aminotransferase (AST), alanine aminotransferase (ALT), and concentrations of calcium, cholesterol, creatinine, C-reactive protein (CRP), lactate, triglycerides and thyroid-stimulating hormone (TSH) in 237,261 patient plasma samples were measured in replicates using routine methods. The performance of duplicate measurements was evaluated in scatterplots with a variable and symmetrical zone of acceptance (A-zone) around the equal line. Two quality markers were established: 1) AZ95: the width of an A-zone at which 95 % of all duplicate measurements were within this zone; and 2) OPM (outliers per mille): the relative number of outliers if an A-zone width of 5 % was applied.

Results: The AZ95 ranges from 3.2% for calcium to 11.5% for CRP and the OPM from 5 (calcium) to 250 (creatinine). Calcium, TSH and cholesterol have an AZ95 of less than 5% and an OPM of less than 50.

Conclusions: Duplicate measurements of a large number of patient samples identify even low frequencies of extreme differences and thereof defined outliers. We suggest two additional quality markers, AZ95 and OPM, to complement description of assay performance and robustness. This approach can aid the selection process of measurement procedures in view of clinical needs.

Key words: clinical chemistry tests; humans; reproducibility of results; analytical quality; quality control

Received: July 05, 2016

Accepted: December 03, 2016

Introduction

Internal quality control schemes (IQC), e.g. Westgard or the Guidelines for Quality Assurance of Medical Laboratory Examinations of the German Medical Association (RilibÄK), have a long tradition in clinical chemistry to monitor the reliability of measurements and have proved to provide a reasonable quality control (1-3). The Clinical & Laboratory Standards Institute (CLSI) has advised procedures to verify the trueness and precision of routine methods (4). The schemes rule that IQC samples are measured at certain intervals, defined either by time or by number of measured patient samples. Therefore, IQC may fail to detect occa-

sional dropouts, i.e. outliers among routine samples, which may cause erroneous clinical decisions in patient care (4-15). To identify outliers, duplicate measurements may be used (16-20), but are often refrained from, due to economic pressure (21,22).

Outliers are differently defined in the literature and their frequencies are reported by various approaches (1,2,4,5,16-19,23-26). Additionally, what is perceived as an outlier may vary depending on the analyte or diagnosis (2,4,16-19,24,27). By describing magnitude and number of differences between duplicate measurements, laboratories and clinicians can define the difference they consider

an outlier and consequently estimate the frequency they will experience (27). If based on a sufficient number, duplicate measurements have the potential to reveal even low frequencies of outliers (4,20). In the present study, we report a large number of duplicate measurements in patient plasma samples under routine conditions in nine commonly used assays. It is an extension of a study published in 2013, where these investigations were performed with one glucose assay (27). To broaden the findings from the glucose study, further frequently used assays were investigated and compared. Since only measurements with valid IQC shall be included, the performance observed in the study reflects the currently accepted frequency and magnitude of extreme differences. Our aim was to detect even low frequencies of outliers and their magnitude in routine assays and to derive new quality markers for assay performance that give more detailed information than the regular IQC. Our approach aims at complementing description of performance and analytical robustness of assays.

Materials and methods

The plasma samples were from an unselected population of hospitalized patients and outpatients at the University Medicine of Greifswald (Germany). During twelve consecutive months, starting in December 2011 the catalytic activity concentrations and concentrations of aspartate aminotransferase (AST), alanine aminotransferase (ALT), and concentrations of calcium, cholesterol, creatinine, C-reactive protein (CRP), lactate, triglycerides and thyroid-stimulating hormone (TSH) were routinely measured in duplicates on a 24/7 basis. Assays were run on three Dimension Vista 1500 instruments using procedures and reagents according to the manufacturer's instruction (Siemens Healthcare Diagnostics, Eschborn, Germany). The instruments were connected to Stream-LAB (Siemens Healthcare Diagnostics, Eschborn, Germany) which allocated the samples to one of the three instruments. Once allocated to a specific instrument, the duplicate measurements were automatically ordered and performed on this same

instrument by two sample aspirations but only the first obtained result was released for patient care.

Only measurements within concentration intervals according to RiliBÄK (published in RiliBÄK's Table B1, column 4: "RiliBÄK applicable concentration intervals of columns 3 and 5"), specific for each analyte, and only measurements with valid IQC according to the RiliBÄK were included in the study (3). TRU-Liquid Monitrol (lot 1AQ104) was used as IQC material for ALT, AST, calcium, cholesterol, creatinine, lactate, Triglycerides; Liquimmune (lot 9LQ105, both Thermo Fisher Scientific, Schwerte, Germany) for TSH and Protein2 (lot 1LQH01, Siemens Healthcare Diagnostics, Eschborn Germany) for CRP. Imprecision was calculated from IQC for each analyte, IQC level (high and low) and RiliBÄK monthly control cycle. Samples were anonymized prior to data collection.

Data analysis was performed using Microsoft Excel® (2010). Ethical approval of the local ethics committee was obtained.

For each assay, differences or agreements of duplicate measurements were summarized in scatter plots including regression and correlation analyses. To categorize the observed differences between the duplicates with respect to their magnitude, a zone surrounding the equal line of the scatter plots was used (Figure 1). This area is often referred to as the A-zone (28). The A-zone (dotted lines Figure 1) is located around the equal line (hatched line Figure 1) of duplicate measurements, with the first measurement on the X-axis and the second measurement on the Y-axis. The A-zone width can be modified symmetrically around the equal line as indicated by the arrows. The triangles represent a duplicate measurement within the chosen A-zone whereas the square represents a duplicate measurement outside the A-zone.

In this study its width was systematically modified and increased to the width of 14% (Figure 2). Differences outside the chosen A-zone were regarded as outliers. Thus, the definition of an outlier depends on the width of the A-zone. A narrow A-zone consequently corresponds to strict limits and would cause small differences to be regarded as outliers whereas a wide A-zone allows large differences.

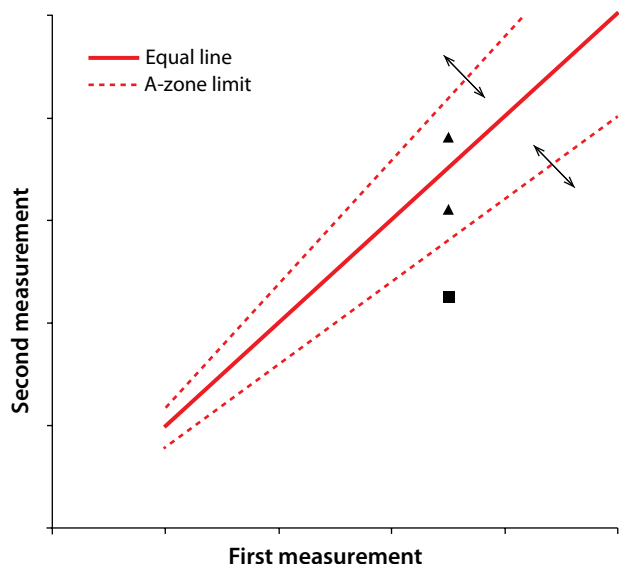


FIGURE 1. The A-zone. The dotted lines around the equal line (hatched line) of duplicate measurements represent the A-zone, the first measurement on the X-axis and the second measurement on the Y-axis. The A-zone width can be modified symmetrically around the equal line as indicated by the arrows. The triangles represent a duplicate measurement within the chosen A-zone whereas the square represents a duplicate measurement outside the A-zone.

At each A-zone width, differences that fell outside this area were counted. The relative numbers of observations outside the various A-zones were plotted against the width of the A-zone (Figure 2). Two quality markers can be derived from this approach and used to describe performance and analytical robustness of assays:

1. AZ95 (A-zone 95%): The width of the A-zone at which 95% of all duplicate measurements are within this zone (Figure 2, the A-zone width can be read from the x-axis where the respective analyte curve crosses the red horizontal line. This horizontal line crosses the y-axis at 50 OPM since this represents 50 out of 1000 i.e. 95%),
2. OPM (outlier *per mille*): The relative number of outliers in *per mille* if an A-zone width of 5% is used to identify outliers (Figure 2, the relative number of outliers can be read from the y-axis where the respective curve crosses the red vertical line).

The common target is 95% of the observations within an A-zone of 5%, which is already described in the CLSI EP 27 guidelines (28) and means in our study a maximum AZ95 of 5% and at the same time a maximum OPM of 50. For the comparison between different analytes it is important to consider, that individual clinical requirements induce different requirements for the AZ95 and the OPM of each analyte.

Results

The number of plasma samples run in duplicates ranged from 1596 for lactate to 73,242 for creatinine (Table 1); in total, 237,261 duplicates were measured. The imprecision calculated from the IQC from low and high levels of the IQC for each analyte is given in Table 1. At an A-zone width of 12% all assays had less than 50 measurements *per mille* outside the A-zone and showed then an as-

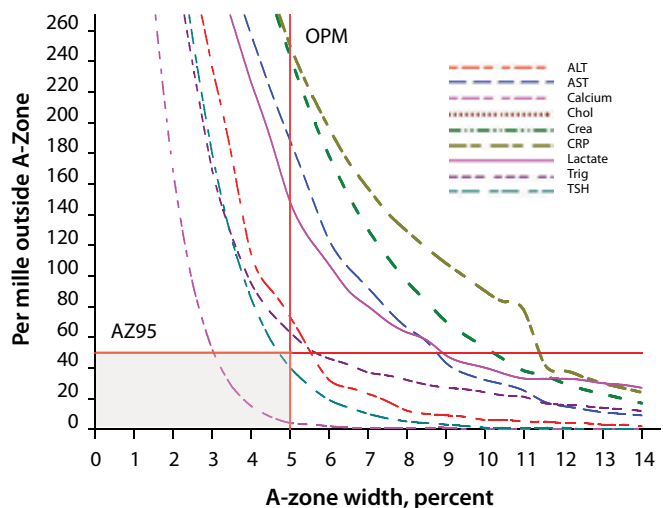


FIGURE 2. Analytical performance of nine assays. The frequency (*per mille*) of observed differences in duplicate measurements, which were outside a given width of the zone of acceptance (A-zone). AZ95, (the width of the A-zone comprising 95% of the observations) for each assay is given on the X-axis where the horizontal red line crosses the curves of the assays. The OPM (the relative number of observations in 1000 observations outside the A-zone) at an A-zone width of 5% can be read from the red vertical line. Horizontal lines crossing assay lines are in the following order: calcium, TSH, cholesterol, ALT, triglycerides, AST, lactate, creatinine and CRP. Assays that cross the shaded area in the left-hand corner comprise 95% of the observations with an A-zone width lower than 5% and have fewer than 50 *per mille* outliers at an A-zone of 5%.

TABLE 1. Basic data on studied assays and imprecision calculated from IQC.

Measurand (Measuring interval according to the manufacturer's instructions)	RilibÄK (column 4) applicable concentration	Number of duplicates	CV (%), (target value IQC low level)	CV (%), (target value IQC high level)
ALT* (6 - 1002 U/L)	30–300	5620	3.1–5.0 (42 U/L)	2.8–3.6 (105 U/L)
AST* (3 – 1002 U/L)	19.8–400.2	11,797	2.1–5.7 (36 U/L)	0.8–1.5 (194 U/L)
Calcium* (1.25 - 3.75 mmol/L)	1.00–6.00	65,077	1.1–2.4 (2.41 mmol/L)	1.0–2.2 (2.85 mmol/L)
Cholesterol* (1.29 - 15.54 mmol/L)	1.3–9.1	12,092	1.7–5.2 (5.1 mmol/L)	1.8–5.4 (2.74 mmol/L)
Creatinine* (9 – 1768 µmol/L)	44–884	73,242	2.2–7.5 (118 µmol/L)	0.9–3.8 (629 µmol/L)
CRP† (3.1 – 190 mg/L)	1–120	27,836	2.4–6.1 (52 mg/L)	2.3–7.0 (12.1 mg/L)
Lactate* (0.1 – 15 mmol/L)	1–10	1596	1.2–4.4 (2.42 mmol/L)	1.2–8.3 (5.71 mmol/L)
Triglycerides* (0.02 - 11.3 mmol/L)	0.68–4.6	17,094	1.3–3.8 (2.35 mmol/L)	1.6–4.5 (1.35 mmol/L)
TSH# (0.005 – 100 mU/L)	0.1–40	23,321	2.0–4.3 (0.27 mU/L)	1.8–3.6 (16.4 mU/L)

*IQC were performed using 1AQ104 TRU-Liquid Monitrol (Thermo Fisher Scientific, Schwerte, Germany).

#IQC Was performed using 9LQ105 Liquimmune (Thermo Fisher Scientific, Schwerte, Germany).

†IQC Was performed using 1LQH01 Protein2 (Siemens Healthcare Diagnostics, Eschborn Germany).

CV – coefficient of variation. Conversion factor from SI Unit to Conventional Unit for enzymes is 60. AST - aspartate aminotransferase; ALT - alanine aminotransferase; CRP – C – reactive protein; TSH - thyroid-stimulating hormone.

ymptotical decrease. In consideration of the asymptotical curve at widen A-zone width we limited the width of the A-zone to 14%.

The number of outliers relative to the A-zone width is shown in Figure 2:

1. AZ95 is read on the X-axis in figure 2 where the horizontal red line crosses the curves of the assays. The AZ95 ranges from 3.2% for calcium to 11.3% for CRP. Triglycerides and ALT have an AZ95 of 5.5%, which is below the values found for lactate, AST, creatinine and CRP, but above those found for calcium, TSH and cholesterol.
2. OPM can be read from the vertical red line in Figure 2. An OPM of 5 *per mille* can be found for calcium and up to 250 *per mille* for CRP and creatinine. Values found for AST, lactate, creatinine and CRP indicate a poorer performance, i.e. a higher relative number of outliers than triglycerides, ALT, cholesterol, TSH and calcium.

Only three out of the investigated nine assays (calcium, TSH, and cholesterol) have

- an A-zone width equal to or lower than 5% and including 95% of the observations and
- equal to or fewer than 50 *per mille* outliers at an A-zone of 5%.

The curves of these assays cross the shaded area in the lower left corner in Figure 2.

Discussion

We used a large number of duplicate measurements to describe the performance and analytical robustness of assays.

We introduce two new quality markers for describing analytical quality (AZ95 and OPM): The AZ95 (width of the A-zone covering 95% of the observations for an assay) was chosen in analogy to the 95% confidence interval and represents the first of

the two suggested quality markers (horizontal red line, Figure 2). It ranges from about 3.2% for calcium to 11.3% for CRP.

In a clinical setting, relative terms may be difficult to handle. Therefore, we translate our findings into absolute terms. When assuming that the first measured value represents a measurement on the equal line, the AZ95 for calcium was found to be 3.2%. Therefore, at a calcium concentration of 2.0 mmol/L, 95% of all duplicate measurements could be expected between 1.94 and 2.06 mmol/L, whereas for a creatinine concentration of 100 μ mol/L it would be between 90 and 110 μ mol/L since its AZ95 was determined at 10.3%. Still, 5% of all measurements will deviate more. For glucose the A-zone width which comprises 95% of all observations was reported in the previous study, which used the same instrument, to be approximately 4% (27). This previously reported performance for glucose is comparable to our findings for calcium.

The second suggested quality marker OPM is the relative number of observations outside the A-zone in *per mille* at an A-zone width of 5% (vertical line, Figure 2). Calcium shows the best performance with an OPM of 5 in 1000 measurements. Translated into clinical terms: at calcium concentration of 2.0 mmol/L \pm 0.1 mmol/L the clinician would have to accept an outlier frequency of 5 *per mille*, i.e. in only 5 cases out of 1000 measurements. For CRP (e.g. at 5.0 mg/L \pm 0.25 mg/L) and creatinine (e.g. at 100 μ mol/L \pm 5.0 μ mol/L) the OPM is 210 and 250 in 1000 measurements, respectively (extrapolated from Figure 2). The initial study reported less than 10 *per mille* outside an A-zone width of 5% for glucose based on 21,000 duplicate measurements and therefore its performance is comparable to calcium in this study (27).

According to the quality markers, AZ95 and OPM used in the present study three assays (calcium, TSH and cholesterol) show the best analytical robustness and performance of all investigated assays with an A-zone width that comprises 95% of the observations that is lower than 5% and fewer than 50 *per mille* outliers at an A-zone of 5%. Both criteria have been reported to be fulfilled also for glucose (27). Deetz *et al.* investigated duplicate

measurements applying College of American Pathologists (CAP)/Clinical Laboratory Improvement Amendments (CLIA) error limits to identify outliers and report 0.2% outliers out of 3000 observations for calcium which corresponds to an AZ95 of about 2.5% (22). This study is in line with our findings for calcium, which showed a very low frequency of extreme differences compared to other assays. Whereas Deetz *et al.* investigated about 3000 observations for calcium, other assays had 100 observations or less (22,29). Onyenekwu *et al.* found 4.9% outliers out of 91 repeats for calcium at critical concentrations also using CAP/CLIA error limits (21). Witte *et al.* aimed to identify outliers in the sense of "errors" defined by a multiple SD e.g. 6 or 7 SD and therefore report a comparatively low frequency of 41 in one million results (0.041 *per mille*) (29).

Due to the heterogeneous approaches, results of different studies cannot be easily compared. In contrast to previous studies, our model represents a flexible approach to search for differences or outliers of various definitions by widening or reducing the A-zone accordingly. Rather than a fixed or statistically based definition of outliers, we evaluated the frequency of extreme differences of diverse magnitudes in relation to a distribution of observed duplicates around an equal line assuming identity of duplicates. In addition to precision and trueness, the frequency of outliers should be considered to describe the analytical quality of an assay (30).

Our findings describe what is presently accepted in clinical practice, i.e. the "state of the art". To facilitate comparability between laboratories and assays we suggest the fixed quality markers AZ95 and OPM in analogy to the 95% confidence interval, but our model also allows for individual adjustments. The results of our study complement performance criteria of assays and may be used for discussing potentials and limitation of assays between clinicians and laboratorians. Furthermore, this approach can aid the selection process of measurement procedures in view of clinical needs.

Due to limited resources, we focused on nine commonly used assays. The number of duplicates was below 10,000 for three assays, which limits the de-

tection of very low frequencies of extreme differences and outliers. These assays showed OPMs of approximately 55 (ALT) and 90 (lactate) which can be sufficiently identified by 5600 and 1500 duplicates, respectively.

In conclusion, duplicate measurements of large numbers of patient samples identify even low frequencies of extreme differences. We suggest two additional quality markers to describe performance and robustness of assays and report what is currently accepted in clinical practice: 1. AZ95: width of an A-zone containing 95% of all duplicate measurements, and 2. OPM: the relative number (in *per mille*) of outliers if an A-zone width of 5% is

used to identify outliers. Out of the investigated nine common assays calcium, TSH, and cholesterol have an A-zone width comprising 95% of the observations that is lower than 5% and have fewer than 50 *per mille* outliers at an A-zone of 5. Our findings complement performance criteria of assays and can aid the selection process of measurement procedures in view of clinical needs.

Acknowledgments

Reagents were partly funded by Siemens Healthcare Diagnostics.

Potential conflict of interest

None declared.

References

1. ISO 15189:2012. *Medical laboratories — Requirements for quality and competence*. Available at: <https://www.iso.org/obp/ui/#iso:std:iso:15189:ed-3:v2:en>. Accessed May 23rd 2016.
2. Westgard JO. "Westgard Rules" and Multirules - Westgard. Available at: <https://www.westgard.com/mltirule.htm>. Accessed May 23rd 2016.
3. Revision of the "Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK". *J Lab Med* 2015;39:26–69.
4. Clinical and Laboratory Standards Institute (CLSI). *User Verification of Precision and Estimation of Bias; Approved Guideline—Third Edition*. CLSI document EP15-A3. Wayne, PA: CLSI; 2014.
5. Kristiansen J. *The Guide to expression of uncertainty in measurement approach for estimating uncertainty: an appraisal*. *Clin Chem* 2003;49:1822–9. <https://doi.org/10.1373/clinchem.2003.021469>.
6. Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. *Clin Chem* 2007; 53:1338–42. <https://doi.org/10.1373/clinchem.2007.088344>.
7. Loh TP, Lee LC, Sethi SK, Deepak DS. Clinical consequences of erroneous laboratory results that went unnoticed for 10 days. *J Clin Pathol* 2013;66:260–1. <https://doi.org/10.1136/jclinpath-2012-201165>.
8. Bashiti O. Accidental Diagnosis of Multiple Myeloma in a 44-Year-Old White Woman due to Erroneous Results via Chemical Analyzers. *Lab Med* 2016;47:e5-e11. <https://doi.org/10.1093/labmed/lmv011>.
9. Rotmensch S, Cole LA. False diagnosis and needless therapy of presumed malignant disease in women with false-positive human chorionic gonadotropin concentrations. *Lancet* 2000;355:712–5. [https://doi.org/10.1016/S0140-6736\(00\)01324-6](https://doi.org/10.1016/S0140-6736(00)01324-6).
10. Cole L, Rinne KM, Shahabi S, Omrani A. False-positive hCG assay results leading to unnecessary surgery and chemotherapy and needless occurrences of diabetes and coma. *Clin Chem* 1999;45:313–4.
11. Mugler K, Lefkowitz JB. False-positive D-dimer result in a patient with Castleman disease. *Arch Pathol Lab Med* 2004;128:328–31.
12. Roller RE, Lahousen T, Lipp RW, Korninger C, Schnedl WJ. Elevated D-dimer results in a healthy patient. *Blood Coag Fibrinol* 2001;12:501–2. <https://doi.org/10.1097/00001721-200109000-00012>.
13. Krahn J, Parry DM, Leroux M, Dalton J. High percentage of false positive cardiac troponin I results in patients with rheumatoid factor. *Clin Biochem* 1999;32:477–80. [https://doi.org/10.1016/S0009-9120\(99\)00044-2](https://doi.org/10.1016/S0009-9120(99)00044-2).
14. Ismail Y, Ismail AA, Ismail AAA. Erroneous laboratory results: what clinicians need to know. *Clin Med* 2007;7:357–61. <https://doi.org/10.7861/clinmedicine.7-4-357>.
15. Ismail AA, Walker PL, Barth JH, Lewandowski KC, Jones R, Burr WA. Wrong biochemistry results: two case reports and observational study in 5310 patients on potentially misleading thyroid-stimulating hormone and gonadotropin immunoassay results. *Clin Chem* 2002;48:2023–9.
16. Pretorius CJ, Dimeski G, O'Rourke PK, Marquart L, Tyack SA, Wilgen U, et al. Outliers as a cause of false cardiac troponin results: investigating the robustness of 4 contemporary assays. *Clin Chem* 2011;57:710–8. <https://doi.org/10.1373/clinchem.2010.159830>.
17. Ungerer JP, Pretorius CJ, Dimeski G, O'Rourke PK, Tyack SA. Falsely elevated troponin I results due to outliers indicate a lack of analytical robustness. *Ann Clin Biochem* 2010;47:242–7. <https://doi.org/10.1258/acb.2010.010012>.

18. Sawyer N, Blennerhassett J, Lambert R, Sheehan P, Vasikaran SD. Outliers affecting cardiac troponin I measurement: comparison of a new high sensitivity assay with a contemporary assay on the Abbott ARCHITECT analyser. *Ann Clin Biochem* 2014;51:476–84. <https://doi.org/10.1177/0004563213499737>.
19. Ryan JB, Southby SJ, Stuart L, Mackay R, Florkowski CM, George PM. Comparison of cardiac TnI outliers using a contemporary and a high-sensitivity assay on the Abbott Architect platform. *Ann Clin Biochem* 2014;51:507–11. <https://doi.org/10.1177/0004563214534637>.
20. Raggatt PR. Duplicates or singletons? An analysis of the need for replication in immunoassay and a computer program to calculate the distribution of outliers, error rate and the precision profile from assay duplicates. *Ann Clin Biochem* 1989;26:26–37. <https://doi.org/10.1177/000456328902600104>.
21. Onyenekwu CP, Hudson CL, Zemlin AE, Erasmus RT. The impact of repeat-testing of common chemistry analytes at critical concentrations. *Clin Chem Lab Med* 2014;52:1739–45. <https://doi.org/10.1515/cclm-2014-0331>.
22. Deetz CO, Nolan DK, Scott MG. An examination of the usefulness of repeat testing practices in a large hospital clinical chemistry laboratory. *Am J Clin Pathol* 2012;137:20–5. <https://doi.org/10.1309/AJCPWPBF62YGEFOR>.
23. Bhat V, Chavan P, Naresh C, Poladia P. The External Quality Assessment Scheme (EQAS): Experiences of a medium sized accredited laboratory. *Clin Chim Acta* 2015;446:61–3. <https://doi.org/10.1016/j.cca.2015.04.007>.
24. Bailey D, Bevilacqua V, Colantonio DA, Pasic MD, Perumal N, Chan MK, et al. Pediatric within-day biological variation and quality specifications for 38 biochemical markers in the CALIPER cohort. *Clin Chem* 2014;60:518–29. <https://doi.org/10.1373/clinchem.2013.214312>.
25. Krouwer JS. Critique of the Guide to the expression of uncertainty in measurement method of estimating and reporting uncertainty in diagnostic assays. *Clin Chem* 2003;49:1818–21. <https://doi.org/10.1373/clinchem.2003.019505>.
26. Buonocore R, Avanzini P, Aloe R, Lippi G. Analytical imprecision of lactate dehydrogenase in primary serum tubes. *Ann Clin Biochem* 2016;53:405–8. <https://doi.org/10.1177/0004563215595644>.
27. Petersmann A, Wasner C, Nauck M, Kallner A. Frequency of Extreme Differences and Clinical Performance of Glucose Concentration Measurements Judged from 21 000 Duplicate Measurements. *Clin Chem* 2013;59:998–1000. <https://doi.org/10.1373/clinchem.2012.197996>.
28. Clinical and Laboratory Standards Institute (CLSI). How to Construct and Interpret an Error Grid for Quantitative Diagnostic Assays; Approved Guideline. CLSI document EP27-A. Wayne, PA: CLSI; 2012.
29. Witte DL, VanNess SA, Angstadt DS, Pennell BJ. Errors, mistakes, blunders, outliers, or unacceptable results: how many? *Clin Chem* 1997;43:1352–6.
30. Burnett RW. Accurate estimation of standard deviations for quantitative methods used in clinical chemistry. *Clin Chem* 1975;21:1935–8.