

ORIGINAL RESEARCH

The quality of veterinary in-clinic and reference laboratory biochemical testing

Mark Rishniw, Paul D. Pion, Tammy Maher

Veterinary Information Network, Davis, CA, USA

Key Words

Clinical chemistry, quality assurance, quality control, sigma metrics, total allowable error

Correspondence

Mark Rishniw, Veterinary Information Network, 777 W Covell Blvd, Davis, CA 95616, USA

E-mail: mrishniw@vin.com

DOI:10.1111/j.1939-165X.2011.00386.x

Background: Although evaluation of biochemical analytes in blood is common in veterinary practice, studies assessing the global quality of veterinary in-clinic and reference laboratory testing have not been reported.

Objective: The aim of this study was to assess the quality of biochemical testing in veterinary laboratories using results obtained from analyses of 3 levels of assayed quality control materials over 5 days.

Methods: Quality was assessed by comparison of calculated total error with quality requirements, determination of sigma metrics, use of a quality goal index to determine factors contributing to poor performance, and agreement between in-clinic and reference laboratory mean results. The suitability of in-clinic and reference laboratory instruments for statistical quality control was determined using adaptations from the computerized program, EZRules3.

Results: Reference laboratories were able to achieve desirable quality requirements more frequently than in-clinic laboratories. Across all 3 materials, > 50% of in-clinic analyzers achieved a sigma metric ≥ 6.0 for measurement of 2 analytes, whereas > 50% of reference laboratory analyzers achieved a sigma metric ≥ 6.0 for measurement of 6 analytes. Expanded uncertainty of measurement and \pm total allowable error resulted in the highest mean percentages of analytes demonstrating agreement between in-clinic and reference laboratories. Owing to marked variation in bias and coefficient of variation between analyzers of the same and different types, the percentages of analytes suitable for statistical quality control varied widely.

Conclusion: These findings reflect the current state-of-the-art with regard to in-clinic and reference laboratory analyzer performance and provide a baseline for future evaluations of the quality of veterinary laboratory testing.

Introduction

Biochemical analysis of serum or plasma is performed routinely in veterinary medicine. Bench-top in-clinic biochemical analyzers have become commonplace in veterinary practice worldwide, and practitioners perform analyses on animal patients as part of diagnostic evaluations, preanesthetic screenings, or wellness examinations. Practices without such equipment typically submit samples for analysis to commercial laboratories.

Regular analysis of analyzer performance is essential in providing believable, reliable, and accu-

rate results. Human medical laboratories in the US are compelled by law to conduct internal quality control (QC) and to participate in external quality assurance (QA) programs.¹ Although similar mandates do not exist in veterinary medicine, commercial reference laboratories and university veterinary hospital laboratories generally conduct internal QC and participate in external QA programs to ensure the high quality of results; however, few private practices participate in such programs. Thus, despite the expanding role of biochemical analyzers in veterinary medicine, little information is available about assessment of their performance. Personal observations suggest that

veterinary clinicians and their staff rarely, if ever, perform rigorous performance evaluations of their equipment.

To our knowledge, there are no reports in the literature assessing the global quality of biochemical testing in veterinary practice. Similarly, there are no comparisons of reference laboratories using different instruments to analyze identical specimens or of performance of in-clinic analyzers compared with that of reference laboratory analyzers measuring identical specimens. Discussions about QA, or the overall approach to ensuring a quality laboratory result that encompasses preanalytical, analytical, and postanalytical aspects, and QC, or the day-to-day practices to ensure analytical quality, and recommendations for various aspects of QA and QC have sporadically appeared in the literature.²⁻⁴ Recently, interest in QA and QC for in-clinic laboratory testing has grown, and QC guidelines are recommended by various veterinary organizations.⁵⁻⁷

In this study, we examined the quality of in-clinic and reference laboratory biochemical testing using commercially available assayed QC materials. We sought to determine whether performance of in-clinic laboratories differed from reference laboratories, whether various brands/models of in-clinic analyzers differed from each other, and whether commercial reference laboratories differed from each other. We examined performance by several statistical methods and included (1) provision of quality requirements expressed as desirable total allowable error (TE_{a_d}), (2) determination of the ability to achieve TE_{a_d} goals, (3) calculation of sigma metrics for in-clinic and reference laboratory analyses, (4) evaluation of analytical components contributing to suboptimal performance, (5) definition of criteria for "in-clinic QC" and suitability of analytes based on these criteria, and (6) agreement of results of means obtained by in-clinic and reference laboratory analyzers.

Materials and Methods

Study design

Twenty veterinarians who were registered for an online course on "Quality Concepts for In-Clinic Biochemistry Testing" offered by the Veterinary Information Network (VIN) participated in the study. Each participant was provided with 3 assayed commercial QC materials (QCMs), each of which represented different concentrations of analytes (ChemtrakH, Liquid Assayed Chemistry Control, Lot numbers CHA11031, 11032, and 11033 or CHA12011, 12012, and 12013,

Microgenics Corporation, Fremont, CA, USA) and instructions for storage and handling of the materials. Nineteen participants analyzed the 3 QCMs for 5 consecutive days with their in-clinic analyzers, and 1 participant analyzed only QCM1). Nineteen of the participants also sent 1 or more QCMs to their usual reference laboratory for testing daily on the same 5 days. Three reference laboratories participated independently in the study and did not have corresponding laboratory data from an in-clinic analyzer for comparison. Their results are included with those for all reference laboratories.

The QCMs were analyzed using the most comprehensive biochemical profile available for the in-clinic analyzers and the standard small animal biochemical profiles for reference laboratories. Each participant provided information about the make and model of the analyzer, analytical methods used by the analyzer, and reference intervals for each analyte. Each participant confirmed that routine maintenance had been conducted according to manufacturer's instructions and that testing was conducted by trained personnel who routinely conduct the laboratory testing within the practice. All participants indicated that their instruments were, to the best of their knowledge, performing well. Similar information was obtained about the analyzers and methodology used by the reference laboratories. A form to record data was supplied to each participant and this was submitted at the end of the 5-day testing period. Raw data were recorded on a spreadsheet and analyzed as detailed below. When > 3 in-clinic analyzers or electrolyte analyzers of the same type, or > 3 reference laboratories of the same company were represented, we compared these by analyzer type or company and also included them in analysis of all in-clinic and reference laboratory analyzers.

Quality requirements (desirable TE_a)

We initially evaluated performance of analyzers by establishing TE_{a_d} quality requirements for each analyte, based on previously published data,⁸ and these methods were replicated for the analytes found in the 3 QCMs used in this study (Table 1). Identical TE_{a_d} was used for both in-clinic and reference laboratory analyzers to allow comparisons between these groups. Calculated total error (TE_c)⁹ was determined for each analyte in each QCM independently, and an analyte was considered as globally meeting quality requirements when $> 60\%$ of analyzers showed $TE_c < TE_{a_d}$. Conversely, results for an analyte were considered globally problematic when $< 60\%$ of

Table 1. Quality requirements for biochemical analytes expressed as desirable total allowable error for 3 levels of commercially available quality control materials.

| Analyte | Assayed Means* | | | TE _{a,d} QCM1 (%) | TE _{a,d} QCM2 (%) | TE _{a,d} QCM3 (%) |
|--------------------------|----------------|------|------|----------------------------|----------------------------|----------------------------|
| | QCM1 | QCM2 | QCM3 | | | |
| Total protein (g/L) | 67.8 | 52.1 | 36.8 | 10 | 10 | 10 |
| Albumin (g/L) | 42.9 | 35.0 | 27.5 | 25 | 25 | 11 |
| Urea (mmol/L) | 5.7 | 14.6 | 23.3 | 25 | 16 | 16 |
| Creatinine (μmol/L) | 88 | 330 | 581 | 17 | 17 | 17 |
| Total bilirubin (μmol/L) | 19 | 68 | 120 | 50 | 50 | 50 |
| Sodium (mmol/L) | 148 | 133 | 119 | 5 | 5 | 5 |
| Potassium (mmol/L) | 2.68 | 4.31 | 5.99 | 13 | 10 | 10 |
| Chloride (mmol/L) | 101 | 92 | 85 | 5 | 5 | 5 |
| Phosphorus (mmol/L) | 0.73 | 1.77 | 2.53 | 33 | 20 | 20 |
| Calcium (mmol/L) | 1.68 | 2.28 | 2.88 | 16 | 14 | 14 |
| Glucose (mmol/L) | 3.3 | 11.4 | 19.8 | 20 | 20 | 20 |
| ALP (U/L) | 42 | 231 | 420 | 25 | 25 | 25 |
| ALT (U/L) | 27 | 99 | 171 | 50 | 50 | 50 |
| Lipase (U/L) | 43.3 | 73.6 | 99.8 | 50 | 50 | 50 |
| Amylase (U/L) | 72 | 229 | 391 | 50 | 50 | 50 |

All assayed means presented were based on data in the manufacturer's package insert for the Olympus AU Series (Olympus, Center Valley, PA, USA). TE_{a,d} was determined from the procedure described by Farr & Freeman.⁸

*Microgenics Corporation (Fremont, CA), ChemtrakH, Liquid Assayed Chemistry Control, Lot numbers CHA11031, 11032, and 11033 and CHA12011, 12012, and 12013.

QCM indicates quality control material; 1–3, level of QCM; TE_{a,d}, desirable total allowable error (quality requirement).

analyzers (either in-clinic, reference laboratory, or both) met quality requirements, ie, when TE_c > TE_{a,d}.

Establishment of target mean values

Target mean values were obtained from the package inserts provided by the QCM manufacturer, and where possible, target values specified for the same instrument and method were used. In those instances when a corresponding instrument was not represented, the target mean was taken from an instrument using the same method. If several instruments used the same method, the method was not represented in the manufacturer's package insert, or methods were indicated to be proprietary, the Olympus AU series (Olympus America Inc., Center Valley, PA, USA) was used as the target instrument, and method for reference laboratories and the Vitros Chemistry system (Ortho Clinical Diagnostics, Raritan, NJ, USA) was used as the target instrument and method for in-clinic analyzers.

Indices of performance and quality

Several variables were calculated first:

1. Absolute bias, which is the absolute difference of the measured mean value and target mean value, and absolute %bias, which is the absolute value of

the % difference between the measured mean value and the target mean value

$$|Bias| = |measured\ mean - target\ mean|$$

$$|\%Bias| = \left| \frac{(measured\ mean - target\ mean)}{target\ mean} * 100 \right|$$

2. Mean, SD, and coefficient of variation (CV) for each analyte in each QCM
3. TE_c

$$TE_c = |\%Bias| + 2CV$$

To better visualize the contributions of the performance characteristics (bias, CV, and TE_c) to the TE_{a,d}, these variables were normalized to the respective TE_{a,d} and normalized ranges were graphed.

$$normalized\ range = \frac{upper\ or\ lower\ variable\ limit}{TE_{a,d}} * 100$$

From these variables, several indices of performance quality were calculated.

1. Indication of whether TE_c < TE_{a,d} ("yes" or "no" outcome). TE_c < TE_{a,d} was interpreted as indicating that the analyzer was suitable for veterinary laboratory testing based on desired interpretation of results. TE_c > TE_{a,d} was considered unsuitable.

2. Sigma metric¹⁰

$$\sigma = \frac{TEa_d - |\%Bias|}{CV}$$

Because the sigma metric has no upper limit, a nominal value of 50 was assigned to the sigma metric to minimize the skewness of the data whenever the analyte-specific CV = 0.0. A sigma metric ≥ 6.0 is considered to represent “world-class quality,” whereas a sigma metric < 3.0 is considered to represent a performance level that is unstable and unsuitable for commercial laboratory practice.¹⁰

3. For those analytes for each QCM that did not achieve a sigma metric ≥ 6.0 , a quality goal index (QGI) was calculated and interpreted as previously reported to determine if the inability to achieve a sigma metric ≥ 6.0 was primarily due to imprecision, inaccuracy, or both.¹¹

$$QGI = \frac{|\%Bias|}{1.5 * CV}$$

The QGI reflects the extent to which both bias and precision meet their respective quality goals without considering goals for CV and bias as separate entities. The QGI is interpreted as follows: QGI < 0.8 = imprecision and QGI > 1.2 = inaccuracy, whereas $0.8 < QGI < 1.2$ = both imprecision and inaccuracy.

We further calculated these indices for groups of in-clinic analyzers and reference laboratory chains that were represented by > 3 participants. These were IDEXX VetTest (IDEXX Laboratories, Inc., Westbrook, ME, USA), IDEXX Catalyst (IDEXX Laboratories), Abaxis VetScan for in-clinic analyzers (Abaxis, Union City, CA, USA), and Abaxis VetScan and IDEXX VetLyte for electrolyte (sodium and potassium) analyzers.

Analyte suitability for statistical “in-clinic QC”

For in-clinic use, a simple quality control rule, instead of a multirule, is desirable. This reduces the complexity of evaluation of data and minimizes the need for training for personnel conducting in-clinic testing. Similarly, restricting the number of QCMs is desirable to minimize the cost of statistical QC. Furthermore, the QC procedure should provide a high probability of error detection (P_{ed}) and low probability of false rejection (P_{fr}) when applied with regard to the actual performance of individual in-clinic instruments. This helps ensure that unacceptable performance is consistently identified when present and that false alerts regarding unacceptable performance are infrequent.

Therefore, we selected a 1_{3s} rule with 1 or 2 levels of QCM ($n = 1$ or $n = 2$) as being the most reasonable format for use in the practice laboratory. P_{ed} was set at $\geq 85\%$ and $\geq 90\%$ for 1 QCM and 2 QCM, respectively. Requirements for analyte bias and CV were obtained from the EZRules3 computer program for QC validation (EZRules3, Version 3; Westgard QC, 2005, <http://www.westgard.com/software.htm>). The P_{fr} achievable with these criteria and the 1_{3s} rule is 0.0%.¹² Specifications were extrapolated from Westgard’s EZRules3 and did not cover all possible combinations of bias and CV that would allow use of the specified QC Options, but provide a simplified reference for these choices (Table 2). Any analyte that did not have bias or CV within the specified criteria represented in this chart was considered unsuitable for in-clinic statistical QC.

Agreement between in-clinic and reference laboratory results

The mean for each of the analytes determined by the in-clinic analyzer was compared with the mean of the same analyte determined by the reference laboratory used by that participant. Four methods were used to determine the range of values on either side of the in-clinic mean within which the reference laboratory method should fall for the results to agree:

1. Expanded uncertainty of measurement (EUM). This was calculated with and without including the bias. This approach considers the square root of the variances inherent in both the in-clinic and reference laboratory means; it is considered likely to be the most inclusive because of consideration of the variances apparent for both the in-clinic and reference laboratories.¹³

$$EUM_{bias} = 2 * \sqrt{|Bias_{RL}|^2 + |Bias_{IC}|^2 + SD_{RL}^2 + SD_{IC}^2}$$

$$EUM_{w/o\ bias} = 2 * \sqrt{SD_{RL}^2 + SD_{IC}^2}$$

2. $Mean_{RL}$ falls within $Mean_{IC} \pm 2 * SD_{IC}$
3. $Mean_{RL}$ falls within $Mean_{IC} \pm 3 * SD_{IC}$
4. $Mean_{RL}$ falls within $Mean_{IC} \pm TEa_d$

Statistical analyses

Three comparisons were made using a chi-squared test: (1) the proportion of analytes from in-clinic and reference analyzers achieving an acceptable TE_c ($TE_c < TEa_d$), a sigma metric ≥ 6.0 , and a sigma

Table 2. Specifications for instrument performance to be suitable for “in-clinic QC” options.

| TE _{ad} (%) | 1 _{3s} Rule with <i>n</i> = 1 | 1 _{3s} Rule with <i>n</i> = 2 | TE _{ad} (%) | 1 _{3s} Rule with <i>n</i> = 1 | 1 _{3s} Rule with <i>n</i> = 2 |
|-------------------------|---|---|-------------------------|---|---|
| | (<i>P</i> _{ed} > 85%, <i>P</i> _{fr} = 0%) B%; CV% | (<i>P</i> _{ed} > 90%, <i>P</i> _{fr} = 0%) B%; CV% | | (<i>P</i> _{ed} > 85%, <i>P</i> _{fr} = 0%) B%; CV% | (<i>P</i> _{ed} > 90%, <i>P</i> _{fr} = 0%) B%; CV% |
| 50 | B < 5.0; CV < 7.5 | B < 5.0; CV < 8.5 | 16 | B < 2.5; CV < 2.2 | B < 2.5; CV < 2.6 |
| 50 | 5.0 < B < 10.0; CV < 7.0 | 5.0 < B < 10.0; CV < 7.7 | 16 | 2.5 < B < 5.0; CV < 1.8 | 2.5 < B < 5.0; CV < 2.1 |
| 50 | 10.0 < B < 15.0; CV < 6.8 | 10 < B < 15; CV < 7 | 16 | 5.0 < B < 10.0; CV < 1 | 5.0 < B < 10.0; CV < 1.1 |
| 50 | 15.0 < B < 20.0; CV < 5.0 | 15.0 < B < 20.0; CV < 5.8 | 16 | 10.0 < B < 12.5; CV < 0.6 | 10.0 < B < 12.5; CV < 0.65 |
| 50 | 20.0 < B < 25.0; CV < 4.0 | 20.0 < B < 25.0; CV < 4.8 | 14 | B < 2.0; CV < 2.0 | B < 2.0; CV < 2.3 |
| 50 | 25.0 < B < 30.0; CV < 3.0 | 25.0 < B < 30.0; CV < 3.8 | 14 | 2.0 < B < 4.0; CV < 1.7 | 2.0 < B < 4.0; CV < 1.9 |
| 33 | B < 5.0; CV < 4.6 | B < 5.0; CV < 5.4 | 14 | 4.0 < B < 6.0; CV < 1.4 | 4.0 < B < 6.0; CV < 1.5 |
| 33 | 5.0 < B < 10.0; CV < 3.9 | 5.0 < B < 10.0; CV < 4.3 | 14 | 6.0 < B < 8.0; CV < 1.0 | 6.0 < B < 8.0; CV < 1.1 |
| 33 | 10.0 < B < 15.0; CV < 3.0 | 10.0 < B < 15.0; CV < 3.3 | 14 | 8.0 < B < 10.0; CV < 0.7 | 8.0 < B < 10.0; CV < 0.75 |
| 33 | 15.0 < B < 20.0; CV < 2.1 | 15.0 < B < 20.0; CV < 2.4 | 14 | 10.0 < B < 12.0; CV < 0.3 | 10.0 < B < 12.0; CV < 0.35 |
| 33 | 20.0 < B < 25.0; CV < 1.3 | 20.0 < B < 25.0; CV < 1.5 | 13 | B < 2.0; CV < 1.8 | B < 2.0; CV < 2.0 |
| 33 | 25.0 < B < 30.0; CV < 0.5 | 25.0 < B < 30.0; CV < 0.6 | 13 | 2.0 < B < 4.0; CV < 1.5 | 2.0 < B < 4.0; CV < 1.6 |
| 25 | B < 5.0; CV < 3.3 | B < 5.0; CV < 3.8 | 13 | 4.0 < B < 6.0; CV < 1.1 | 4.0 < B < 6.0; CV < 1.2 |
| 25 | 5.0 < B < 10.0; CV < 2.6 | 5.0 < B < 10.0; CV < 2.8 | 13 | 6.0 < B < 8.0; CV < 0.7 | 6.0 < B < 8.0; CV < 0.85 |
| 25 | 10.0 < B < 15.0; CV < 1.6 | 10.0 < B < 15.0; CV < 1.9 | 11 | B < 2.0; CV < 1.5 | B < 2.0; CV < 1.7 |
| 25 | 15.0 < B < 20.0; CV < 0.8 | 15.0 < B < 20.0; CV < 0.9 | 11 | 2.0 < B < 4.0; CV < 1.2 | 2.0 < B < 4.0; CV < 1.3 |
| 20 | B < 2.5; CV < 2.8 | B < 2.5; CV < 3.3 | 11 | 4.0 < B < 6.0; CV < 0.8 | 4.0 < B < 6.0; CV < 0.9 |
| 20 | 2.5 < B < 5.0; CV < 2.5 | B < 5.0; CV < 2.8 | 11 | 6.0 < B < 8.0; CV < 0.5 | 6.0 < B < 8.0; CV < 0.55 |
| 20 | 5.0 < B < 10.0; CV < 1.7 | 10.0 < B < 15.0; CV < 1.9 | 10 | B < 2.0; CV < 1.3 | B < 2.0; CV < 1.5 |
| 20 | 15.0 < B < 17.5; CV < 0.4 | 15.0 < B < 17.5; CV < 0.45 | 10 | 2.0 < B < 4.0; CV < 1 | 2.0 < B < 4.0; CV < 1.1 |
| 17 | B < 2.0; CV < 2.6 | B < 2.0; CV < 2.8 | 10 | 4.0 < B < 6.0; CV < 0.6 | 4.0 < B < 6.0; CV < 0.7 |
| 17 | 2.0 < B < 4.0; CV < 2.2 | 2.0 < B < 4.0; CV < 2.5 | 10 | 6.0 < B < 8.0; CV < 0.3 | 6.0 < B < 8.0; CV < 0.35 |
| 17 | 4.0 < B < 6.0; CV < 1.8 | 4.0 < B < 6.0; CV < 2.1 | 5 | B < 1.0; CV < 0.6 | B < 1.0; CV < 0.7 |
| 17 | 6.0 < B < 8.0; CV < 1.5 | 6.0 < B < 8.0; CV < 1.7 | 5 | 1.0 < B < 2.0; CV < 0.5 | 1.0 < B < 2.0; CV < 0.6 |
| 17 | 8.0 < B < 10.0; CV < 1.2 | 8.0 < B < 10.0; CV < 1.3 | 5 | 2.0 < B < 3.0; CV < 0.3 | 2.0 < B < 3.0; CV < 0.35 |
| 17 | 10.0 < B < 12.0; CV < 0.8 | 10.0 < B < 12.0; CV < 0.9 | 5 | 3.0 < B < 4.0; CV < 0.1 | 3.0 < B < 4.0; CV < 0.15 |

Adapted from Westgard's EZRules3, Version 3, Westgard QC, 2005, <http://www.westgard.com/software.htm>.

TE_{ad} indicates desirable total allowable error; *P*_{ed}, probability of error detection; *P*_{fr}, probability of false rejection; B, absolute bias; CV, coefficient of variation.

metric < 3.0; (2) the proportion of analytes from reference laboratories operated by IDEXX and Antech Diagnostics (Irvine, CA, USA) achieving an acceptable TE_c, a sigma metric ≥ 6.0, and a sigma metric < 3.0; and (3) the proportion of analytes from the IDEXX VetTest, the IDEXX Catalyst, Abaxis VetScan analyzers, and IDEXX VetLyte analyzers achieving an acceptable TE_c, a sigma metric ≥ 6.0, and a sigma metric < 3.0.

Results

Participant information

Each participant analyzed between 11 and 15 analytes for each QCM. In some instances, participants used different analyzers to measure different analytes, eg, electrolytes with 1 analyzer and other analytes with another analyzer. Analytes measured included concentrations of total protein, albumin, urea, creatinine, total bilirubin, sodium, potassium, chloride, phosphorus, calcium, and glucose, and activities of alkaline

phosphatase (ALP), alanine aminotransferase (ALT), lipase, and amylase.

In-clinic analyzers used by participants in this study were the IDEXX VetTest (4), IDEXX Catalyst Dx (4), Abaxis VetScan (4), Heska SpotChem (Heska Corporation, Loveland, CO, USA) (2), Konelab 20 XT (Thermo Scientific, Waltham, MA, USA) (2), Medica EasyVet (Medica, Newtonville, MA, USA) (1), Hemagen Analyst III (Hemagen Diagnostics Inc., Columbia, MD, USA) (1), Medica Easy RA (Medica) (1), and ABX Horiba Pentra 400 (Horiba Medical, Irvine, CA, USA) (1). In-clinic electrolyte analyzers were the IDEXX VetLyte (7), Abaxis VetScan (4), Medica EasyLyte Plus (3), I-stat (Abbott Point of Care Inc., Princeton, NJ, USA) (2), Konelab 20XT (2), Nova Electrolyte Analyzer (Nova Biomedical, Waltham, MA, USA) (1), and ABX Horiba Pentra 400 (1). Commercial reference laboratories to which participants sent samples and the analyzers used were IDEXX Delta BC, Canada/Olympus; IDEXX Houston/Olympus; IDEXX Irvine/Olympus; IDEXX Sacramento/Olympus; IDEXX Massachusetts/

Olympus; IDEXX New Jersey/Olympus; IDEXX Portland/Olympus; Antech Chicago/Olympus (2); Antech New York Metro/Olympus (2); Antech Houston/Olympus; Antech Lake Success, NY/Olympus;(2); Antech Long Island, NY/Olympus; Antech San Jose/Olympus; True North Veterinary Diagnostics, Langley, BC/Siemens Dimension Max (Siemens USA, Deerfield, IL, USA); Animal Health Laboratory, University of Guelph/Roche Cobas 6000 (F. Hoffmann-La Roche Ltd, Basel Switzerland); and University of Agricultural Science, Uppsala, Sweden/Konelab 30 (2). Reference laboratories participating independently of an in-clinic laboratory included Phoenix Central Laboratory, Everett, WA/Siemens Advia (Siemens USA); Cornell University Clinical Pathology Laboratory/Roche Mod P (Roche-Hitachi, Indianapolis, IN, USA); and True North Diagnostics, Langley, BC/Siemens Dimension Max, which analyzed samples submitted by clinicians participating in the study and performed an independent analysis of all 3 QCM.

Quality indices

Bias and CV

The ranges of bias (Figure 1) and CV (Figure 2) as a proportion of the TE_{a_d} for each analyte for in-clinic analyzers for QCM1, QCM2, and QCM3 were calculated. Note that for both bias and CV, the lower limit was zero or > zero, depending on the analyte.

TE_{a_d} (quality) requirements

Reference laboratories achieved a $TE_c < TE_{a_d}$ more often than did in-clinic analyzers (%bias + 2CV: 377/457 [82%] analytes vs 571/819 [70%] analytes, respectively, $P = .0002$) (Table 3). Neither of the reference laboratory groups that we compared (IDEXX and Antech) achieved $TE_c < TE_{a_d}$ more or less often (%bias + 2CV: 101/122 [83%] vs 153/179 [85%], respectively, $P = .64$). None of the 3 models of in-clinic analyzers that we compared, the IDEXX VetTest, IDEXX Catalyst, and Abaxis VetScan, achieved $TE_c < TE_{a_d}$ more or less often than any other (%bias + 2CV: 107/164 [65%] vs 128/180 [71%] vs 116/150 [77%], respectively, $P = .06$).

All reference laboratories achieved quality requirements ($TE_c < TE_{a_d}$) using bias + 2CV criteria for 5 analytes within specific QCMs: albumin (QCM1), potassium (QCM1, QCM2, QCM3), amylase (QCM1, QCM3), urea (QCM2), and glucose (QCM3). Analytes within specific QCMs for which < 60% of reference laboratories met quality requirements included albumin (QCM3), creatinine (QCM1), total bilirubin

(QCM1), and lipase (QCM2). Overall, $\geq 60\%$ of reference laboratories achieved quality requirements ($TE_c < TE_{a_d}$) for 40/45 (89%) analytes when using % bias + 2CV criteria.

All in-clinic analyzers achieved quality requirements ($TE_c < TE_{a_d}$), using %bias + 2CV, for only a single analyte within a single QCM: calcium (QCM1). Those analytes for which < 60% of in-clinic laboratories met quality requirements were ALP (QCM1), creatinine (QCM1, QCM3), chloride (QCM1, QCM3), lipase (QCM1, QCM2, QCM3), and amylase (QCM2). Overall, $\geq 60\%$ of in-clinic laboratories achieved quality requirements ($TE_c < TE_{a_d}$) for 33/45 (73%) analytes.

Sigma metrics

Reference laboratory analyzers achieved a sigma metric ≥ 6.0 for a greater proportion of analytes than did in-clinic analyzers (231/457 [51%] vs 366/832 [44%], respectively, $P = .03$) (Table S1). Similarly, reference laboratory analyzers achieved a sigma metric < 3.0 for a lesser proportion of analytes than did in-clinic analyzers (130/457 [28%] vs 297/832 [36%], respectively, $P = .01$). There was no difference between the 2 main commercial reference laboratory groups (IDEXX and Antech) in the proportion of analytes achieving a sigma metric ≥ 6.0 (87/179 [49%] vs 60/122 [49%], respectively, $P = .92$) or the proportion of analytes achieving a sigma metric < 3.0 (52/179 [29%] vs 39/122 [29%], respectively, $P = .82$).

IDEXX Catalyst analyzers had a larger proportion of analytes achieving a sigma metric ≥ 6.0 than did either IDEXX VetTest or Abaxis VetScan analyzers (92/144 [63%] vs 60/125 [48%] vs 77/156 [49%], respectively, $P = .01$). IDEXX VetTest analyzers had a larger proportion of analytes achieving a sigma metric < 3.0 than did either IDEXX Catalyst or Abaxis VetScan analyzers (60/125 [48%] vs 33/144 [23%] vs 40/156 [26%], $P = .004$).

The percentages of in-clinic analyzers and reference laboratories with sigma metric < 6.0 and the percentages of analyzers within which the QGI identified inaccuracy, imprecision, or both inaccuracy and imprecision as being the primary contributors to failure to achieve world-class performance, defined as > 6.0 sigma, were examined (Table S1). More than 50% of in-clinic analyzers achieved a sigma metric ≥ 6.0 for phosphorus and glucose with all 3 QCM. Chloride analysis was the most problematic for in-clinic analyzers, with only 7% of analyzers achieving a sigma metric ≥ 6.0 for any of the QCMs and > 75% of analyzers demonstrating inaccuracy of measurement with any of the QCM. Conversely, there

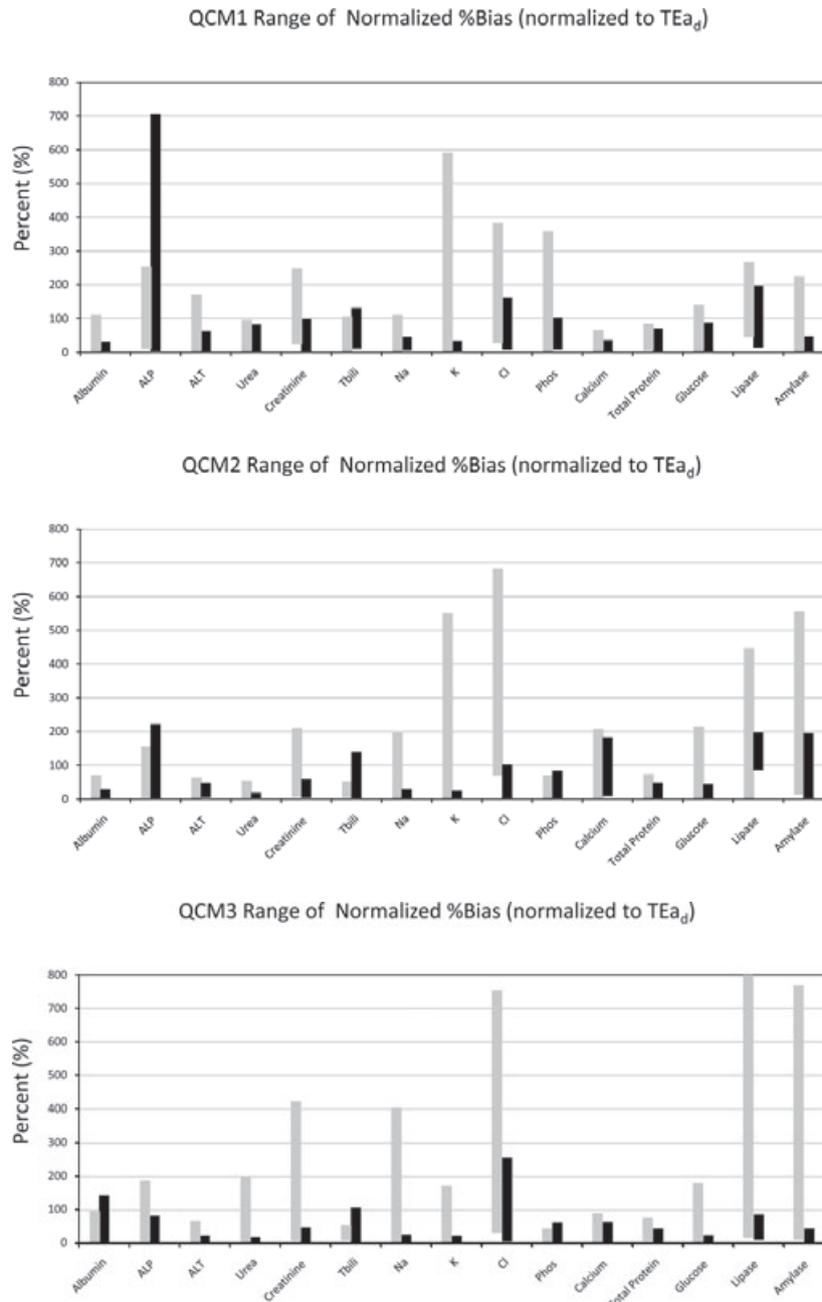


Figure 1. Range of %bias in the measurement of concentrations or activities of analytes in 3 levels of quality control materials (QCM) normalized to total allowable error (TE_{a,d}) for in-clinic (gray bars) and reference laboratory (black bars) analyzers. Median values are not displayed and are found midway along the normalized range for each analyte.

were 6 analytes for which > 50% of reference laboratories achieved a sigma metric ≥ 6.0 across all 3 QCM: ALT, potassium, phosphorus, calcium, glucose, and amylase. However, chloride analysis was similarly the most problematic for reference laboratory analyzers, with a maximum of 10% of reference laboratory analyzers achieving a sigma metric ≥ 6.0 for any of

the QCMs and > 50% of reference laboratory analyzers demonstrating imprecision of measurement with any of the QCM.

When comparing the 3 in-clinic analyzers for which sufficient data were provided across all 3 QCMs, the IDEXX VetTest performed worse than either the IDEXX Catalyst or Abaxis Vetscan. The IDEXX VetTest

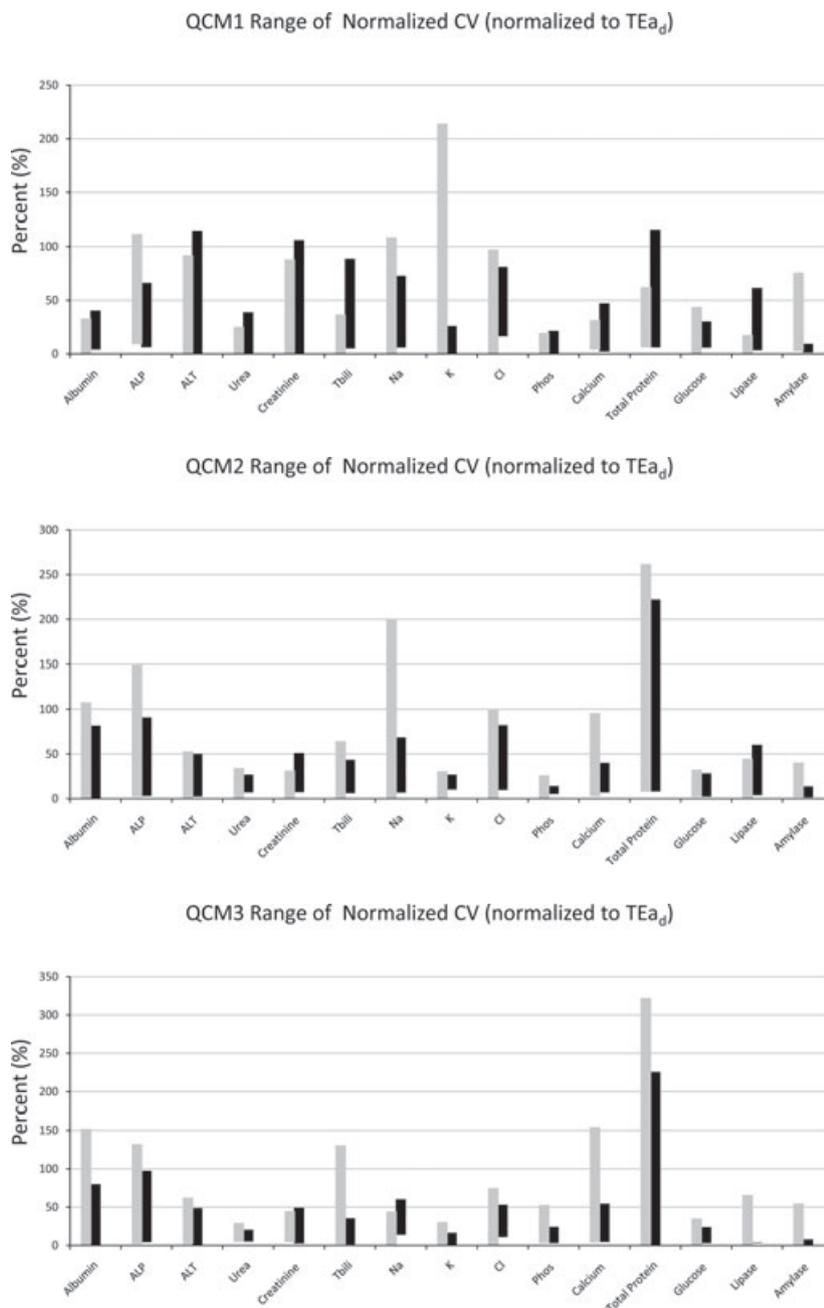


Figure 2. Range of CV in the measurement of concentrations or activities of analytes in 3 levels of quality control materials (QCM) normalized to total allowable error (TE_a) for in-clinic (gray bars) and reference laboratory (black bars) analyzers. Median values are not displayed and are found midway along the normalized range for each analyte.

achieved a sigma metric ≥ 6.0 for (1) albumin in 2/12 (17%) analyses, whereas both the IDEXX Catalyst and Abaxis VetScan achieved a sigma metric ≥ 6.0 in 6/12 (50%) analyses ($P = .17$); (2) total bilirubin in 4/12 (33%) analyses, compared with 11/12 (92%) and 8/12 (67%) analyses for the IDEXX Catalyst and Abaxis VetScan, respectively ($P = .007$); (3) urea in 4/12 (33%) analyses, compared with 10/12 (83%) and 7/12 (58%)

analyses for the IDEXX Catalyst and Abaxis VetScan, respectively ($P = .04$); and 40 glucose in 5/12 (42%) analyses, compared with 12/12 (100%) and 10/12 (83%) analyses for the IDEXX Catalyst and Abaxis VetScan, respectively ($P = .003$). Conversely, the IDEXX VetLyte achieved a sigma metric ≥ 6.0 for potassium in 11/24 (46%) analyses, compared with 1/12 (8%) for the Abaxis VetScan ($P = .03$). None of the

Table 3. Percentages of in-clinic and reference laboratories able to achieve desirable total allowable error (quality) requirements for up to 15 biochemical analytes.

| Analyte | QCM1 | | | | QCM2 | | | | QCM3 | | | |
|-----------------|--------|---|--------|---|--------|---|--------|---|--------|---|--------|---|
| | IC (n) | % IC with TE _{a_c} < TE _{a_d} | RL (n) | % RL with TE _{a_c} < TE _{a_d} | IC (n) | % IC with TE _{a_c} < TE _{a_d} | RL (n) | % RL with TE _{a_c} < TE _{a_d} | IC (n) | % IC with TE _{a_c} < TE _{a_d} | RL (n) | % RL with TE _{a_c} < TE _{a_d} |
| Total protein | 20 | 70 | 13 | 85 | 20 | 70 | 10 | 80 | 19 | 74 | 11 | 91 |
| Albumin | 19 | 89 | 13 | 100 | 19 | 95 | 10 | 90 | 19 | 42 | 11 | 36 |
| Urea | 20 | 90 | 13 | 92 | 20 | 90 | 10 | 100 | 20 | 80 | 11 | 100 |
| Creatinine | 20 | 10 | 13 | 54 | 20 | 50 | 10 | 70 | 20 | 55 | 11 | 82 |
| Total bilirubin | 16 | 63 | 12 | 42 | 17 | 94 | 9 | 67 | 17 | 88 | 10 | 90 |
| Sodium | 20 | 80 | 13 | 69 | 20 | 80 | 10 | 90 | 20 | 65 | 11 | 91 |
| Potassium | 20 | 65 | 13 | 100 | 20 | 75 | 10 | 100 | 20 | 85 | 11 | 100 |
| Chloride | 15 | 7 | 13 | 69 | 15 | 7 | 10 | 60 | 16 | 6 | 11 | 64 |
| Phosphorous | 18 | 83 | 12 | 92 | 18 | 72 | 9 | 89 | 18 | 94 | 10 | 90 |
| Calcium | 18 | 100 | 12 | 92 | 18 | 67 | 9 | 78 | 18 | 72 | 10 | 80 |
| Glucose | 20 | 85 | 13 | 92 | 20 | 95 | 10 | 90 | 20 | 95 | 11 | 100 |
| ALP | 18 | 39 | 13 | 62 | 18 | 78 | 9 | 78 | 18 | 72 | 11 | 82 |
| ALT | 20 | 65 | 13 | 92 | 20 | 95 | 10 | 90 | 20 | 95 | 11 | 91 |
| Lipase | 12 | 33 | 6 | 67 | 11 | 55 | 2 | 50 | 11 | 55 | 2 | 100 |
| Amylase | 17 | 76 | 6 | 100 | 17 | 47 | 4 | 75 | 17 | 71 | 5 | 100 |

TE_c = bias + 2CV.

QCM indicates quality control material; 1–3, level of QCM; TE_{a_d}, desirable total allowable (quality requirement); IC, in-clinic participants; RL, reference laboratory participants; TE_c, calculated total error; TE_{a_d}, desirable total allowable error; B, absolute bias; CV, coefficient of variation.

3 analyzers were able to achieve a sigma metric ≥ 6.0 for creatinine with any consistency: for IDEXX VetTest, 4/12 (33%); for IDEXX Catalyst, 3/12 (25%); and for Abaxis VetScan, 1/12 (8%) analyses. The IDEXX VetLyte was able to achieve a sigma metric ≥ 6.0 for chloride in only 2/24 (8%) analyses. The IDEXX VetTest and IDEXX Catalyst achieved a sigma metric ≥ 6.0 for total protein for 3/12 (25%) and 4/12 (33%) analyses, respectively, compared with 10/12 (83%) analyses for the Abaxis VetScan ($P = .007$). The IDEXX VetLyte achieved a sigma metric < 3.0 for chloride in 23/25 (92%) analyses. The IDEXX VetTest achieved a sigma metric < 3.0 more often for total protein more often than either the IDEXX Catalyst or Abaxis VetScan analyzers (9/13 [69%] vs 4/12 [33%] and 0/12 [0%]) analyses, $P = .001$). In addition, the IDEXX VetTest achieved a sigma metric < 3.0 for glucose more often than either the IDEXX Catalyst or Abaxis VetScan analyzers (5/13 [33%] vs 0/12 [0%] and 0/12 [0%], $P = .002$). Finally, each of these 3 in-clinic analyzers (the IDEXX VetTest, IDEXX Catalyst, and Abaxis VetScan) achieved a sigma metric < 3.0 for creatinine in $> 7/12$ (58–75%) analyses.

When comparing the 2 reference laboratories for which sufficient data were provided across all 3 QCMs (Antech and IDEXX), no differences were found between these laboratories for any of the analytes achieving a sigma metric ≥ 6.0 . Both Antech and IDEXX failed to achieve a sigma metric ≥ 6.0 for

either sodium (1/14 [7%] and 2/9 [22%], respectively) or chloride (1/14 [7%] and 0/9 [0%]) in most analyses. No differences were found between Antech and IDEXX laboratories for analytes achieving sigma metrics < 3.0 .

Quality goal index

When analytes on individual analyzers were unable to achieve a sigma metric ≥ 6.0 , the QGI was calculated to determine the source of the error as inaccuracy, imprecision, or both (Table S1). The inaccuracy of in-clinic laboratories was 2.51 times more common than either imprecision or the combination of imprecision and inaccuracy (4999 vs 1999 vs 1966, respectively) as the cause of the error. For reference laboratories imprecision was 6.56 times more common than inaccuracy (9677 vs 1474, respectively) and 15 times more common than the combination of both inaccuracy and imprecision (646) as the cause of the error.

Agreement between analyte means of in-clinic and reference laboratories

Agreement between analyte means obtained by in-clinic and reference laboratories using 5 methods of determining agreement ($EUM_{w/bias}$, $EUM_{w/o\ bias}$, reference laboratory means within ± 2 in-clinic SD from in-clinic mean, reference laboratory mean within ± 3 in-clinic SD from in-clinic mean, and reference

mean within $\pm TE_{a_d}$ from in-clinic mean) was analyzed (Table S2). Analysis using $EUM_{w/bias}$ and $\pm TE_{a_d}$ methods resulted in the highest mean percentages of analytes demonstrating agreement between in-clinic and reference laboratories (89% and 81%, respectively). Analytes with poor agreement across the range of analyte concentrations or activities included albumin, creatinine, urea, sodium, phosphorus, calcium, ALP, and ALT. Total protein, glucose, chloride, potassium, and total bilirubin showed good agreement across the range of analyte concentrations.

Ranges of bias, CV, and TE_c (bias + 2CV) for in-clinic and reference laboratory groups

The ranges of bias, CV, and TE_c varied relatively widely among in-clinic analyzers of the same and different types (Tables 4 and S3) and between the reference laboratories (Tables 4 and S4). These ranges varied across the values represented by the 3 levels of QCMs. In general, the IDEXX Catalyst and Abaxis VetScan had smaller ranges for these variables compared with the IDEXX VetTest (Table S3). The ranges of bias, CV, and TE_c normalized to TE_{a_d} for all 3 QCMs were calculated (Figures 1–3).

Suitability for “in-clinic QC” options

Mean, median, maximum, and minimum percentages of analytes suitable for statistical QC according to the definition of “in-clinic QC” and the 2 QC options were calculated (Table S5). The highest mean and median percentages of analytes suitable for statistical QC were obtained for QCM2 and QCM3 for both in-clinic and reference laboratories. Some in-clinic analyzers and reference laboratories had large percentages of analytes (up to 80% and 86% of analytes, corresponding to 12 of 15 analytes and 13 of 15 analytes, for in-clinic and reference laboratory analyzers, respectively) suitable for statistical QC. However, there were some in-clinic and reference laboratories for which very few analytes were suitable for statistical QC based on these criteria, with the minimum percentages of analytes suitable for statistical QC falling as low as 7% and 0% for in-clinic and reference laboratories, respectively.

Discussion

Our study provides a comprehensive assessment of in-clinic biochemistry analyzer performance and an independent assessment of biochemistry analyzer

performance of commonly used US reference laboratories. Our results suggest a great heterogeneity of performance of similar in-clinic analyzers, with many individual in-clinic analyzers failing to meet “world-class performance” standards for a range of analytes, unbeknownst to the clinician performing the analysis. Such inadequacies in analyzer performance might impact clinical decision-making, which is based on the assumption that results provided on a specific patient can be believed.

TE_{a_d} is an expression of the total amount of error tolerable based on the desired interpretation of a laboratory test result and represents all potential sources of error in a measurement that could have an impact on clinical decisions that are based on a test result. TE_{a_d} varies depending on the analyte in question and the clinical decision threshold(s) of that analyte. We determined quality requirements in the form of TE_{a_d} for various analytes in this study based on a prior study that used the veterinary literature and expert group opinion⁸; these TE_{a_d} provide a standard by which instrument performance can be judged. Determination of quality requirements presents a challenge for any veterinary laboratory. There are only limited recommendations for veterinary biochemical testing based on biologic variation, current performance capabilities, and expert opinions^{2,14,15} compared with those in human laboratory medicine.^{16–25} Challenges encountered with establishment of TE_{a_d} , a concept relatively new to veterinary medicine, include that clinical decision thresholds vary by species and clinician experience and often are not clearly defined in the literature. Furthermore, decision limits are somewhat subjectively defined and will vary depending on the way the laboratory values are used, eg, to evaluate healthy vs diseased animals or referral vs primary care animal patients). Finally, TE_{a_d} is dependent on the intrinsic instrument performance and may need adjustment or adaptation to specific analyzers that cannot meet more commonly adopted TE_{a_d} . The quality requirements presented in this report provide a standard for comparison of in-clinic and reference laboratory performance whether or not they are suitable for every veterinary laboratory or situation. TE_{a_d} was based on stringency criteria for in-clinic analyzers and then was used for both in-clinic and reference laboratory analyzers to fairly compare performance, given the anticipated superior equipment and quality assurance programs of most reference laboratories; however, a more stringent set of TE_{a_d} may be more appropriate for evaluation of reference laboratory performance. If more stringent criteria had been

Table 4. Ranges of coefficient of variation, bias, and calculated total error for concentrations or activities of 15 analytes measured in 3 levels of quality control materials for all in-clinic and reference laboratories.

| Analyte | QC Material | All In-Clinic Laboratories | | | All Reference Laboratories | | |
|-----------------|-------------|------------------------------|--------------|---------------------------|------------------------------|-------------|---------------------------|
| | | Coefficient of Variation (%) | Bias (%) | TE _c (B + 2CV) | Coefficient of Variation (%) | Bias (%) | TE _c (B + 2CV) |
| Total protein | QCM1 | 0.61–6.23 | 0.43–8.47 | 3.51–21.23 | 0.61–11.58 | 0.14–7.08 | 2.54–34.88 |
| | QCM2 | 0.77–26.18 | 0.19–7.40 | 3.47–83.55 | 0.80–22.21 | 0.18–4.85 | 3.15–69.35 |
| | QCM3 | 0.00–32.22 | 0.38–7.44 | 0.76–99.97 | 0.00–22.57 | 0.00–4.62 | 0.76–69.75 |
| Albumin | QCM1 | 0.00–8.29 | 0.69–27.98 | 2.10–37.19 | 1.01–10.13 | 0.23–8.01 | 3.37–30.63 |
| | QCM2 | 0.00–26.96 | 0.41–17.71 | 2.52–84.74 | 0.00–20.46 | 0.28–7.40 | 0.28–62.21 |
| | QCM3 | 0.00–38.07 | 0.00–23.48 | 0.00–117.01 | 0.00–20.01 | 0.00–36.14 | 0.00–60.37 |
| Urea | QCM1 | 0.00–6.30 | 0.12–23.88 | 3.95–39.45 | 0.00–9.72 | 0.56–20.66 | 0.52–38.31 |
| | QCM2 | 0.00–8.59 | 0.51–13.68 | 3.95–39.45 | 1.12–6.76 | 0.43–4.64 | 5.45–22.34 |
| | QCM3 | 0.72–7.36 | 0.32–49.52 | 5.88–58.76 | 0.82–5.12 | 0.43–4.97 | 3.02–18.64 |
| Creatinine | QCM1 | 0.00–15.00 | 4.13–42.22 | 11.14–70.74 | 0.00–18.05 | 0.00–16.73 | 0.00–70.16 |
| | QCM2 | 0.00–5.34 | 0.97–35.60 | 6.68–37.60 | 1.24–8.68 | 0.72–10.00 | 7.86–30.92 |
| | QCM3 | 0.74–7.64 | 0.84–21.60 | 5.06–39.00 | 0.45–8.36 | 0.19–8.32 | 3.51–29.80 |
| Total bilirubin | QCM1 | 0.00–18.45 | 0.06–53.13 | 0.06–101.45 | 2.59–44.41 | 5.14–66.08 | 26.76–183.5 |
| | QCM2 | 0.00–32.15 | 2.29–26.37 | 12.33–98.73 | 3.03–21.79 | 1.16–69.97 | 20.37–135.3 |
| | QCM3 | 0.69–65.37 | 3.18–26.47 | 9.51–201.10 | 0.63–17.79 | 1.27–54.06 | 9.33–107.42 |
| Sodium | QCM1 | 0.00–5.44 | 0.00–5.60 | 0.00–17.22 | 0.29–3.64 | 0.27–2.30 | 1.72–11.33 |
| | QCM2 | 0.00–9.99 | 0.00–9.86 | 0.75–39.84 | 0.34–3.43 | 0.00–1.50 | 1.16–11.81 |
| | QCM3 | 0.00–55.91 | 0.00–20.16 | 0.00–187.89 | 0.69–3.02 | 0.00–1.34 | 2.24–9.39 |
| Potassium | QCM1 | 0.00–27.84 | 0.00–76.95 | 2.62–95.79 | 0.00–3.39 | 0.38–4.48 | 2.36–13.87 |
| | QCM2 | 0.00–4.00 | 0.11–71.62 | 0.22–82.79 | 1.02–3.51 | 0.24–3.24 | 3.84–12.63 |
| | QCM3 | 0.00–3.97 | 0.01–22.15 | 0.17–33.61 | 0.00–2.14 | 0.00–3.06 | 0.00–7.94 |
| Chloride | QCM1 | 0.00–4.87 | 1.36–19.17 | 5.46–26.74 | 0.83–4.06 | 0.40–8.12 | 3.20–20.31 |
| | QCM2 | 0.00–4.96 | 3.49–34.13 | 5.45–49.02 | 0.49–4.12 | 0.00–5.14 | 2.44–17.49 |
| | QCM3 | 0.00–3.75 | 1.42–37.70 | 4.34–44.05 | 0.54–2.66 | 0.34–12.86 | 2.22–18.51 |
| Phosphorus | QCM1 | 0.00–6.47 | 0.02–118.08 | 1.77–127.06 | 0.00–7.08 | 2.65–33.78 | 7.67–45.81 |
| | QCM2 | 0.00–8.60 | 0.06–23.28 | 3.10–37.13 | 1.04–4.70 | 0.45–27.80 | 4.65–38.46 |
| | QCM3 | 0.57–17.51 | 0.11–13.86 | 3.25–61.69 | 0.59–8.03 | 0.63–20.96 | 3.25–44.21 |
| Calcium | QCM1 | 0.67–5.06 | 0.10–10.57 | 3.69–17.60 | 0.33–7.56 | 0.00–5.80 | 2.00–25.03 |
| | QCM2 | 0.35–15.34 | 0.37–33.04 | 3.33–46.38 | 0.98–6.42 | 1.27–29.28 | 5.47–40.75 |
| | QCM3 | 0.52–24.72 | 0.05–14.03 | 2.45–77.17 | 0.64–8.74 | 0.17–10.45 | 3.92–35.05 |
| Glucose | QCM1 | 0.00–8.76 | 0.13–28.18 | 2.70–32.77 | 1.18–6.06 | 0.42–17.65 | 4.86–27.32 |
| | QCM2 | 0.40–6.48 | 0.20–42.87 | 1.82–45.22 | 0.43–5.68 | 0.00–8.95 | 2.34–25.98 |
| | QCM3 | 0.16–7.06 | 0.22–35.74 | 2.92–41.72 | 0.58–4.78 | 0.08–5.05 | 2.92–19.37 |
| ALP | QCM1 | 2.28–27.96 | 2.65–63.51 | 10.22–143.93 | 1.57–16.57 | 0.95–176.59 | 10.45–190.4 |
| | QCM2 | 0.56–37.24 | 1.01–38.79 | 4.33–148.29 | 0.76–22.77 | 0.35–55.53 | 7.66–80.15 |
| | QCM3 | 0.48–33.05 | 0.78–46.52 | 4.14–130.92 | 1.08–24.37 | 0.35–55.53 | 4.14–88.71 |
| ALT | QCM1 | 0.00–46.01 | 1.13–84.91 | 4.08–38.31 | 0.00–57.34 | 0.32–31.51 | 7.86–203.52 |
| | QCM2 | 0.00–26.43 | 0.00–31.67 | 8.44–85.59 | 1.18–25.13 | 2.24–24.02 | 5.79–88.56 |
| | QCM3 | 0.81–31.18 | 0.70–32.27 | 5.21–97.98 | 0.45–24.40 | 2.20–12.10 | 5.56–85.30 |
| Lipase | QCM1 | 0.00–8.88 | 22.44–133.52 | 33.78–138.99 | 1.67–30.84 | 6.70–98.66 | 16.66–191.2 |
| | QCM2 | 0.00–22.41 | 0.65–223.32 | 3.30–237.08 | 1.93–30.11 | 42.79–99.1 | 48.57–189.4 |
| | QCM3 | 0.00–32.93 | 7.05–1263.00 | 16.91–1277.00 | 1.68–2.08 | 5.81–44.00 | 10.86–50.00 |
| Amylase | QCM1 | 1.23–37.91 | 0.84–112.00 | 4.81–183.91 | 0.62–4.67 | 1.12–23.68 | 2.97–37.69 |
| | QCM2 | 0.77–20.25 | 6.47–278.26 | 10.02–283.57 | 0.68–6.93 | 0.09–97.94 | 3.72–118.73 |
| | QCM3 | 0.13–27.39 | 4.11–384.00 | 8.10–138.00 | 0.13–3.96 | 0.15–22.00 | 3.55–28.00 |

The target mean used to calculate bias was derived from the manufacturer's package insert for either the same instrument or the same method used by participant analyzers.

TE_c = bias + 2CV.

TE_c indicates calculated total error; B, absolute bias; CV, coefficient of variation; QCM, quality control material; 1–3, level of QCM.

applied to reference laboratory data, performance would have probably been worse. Additional studies of current instrument capability and surveys of

expert opinion may be required to establish acceptable global quality goals for laboratory testing in veterinary medicine.

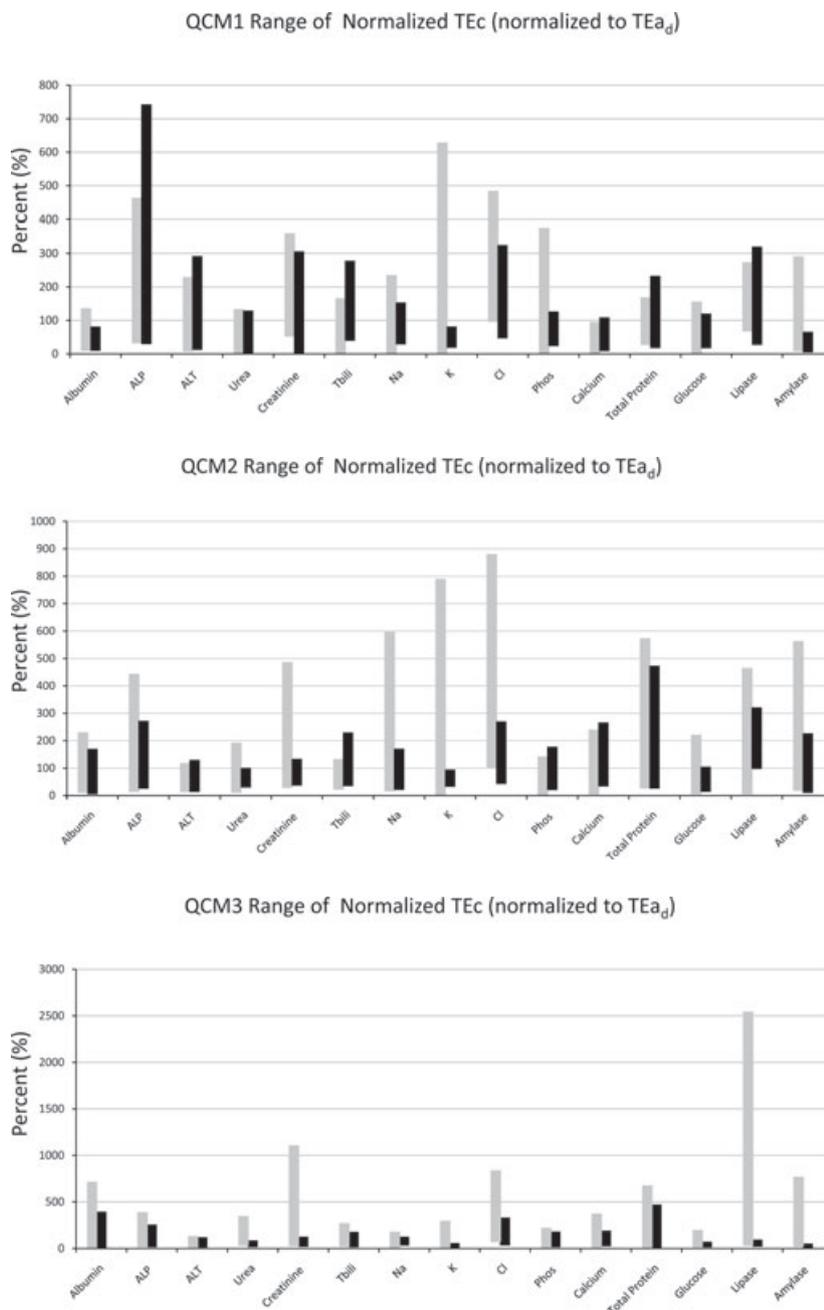


Figure 3. Range of total calculated error (TE_c) in the measurement of concentrations or activities of analytes in 3 levels of quality control materials (QCM) normalized to total allowable error (TE_{a_d}) for in-clinic (gray bars) and reference laboratory (black bars) analyzers. Median values are not displayed and are found midway along the normalized range.

Overall, the quality of in-clinic and reference laboratory testing apparent in this study is reflected by the ability of the majority of in-clinic laboratory and reference laboratory analyzers to achieve TE_c < TE_{a_d} for 73% and 91%, respectively, of analytes measured. When more stringent quality assessment was applied, world-class performance was achieved routinely across

all 3 QCMs for only 24% and 44% of analytes with in-clinic and reference laboratory analyzers, respectively. Reference Laboratory analyzers were less likely to have unstable performance (reflected by a sigma metric < 3.0) than in-clinic analyzers, although measurement of individual analytes was problematic for many analyzers. The superior performance of reference

laboratory analyzers, with more stringent quality assurance protocols and more sophisticated equipment, is expected when compared with smaller, less sophisticated in-clinic instruments that are not calibrated daily. IDEXX VetTest analyzers tended to perform worse than either IDEXX Catalyst or Abaxis VetScan analyzers. However, no differences were found in performance of the 2 reference laboratories that we were able to compare.

Both TE_{a_d} and TE_c rely on estimates of bias and CV; thus, inaccuracies in either bias or CV will be reflected in inaccurate estimates of TE_{a_d} and TE_c . Despite some differences in methods between the in-clinic analyzers and those used to derive the target means for each analyte, both of which are used to calculate bias, assessment of bias in this study is likely to be more accurate than assessment of CV, as the assayed control materials provided a standard for comparison and calculation of bias. CVs are expected to have a wider confidence interval, reflecting the small number of data points used in its calculation. The TE_c is the least confident estimate, as it is based on the combination of both %bias and $2CV$. However, we are confident that our results give an accurate reflection of the contributions of bias and CV to TE_c based on analysis of each of these separately and the findings from sigma metric and QGI calculations.

Bias should be estimated by comparing the mean concentration or activity of an analyte obtained from an in-clinic analyzer with the target mean obtained from the identical instrument as provided by the manufacturer of the assayed QCM for each QCM batch. Unfortunately, no target means exist for veterinary in-clinic analyzers, because manufacturers of assayed QCMs do not evaluate these instruments. Thus, the bias observed for in-clinic analyzers in this study may be lower if target means using identical analyzers had been provided by the QCM manufacturer.

The reference laboratories that participated in this study probably would have obtained different estimates of CV, bias, and TE_c based on their internal evaluation of quality control materials. However, results of this study provide an independent source of data that reflect quality of testing and comparison with in-clinic analyses and represent a “snapshot” of laboratory performance, similar to that obtained by analyzing an individual sample from either an animal patient or as part of a proficiency testing/quality assurance program. Confidence in the accuracy and validity of such a “snapshot” can best be obtained by serial analyses over time to truly determine the “health status” of the animal patient or veterinary laboratory.

In using $TE_c < TE_{a_d}$ as the indicator of performance, the magnitude of the difference between TE_c and TE_{a_d} for each analyzer was not considered; rather, the difference was simply evaluated as a dichotomous variable (pass/fail). It is possible that analyzers for which TE_c was marginally $> TE_{a_d}$ for an analyte still provide results with a tolerable degree of error. Furthermore, we used QCMs with varying concentrations of analytes. With many analytes, such as ALT, bilirubin, and creatinine, errors associated with low activities or concentrations are of little clinical importance; for example, for ALT of 15 U/L, an error of even 100% is clinically irrelevant. However, error \geq the upper reference limit may be clinically important, and an error of 100% at an ALT of 150 U/L is intolerable. In general, concentrations or activities of analytes in QCM1 were low, and performance with this QCM was generally worse than with either QCM2 or QCM3. Thus, we may have underestimated the true clinical performance of both in-clinic and reference laboratory analyzers. Based on our findings, reference laboratories and clinics that use statistical QC methods in their analysis of performance are best served in most cases using QCMs that include higher concentrations or activities of analytes. However, before using statistical QC to assess performance, each in-clinic or reference laboratory must determine the suitability of evaluating performance in this way.

World-class performance as judged by sigma metrics was achievable for both in-clinic and reference laboratories for multiple analytes. However, for some analytes, performance was categorized as likely to be unstable and unsuitable for clinical interpretation; improved performance would increase confidence in interpretation and medical decision-making. For analytes that fell between world-class and unacceptable performance, the QGI was evaluated to determine the source of the error. It is apparent that both in-clinic and reference laboratory instruments could benefit from decreases in inaccuracy and imprecision. When inaccuracy is identified as the most limiting factor for in-clinic instruments to achieve a sigma metric ≥ 6.0 , manufacturers are encouraged to improve software updates by use of materials traceable to a common international standard for calibration and to improve performance across a range of clinically important values. When imprecision is identified as the most limiting factor for reference laboratories to achieve a sigma metric ≥ 6.0 , the reference laboratories should improve precision, because incorporating a simple correction factor for a specific analyzer is not possible when imprecision is the cause of suboptimal performance. Improvements in accuracy and precision will

be helpful for analysis of many analytes by both in-clinic and reference laboratories.

In-clinic analyzers of the same and different types displayed a high degree of variation. Similarly, individual reference laboratories displayed considerable variation both across all reference laboratories and even across laboratories belonging to the same corporation (ie, IDEXX or Antech). Variation among reference laboratories using the same or different methods for various analytes is reflected in results obtained from external quality assurance/proficiency testing programs. Many reference laboratories invest considerable time and effort in evaluation of external quality assurance/proficiency testing results and in review of internal quality control performance to minimize bias and imprecision. This minimization can be achieved in a number of ways, including calibration adjustments, correction factors, instrument maintenance and cleaning, and use of particular reagents or chemical methods in measuring various analytes. These options are not available for most in-clinic instruments and may not be recommended when personnel have limited training in laboratory instrumentation and quality control. Therefore, improvements in instrument performance and minimization of variation among instruments rest with the manufacturers. In human and some veterinary laboratories, information about instrument performance is provided by the manufacturer, and careful evaluation of an instrument is undertaken prior to its use in analyzing patient samples. However, manufacturers of in-clinic biochemical analysers for veterinary testing currently do not provide minimum performance specifications for the instruments, and, thus, the preliminary information about their performance and establishment of performance benchmarks provided by this study may be useful.

Owing to the high degree of variation among in-clinic and reference laboratory analyzers of the same and different types, biochemical analysis is often repeated when animal patients are presented to referral facilities, which permits the referral clinician to have confidence in the laboratory results. For both human and animal laboratory testing, providing comparable results of similar quality across multiple laboratories is a significant challenge that is currently being addressed in human testing in several countries.^{26–29} Comparing results or using common reference intervals across multiple laboratories is a complex process that is not widely appreciated in veterinary medicine. Discussions with veterinary practitioners by the authors suggest that assessments of veterinary in-clinic instrument performance are not routinely provided by manufacturers, and reference intervals generated by

manufacturers are usually applied by practitioners without validating transference of the intervals or considering the number and characteristics of reference individuals used to establish the intervals.

Although in many cases intuitive assessment of results obtained by the in-clinic and reference laboratory will determine whether the results are roughly comparable, it is important to note that multiple methods were used in the study; caution is needed when comparing results obtained by different methods in the in-clinic and reference laboratory, and knowledge of the methods used for analysis is imperative. Most participants in this study were not aware of the methods used by the in-clinic or reference laboratory analyzers; in some cases, information about methodology was either not available or was proprietary. However, details of methodology should be readily available to instrument users to facilitate data comparisons and determine suitable peer groups for evaluation of performance in external quality assurance programs. EUM with and without bias was included, as it accounts for variances contributed by both the in-clinic and reference laboratory analyzers and probably reflects the most accurate assessment of agreement. However, such calculations are unlikely to be attempted by in-clinic laboratories owing to their complexity and because information about bias and SD for analysis of analytes by reference laboratories is unlikely to be readily available. Use of $\pm 2SD$ or $\pm 3SD$ as a means to determine the range of values in agreement is a relatively simple method, and such analysis is commonly used for assessment of performance and agreement with a peer group in external quality assurance and proficiency testing; however, more results were found to agree with $\pm TEa_d$ when this was used as the method of assessing agreement.

When evaluating reference laboratory performance, a minimum of 20 data points is usually recommended to determine bias, SD, and CV, as this will produce narrower confidence limits than data generated over 5 days, which was the design of this study. However, we sacrificed the slight improvement in these estimates that may have been achieved over 20 days for the willingness of in-clinic laboratories to invest the time, effort, and expense needed to characterize instrument performance for clinical use and capacity to generate meaningful data. Further evaluation of animal specimens from the species of interest over a range of clinically important values will also be needed for full characterization of instrument performance. However, use of assayed QCMs as a means to evaluate instrument performance provides a readily available source of well-characterized material that is

stable over a reasonable period of time for conducting such a study.

Measurement of several analytes was problematic for many analyzers and might require additional attention. The poor performance of nearly all laboratories in measuring chloride was an unexpected finding. We have no explanation for the universally poor performance of chloride measurements in this study, but suggest that reference and in-clinic laboratories carefully investigate the performance of their chloride testing.

This study represents the current state-of-the-art in analyzing performance of in-clinic and reference laboratory analyzers and provides a baseline for future comparisons and ongoing improvements in the quality of veterinary laboratory testing. The findings support the need for further improvement in the quality of veterinary biochemical testing by in-clinic and reference laboratories.

Acknowledgments

The authors recognize the advice and guidance of Dr. Kathleen Freeman and thank the VIN Foundation for funding this study. We thank the participants in this study, listed in alphabetical order: Bohn S, Bufanio K, Caffarini J, Carreras T, Chody J, Doherty J, Grus W, Hensley E, Heubner D, Hill M, Hillstrom A, Kreeger A, Lewis J, Lundgren L, Mandocdoc PB, Oberg J, Pederson S, Pinerd W, Schengrund K, Stokol T, Thompson S, and Upton G. We also thank Fisher Scientific for consultation regarding laboratory performance and help in shipping the quality control materials to the participants.

Disclosure: Mark Rishniw is a consultant for Merial, and Mark Rishniw, Paul Pion, and Tammy Maher are employees of the Veterinary Information Network (VIN). This study was funded by the VIN Foundation.

References

1. CLIA-88 proficiency testing (PT) criteria for acceptable performance. *Federal Register*. 1992;57:7002–7186.
2. Freeman KP, Evans EW, Lester S. Quality control for in-hospital veterinary laboratory testing. *J Am Vet Med Assoc*. 1999;215:928–929.
3. Weiser MG, Thrall MA. Quality control recommendations and procedures for in-clinic laboratories. *Vet Clin North Am Small Anim Pract*. 2007;7:237–244.
4. Mitzner BT. It's time for in-house quality assurance [Guest Editorial]. *J Am Anim Hosp Assoc*. 2002;28:12–13.
5. Royal College of Veterinary Surgeons Practice Standards Scheme Manual. Available at: <http://www.rcvsv.org.uk/document-library/practice-standards-manual/>. Accessed November 21, 2011.
6. The College of Veterinarians of Ontario Guideline: Ordering, Performing and Interpreting Laboratory Tests in Veterinary Clinical Practice. Available at: <http://www.cvo.org/uploadattachments/Laboratorytestsguideline.pdf>. Accessed November 21, 2011.
7. American Animal Hospital Association, Accreditation. Available at: <http://www.aahanet.org/accreditation/index.aspx>. Accessed November 21, 2011.
8. Farr AJ, Freeman KP. Quality control validation, application of sigma metrics, and performance comparison between two biochemistry analyzers in a commercial veterinary laboratory. *J Vet Diagn Invest*. 2008;20:536–544.
9. Westgard JO. *Basic Method Validation*. 3rd ed. Madison, WI: Westgard QC, Inc.; 2008:130–131.
10. Westgard JO. *Six Sigma Quality Design and Control*. Madison, WI: Westgard QC, Inc.; 2001:49.
11. Parry DM. The quality goal index – its use in benchmarking and improving sigma quality performance of automated analytic tests. Guest Essay. Available at: <http://www.westgard.com/quality-goal-index.htm>. Accessed November 21, 2011.
12. Westgard JO, Groth T, Aronsson T, Falk H, de Verdier CH. Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection. *Clin Chem*. 1977;23:1857–1867.
13. White G. Basics of estimating measurement uncertainty. *Clin Biochem Rev*. 2008;29(Suppl):S53–S60.
14. Freeman KP, Gruenewaldt J. Quality control validation in veterinary laboratories. *Vet Clin Pathol*. 1999;28:150–155.
15. Jensen AL, Kielgaard-Hansen M. Method comparison in the clinical laboratory. *Vet Clin Pathol*. 2006;35:276–286.
16. Fraser CG, Hyltoft-Petersen P, Ricos C, Haeckel R. Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry. *Eur J Clin Chem Biochem*. 1992;30:311–317.
17. Westgard JO, Hyltoft-Petersen P, Wiebe DA. Laboratory process specifications for assuring quality in the U.S. National Cholesterol Education Program (NCEP). *Clin Chem*. 1991;37:656–661.
18. Westgard JO, Wiebe DA. Cholesterol operational process specifications for assuring the quality required by CLIA proficiency testing. *Clin Chem*. 1991;37:1938–1944.
19. Westgard JO, Seehaver JJ, Barry PL. Allowable imprecision for laboratory tests based on clinical and analytical test outcome criteria. *Clin Chem*. 1994;40:1909–1914.
20. Westgard JO, Seehaver JJ, Barry PL. European specifications for imprecision and inaccuracy compared with

- operating specifications that assure the quality required by U.S. CLIA proficiency testing criteria. *Clin Chem.* 1994;40:1228–1232.
21. Skendzel LP, Barnett RN, Platt R. Medically useful criteria for analytic performance of laboratory tests. *Am J Clin Pathol.* 1985;83:200–205.
 22. Fraser CG. Biological variation in clinical chemistry. An update: collated data, 1988–1991. *Arch Pathol Lab Med.* 1992;116:916–923.
 23. Fraser CG. The application of theoretical goals based on biological variation data in clinical chemistry. *Arch Pathol Lab Med.* 1988;112:404–415.
 24. Hyltoft-Petersen P, Ricos C, Stockl D, et al. Proposed guidelines for the internal quality control of analytical results in the medical laboratory. Discussion paper from the members of the external quality assessment (EQA) working group A on the analytical quality goals in laboratory medicine. *Eur J Clin Chem Biochem.* 1996;34:983–999.
 25. Rilibak-German Guidelines for Quality. Available at: <http://www.westgard.com/rilibak.htm>. Accessed November 21, 2011.
 26. Jones GR, Baker A, Tate J, Lim CF, Robertson K. The case for common reference intervals. *Clin Chem Rev.* 2004;25:99–104.
 27. Peterson H, Rustad P. Prerequisites for establishing common reference intervals. *Scand J Clin Lab Invest.* 2004;64:285–292.
 28. Horowitz GL, Altaie S, Boyd J, et al. Clinical and Laboratory Standards Institute (CLSI). Defining, establishing and verifying reference intervals in the clinical laboratory; approved guideline – 3rd ed. CLSI document C28-A3 2008;28(3). Available at: <http://www.clsi.org/source/orders/free/c28-a3.pdf>. Accessed November 21, 2011.
 29. Ceriotti F. Prerequisites for use of common reference intervals. *Clin Biochem Rev.* 2007;28:115–121.

Appendix: Glossary of Terms

1. Total Allowable Error (TE_a)

A quality characteristic of a test or method that is based on the performance characteristics of imprecision and bias. It represents the overall or total error that may occur in a test result due to both the imprecision (random error) and inaccuracy (systematic error) of the measurement procedure. Total error describes the quality of a test result, providing a worst case estimate of how large the error might be in a single measurement. The concept of total error may be used to: (1) set quality requirements (**desirable total allowable**

error) for test performance and (2) calculate from measurements, obtained from various instruments and methods, a quantitative measurement of test or method performance (**calculated total error**).

2. Desirable Total Allowable Error (TE_{a,d})

There are a number of ways in which TE_{a,d} can be determined, including studies of clinical outcomes based on test results, calculations based on biologic variation within and between individuals of various species, expert opinions, state-of-the-art analysis with regard to instrument and method performance, and information from providers of external quality assurance/proficiency testing programs.

In this study, TE_{a,d} was determined based on the most stringent interpretive needs for test results in dogs, cats, and horses based on a review of the literature and expert opinions of clinical pathologists in a large commercial laboratory. TE_{a,d} provided a quality requirement (quality specification) as a standard by which to judge the performance of various instruments and methods in this study. TE_{a,d} is an analytical quality requirement that sets a limit for both the imprecision (random error) and bias (systematic error) which are tolerable in a single measurement or single test result. The term **tolerance limit** is also used to describe TE_{a,d} for a particular test result.

3. Total Error Calculated (TE_c)

TE_c was defined as absolute bias + Z*SD or % absolute bias + Z*CV, where bias is the estimate of systematic error, SD is the estimate of random error, CV is the coefficient of variation, and Z is the multiplier that represents the desired confidence level. Different recommendations have been made for the Z value, ranging from 2 to 6. Most commonly, a Z value of 2 is commonly used to represent an approximately 95% probability level. The intended use of TE_c is to describe the maximum error that might occur in a test result obtained from a measurement procedure. It provides a measure of quality that can be compared with TE_{a,d}. Acceptable performance of a test or method is represented by TE_c < TE_{a,d}.

4. Quality Control Material (QCM)

A specimen or solution that is analyzed solely for quality control purposes (rather than for calibration). Most often these are commercially available products that are stable for a defined period of time. In this study, **assayed quality control materials** were used,

meaning the product includes a listing of expected values and ranges for the analytes included in the control material, assayed by various methods and instrument systems. The assayed values of analytes for 3 QCM at 3 different levels, representing a range of analyte results, were used as the basis for determining **bias** in this study.

5. Bias (Inaccuracy, Systematic Error, “Trueness”)

Bias is a performance characteristic of an instrument or method that refers to the agreement between a measurement and the correct or true value. The term “systematic error” is also used because this type of error causes test values to be systematically high or low. The term “stable inaccuracy” is used to further clarify that this is the performance expected when the measurement procedure is working properly. The correct or “true value” of an analyte may be defined by an internationally recognized standard or reference material, use of an assayed material, or comparison with the mean of a peer group using the same instrument or method.

Bias may be positive or negative or described as an absolute value. It may be constant or vary across the range of results of clinical importance. In this study, bias was determined by comparison of the mean results obtained by the participating laboratories with that of the assayed value provided by the manufacturer of the QCM for the same instrument or method. The formula used for calculation of bias in this study was:

$$|Bias| = |measured\ mean - target\ mean|$$

$$|\%Bias| = \left| \frac{(measured\ mean - target\ mean)}{target\ mean} * 100 \right|$$

6. Imprecision (Random Error)

Imprecision is a performance characteristic of an instrument or method that represents the agreement between replicate measurements. It is usually estimated by calculating an SD or CV. The term “random error” may also be used to describe this type of error because values are randomly higher or lower than the expected or average value. The terms “stable imprecision,” “stable random error,” and “inherent imprecision” are all used to indicate the performance expected when the measurement procedure is working properly.

7. Sigma Metric

Sigma metric is a quality characteristic (numeric value) that characterizes method performance in terms of the

number of SDs or sigmas that fit within the tolerance limit or quality requirement of a test. A sigma metric of < 3 is considered to represent unstable and unreliable performance that is not suitable for routine laboratory testing, whereas a sigma metric of ≥ 6.0 is considered to represent “world-class quality”. The sigma metric provides a benchmark for performance of an instrument or method. The formula for calculation of the sigma metric (σ) is:

$$\sigma = \frac{TEa_d - |\%Bias|}{CV}$$

8. Quality Goal Index (QGI)

QGI is a mathematical calculation that expresses the relative extent to which both bias and precision meet their respective quality goals. The quality goals chosen for use in this expression are $1.5 * TE_a/6$ for bias and $TE_a/6$ for precision based on their widespread use in the literature on 6 sigma methodology. This index is useful for determining whether imprecision, inaccuracy, or both are contributing to performance of < 6 sigma. The formula for calculation of QGI is:

$$QGI = \frac{|\%Bias|}{1.5 * CV}$$

The QGI is higher for a test application when bias exceeds its accuracy goal and imprecision meets its precision goal, and QGI is lower when bias meets its accuracy goal and imprecision exceeds its precision goal. The criteria used in this study for interpreting QGI when test applications fell short of 6 sigma quality were as follows:

QGI < 0.8 reflects imprecision.

QGI = 0.8–1.2 reflects a combination of both imprecision and inaccuracy.

QGI > 1.2 reflects inaccuracy.

9. Statistical Quality Control (Statistical QC)

An approach to control of laboratory systems in which statistics are applied to determine whether observed performance is within the expected limits of inherent variation within the system. The process of statistical QC in the medical laboratory involves use of 1 or more control materials to determine if the reagents, instrument, and operator are performing within specified limits that reflect the inherent stable variation of a test/method, as defined by the desirable total allowable error. The number of control materials and control rules applied to determine acceptability or unacceptability of the results are determined by the process of

QC validation and are designed to provide a high probability of error detection (P_{ed}) and a low probability of false rejection (P_{fr}).

10. Expanded Uncertainty of Measurement (EUM)

EUM is an estimate of the interval of values within which the true value of a measurement is believed to lie, with a stated probability. It provides a quantitative indication of the reliability of a measurement (laboratory result). It can be calculated with and without inclusion of bias. Most reference laboratory instruments can be adjusted by calibration or correction factors to minimize or eliminate bias, so that bias can be eliminated from the calculation. A coverage factor of 2 within the equation gives a probability of 95% for inclusion of the true value of a measurement in the calculated interval. The formula for calculation of the EUM is:

$$EUM_{bias} = 2 * \sqrt{|Bias_{RL}|^2 + |Bias_{IC}|^2 + SD_{RL}^2 + SD_{IC}^2}$$

$$EUM_{w/o\ bias} = 2 * \sqrt{SD_{RL}^2 + SD_{IC}^2}$$

EUM can be used to determine the acceptable range of values within which results should fall when 2 instruments or methods are being compared, taking into account the variances of each instrument or method. It was for this purpose that EUM was used in this study.

11. Quality Requirement (Quality Specifications, Quality Goals)

Quality requirement is a description of the quality that must be attained for a laboratory test to be clinically useful. Goals for analytical quality can be defined in quantitative terms, typically as TE_{a,d}, as used in this study. Some approaches may provide separate specifications for maximum bias and SD or CV. Some planning models include, in addition to stable imprecision and stable bias, allowances for preanalytical factors and a “safety margin” for unstable analytical performance. From the clinical standpoint, quality requirement is defined as the amount of error or variation in a

test result that causes a change in its interpretation and a corresponding change in the medical actions taken with regard to diagnosis, additional investigation, or monitoring of an animal patient.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Percentage of participants achieving > 6 sigma performance for analytes in each of 3 quality control materials and percentage of participants not achieving ≥ 6 sigma performance classified according to the quality goal index.

Table S2. Percentage of analytes showing agreement between means of results from in-clinic analyzers and means obtained by their reference laboratories using the expanded uncertainty of measurement, with and without bias, ± 2SD from the in-clinic mean, ± 3SD from the in-clinic mean, and ± desirable total allowable error from the in-clinic mean.

Table S3. Ranges of coefficient of variation, bias, and calculated total error obtained for concentrations and activities of 12 analytes measured in 3 levels of quality control materials for in-clinic analyzer groups.

Table S4. Ranges of coefficient of variation, bias, and calculated total error obtained for concentrations and activities of 12 analytes measured in 3 levels of quality control materials for reference laboratory groups.

Table S5. Mean, median, maximum, and minimum percentage of analytes suitable for statistical quality control based on “in-clinic QC” criteria across all 3 levels of quality control materials for in-clinic and reference laboratories.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.