AA**V**LD

Validation study of canine serum cortisol measurement with the Immulite 2000 Xpi cortisol immunoassay

Jérémie Korchia,[1] [iD] Kathleen P. Freeman [iD]

Abstract. We report the results of validation of canine serum cortisol determination with the Immulite 2000 Xpi cortisol immunoassay (Siemens), with characterization of precision (CV), accuracy (spiking-recovery [SR] bias), and observed total error (TEo = bias + 2CV) across the reportable range, specifically at the most common interpretation thresholds for dynamic testing. Imprecision increased at increasing rate with decreasing serum cortisol concentration and bias was low, resulting in increasing TEo with decreasing serum cortisol concentration. Inter-laboratory comparison study allowed for determination of range-based bias (RB) and average bias (AB). At 38.6 and 552 nmol/L (1.4 and 20 μg/dL), between-run CV was 10% and 7.5%, respectively, and TEo ~30% and ~20%, respectively (TEo remained similar regardless of the considered bias: SR, RB, or AB). These analytical performance parameters should be considered in the interpretation of results and for future expert consensus discussions to determine recommendations for allowable total error (TEa). Importantly, the commonly used thresholds for interpretation of results were determined ~40 y ago with different methods of measurements and computation, hence updating is desirable. Quality control material (QCM) had between-run imprecision of 4% for QCM1 and 7% for QCM2; the bias was minimal for both levels. Acceptable QC rules are heavily dependent on the desired TEa for the QCM system (TEa_{QCM}), itself limited by the desired clinical TEa. At low TEa_{QCM} (20–33%), almost no rules were acceptable, whereas at high TEa_{QCM} (50%), almost all rules were acceptable; further investigation is needed to determine which TEa_{QCM} can be guaranteed by simple QC rules.

Keywords: accuracy; bias; cortisol; Cushing; dogs; endocrinology; precision; total error.

Quantitative medical laboratory test interpretation requires knowledge about assay performance. A commonly used goal for test performance is allowable total error (TEa). In human medicine, many measurands have a consensus TEa established or reported by regulatory governing bodies.[50,53-55] In veterinary medicine, expert consensus TEa recommendations for hematology[35] and biochemistry[23] are available from the American Society of Veterinary Clinical Pathology (ASVCP). To date, there are no published consensus TEa goals for veterinary endocrinology.

According to the 2019 ASVCP guidelines,[1] TEa is defined as "a quality goal that sets a limit for combined imprecision (random error) and bias (inaccuracy, or systematic error) that is tolerable in a single measurement to ensure clinical usefulness." TEa represents a consensus about the level of allowed combined errors to assure sufficient quality to achieve a given clinical interpretation. TEa corresponds with a concept, as opposed to a measurement, such as observed total error (TEo), calculated from the coefficient of variation (CV) and the bias by the formula: TEo = 2CV + bias. TEa should be determined before quality assessment of a method, and can be derived mathematically from clinical decision limits, or *interpretation threshold* (IT). A method can be considered as a candidate for use when TEo < TEa. When a published recommendation for TEa is not available, knowledge of TEo (based on the analytical performance that is technically achievable) for the major IT may encourage and guide a future expert consensus discussion to determine recommendations for TEa.[23,35]

Importantly, the IT may depend on the assay method, which has varied greatly since the first canine blood cortisol studies. One of the earliest methods used in veterinary medicine[15,31] to measure cortisol in blood (heparinized plasma) for dexamethasone suppression and adrenocorticotropic hormone (ACTH) stimulation consisted of a fluorometric assay borrowed from human medicine: the fluorometric method of Mattingly (FMM)[31]; this method was not an immunoassay, given that it did not involve any antibody in its reagents. Instead of a tracer carrying the fluorescent signal, this method relied on the specific fluorescence of

Texas A&M Veterinary Medical Diagnostic Laboratory, Texas A&M University, College Station, TX, USA (Korchia); SYNLAB-VPG/Exeter, Exeter, United Kingdom (Freeman).

[1]Corresponding author: Jérémie Korchia, Department of Clinical Pathology, Texas A&M Veterinary Medical Diagnostic Laboratory, Texas A&M University, College Station, TX 77843, USA.
jeremie.korchia@yahoo.fr

1-hydroxycorticoids in concentrated sulfuric acid. In one of those studies,[32] cortisol was also measured by radioimmunoassay (RIA) for diurnal variation investigation, a technique first described 4 y earlier. From the 1980s to the early 1990s, various RIAs for canine blood cortisol were validated in dogs, on serum[17,41] or EDTA,[57] heparinized,[17,51] or non-specified plasma,[25] mostly relying on dilutional linearity, spiking-recovery, cross-reactivity, intra- and inter-assay precision, and analytical sensitivity studies. Plasma cortisol was also validated in cats with an RIA[56] (EDTA) and with an enzyme immunoassay[47] (heparinized).

During the same period in the 1980s, another assay for canine plasma cortisol was developed, validated, and used for clinical studies: the competitive protein-binding method (CPBM).[9-13,39] Similarly to the FMM, the CPBM does not use antibodies in its reagents and is not an immune method, but similarly to the RIA, the signal measured is radioactivity. Limited studies have found a lack of agreement[9] between the CPBM and the FMM, and some agreement[39] between the CPBM and the RIA. We are not aware of a comparison between the CPBM and a chemiluminescent immunoassay for canine blood cortisol.

In the early 1990s, plasma cortisol was validated in dogs with an enzyme immunoassay relying on the same criteria as above except with dilutional linearity.[14] In the late 1990s, the first chemiluminescent enzyme immunoassay (CLEIA) was validated for canine (and feline and equine) serum cortisol on the Immulite 1, the first version of the Immulite (Diagnostic Products Corporation; since purchased by Siemens).[42] The Immulite 1 demonstrated dilutional linearity from 27.0–530 nmol/L (1–19.2 μg/dL), a maximal inter-assay imprecision at 29 nmol/L (1 μg/dL) of 29%, and a linear regression of y = 0.8x + 14.6 (nmol/L) compared to a validated RIA. In 1997, EDTA plasma cortisol in dogs (and cats and horses) was validated for a similar CLEIA on the next version of the Immulite (Immulite 1000; Siemens), using instrument performance parameters (linearity, recovery, and precision) and compared with a validated RIA.[46] In that study, canine plasma cortisol recovery was >90% regardless of the concentration, and intra- and inter-assay precision was <20% from 0.5–25 μg/dL. In 2007, a comparison study[45] found a strong correlation for canine serum cortisol between a validated RIA and an Immulite (Siemens; model not specified); the Immulite tested higher than the RIA (difference nonsignificant >400 nmol/L [14.5 μg/dL], and significant <100 nmol/L [3.6 μg/dL] especially when <40 nmol/L [1.45 μg/dL]). A 2018 clinical study[36] used a non-Immulite chemiluminescent immunoassay for canine serum cortisol. The study did not directly reference a validation study; the provided reference Wenger-Riggenbach, et al.[52] was a validation for canine saliva cortisol, showing a correlation with plasma cortisol, but not providing validation for canine plasma cortisol measurement.

Guidelines for interpretation of serum cortisol results based on single sample results, or those obtained from dynamic testing protocols, are available from the literature[2,3,38]; those guidelines were not established according to classical techniques recommended for reference interval (RI) determination commonly used in biochemistry and hematology, but were rather elaborated by experts in the field based on empirical observations.[9,10] Our objective was not to redefine IT with the method validated here, as this would require a properly designed clinical study (see discussion). Our objectives were to: 1) validate canine serum cortisol measurement with the Immulite 2000 Xpi; 2) characterize test performance parameters (CV, bias, TEo) across the reportable concentration range; 3) evaluate analytical performance specifically at commonly used serum cortisol IT[11] of 38.6 nmol/L (1.4 μg/dL) and 552 nmol/L (20 μg/dL), in dynamic testing for canine hyperadrenocorticism (HAC); and 4) validate quality control (QC) rules for quality control materials (QCM) to ensure the ongoing ability to evaluate and demonstrate stable system performance on the Immulite 2000 Xpi.

Materials and methods

Study overview

We designed our study to validate canine serum cortisol immunoassay with the Immulite 2000 Xpi (Siemens), using 7 of the 9 immunoassay validation studies recommended by the guidelines[1] of the Quality Assurance and Laboratory Standard committee of the ASVCP. The immunoassay validation protocol was as follows: reportable range determination, within-run replication study, between-run replication study, recovery study, detection limit study, inter-laboratory comparison study, and QC rule validation study.

We did not perform interference studies, and we did not determine RIs. Limited information about possible interferents is available from the reagent manufacturer's product insert. We expanded the final step of QC rule determination by investigating performance not only on the 2 QCM levels, but also at the 2 relevant serum ITs.

Our approach was conceived as 3 complementary phases, consisting of 1) a spiking-recovery phase in some "cortisol-free" canine serum, 2) an interlaboratory comparison phase with serum patient samples, and 3) a QC phase with QCM data from the Texas A&M Veterinary Medical Diagnostic Laboratory (TVMDL; College Station, TX, USA) over one month (April 2019) and with data from the spiking-recovery phase at both IT levels over one week (Fig. 1). Utilization of leftover samples was undertaken with owners' consent, according to the appropriate TVMDL policy.

Immunoassay

The Immulite 2000 cortisol immunoassay (Siemens), a chemiluminoassay for cortisol, is a competitive heterogeneous phase assay, using a surface-bound capture anti-cortisol leporine polyclonal antibody and cortisol–alkaline phosphatase
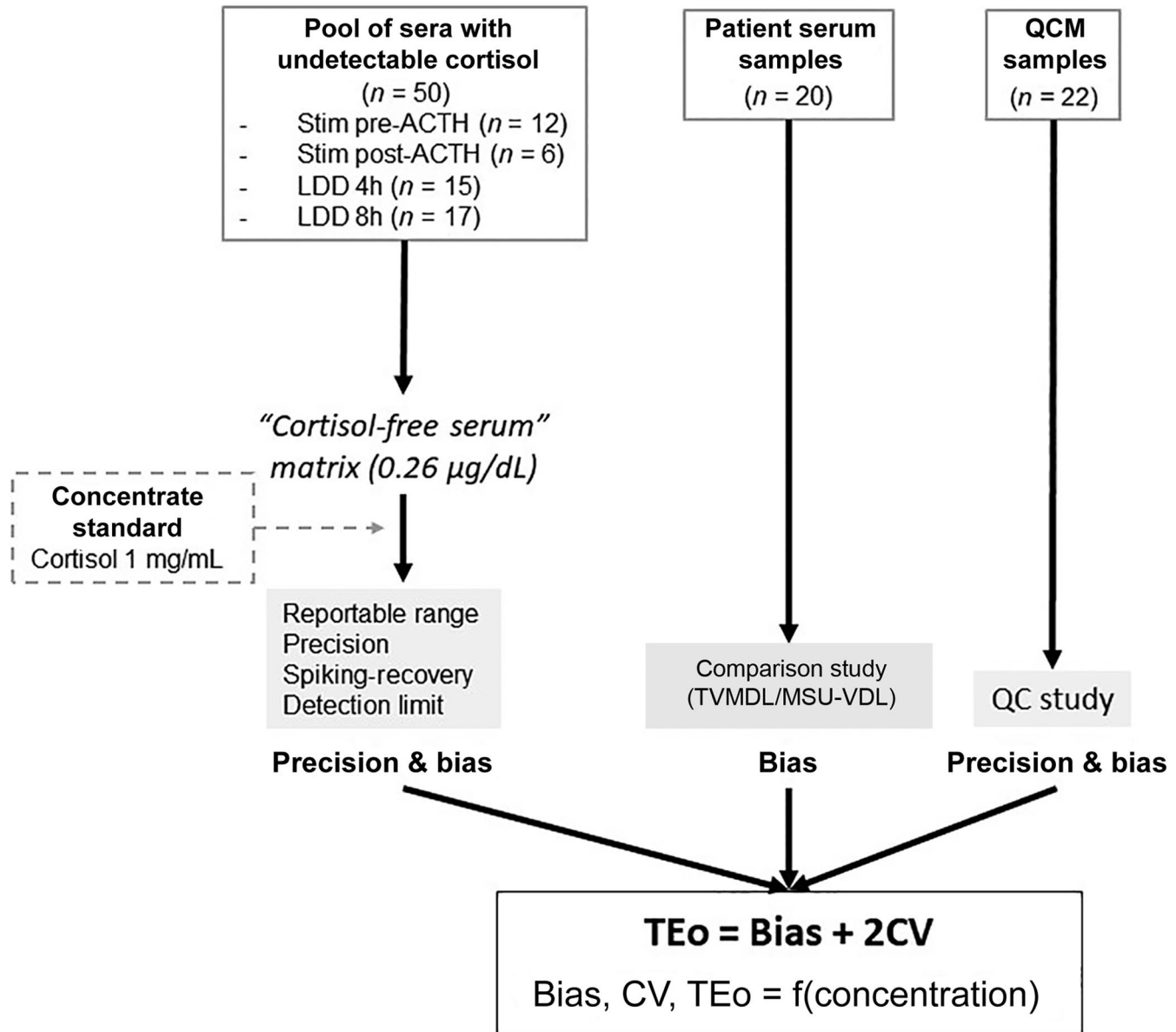
Figure 1. Canine serum cortisol validation study overview: a spiking-recovery phase, a between-laboratory comparison phase, and a quality control material (QCM) phase. LDD = low-dose dexamethasone suppression test; MSU-VDL = Veterinary Diagnostic Laboratory of Michigan State University; TEo = total observable error; TVMDL = Texas A&M Veterinary Medical Diagnostic Laboratory.

as tracer. This assay was validated by the manufacturer for human serum cortisol. It cross-reacts with prednisone (and prednisolone metabolized to prednisone). The cortisol molecule is identical in humans and in dogs,[37] hence the use of this immunoassay is appropriate in dogs.

The Immulite analyzers are de facto configured to not report results outside the predefined reportable range of 27.6–1,380 nmol/L (1–50 μg/dL); when a result is measured below or above those limits, the provided results are < or > the limit, respectively. To allow measurement for concentrations <27.6 nmol/L (<1 μg/dL) for the serum matrix constitution as well as for the detection limit study, and potentially

>1,380 nmol/L (>50 μg/dL) for the spiked L10, the Immulite 2000 Xpi was used in calibration verifier mode. The samples of interest were then scheduled as "Cal Verifier" on the worklist.

Serum matrix constitution

The matrix was selected to consist of serum with a cortisol concentration as close as possible to zero. Excess canine serum samples from ACTH stimulation tests (ACTHST) and low-dose dexamethasone suppression tests (LDDST) were collected from February 2019 to May 2019, and frozen at

Table 1. Serum cortisol dilutions used for reportable range/linearity, precision, recovery, and detection limit studies.

Level	Concentration, nmol/L (μg/dL)	Starting solution	Diluted with:	Studies*
L10: high pool	1,380 (50)	500 μL of FIS	9.5 mL of L1	Recovery
L9	1,035 (37.5)	750 μL of L10	250 μL of L1	Recovery
L8	552 (20)	400 μL of L10	600 μL of L1	Recovery
L7	345 (12.5)	250 μL of L10	750 μL of L1	Recovery
L6	172 (6.25)	125 μL of L10	875 μL of L1	Recovery
L5	69 (2.5)	50 μL of L10	950 μL of L1	Recovery
L4	38.6 (1.4)	28 μL of L10	972 μL of L1	Recovery
L3	13.8 (0.5)	100 μL of SIS	900 μL of L1	DLS
L2	6.9 (0.25)	50 μL of SIS	950 μL of L1	DLS
L1: low pool	Very low (undetectable)	Matrix	—	DLS
L0 (saline)	Blank	Saline	—	DLS

Dash (—) = no dilution; DLS = detection limit study; FIS = first intermediate solution (1,000 μg/dL); SIS = second intermediate solution (5 μg/dL).
* Linearity and precision were performed for all dilution levels.

−74°C until thawed for the spiking-recovery study. Inclusion criteria corresponded with undetectable cortisol concentration (<1 μg/dL) using the Immulite 2000 Xpi in use at TVMDL during that time, and no more than slight hemolysis and/or lipemia. The vast majority of the selected samples had no hemolysis, lipemia, or icterus; only rare samples had slight hemolysis or slight lipemia; samples with more than slight hemolysis or lipemia were discarded.

A total of 50 samples were selected and consisted of pre-ACTHST samples ($n = 12$), post-ACTHST samples ($n = 6$), 4 h post-LDDST samples ($n = 15$), and 8 h post-LDDST samples ($n = 17$). The samples came from 31 test protocols (11 ACTHST and 20 LDDST). The 31 corresponding dogs consisted of 4 intact males, 8 castrated males, and 17 spayed females (the sex was unavailable for 2 dogs). The mean age was 9 y old (which increased to 9.5 y old when the 3 dogs assessed for hypoadrenocorticism were removed). Breeds consisted of: 3 Dachshunds, 2 Maltese, 2 West Highland Whites, 2 Beagles (including 1 mixed), 2 Chihuahuas (including 1 mixed), 2 Labrador Retrievers (including 1 mixed), 14 dogs of 14 different single breeds, and 2 mixed-breed dogs without further specifications; the breed was unavailable for 2 dogs.

On the day of thawing, samples were immediately pooled and homogenized (vortexed for 10 s) to form the serum matrix pool, and endogenous cortisol concentration was measured in real time in duplicate; after immediate on-site confirmation of undetectable cortisol (<1 μg/dL) using the Immulite 2000 Xpi, the spiking-recovery study was performed. A sample of serum matrix was also sent out to a reference laboratory, the Michigan State University Veterinary Diagnostic Laboratory (MSU-VDL; Lansing, MI, USA), using an Immulite 2000 Xpi as well, for confirmation in duplicates with results obtained of 8.6 and 6.6 nmoL/L (0.31 and 0.24 μg/dL). Quantification of the endogenous cortisol concentration of the serum matrix was determined by using the calibration verifier manual option of the Immulite 2000 Xpi. The endogenous cortisol concentration of the

serum matrix, determined by averaging 4 daily replicates over 5 consecutive days ($n = 20$), was judged significant (0.27 μg/dL); consequently, the endogenous serum cortisol was taken into account (subtracted from the calculations) for the spiking-recovery computations. Because the spiking-recovery across the full concentration range (L0–L10) was only available for the first day ($n = 4$), the endogenous cortisol concentration in the serum matrix was also calculated from the same data base (day 1, $n = 4$) and was fairly similar (0.26 μg/dL). Thus, 0.26 μg/dL was removed from the recovery concentration of each level before computing the recovery percentages.

Cortisol spiking

The pooled sera matrix was spiked with a standard cortisol concentrate (Cerilliant), or certified reference material at 1 mg/mL (or 100,000 μg/dL or 2,759,000 nmol/L), to follow a defined dilution protocol (Table 1), and the results were used to assess linearity, precision, recovery, and detection limits. The spiking-recovery study was performed within 1 wk, with samples conserved by refrigeration (4°C).

Choice of level concentration. The spiking concentrations were chosen based on the relevant values for serum cortisol testing. Because the upper limit of the reportable range stated within the package insert of the Immulite 2000 Xpi is 1,380 nmol/L (50 μg/dL), and because there was no clinical interest in validating linearity beyond this point, the highest level was set as such. Because of their critical clinical relevance as IT for ACTHST and/or LDDST, 552 nmol/L (20 μg/dL) and 38.6 nmol/L (1.4 μg/dL) were included. Thus, the spiking scheme included: 50, 37.5, 20, 12.5, 6.25, 2.5, 1.4, 0.5, and 0.25 μg/dL levels (the last two were primarily used in the detection limit study). All levels were prepared separately to avoid carryover and amplification of errors that may occur with transfers when making multiple dilutions.

Dilutions prepared. A first intermediate solution of 5 mL at 27,600 nmol/L (1,000 μg/dL) of cortisol was prepared to allow for preparation of L10 at 1,380 nmol/L (50 μg/dL). Then, L10 and L1 (pooled sera matrix) were used to prepare L9, L8, L7, L6, L5, and L4. A second intermediate solution of 1 mL at 138 nmol/L (5 μg/dL) of cortisol was prepared from mixing 100 μL of L10 at 1,380 nmol/L (50 μg/dL) with 900 μL of L1; L3 and L2 were prepared by mixing this second intermediate solution with L1 (Table 1).

Reportable range study

Four within-run replicates of each level (L0–L10) were performed on day 1, and the mean of each level was calculated. Measured means (y-axis) were plotted against the spiked concentrations (x-axis) on a function graph, after what an ordinary least-squares simple linear regression (hereafter, "simple linear regression" or "linear regression") was performed (Excel 2016; Microsoft).

Within-run replication study

We calculated the CV of the 4 replicates of each spiked level used for the reportable range study to provide an estimate of the within-run precision across the reportable testing range. The CVs were plotted as a function of the cortisol concentration, and trendlines were generated (Excel 2016). Moreover, the comprehensive within-run precision ($n = 20$, intra-run) was assessed for the levels of greatest interest: 552 nmol/L (20 μg/dL; L8) and 38.6 nmol/L (1.4 μg/dL; L4).

Between-run replication study

The between-run CVs were calculated for the levels of greatest interest (L8 at 552 nmol/L [20 μg/dL] and L4 at 38.6 nmol/L [1.4 μg/dL]) as 20 replicates (4 replicates each day for 5 consecutive days), and for 2 QCM levels based on the QCM data from one month (April 2019; single reagent lot). The QCM was the K9CON (Immulite Systems Control; Siemens), the target value of level 1 (QCM1) was 193 nmol/L (7.0 μg/dL), and the target value of level 2 (QCM2) was 389 nmol/L (14.1 μg/dL). QCM1 and QCM 2 were measured once daily for 22 d (over the month of April 2019, with a single QCM lot).

Recovery study

The recovery percentage for each level from L2 to L10 (noted Lx in the formula below) was calculated as:

$$\text{Recovery } (\%) = \left(\left[\left(\text{Lx}_{recovered} - \text{L1}\right) - \text{Lx}_{spiked}\right] / \text{Lx}_{spiked}\right) \times 100\%.$$

The spiking-recovery (SR) bias was then calculated as the recovery percentage minus 100% and plotted on a function graph.

Detection limit study

We explored the detection limits by measuring the blank (L0; saline), the non-spiked matrix (L1), as well as spiked levels L2 and L3, 4 times a day for 5 consecutive days. Similar data (4 replicates a day for 5 consecutive days) for L4 were also available from the between-run replication study.

The limit of blank (LOB), determining the highest measurable cortisol concentration in the blank (L0), was determined according to the formula:

$$\text{LOB} = \text{mean} + 1.65 \text{ SD}.$$

The limit of detection (LOD), determining the lowest measurable serum cortisol concentration without precision requirements, was determined according to the formula:

$$\text{LOD} = \text{mean} + 1.65 \text{ SD}.$$

The limit of quantification (LOQ), determining the lowest measurable serum cortisol concentration with precision requirements, was determined according to the formula:

$$\text{LOQ} = \text{mean} + 2 \text{ SD}.$$

Inter-laboratory comparison study

The comparison study was performed by selecting excess patient samples over 2 wk, tested at TVMDL for serum cortisol, and stored frozen at −74°C. Samples were sent out (overnight and refrigerated) in one batch to a reference laboratory (MSU-VDL) that also uses the Immulite 2000 Xpi. Purposefully, 20 samples were selected as 4 samples belonging to 5 different, increasing concentration ranges (1.25–2.42 μg/dL; 3.77–5.75 μg/dL; 10.2–11.3 μg/dL; 14.8–17.9 μg/dL; 23.8–34.1 μg/dL) within the reportable range (1–50 μg/dL) to allow for assessment of the average bias (AB) between institutions, and for assessment of the range-based bias (RB) across various concentration levels. Of note, the non-spiked serum matrix of the spiking-recovery study was added as a 21st sample for the comparison study, given that it was measured at TVMDL with a cortisol concentration different from zero (~0.26 μg/dL). The comparison instrument/method (Immulite 2000 Xpi) at the reference laboratory (MSU-VDL) was the same as our instrument/method.

A comparison graph plotting the tested method on the y-axis against the reference method on the x-axis was performed. A Passing–Bablok regression, for which some level of error is expected in both compared methods,[5] was performed (www.acomed-statistik.de). Simple linear regression was also performed (Excel 2016) for comparison.

A Bland–Altman comparison plot was manually performed (Excel 2016; y-axis = difference from our result minus the reference laboratory result; x-axis = mean of our result and the reference laboratory result). First, the line of

Table 2. Types of computed observed total error function of the different coefficient of variations (rows) and biases (columns). Within-run CV for QCM is not applicable: we did not investigate within-run precision for QCM, given that their typical use is between-run.

TEo computations	Spiking-recovery bias	Average bias	Range-based bias	QCM between-run bias
Within-run CV (used for identifying the trend of TEo across the concentration range)	TEo_{SR} (L2–L10)	TEo_{AB} (L2–L10)	TEo_{RB} (equivalent L4, L6, L7, L8, L9*)	NA
Between-run CV (used for QC rule validation)	TEo_{SR} (L4, L8)	TEo_{AB} (L4, L8)	TEo_{RB} (equivalent L4, L8*)	TEo_{QCM} (QCM1, QCM2)

AB = average bias; CV = coefficient of variation; Lx = level x (see dilution Table 1); NA = not applicable; QCM = quality control material; RB = range-based bias; SR = spiking-recovery bias; TEo = observed total error.
* Range-based bias from groups of the comparison study of the closest ranges from those spiked levels.

the mean of the differences (M) was traced, surrounded by its 95% confidence intervals (95% CIs) calculated as[6,20]

$$M \pm t \times SE,$$

where t was the t value, taken from the t-distribution table, and SE was the standard error, calculated as $\sqrt{(SD^2/n)}$, for which SD was the standard deviation of the differences. The t value for serum (95%, n-1 = 20 degrees of freedom) was 1.725.

Then, the normality of the differences was investigated with a D'Agostino–Person normality test, with significance set at a p value threshold of 0.3[29] (see discussion). After verification of normality of the differences in serum ($p = 0.447$), agreement limits were calculated as[6,18]:

$$M \pm 1.96\ SD.$$

The 95% CI surrounding the agreement limits were calculated as[6,20]:

$$(UAL\ or\ LAL) \pm t\ SE,$$

where UAL = upper agreement limit and LAL = lower agreement limit; t was the t value, taken from the t-distribution table (1.725 for serum cortisol); and SE was the standard error, calculated as $\sqrt{(3 \times SD^2/n)}$, for which SD was the standard deviation of the differences.

Observed total error computation

TEo (%) was calculated according to the formula: $2 \times CV(\%)$ + absolute bias (%). Four types of TEo were calculated depending on the considered type of bias: TEo_{SR} (spiking-recovery), TEo_{AB} (average bias), TEo_{RB} (range-based bias), and TEo_{QCM} (quality control material; Table 2).

TEo_{SR} was calculated from within-run ($n = 4$, intra-run) precision and bias, across the entire concentration range (L2–L10), and plotted on a graph. TEo_{SR} was also calculated with a longer term precision ($n = 20$, between-run: 4 samples a day for 5 d) for L4 and L8 only; for the latter calculations given the increase in imprecision between within-run and

between-run data, the bias remained calculated from the within-run study ($n = 4$, intra-run) to not add some error from imprecision, generated by sample conservation, into the bias computation.

TEo_{AB} was calculated from the average bias (AB) of the comparison study, and either from the within-run precision (L2–L10) or from the between-run precision (L4 and L8).

TEo_{RB} was calculated from either the within-run (L4, L6–L9) or the between-run (L4 and L8) precision, and from the range-based bias (RB) of the comparison study. The RB consisted of the bias observed between TVMDL and the reference institution (MSU-VDL), in a subset of samples grouped by similar concentration ranges. The 20 serum samples for the comparison study were prospectively chosen, allowing use of 4 samples in 5 different concentration ranges (group 1: 1.25–2.42; group 2: 3.77–5.75; group 3: 10.2–11.3; group 4: 14.8–17.9; and group 5: 23.8–34.1 μg/dL). Thus, for example, TEo_{RB} for L4 used the bias of group 1, and TEo_{RB} for L8 used the bias of group 4.

TEo_{QCM} was calculated at both QCM levels from the between-run CV and bias (compared to target values provided by the manufacturer) observed on the data over one month (April 2019, $n = 22$, one QCM lot).

Quality control rule validation study

Usable QC rules were explored for both QCM levels (QCM1 and QCM2) with regular QCM data from TVMDL over one month (April 2019), as well as for the 2 critical concentration levels (spiked L4 and L8 used as QCM). We determined the acceptable QC rules manually, from normalized operational process specifications (OPSpec) charts, plotting each operational point determined from the bias (as a % of TEa) on the y-axis, and the CV (as a % of TEa) on the x-axis.

Concretely, we first summarized CVs and biases as well as the resulting TEo. The considered CVs were always the between-run CVs ($n = 20$ over 5 d for L4 and L8, $n = 22$ over 1 mo for QCMs). For L4 and L8, the elected bias was the within-run spiking-recovery bias ($n = 4$, within-run) to limit the impact of sample conservation on the bias. For QCMs,
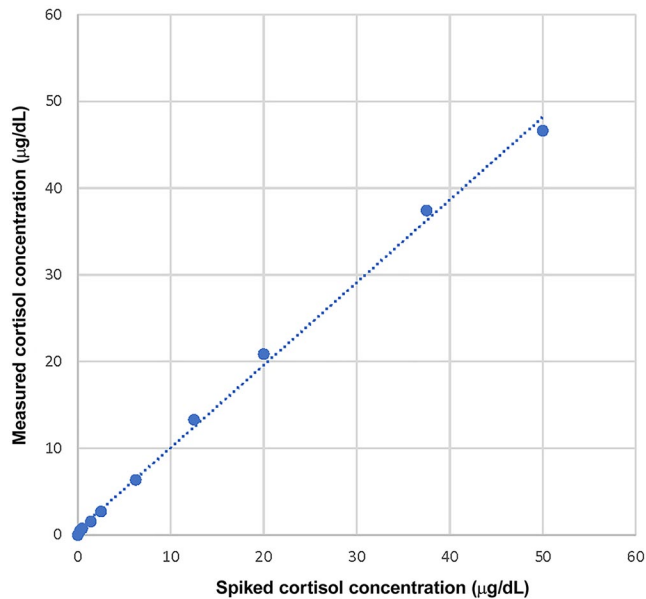
Figure 2. Reportable range study, with simple linear regression for within-run spiking-recovery canine serum cortisol. The simple linear regression yields a slope of 0.953 (close to 1), an intercept of 0.538 (close to 0). The coefficient of determination R^2 of 0.997 (close to 1) shows that the assessed range is sufficient.[1] Linearity was confirmed between 0.25 and 50 μg/dL.

the bias was considered compared to the target values provided by the manufacturer. We investigated the sigma metric and the acceptable QC rules:

- at low TEa (slightly higher than TEo) and at arbitrarily chosen high TEa (50%),
- at high probability of error detection (P_{ed}; 90%) and arbitrarily chosen low P_{ed} (50%),
- at $N = 2$ QC measurements (2 levels analyzed once; results provided: all of the results of this study are for $n = 2$) or $n = 4$ QC measurements (2 levels in duplicate; results not provided),

to draw conclusions about the influence of those parameters on acceptable QC rules. The probability of false rejection (P_{fr}) was fixed per QC rule in all QC scenarios:

1_{2S}: $P_{fr} = 0.09$
$1_{2.5S}$: $P_{fr} = 0.03$
$1_{3S}/2_{2S}/R_{4S}$: $P_{fr} = 0.01$
1_{3S}: $P_{fr} = 0.00$
$1_{3.5S}$: $P_{fr} = 0.00$

For the low level of TEa, when TEo was <20% (L8, QCM1, QCM2), we set TEa at 20%; when TEo was >20%, we set TEa at 33% (L4). The high level of TEa was set at 50%. Acceptance or rejection of candidate QC rules (1_{2S}, $1_{2.5S}$, $1_{3S}/2_{2S}/R_{4S}$, 1_{3S}, $1_{3.5S}$) was determined manually on

OPSpec charts, in each conditional scenario, depending on the position of the operating points compared to QC rule curves.

Results

Serum matrix constitution

The pooled samples of the serum matrix were each measured as undetectable in the regular setting (<27.6 nmol/L [<1 μg/dL]) before freezing. Once thawed and pooled to form the serum matrix, the serum matrix was measured in quadruplicate immediately (day 1) at a mean of 7.2 nmol/L (0.26 μg/dL), in quadruplicate during the 5 consecutive days of the study (day 1–5, $n = 20$) at a mean of 7.4 nmol/L (0.27 μg/dL), and also sent out on day 1 to the reference institution (MSU-VDL) for measurement in duplicate: 8.6 nmol/L (0.31 μg/dL) and 6.6 nmol/L (0.24 μg/dL); mean = 7.6 nmol/L (0.28 μg/dL).

Reportable range study

Linearity appeared excellent based on visual examination of the graph (Fig. 2) and on the linear regression characteristics. Indeed, the linear regression yielded a slope of 0.953 (close to 1), an intercept of 0.538 (close to 0), and a coefficient of determination R^2 of 0.997 (close to 1). Linearity was thus confirmed between 6.9 and 1,380 nmol/L (0.25 and 50 μg/dL). The reportable range of 27.6–1,380 nmol/L (1–50 μg/dL) provided by the manufacturer is adequate.

Within-run replication study

CV increased at an increasing rate (trendline: power function) with decreasing serum cortisol concentration, being ~2, 4, 8, and 20%, at 1,380, 552, 38.6, and 6.9 nmol/L (50, 20, 1.4, and 0.25 μg/dL), respectively; the shift of the curve happened ~38.6 nmol/L (1.4 μg/dL) with a CV of 7.97% (Table 3; Fig. 3A). In other words, imprecision increased at increasing rate with decreasing serum cortisol concentration.

Between-run replication study

Between-run CV was barely <10% for L4, ~7.5% for L8, ~4% for QCM1, and 7% for QCM2 (Table 4). For spiked samples, CV respected the following relationships, as expected to occur:

- Between-run precision CV > within-run precision CV
- CV from L8 (higher cortisol concentration) < CV from L4 (lower cortisol concentration)

For QCM, it is unclear why QCM1 between-run CV was lower than QCM2 between-run CV. No outliers or errors were identified on the Levey–Jennings charts. However,

Table 3. Results from the linearity, within-run precision (day 1, $n = 4$), and recovery studies across the canine serum cortisol reportable range.

Level	Spiked cortisol concentration, nmol/L (µg/dL)	Serum Mean measured concentration, nmol/L (µg/dL)	Within-run CV (%)	Recovery (%)
L10	1,380 (50)	1,280 (46.4)	1.81	92.7
L9	1,035 (37.5)	1,026 (37.2)	1.49	98.3
L8	552 (20)	568 (20.6)	3.98	103.1
L7	345 (12.5)	359 (13.0)	3.19	104.3
L6	172 (6.25)	169 (6.11)	7.02	97.7
L5	69 (2.5)	68 (2.45)	1.95	98.1
L4	38.6 (1.4)	36.1 (1.31)	7.97	93.5
L3	13.8 (0.5)	13.4 (0.486)	7.12	97.3
L2	6.9 (0.25)	6.5 (0.234)	18.4	93.6
L1 (matrix)	0	7.1 (0.259*)	12.7	NA
L0 (blank)	0	0.02 (0.00075)	200	NA

NA = not applicable.
* The intrinsic cortisol concentration of the serum matrix (0.259 µg/dL) was considered for the recovery and then for the bias computations.

QCM daily values for serum QCM1 were clustered around the target values, whereas serum QCM2 values were scattered within the mean ± manufacturer 2SD defining the acceptable range, supporting a between-run CV truly lower for QCM1 than for QCM2.

Recovery study

The SR bias ($n = 4$, within-run) was minimal across the entire range and adopted a bell shape (trendline: polynomial) from 50 µg/dL (–7.3%) to 12.5 µg/dL (4.3%) to 0.25 µg/dL (–6.4%) (Table 3; Fig. 3B).

Detection limit study

The LOB on L0 (saline) was 1.66 nmol/L (0.06 µg/dL). The LOD determined on L1 (the solution of lowest available concentration) was 9.10 nmol/L (0.33 µg/dL). The LOQ on L1 was 9.38 nmol/L (0.34 µg/dL), associated with a CV of 13.7%; no bias can be determined on L1, therefore no TEo could be calculated. When LOQ was determined from L2, it was 16.6 nmol/L (0.60 µg/dL), with corresponding CV of 13% and corresponding TEo of 43%.

Inter-laboratory comparison study

The AB was lower than any RB (Table 5). The Passing–Bablok regression was excellent, with a slope of 0.981 (0.937–1.023: 95% CI including 1) and an intercept of −0.122 (−0.443 to 0.176: 95% CI including 0; Fig. 4). The simple linear regression obtained very similar results (figure not shown), with a slope of 0.966 and an intercept of 0.062; coefficient of determination $R^2 = 0.994$ (indicating that a sufficient range of values was evaluated). In the Bland–Altman comparison graph (Fig. 5), the mean of differences with its 95% CI were −9.66 [95% CI = −17.66 to –1.66] nmol/L (−0.35 [95% CI = −0.64 to 0.06] µg/dL). The normality of the differences could be verified by the D'Agostino–Person normality test ($p = 0.447$; >0.3 taken as an interpretation threshold[29]: see discussion), therefore the agreement limits could be generated with their 95% CI (in brackets): −51.6 [−65.4 to −37.5] nmol/L (−1.87 [−2.37 to −1.36] µg/dL) and 32.0 [18.2–46.1] nmol/L (1.16 [0.66–1.67] µg/dL).

Observed total error computation

The within-run CV increased with decreasing serum cortisol concentration (6; Fig. 3A), whereas the SR bias remained relatively constant and small (Table 6; Fig. 3B). RB was relatively similar to the SR bias (Table 6), and the TEo generated from each of the 3 different biases (SR, AB, and RB) across the cortisol concentration range were relatively similar (Table 6). Similar to the pattern of CV, TEo_{SR} increased at an increasing rate with decreasing of the serum cortisol concentration (Fig. 3C).

The between-run CV for L4 and L8 (bias remaining within-run; Table 6), the between-run CV and bias for both QCM levels, and the resulting TEo, are provided (Table 7). For L4 and L8, TEo_{SR}, TEo_{RB}, and TEo_{AB} were roughly equivalent, being ~30% at 1.4 µg/dL and ~20% at 20 µg/dL (Table 8).

QCM precision, bias, and TEo were respectively 4.1%, 0.3%, and 8.5% at the lower level QCM1 (7 µg/dL), and 7%, 0.5%, and 14.5% at the higher level QCM2 (14 µg/dL). Of note, the QCM bias determined against target values provided by the manufacturer was minimal (0.3% and 0.5%).

Quality control rule validation study

One example of utilization of a normalized OPSpec chart for manual determination of the QC rules is provided (Fig. 6).
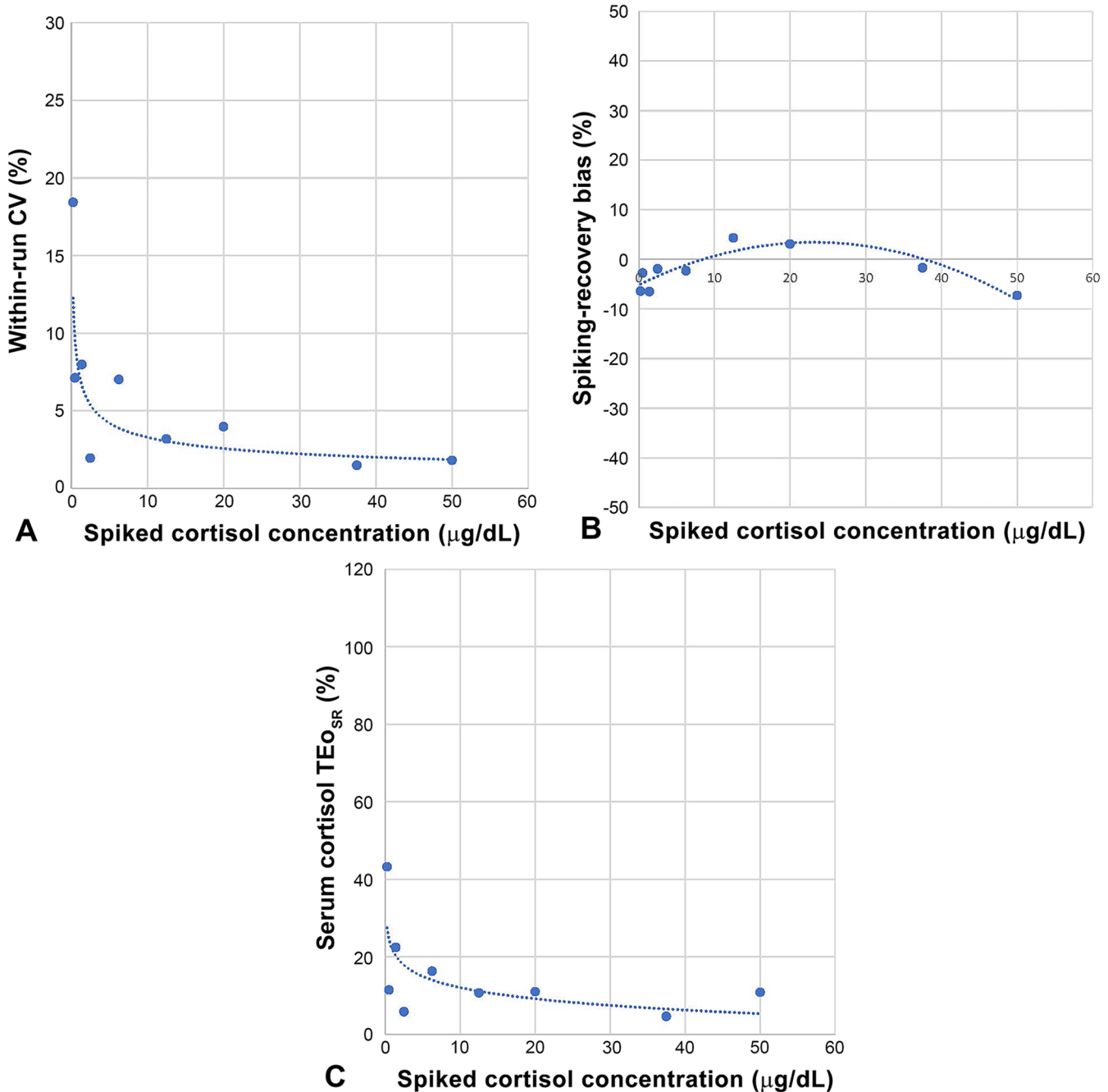
Figure 3. A. Within-run precision = f [serum cortisol]. Evolution of within-run ($n = 4$) CV (%) across the serum cortisol concentration. Spiked cortisol levels for the linearity study were run in quadruplicate, allowing assessment of the CV evolution across the concentration: CV remained low (1.5–7%) from 1,380 nmol/L (50 μg/dL) to 69 nmol/L (2.5 μg/dL), before increasing at an increasing rate at lower concentrations. **B.** Serum cortisol spiking-recovery bias = f [serum cortisol]. Evolution of the spiking-recovery bias (%) across the serum cortisol concentration. Spiked cortisol levels for the linearity study were run in quadruplicate, and the mean was used to calculate the recovery and the bias percentages. The spiking-recovery bias remained minimal (–7% to +4%) across the reportable range. **C.** Serum cortisol TEo$_{SR}$ = f [serum cortisol]. Evolution of TEo (%) across the serum cortisol concentration. Within-run CV (from Fig. 3A) and spiking-recovery bias (from Fig. 3B) were combined to calculate TEo across the serum cortisol concentration.

Almost no rules were acceptable at high P$_{ed}$ (90%) and low TEa, regardless of the considered level, with the exception of QCM1 (for which 1$_{2S}$, 1$_{2.5S}$, and 1$_{3s}$/2$_{2s}$/R$_{4s}$ were acceptable thanks to a minimal TEo of 8.5%). For all levels, decreasing P$_{ed}$ from 90% to 50% resulted in an increased proportion of acceptable QC rules; however, this proportion of newly

Table 4. Within-run and between-run precision for 2 spiked serum cortisol levels of clinical interest in dogs and 2 levels of quality control material.

Precision	n	Days	Level: nmol/L (µg/dL)	CV (%)
Within-run (spiked)	20	1	L4: 38.6 (1.4)	7.5
	20	1	L8: 552 (20)	4.7
Between-run (spiked)	20	5	L4: 38.6 (1.4)	9.5
	20	5	L8: 552 (20)	7.4
Between-run (QCM)	22	22	QCM1: 193 (7.0)	4.1
	22	22	QCM2: 389 (14.1)	7

CV = coefficient of variation; Lx = level x (see dilution Table 1); n = numbers of repeats; QCM = quality control material.

Table 5. Inter-laboratory comparison study results for canine serum cortisol: range-based bias and average bias.

Increasing concentration groups (1–5)	Serum cortisol, nmol/L (µg/dL)		Range-based bias
	TVMDL	MSU-VDL	
Group 5	941 (34.1)	946 (34.3)	−3.7%
	704 (25.5)	745 (27.0)	
	679 (24.6)	737 (26.7)	
	657 (23.8)	668 (24.2)	
Group 4	494 (17.9)	527 (19.1)	−4.5%
	480 (17.4)	464 (16.8)	
	436 (15.8)	461 (16.7)	
	408 (14.8)	452 (16.4)	
Group 3	312 (11.3)	279 (10.1)	4.1%
	303 (11.0)	284 (10.3)	
	298 (10.8)	298 (10.8)	
	281 (10.2)	287 (10.4)	
Group 2	159 (5.75)	167 (6.12)	−3.7%
	153 (5.54)	152 (5.51)	
	108 (3.92)	118 (4.28)	
	104 (3.77)	105 (3.80)	
Group 1	67 (2.42)	67 (2.43)	−13.8%
	49 (1.76)	59 (2.14)	
	37 (1.35)	49 (1.78)	
	34.5 (1.25)	42 (1.52)	
Average bias	−2.9%		

The 20 serum samples for the between-laboratory comparison study were chosen as 5 clustering sets of 4 samples to investigate the range-based bias; availability of samples also influenced the selection. Samples were tested fresh at Texas Veterinary Medical Diagnostic Laboratory (TVMDL) with the Immulite 2000 Xpi (Siemens), frozen at −80°C for <2 wk, and then sent overnight to Michigan State University Veterinary Diagnostic Laboratory (MSU-VDL) to be tested in one batch with the Immulite 2000 Xpi.

acceptable QC rules remained very limited. On the other hand, for all levels, increasing TEa to 50% resulted in a prominently increased proportion of acceptable QC rules. This illustrates the major impact of the chosen TEa level versus the minor impact of the chosen P_{ed} (Table 9).

Discussion

We validated canine serum cortisol on the Immulite 2000 Xpi, and characterized the CV, bias, and TEo function of concentrations spanning the reportable range, especially at commonly used IT for the ACTHST and the LDDST, for kit lot numbers <550. For lots ≥550, the anti-cortisol capture antibody was modified by Siemens in 2020, and has shown significant negative bias in canine serum,[21] whereas not affecting the measurements in human samples. The average bias between the kits was −23.1%, and reached maxima from −55.8% to −133.3% depending on the considered concentration ranges. This observation (November 2020) led to the commercialization (January 2021) by Siemens of a "veterinary cortisol kit" containing the new antibody but also integrating a correction formula (historical value = (1.1 × new antibody kit value) + 4.14 nmol/L). After correction, the average bias between the kits dropped to −8.6%, which was deemed acceptable, but reached maxima from −27.4% to −191.5% depending on the considered concentration ranges, therefore
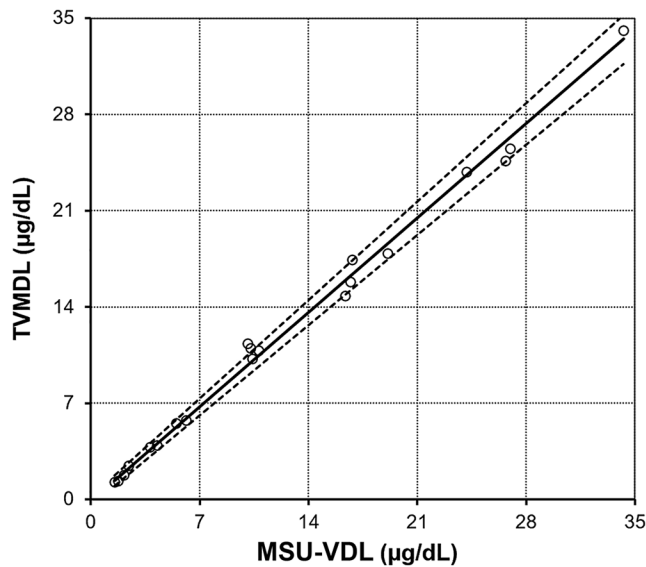
Figure 4. Inter-laboratory comparison study for canine serum cortisol ($n = 20$), with Passing–Bablok regression (solid line) and its 95% CI (dotted lines). MSU-VDL = Veterinary Diagnostic Laboratory of Michigan State University; TVMDL = Texas A&M Veterinary Medical Diagnostic Laboratory.
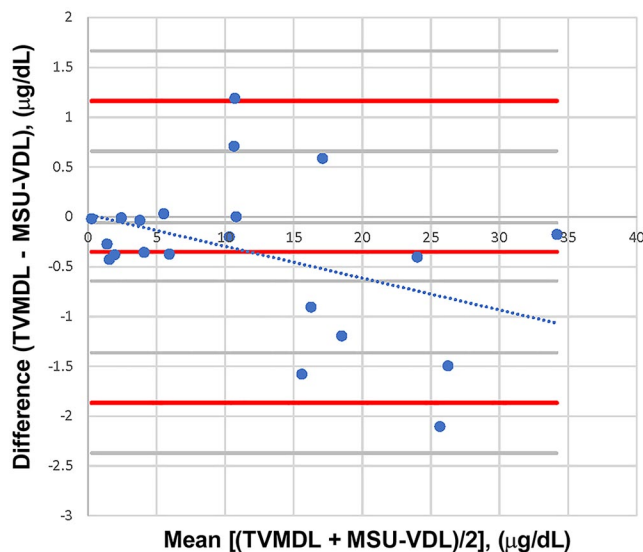

Figure 5. Bland–Altman comparison for the inter-laboratory canine serum cortisol comparison study ($n = 21$), with simple linear regression of the differences between laboratories. Both laboratories (TVMDL and MSU-VDL) used the same method of measurement: Immulite 2000 Xpi. There was one freezing cycle added before measurement by the MSU-VDL. The simple linear regression of the difference followed the equation: $y = -0.0319x + 0.0236$ (dotted line). Thus, there was nearly no constant bias; the proportional bias was negligible. Indeed, the latter was minimal at low concentration, and it would not impact clinical interpretation at higher concentration. MSU-VDL = Veterinary Diagnostic Laboratory of Michigan State University; TVMDL = Texas A&M Veterinary Medical Diagnostic Laboratory.

occasional significantly negatively biased samples could not be prevented. Further characterization of the analytical and clinical performance of the new kits (lot numbers ≥550) is warranted.

The reported ITs of 38.6 and 552 nmol/L (1.4 and 20 μg/dL, respectively) were established in 1983[11] with a non-immune radioassay (CPBM), not in use anymore. It is not fully characterized if these are the most adequate ITs, meaning optimizing the sensitivity and the specificity according to a ROC curve, with the Immulite 2000 Xpi data. Canine serum cortisol measurement assays have changed over the decades, and those changes should be taken into account when transferring IT from one method to another, for ACTHST (Table 10) as for LDDST (Table 11); 2 main studies on high-dose dexamethasone suppression test (HDDST),[10,16] which contributed to the determination of its currently used IT, have been included because the IT is the same for the LDDST (38.6 nmol/L or 1.4 μg/dL at 8 h post-dexamethasone injection).

Earlier studies did not provide an interpretation threshold, sensitivity, and specificity. They rather provided the mean, SD, and range at each time in each population. Studies later started providing a threshold, mostly consisting of the mean plus 2 SD or 3 SD from a limited population of healthy dogs, and without verifying the normality of the results. Later, the sensitivity alone, and finally the specificity, were added. In between, there was a switch of assays from non-immune methods to immune methods, first RIA, and then enzyme and chemiluminescent assays. There was also a trend to not include a control population anymore, but rather to quote former studies, with various degrees of accuracy. Much more recently, studies started to compare dogs with HAC to diseased dogs or dogs suspected of HAC but not having HAC, rather than to healthy dogs, challenging the specificity of the assays. We believe that the recent approach of the interpretation threshold with ROC curves[22] comparing dogs with HAC with suspect or diseased dogs is better. For example, a recent study[36] found an optimal interpretation threshold of 684–717 nmol/L (24.8–26.0 μg/dL) for post-ACTH cortisol with a chemiluminescent immunoassay different from the Immulite 2000 Xpi; it is not known if the Immulite 2000 Xpi would yield a similar threshold. The TEo we determined for 20 μg/dL in our study would likely be a good approximation of TEo at this threshold on the Immulite 2000 Xpi. For the LDDST, the most commonly used interpretation threshold at 8 h was determined in 1983 by an assay that was neither immune nor chemiluminescent, and on relatively limited populations (33 HAC and 22 control dogs)[11]; thus, a study with ROC curve determination of the optimal threshold with the Immulite 2000 Xpi is highly desirable.

A sensitivity or a specificity by itself is nearly meaningless and should at the very minimum be provided with the associated IT and the associated assay method. Ideally, it should mention the source of the IT (upper

Table 6. Coefficient of variation (within-run), bias, and observed total error results across the canine serum cortisol concentration range.

Serum cortisol		Precision (%)	Bias (%)			TEo = Bias + 2CV (%)		
Level	nmol/L (µg/dL)	CV (within-run)	SR	RB	AB	TEo$_{SR}$	TEo$_{RB}$	TEo$_{AB}$
L10	1,380 (50)	1.81	−7.27	NA	−2.93	10.9	NA	6.5
L9	1,035 (37.5)	1.49	−1.68	−3.72		4.6	6.7	5.9
L8	552 (20)	3.98	3.08	−4.47		11.0	12.5	10.9
L7	345 (12.5)	3.19	4.33	4.10		10.7	10.5	9.3
L6	172 (6.25)	7.02	−2.30	−3.70		16.3	17.7	17.0
L5	69 (2.5)	1.95	−1.94	NA		5.8	NA	6.8
L4	38.6 (1.4)	7.97	−6.50	−13.8		22.4	29.7	18.9
L3	13.8 (0.5)	7.12	−2.75	NA		11.5	NA	17.2
L2	6.9 (0.25)	18.4	−6.40	NA		43.3	NA	39.8

AB = average bias; CV = coefficient of variation; Lx = level x (see dilution Table 1); NA = not applicable; RB = range-based bias; SR = spiking-recovery bias; TEo = observed total error. The 3 types of bias allow computation of 3 types of TEo across the serum cortisol concentration. For the RB bias (from the interlaboratory comparison study), not all the levels (initially determined in the spiking-recovery study) could be investigated because of sample availability. RB bias and TEo$_{RB}$ are provided when available. The low within-run CV for L5, combined with the minimal SR bias (middle of a flat bell: Fig. 3B), results in a low, potentially underestimated, TEo at this concentration level.

Table 7. Coefficient of variation (between-run), bias, and observed total error results for 2 clinically relevant canine serum cortisol concentrations and the 2 quality control material levels.

Spiked serum samples								
Cortisol level	Target values, nmol/L (µg/dL)	Precision (%) Between-run CV	Bias (%) SR	RB	AB	TEo (%) TEo$_{SR}$	TEo$_{RB}$	TEo$_{AB}$
L4	38.6 (1.4)	9.53	−6.50	−13.8	−2.93	25.6	32.8	22.0
L8	552 (20)	7.42	3.08	−4.47		17.9	19.3	17.7

QCM				
Cortisol level	Target values,* nmol/L (µg/dL)	Precision (%) Between-run CV	Bias (%)	TEo$_{QCM}$ (%)
QCM1	193 (7.0)	4.08	−0.27	8.5
QCM2	389 (14.1)	7.01	−0.52	14.5

AB = average bias; CV = coefficient of variation; Lx = level x (see dilution Table 1); QCM = quality control material; RB = range-based bias; SR = spiking-recovery bias; TEo = observed total error.
* The serum QCMs had target values provided by the manufacturer for canine serum cortisol.

Table 8. Variability of the canine serum cortisol interpretation thresholds considering the between-run coefficient of variation only and considering the observed total error.

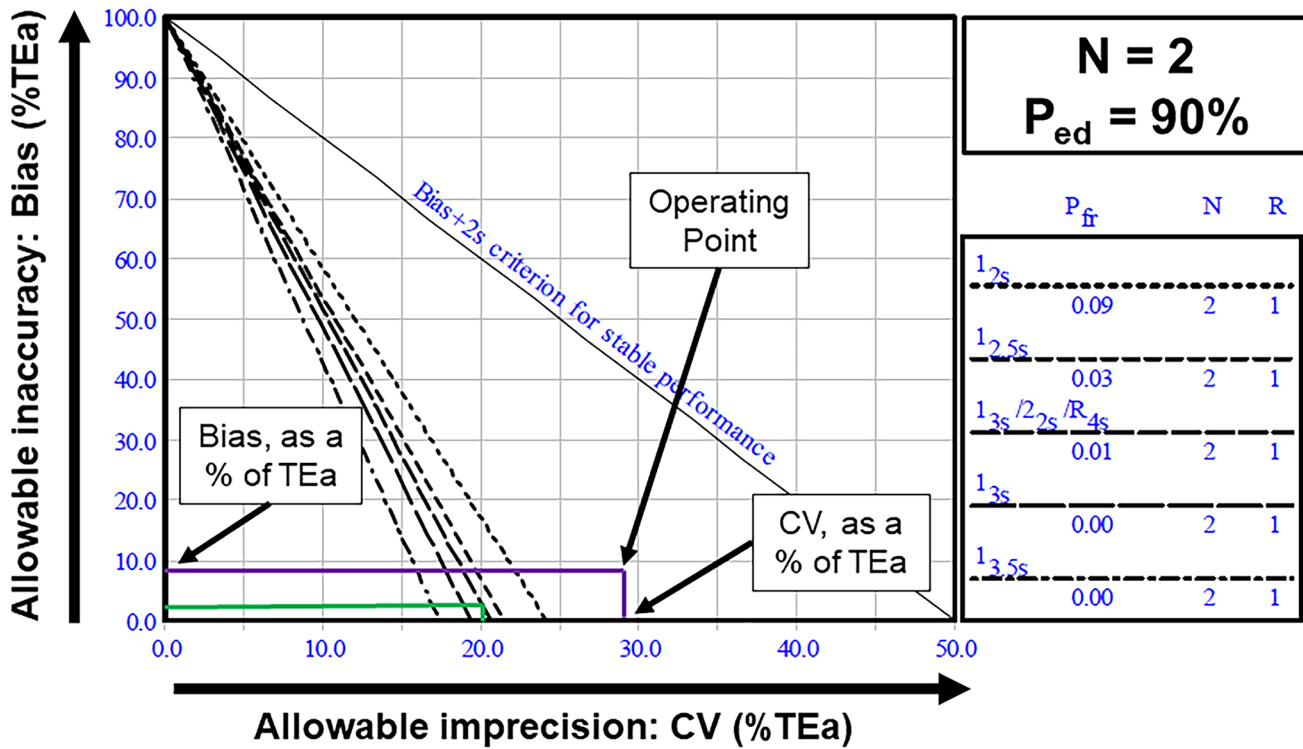| Interpretation threshold, nmol/L (µg/dL)	Between-run CV: resulting interval,* nmol/L (µg/dL)	TEo: resulting interval,† nmol/L (µg/dL)
38.6 (1.4)	9.53%: 31–46 (1.13–1.67)	TEo$_{SR}$ ≈ TEo$_{RB}$ ≈ TEo$_{AB}$ ≈ 30%: 27–50 (0.98–1.82)
552 (20)	7.42%: 470–634 (17.0–23.0)	TEo$_{SR}$ ≈ TEo$_{RB}$ ≈ TEo$_{AB}$ ≈ 20%: 442–662 (16–24)

AB = average bias; CV = coefficient of variation; RB = range-based bias; SR = spiking-recovery bias; TEo = observed total error.
* The resulting interval around the concentration levels corresponds with the 95% probability estimate range of cortisol concentrations expected using 2CV. It accounts for precision only.
† The resulting interval around the concentration levels corresponds with the concentration ±TEo, with TEo = bias + 2CV. It accounts for precision and accuracy.

limit of the range for normal dogs, mean + x SD for normal dogs, ROC curve analysis, etc.) and the nature of the considered population(s) (healthy, dogs with HAC, dogs suspected of but not having HAC [HAC-S], dogs with non-adrenal illness, etc.) with the number of individuals (Tables 10, 11). Following this good practice,

Figure 6. Example of utilization of a normalized operational process specifications (OPSpec) chart, at $N = 2$ levels of QCM, and P_{ed} of the system = 90%, for QC rule validation of L4 (38.6 nmol/L [1.4 μg/dL]) in green, and QCM1 (193 nmol/L [7.0 μg/dL]) in purple. In this example, we are considering the averaged bias (from the comparison study) for QC rule determination of L4 used as a QCM (L4$_{AB}$). First, CV and bias are determined, and TEo is calculated. Then, a given level of TEa is chosen. Here, when TEo was <20%, TEa was chosen as 20% (QCM1); when TEo was >20%, TEa was chosen as 33% (L4$_{AB}$). Then, CV and bias are expressed as % of TEa ("normalized"), and those values are plotted on the x-axis and y-axis of the OPSpec chart, respectively, as the "operating point." All the QC rules to the right of the operating point are candidates for use with the specified P_{ed} and P_{fr}; all the QC rules left from the operating point are not. Thus, there is no candidate QC rule for L4$_{AB}$. For QCM1, $1_{2.5s}$ or $1_{3s}/2_{2s}/R_{4s}$ would be candidate QC rules; 1_{2s} should be avoided because of the P_{fr} of 0.09 (see discussion). AB = averaged bias; CV = coefficient of variation; L4 = level 4 in this study, corresponding with 1.4 μg/dL; N = number of levels tested for QC; OPSpec = operational process specifications; P_{ed} = probability of error detection; P_{fr} = probability of false rejection; QCM = quality control material; R = number of repeats for each QC level; TEa = allowable total error; TEo = observed total error.

the commonly stated sensitivities of the ACTHST and the LDDST from 1983[11] would be provided as:

- Sensitivity of 83% of the post-ACTH cortisol (0.25 mg/dog IM, 1 h) for HAC associated with an IT of 552 nmol/L (20 μg/dL) determined by the mean + 3 SD of 64 control tests from 16 control dogs, when using a CPBM (non-immune radioassay) on heparinized plasma.
- Sensitivity of 92% of the LDDST (0.01 mg/dog IV, 8 h) for HAC associated with an IT of 39 nmol/L

(1.4 μg/dL) determined by the mean + 3 SD of 22 control dogs, when using a CPBM (non-immune radioassay) on heparinized plasma.

It is critical to consider thresholds as *interpretation* thresholds, and not as *diagnostic* thresholds. Indeed, TEo intervals account only for the analytical variability of the result, which is not the only aspect to take into account for test result interpretation. For example, a post-ACTH serum cortisol result clearly beyond the upper limit of the TEo interval for 20 μg/dL (e.g., a result of 28 μg/dL) could still be

Table 9. Quality control rule validation for 2 relevant cortisol concentrations (L4 = 38.6 nmol/L = 1.4 µg/dL; L8 = 552 nmol/L = 20 µg/dL) and both quality control material levels in canine serum.

Serum cortisol	High $P_{ed} \Leftrightarrow P_{ed}$ 90%					Low $P_{ed} \Leftrightarrow P_{ed}$ 50%				
	Level	TEo (%)	TEa (%)	σ	Candidate QC rules	Level	TEo (%)	TEa (%)	σ	Candidate QC rules
Low TEa	L4$_{(SR)}$	25.6	33	2.8	None	L4$_{(SR)}$	25.6	33	2.8	None
	L4$_{(RB)}$	32.8	33	2.0	None	L4$_{(RB)}$	32.8	33	2.0	None
	L4$_{(AB)}$	22.0	33	3.2	None	L4$_{(AB)}$	22.0	33	3.2	$\mathbf{1_{2s}}$
	L8$_{(SR)}$	17.9	20	2.3	None	L8$_{(SR)}$	17.9	20	2.3	None
	L8$_{(RB)}$	19.3	20	2.1	None	L8$_{(RB)}$	19.3	20	2.1	None
	L8$_{(AB)}$	17.8	20	2.3	None	L8$_{(AB)}$	17.8	20	2.3	None
	QCM1	8.5	20	4.8	1_{2s}; $1_{2.5s}$; $1_{3s}/2_{2s}/R_{4s}$	QCM1	8.5	20	4.8	**All**
	QCM2	14.5	20	2.8	None	QCM2	14.5	20	2.8	None
High TEa	L4$_{(SR)}$	25.6	50	4.6	$\mathbf{1_{2s}}$	L4$_{(SR)}$	25.6	50	4.6	1_{2s}; $\mathbf{1_{2.5s}}$; $\mathbf{1_{3s}/2_{2s}/R_{4s}}$; $\mathbf{1_{3s}}$
	L4$_{(RB)}$	32.8	50	3.8	None	L4$_{(RB)}$	32.8	50	3.8	$\mathbf{1_{2s}}$; $\mathbf{1_{2.5s}}$
	L4$_{(AB)}$	22.0	50	4.9	$\mathbf{1_{2s}}$; $\mathbf{1_{2.5s}}$; $\mathbf{1_{3s}/2_{2s}/R_{4s}}$	L4$_{(AB)}$	22.0	50	4.9	**All**
	L8$_{(SR)}$	17.9	50	6.3	**All**	L8$_{(SR)}$	17.9	50	6.3	All
	L8$_{(RB)}$	19.3	50	6.1	**All**	L8$_{(RB)}$	19.3	50	6.1	All
	L8$_{(AB)}$	17.8	50	6.3	**All**	L8$_{(AB)}$	17.8	50	6.3	All
	QCM1	8.5	50	12.1	**All**	QCM1	8.5	50	12.1	All
	QCM2	14.5	50	7.1	**All**	QCM2	14.5	50	7.1	All

AB = average bias; Lx = level x (see dilution Table 1); All = includes all 4 QC rules (1_{2s}; $1_{2.5s}$; 1_{3s}; $1_{3s}/2_{2s}/R_{4s}$); P_{ed} = Probability of error detection (by the QCM); QCM = quality control material; RB = range-based bias; SR = spiking-recovery bias; TEo = total allowable error; TEo = observed total error. Bolded rules are those added when switching from [Low TEa and P_{ed} 90%] to [Low TEa and P_{ed} 50%], from [Low TEa and P_{ed} 90%] to [High TEa and P_{ed} 90%], and from [High TEa and P_{ed} 90%] to [High TEa and P_{ed} 50%]. We investigated the sigma metric and the candidate QC rules at "low" TEa (slightly higher than TEo), voluntarily "high" TEa (arbitrarily chosen at ≥50% when TEo was >50%), "low" P_{ed} (arbitrarily chosen at 50%, even if such a low P_{ed} is contraindicated), and at voluntarily "high" P_{ed} (chosen at the minimal recommended level of 90%). The goal was to illustrate the limited influence of varying P_{ed} contrasting with the large influence of varying TEa on candidate QC rules.

a true positive (Cushing disease) or a false positive (e.g., a chronic, consequent inflammation generating adrenal hyper-reactivity). In other words, passing beyond the variability associated with TEo only assures you the ability to eliminate the analytical variability for your interpretation (the *test* is "truly positive"), but it does not assure you that your *animal* is truly positive for the disease investigated by the test. The final interpretation is performed according to the predictive values, positive and negative, which are a function of the theoretical prevalence of the disease in a population composed of animals presenting as your patients (age, breed, sex, clinical signs, biochemistry, etc.), otherwise called pre-test probability. The more a dog accumulates support for Cushing disease, the more a positive result is likely to be truly positive, and inversely. Importantly, the goal of our study was not to assess the relevance of the 2 ITs themselves, but rather to analyze to what extent cortisol results could be relied upon according to the TEo related to those ITs.

One strength of our study resides in the fact that we looked at CV, bias, and TEo across the entire reportable concentration range, and especially at the 2 levels of most clinical significance (1.4 and 20 µg/dL), rather than at single average means from samples of different concentrations. Because the CV and thus TEo demonstrated marked variations across the cortisol concentration range, the degree of variation should be taken into account in the interpretation of results.

Another strength of our study is the investigation of 3 types of biases: SR, RB, and AB, which bring different infor-

mation and have different properties. SR bias shows the intrinsic bias of the immunoassay compared to the calculated spiking concentrations. RB shows the bias existing between 2 laboratories at a given serum cortisol concentration level. AB shows the average bias between 2 laboratories; because it averages all of the RB (some positive and some negative), it artifactually minimizes the estimated bias, and thus gives a false impression of excellence. In a context in which the analytical specificity is good, RB is probably the most relevant type of bias in a clinical setting, given that interpretation of results should be the same across laboratories; clinical results are rarely compared to SR performance, yet it is relevant to quantify how far from a "true value" a method can be. The SR bias becomes more important when the analytical specificity is challenged, as for example in the measurement of urine cortisol in the presence of multiple cross-reacting metabolites (urinary corticoids).[27] In any case, because the 3 resulting TEo (TEo$_{SR}$, TEo$_{RB}$, and TEo$_{AB}$) were similar for L4 and very similar for L8, we can simplify the concept in serum and state that TEo for L4 is ~30% and TEo for L8 is ~20%. This was definitely not the case in the urine matrix.[27] The reportable range of an assay is the range in which the measurand can be measured with acceptable accuracy and precision. Linearity is not mandatory, but it remains highly desired, and is often used to justify the reportable range of an assay. Demonstration of linearity by dilutions will actually be conditioned by the linearity of the calibration curve. The more linear is the calibration curve, the more comfortable it

Table 10. Chronologic and technical evolution of the post–adrenocorticotropic hormone stimulation interpretation threshold for the diagnosis of hyperadrenocorticism in dogs.

Ref.	Sample medium	Cortisol assay	ACTH stimulation	Cushing population	Compared population	IT	Source of IT	Se (%)	Sp (%)
32	Plasma (not specified)	FMM (non-immune)	Cortrosyn, 0.25 mg/dog IV Post (90 min)	21 PDH	18 healthy kennel dogs	Not provided / Deducible IT from provided data: 800–850 nmol/L (29–31 µg/dL) • Post-ACTH range normal dogs: 330–800 nmol/L • Post-ACTH cortisol range PDH: 850–3,420 nmol/L	32	—	—
15	Plasma (heparinized)	FMM (non-immune)	Cortrosyn, 0.25 mg/dog IM Pre & post (1 h)	7 PDH, 2 AT	21 normal dogs	Not provided / Deducible IT from provided data: 1,002–1,048 nmol/L (36.3–38 µg/dL) • Post-ACTH normal dogs: mean + 2SD = 1,002 nmol/L • Post-ACTH range PDH: 1,048–2,690 nmol/L	15	78†	100‡
33	Plasma (not specified)	RIA	Cortrosyn, 0.25 mg/dog IV Post (90 min)	9 AT	21 healthy kennel dogs	Not provided / Deducible IT from provided data: 578 nmol/L (21 µg/dL) • Post-ACTH normal dogs: mean + 2SD = 578 nmol/L	33	—	—
9	Plasma (not specified)	FMM (non-immune)	Cortrosyn, 0.25 mg IM Pre & post (1 h)	5 AT	None	1,002 nmol/L (36.3 µg/dL)	Quoted study[15]; 21 normal dogs IT = mean + 2SD	—	—
	Plasma (not specified)	CBPM (non-immune radioassay)	Cortrosyn, 0.25 mg IM Pre & post (1 h)	1 AT	None	544 nmol/L (19.7 µg/dL) • (supposedly the post-ACTH: mean + 2SD from some healthy dogs)	No source (quoted study does not match quoted numbers)	—	—
39	Plasma (heparinized)	RIA (also uses CBPM, and finds an agreement)	ACTH Gel, 20 units/dog IM Pre & post (2 h)	22 AT	21 control dogs	414 nmol/L (15 µg/dL) • Post-ACTH range normal dogs: 127–408 nmol/L (4.6–14.8 µg/dL) • Post-ACTH normal dogs: mean + 3SD = 433 nmol/L (15.7 µg/dL)	39	59	100‡
11	Plasma (heparinized)	CBPM (non-immune radioassay)	Cortrosyn, 0.25 mg/dog IM Pre & post (1 h)	26 PDH, 7 AT, 31 HAC (unknown location)	64 control samples from 16 control dogs	552 nmol/L (20 µg/kg) • Post-ACTH normal dogs: mean + 3SD = 538 nmol/L (19.5 µg/dL)	11	83 HAC 88 PDH 57 AT	—
12	Plasma (heparinized)	CBPM (non-immune radioassay)	Cortrosyn, 0.25 mg/dog IM Pre & post (1 h)	15 PDH, 6 AT	None	469 nmol/L (17.0 µg/dL) • (supposedly the post-ACTH: mean + 3SD from 26 healthy dogs)	12	76 HAC 87 PDH 50 AT	—
7	Plasma (EDTA)	RIA	ACTH gel (8h post LDDST 0.01 mg/kg IV), 2.2 U/kg IM Post (2 h)	None	14 clinically normal dogs; 33 chronically ill dogs	502 nmol/L (18.2 µg/dL) • Post-ACTH normal dogs: mean + 2SD = 502 nmol/L (18.2 µg/dL)	7	—	64
13	Plasma (heparinized)	CBPM (non-immune radioassay)	Cortrosyn, 0.25 mg/dog IM Pre & post (1 h)	14 PDH, 3 AT	None	469 nmol/L (17.0 µg/dL) • (supposedly the post-ACTH: mean + 3SD from 26 healthy dogs)	No source (quoted studies do not match quoted numbers)	88	—
43	Plasma (heparinized)	CBPM (non-immune radioassay)	Cortrosyn, 0.25 mg/dog IM Pre & post (1 h)	26 PDH, 41 AT	None	552 nmol/L (20 µg/dL)	Quoted study[11]: 64 control samples from 16 control dogs, IT = mean + 3SD	73 HAC 88 PDH 63 AT	—
26	Plasma (not specified)	Enzyme immunoassay	Cortrosyn, 0.25 mg/dog IM Pre & post (1 h)	43 PDH	None	552 nmol/L (20 µg/dL)	Quoted study[11]: 64 control samples from 16 control dogs, IT = mean + 3SD	79	—
24	Serum	RIA	Cosyntropin (8h post LDDST 0.015 mg/kg IV), 10 µg/kg IV Post (1 h)	20 PDH	59 NAI	≈500 nmol/L (≈18 µg/dL§)	No source	80	86
49	Plasma (not specified)	Assay not specified	ACTH gel, 2.2 U/kg IM Post (2 h)	22 HAC	33 HAC-S	552 nmol/L (20 µg/dL)	IT established by the endocrine diagnostic laboratory at the teaching hospital (unpublished data)	95	91
34	Serum	RIA	Cortrosyn, 5 µg/kg IM Post (1 h)	32 HAC	29 HAC-S	M: 298 nmol/L (10.8 µg/dL) CM: 417 nmol/L (15.1 µg/dL) F: 483 nmol/L (17.6 µg/dL) SF: 486 nmol/L (17.5 µg/dL)	Quoted study[18]: RI for post-ACTH cortisol in healthy dogs per sex per sexual status	84	59
4	Plasma (heparinized)	CLIA: Immulite 1000 and 2000 (Siemens)	Synacthen, 125 µg/dog <5 kg IM 250 µg/dog >5 kg IM Pre & post (1 h)	59 HAC	52 NAI	599 nmol/L (21.7 µg/dL)	No source	71.2	79
36	Serum	CLIA: Elecsys/Cobas (Roche)	Tetracosactide, 125 µg/dog <5 kg IM 250 µg/dog >5 kg IM Pre & post (1 h)	36 PDH	18 NAI	ROC curve: • Either 684 nmol/L (24.8 µg/dL) • Or 717 nmol/L (26.0 µg/dL) (when threshold increases: Se decreases, and Sp increases)	36	86 81	94 100

Dash (—) = Se or Sp not provided; ACTH = adrenocorticotropic hormone; AT = dogs with adrenal tumor responsible for hyperadrenocorticism; CBPM = competitive binding protein method; CLIA = chemiluminescent immunoassay; CM = castrated males; F = intact females; FMM = fluorometric method of Mattingly; HAC = dogs with hyperadrenocorticism; HAC-S = dogs suspected of hyperadrenocorticism confirmed to not have hyperadrenocorticism; IT = interpretation threshold; LDDST = low-dose dexamethasone suppression test; M = intact males; NAI = dogs with non-adrenal illness; PDH = dogs with pituitary-dependent hyperadrenocorticism; RI = reference interval; RIA = radioimmunoassay; ROC = receiver operating characteristic; Se = sensitivity; SF = spayed females; Sp = specificity.

* Articles in which the LDDST is also investigated.
† The sensitivity is not stated in the article but can be retrieved with the data of the article and is also stated as such by later reviews.
‡ Importantly, when the threshold is defined according to data in the article itself, if the threshold is defined so that it is higher than the highest healthy dogs, the specificity is 100% by a mathematical effect. It does not mean that the test does not have false positives. In a different study with a different healthy control population, or, even better, a compared population of diseased dogs or HAC suspects, the specificity becomes typically imperfect.
§ No IT was provided in the text; a figure contained a shaded area picturing the "reference range values," from which the threshold could be roughly inferred.

Table 11. Chronologic and technical evolution of the post-dexamethasone suppression interpretation threshold for Cushing disease diagnosis (low dose) and localization (high dose) in dogs.

Ref.	Sample medium	Cortisol assay	Dexamethasone suppression test†	Cushing population	Compared population	IT	Source of the IT	Se (%)	Sp (%)
32	Plasma (not specified)	FMM (non-immune)	LDDST (0.01 mg/kg IV)	22 PDH	24 healthy kennel dogs	140 nmol/L (5.1 μg/dL) • 8h post-dex range normal dogs: 15–130 nmol/L • 8h post-dex range PDH: 150–520 nmol/L	32	100‡	—
33	Plasma (not specified)	RIA	LDDST (0.01 mg/kg IV)	8 AT	21 healthy kennel dogs	Not provided Deducible IT from provided data: 41–65 nmol/L (1.5–2.4 μg/dL) • 8h post-dex range normal dogs: 0–40 nmol/L • 8h post-dex range AT: 66–290 nmol/L	33	100‡	—
11	Plasma (heparinized)	CBPM (non-immune radioassay)	LDDST (0.01 mg/kg IV) (Post 8h only)	26 PDH, 7 AT, 31 HAC (unknown location)	22 control dogs	38.6 nmol/L (1.4 μg/kg) • 8h post-dex normal dogs: mean + 3SD = 37.8 nmol/L (1.37 μg/dL)	11	92 HAC 96 PDH 100 AT	—
10	Plasma (heparinized)	CBPM (non-immune radioassay)	HDDST (0.1 mg/kg IV) (Post 8h only)	10 PDH, 3AT	20 control dogs	8h >50% baseline • 8h post-dex all normal dogs <50% baseline • better than 8h > (mean + 3SD) of healthy dogs: 27.6 nmol/L (1 μg/dL)	10	92§	—
12	Plasma (heparinized)	CBPM (non-immune radioassay)	LDDST (0.01 mg/kg IV)	15 PDH, 6 AT	None	38.6 nmol/L (1.4 μg/dL)	Quoted study[11]	100	—
7	Plasma (EDTA)	RIA	LDDST (0.01 mg/kg IV)	None	33 NAI 14 clinically normal dogs	37 nmol/L (1.34 μg/dL) • 8h post-dex normal dogs: mean + 2SD = 37 nmol/L	7	—	52
13	Plasma (heparinized)	CBPM (non-immune radioassay)	LDDST (0.01 mg/kg IV)	14 PDH, 3 AT	None	38.6 nmol/L (1.4 μg/dL)	Quoted study[11]	94	—
44	Plasma (not specified)	RIA	LDDST (0.01 mg/kg IV)	129 HAC	37 HAC-S	40 nmol/L (1.45 μg/dL)	No source	85	73
30	Plasma (heparinized)	Enzyme immunoassay	LDDST (0.01 mg/kg dexamethasone phosphate IV)	14 PDH, 4 AT	5 healthy control dogs	38.6 nmol/L (1.4 μg/dL)	Quoted study[10]‖	100	100
			LDDST (0.015 mg/kg dexamethasone polyethylene glycol IV)	14 PDH, 4 AT	5 healthy control dogs	38.6 nmol/L (1.4 μg/dL)	Quoted study[10]‖	89 HAC 86 PDH 100 AT	100
43	Plasma (heparinized)	CBPM (non-immune radioassay)	LDDST (0.01 mg/kg IV)	28 AT, 26 PDH	22 healthy dogs	38.6 nmol/L (1.4 μg/dL)	Quoted study[10]‖	98 HAC 100 AT 96 PDH	—
26	Plasma (not specified)	Enzyme immunoassay	LDDST (0.01 mg/kg IV)	33 PDH	None	38.6 nmol/L (1.4 μg/dL)	Quoted study[11]	97	—
24	Serum	RIA	LDDST (0.015 mg/kg IV)	20 PDH	59 NAI	About 30 nmol/L.# (~1.1 μg/dL.#)	No source	100	44
16	Plasma (heparinized)	Enzyme immunoassay	LDDST (0.01 mg/kg IV) Post 4h and 8h	181 PDH, 35 AT	None	For HAC diagnosis: 38.6 nmol/L (1.4 nmol/L) For HAC localization: 4h >1.4 μg/dL. Or 4h or 8h >50% baseline	For HAC diagnosis: quoted study[11] For HAC localization: article itself	61§	100§
			HDDST (0.1 mg/kg IV) Post 4h and 8h	181 PDH, 35 AT	35 AT	4h or 8h: >1.4 μg/dL or >50% baseline	16	75§	95§
49	Plasma (not specified)	Assay not specified	LDDST (0.01 mg/kg IV)	28 HAC	10 NAI	41.4 nmol/L (1.5 μg/dL)	IT established by the endocrine diagnostic laboratory at the teaching hospital (unpublished data)	96	70
4	Plasma (heparinized)	CLIA: Immulite 1000 and 2000 (Siemens)	LDDST (0.015 mg/kg IV) Post 4h and 8h	59 HAC	64 NAI	4h or 8h >27.6 nmol/L (1 μg/dL)	No source	96.6	67.2
36	Serum	CLIA: Elecsys/Cobas (Roche)	LDDST (0.01 mg/kg IV)	LDDST only for selection of HAC cases, not part of the investigation (focused on the best ACTH threshold)		8h >30.3 nmol/L (1.1 μg/dL)	IT based on the laboratory's RI for these tests and previous reports	—	—

Dash (—) = Se or Sp not provided; ACTH = adrenocorticotropic hormone; AT = dogs with adrenal tumor responsible for hyperadrenocorticism; CBPM = competitive binding protein method; CLIA = chemiluminescent immunoassay; Dex = dexamethasone; F = intact females; FMM = fluorometric method of Mattingly; HAC = dogs with hyperadrenocorticism; HAC-S = dogs suspected of hyperadrenocorticism confirmed to not have hyperadrenocorticism; HDDST = High dose dexamethasone suppression test; IT = interpretation threshold; LDDST = Low dose dexamethasone suppression test; M = intact males; NAI = dogs with non-adrenal illness; NM = neutered males; PDH = dogs with pituitary-dependent hyperadrenocorticism; RI = reference interval; RIA = radioimmunoassay; ROC = receiver operating characteristic; Se = sensitivity; SF = spayed females; Sp = specificity.

* Articles in which the ACTH stimulation test is also investigated.
† When hours are not specified, it is a suppression 8h post-dexamethasone injection. It does not necessarily mean that no previous time points (3 and/or 4h) have not been performed in the studies, but we do not report them. When we chose to report the additional time point at 4h, it is associated with an important conclusion in the studies.
‡ The sensitivity of the LDDST for HAC was not stated in the article, but could be computed from the provided data, and is also stated as such by later reviews.
§ Those Se and Sp are not for the diagnosis of HAC; they are for the localization of HAC, provided that HAC has already been diagnosed. In this context, a "positive" dog is a dog that suppresses cortisol after dexamethasone injection (LDDST or HDDST), whereas a "negative" dog does not suppress. Thus, imperfect sensitivity (presence of false positives) corresponds with PDH, which do not suppress, and imperfect specificity (presence of false negatives) corresponds with AT, which suppress.
‖ Those studies are wrongly referencing the Feldman's study on HDDST[10] to support their choice of IT for the LDDST. They intended to reference another study.[11] The mistake is not rare, as both studies from Feldman were published in 1983 and in the same journal.
No IT was provided in the text; a figure contained a shaded area picturing the "reference range values," from which the threshold could be only roughly inferred.

is to manage the assay. For competitive immunoassays (frequent in endocrinology, among which the cortisol assay we used in our study), because the measured fraction (tracer) is the displaced fraction (as opposed to the bound fraction in non-competitive immunoassays), the graph signal = f(concentration) is typically sigmoid. To straighten the curve, one or several functions are used on the axes, in an algorithm fixed for the life of the assay once it has been validated. For example, the signal on the y-axis is often transformed by the Logit function (log[y/(1-y)], which is the opposite of a sigmoid. The concentration on the x-axis may be transformed by the Log function (log[x]) to further straighten the curve, in a common algorithm for competitive immunoassays called "Logit-Log." Unlike radioimmunoassays, the Immulites do not allow visual assessment of the calibration curve and the algorithms; instead, 2 adjustors are used to adjust a calibration curve preregistered in the software. In any case, the canine serum cortisol linearity that we obtained with the Immulite 2000 Xpi was excellent, mirroring the inaccessible calibration curve of the assay, and confirmed that the reportable range of 27.6–1,380 nmol/L (1–50 µg/dL) provided by the manufacturer was adequate.

Samples measured outside the manufacturer reportable range spontaneously are reported as < or > the lower or upper limit, respectively. The Immulite 2000 and 2000 Xpi offer 2 options to measure samples outside of the manufacturer's reportable range: the calibration verifier mode (CVM) and the range change software (RCS). They generate, in theory, the same result, but have very different contexts of use. The CVM eliminates the limits of the reportable range (thus with no need of setting up a new acceptable range). It is not typically used in veterinary medicine. It is not typically used for patient testing in the human field either. It is most often used by human laboratories performing linearity testing: facilities that perform human testing are required by most accrediting agencies to verify the reportable range every 6 mo. Siemens sells a product called Calibration Verification Modules that customers can use to fulfill this requirement, which is where the term "verifier" came from. On the other hand, the RCS is a CD-ROM provided by Siemens exclusively for veterinary customers, which uploads new software on the Immulite computer. This software creates a file from which the user can manually set up new reportable ranges at will for each measurand. It is the responsibility of the veterinary laboratory to document acceptability of the new reportable range. Modification of the reportable range must be done measurand by measurand, and for each new lot, as the reportable range comes back to its original setting at each new lot.

We performed the detection limit study to verify the manufacturer's decision of defining 27.6 nmol/L (1 µg/dL) as the lower limit, as well as to characterize the precision of results that we report below this limit. LOQ is the most useful value, as it quantifies the precision of the lower limit. We determined that the precision remained ~13% as low as 16.6 nmol/L

(0.60 µg/dL) and 9.4 nmol/L (0.34 µg/dL), which is acceptable for such low concentrations.

Because of the critical importance of cortisol interpretation at 38.6 and 552 nmol/L (1.4 and 20 µg/dL, respectively), these levels were further investigated with a between-run replication study. Terms other than within-run (e.g., "between-run," "long-term," or "reproducibility") should always be defined in publications because they mean different designs for different sources.[1,40,48] "Between-run" is the vaguer version and encompasses everything that is not within-run. We chose this denomination to avoid overinterpretation of our results. In the 2019 ASVCP guidelines,[1] "between-run," long-term," and "reproducibility" are used interchangeably, and defined as 20 measurements over at least 20 d for QCMs, and 20 measurements divided as 4 daily measurements over 5 d, which is what we did in our study. On the other hand, meteorologists,[48] define "intermediate precision" as >90–100 d. Similarly, the International Organization for Standardization (ISO) defines[19] "intermediate precision" as measurements over months, and uses it as a synonym for within-laboratory reproducibility, as opposed to between-laboratory reproducibility.

The within-run and between-run CV(%) should not represent >25% and 33% of TEa, respectively.[1] This is easily achievable for methods with high precision, as, for example, for biochemistry measurands. Immunoassays have lower precision, therefore we may or may not be able to respect this guideline, depending on the elected TEa; we anticipate that for some immunoassays, the CV may represent a higher proportion of the TEa budget.

The analytical performances of the Immulite 2000 Xpi at both ITs can be regarded with the between-run CV only, or with TEo. Taking the example of the 38.6 nmol/L (1.4 µg/dL) IT, the nearly 10% between-run CV signifies that a sample of exactly 38.6 nmol/L (1.4 µg/dL) has a 95% probability of being measured within 38.6 nmol/L ± 2 SD, which is 31–46 nmol/L (or within 1.4 µg/dL ± 2SD, which is 1.1–1.7 µg/dL). It is interesting to consider, for example, one measurement at 1.2 µg/dL and one measurement at 1.6 µg/dL as not essentially being different. Moreover, because both of those ITs were determined ~40 y ago with a different method,[11] the bias needs to be considered for interpretation of the results because imprecision alone will not account for the technical difference between methods or laboratories. In theory, the bias could be addressed either with the determination of an IT within the laboratory (which is actually removing the bias component, but is virtually never done because of the complexity and cost of such studies) or addressed at the level of the interpretation of results by considering TEo instead of the between-run CV only for significance. The bias in our study is especially interesting, given that it results from measurement differences between 2 Immulites 2000 Xpi, and thus is an example of the lowest inter-laboratory bias that one could expect. In an ideal situation, results should be comparable across

laboratories. If the bias is eliminated or minimized, RIs determined for a specific instrument and method, or even across different instruments and methods, could be generated. This is not currently the case for state-of-the art performance, and different biases exist even for the same instruments within different laboratories[8] (and non-published observations from SYNLAB-VPG). Then, taking the bias into consideration appears clearly justified when interpreting the results compared to IT according to TEo. Thus, when adding the bias (SR, RB, or AB) component, and thus considering TEo (TEo_{SR}, TEo_{RB}, or TEo_{AB}) of ~30% in our study with the Immulite 2000 Xpi, the interval of 95% probability of measurement becomes 38.6 nmol/L ± TEo, which is 27–50 nmol/L (or 1.4 µg/dL ± TEo, which is 0.98–1.82 µg/dL). The same reasoning can be applied to the IT of 552 nmol/L (20 µg/dL) with a CV of ~7.5% and TEo of ~20%, for which 552 nmol/L ± 2 SD generates an interval of 469–635 nmol/L (17–23 µg/dL), and 552 nmol/L ± TEo generates an interval of 442–662 nmol/L (16–24 µg/dL).

We used the Bland–Altman method for our between-laboratory comparison study. The mean of the differences of 9.7 nmol/L (−0.35 µg/dL) suggested a minimal negative bias in our facility, with no clinical significance. Agreement limits were computed after the normality of the differences between methods was demonstrated; we elected a threshold for the *p* value of the D'Agostino–Person normality test of 0.3 rather than 0.05, given that a recent study demonstrated in simulated clinical settings that the *p* value with optimal sensitivity and specificity according to a ROC curve was 0.18 at *n* = 30, and could go up to 0.29 with the Anderson–Darling method under the same conditions.[29] We believe that in a clinical setting at low *n*, it is more reasonable to increase the threshold for normality from 0.05 to 0.3 in order to limit false positives (type I error). Agreement limits were determined as −51.6 to +32.0 nmol/L (−1.87 to +1.16 µg/dL), then with no clinical significance. Admittedly those variations could have a clinical significance at low concentration; however, the visual analysis of the Bland–Altman graph revealed that differences between methods were close from zero up to 27.6 nmol/L (1 µg/dL), and that the limited negative bias was observed mostly at high concentration, which would not result in a change in the interpretation.

The original aspect of our study is to consider 2 spiked samples (L4 and L8) to model patient samples at 2 clinically relevant serum cortisol concentrations, and to consider their use as 2 QCM levels. When considered as QCM levels, these 2 samples each provide biases (SR, AB, RB) and CV allowing calculation of the corresponding TEo. The 3 investigated types of bias (SR, AB, RB) yield relatively close TEo (25.6%, 22.0%, and 32.8% for L4; 17.9%, 17.7%, and 19.3% for L8). At P_{ed} >90%, the acceptable QC rules are roughly the same within each level (L4 and L8) regardless of the considered type of bias (except $L4_{AB}$: see next paragraph), whereas the chosen TEa is critical in determining the set of acceptable QC rules. This emphasizes the major influence of the chosen

TEa on the resulting acceptable QC rules, further illustrating the direct relationship between the usable QC rules and the quality goal represented by TEa.

For $L4_{AB}$, more QC rules were acceptable given a low AB achieved by averaging multiple RB and thus providing an illusion of a better performing test. Especially in endocrinology, we believe that the analytical performance (precision and accuracy) should be assessed at clinically relevant concentrations to avoid the pitfall of error averaging that may occur when errors vary with concentrations of clinical interest.

The CV for both commercial QCM levels measured over one month is lower than the CV of the spiked samples L4 and L8 measured over only a week. The existence of target values from the manufacturer allows for a straightforward computation of the bias, which is minimal for both levels (0.3% and 0.5%). Both low CV and low bias illustrate the desirable stability properties of QCMs. The extensive QC rule validation study highlights the small influence of P_{ed} and the major influence of TEa on acceptable QC rules. Decreasing P_{ed} below 90% is highly discouraged. Maintaining high P_{ed} is crucial for QC, and "doesn't cost much" in terms of QC rules. At an arbitrarily chosen low TEa, almost no rules were acceptable for any level (with the exception of QCM1 as a result of especially low TEo); at voluntarily high TEa, all rules were acceptable for almost all levels (with the exception of L4 as a result of especially high TEo). Given the major influence of TEa on QC rule candidates, increasing TEa appears to be the best way of increasing QC rule availability. Of course, TEa cannot be increased indefinitely and needs to remain below the clinically useful TEa. From this equation came the idea of reversing the QC rule validation approach, optimizing TEa just to the needed level to accept a simple QC rule, but not more.[28] Of note, for all levels, the 1_{2S} rule was disregarded because of excessive P_{fr} (0.09), resulting in excessive occurrence of false rejections, and thus being time and cost prohibitive. For similar practical reasons, even if performance was improved, scenario with *N* = 4 control levels (results not provided) or with more complex multi-rules (results not provided) were disregarded as impractical for use in veterinary laboratories.

In light of these observation, we cannot yet provide recommendations for the use of specific QC rules, given that those mostly depend on the elected TEa for which there is currently no consensus in veterinary medicine. However, we are proposing the use of a "reverse approach" to determine the amount of error for which various QC rules can provide a high P_{ed} and low P_{fr} in another paper in this series.[28]

We took extensive precautions to minimize limitations in our study (pre- and post-matrix constitution measurements, removal of not insignificant cortisol in the serum matrix, multiple determination of TEo and QC rules with various materials at various concentrations, etc.); however, some limitations could not be overcome and require acknowledgment. One limitation is the absence of clinical investigation

from our study about the relevance of the 38.6 and 552 nmol/L (1.4 and 20 µg/dL, respectively) IT for serum cortisol in dogs with this measurement method. We aimed specifically to quantify the amount of random error, systematic error, and total error associated with those commonly used IT with the commonly used method of the Immulite 2000 Xpi. We also did not characterize interferences caused by hemolysis, lipemia, or icterus. Finally, we reported results with a single instrument; to make recommendations for QC validation purposes, one would need to know that these goals are achievable consistently over time and across analyzers.

Acknowledgments

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iDs

Jérémie Korchia (iD) https://orcid.org/0000-0002-9344-6639

Kathleen P. Freeman (iD) https://orcid.org/0000-0003-1796-0158

References

1. Arnold JE, et al. ASVCP guidelines: principles of quality assurance and standards for veterinary clinical pathology (version 3.0): developed by the American Society for Veterinary Clinical Pathology's (ASVCP) Quality Assurance and Laboratory Standards (QALS) Committee. Vet Clin Pathol 2019;48:542–618.
2. Behrend EN, et al. Diagnosis of canine hyperadrenocorticism. Vet Clin North Am Small Anim Pract 2001;31:985–1003.
3. Behrend EN, et al. Diagnosis of spontaneous canine hyperadrenocorticism: 2012 ACVIM consensus statement (small animal). J Vet Intern Med 2013;27:1292–1304.
4. Bennaim M, et al. Evaluation of individual low-dose dexamethasone suppression test patterns in naturally occurring hyperadrenocorticism in dogs. J Vet Intern Med 2018;32: 967–977.
5. Bilić-Zulle L. Comparison of methods: Passing and Bablok regression. Biochem Med (Zagreb) 2011;21:49–52.
6. Bland JM, et al. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307–310.
7. Chastain CB, et al. Evaluation of the hypothalamic pituitary-adrenal axis in clinically stressed dogs. J Am Anim Hosp Assoc 1986;22:435–442.
8. Farr AJ, et al. 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. Feldman EC. Effect of functional adrenocortical tumors on plasma cortisol and corticotropin concentrations in dogs. J Am Vet Med Assoc 1981;178:823–826.
10. Feldman EC. Distinguishing dogs with functioning adrenocortical tumors from dogs with pituitary-dependent hyperadrenocorticism. J Am Vet Med Assoc 1983;183:195–200.
11. Feldman EC. Comparison of ACTH response and dexamethasone suppression as screening tests in canine hyperadrenocorticism. J Am Vet Med Assoc 1983;182:506–510.
12. Feldman EC. Evaluation of a combined dexamethasone suppression/ACTH stimulation test in dogs with hyperadrenocorticism. J Am Vet Med Assoc 1985;187:49–53.
13. Feldman EC. Evaluation of a six-hour combined dexamethasone suppression/ACTH stimulation test in dogs with hyperadrenocorticism. J Am Vet Med Assoc 1986;189: 1562–1566.
14. Feldman EC, et al. Plasma cortisol response to ketoconazole administration in dogs with hyperadrenocorticism. J Am Vet Med Assoc 1990;197:71–78.
15. Feldman EC, et al. The synthetic ACTH stimulation test and measurement of endogenous plasma ACTH levels: useful diagnostic indicators for adrenal disease in dogs. J Am Anim Hosp Assoc 1978;14:524–531.
16. Feldman EC, et al. Use of low- and high-dose dexamethasone tests for distinguishing pituitary-dependent from adrenal tumor hyperadrenocorticism in dogs. J Am Vet Med Assoc 1996;209:772–775.
17. Foster LB, et al. Single-antibody technique for radioimmunoassay of cortisol in unextracted serum or plasma. Clin Chem 1974;20:365–368.
18. Frank LA, et al. Steroid hormone concentration profiles in healthy intact and neutered dogs before and after cosyntropin administration. Domest Anim Endocrinol 2003;24:43–57.
19. Fuentes-Arderiu X. Glossary of ISO terms. Westgard QC. [cited 2020 Aug 23]. https://www.westgard.com/isoglossary. htm
20. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25:141–151.
21. Graham P. Preliminary report on impact of Immulite 2000 cortisol antibody change October 2020 including evaluation of manufacturer recommended adjustment factors.[cited 2020 Nov 3]. https://www.esve.org/news/20 20/20201109cortisolmeasurement_PreliminaryReport_ Immulite2000Impact.pdf
22. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 2013;4:627–635.
23. Harr KE, et al. ASVCP guidelines: allowable total error guidelines for biochemistry. Vet Clin Pathol 2013;42: 424–436.
24. Kaplan AJ, et al. Effects of disease on the results of diagnostic tests for use in detecting hyperadrenocorticism in dogs. J Am Vet Med Assoc 1995;207:445–451.
25. Kemppainen RJ, et al. Use of a low dose synthetic ACTH challenge test in normal and prednisone-treated dogs. Res Vet Sci 1983;35:240–242.

26. Kipperman BS, et al. Pituitary tumor size, neurologic signs, and relation to endocrine test results in dogs with pituitary-dependent hyperadrenocorticism: 43 cases (1980–1990). J Am Vet Med Assoc 1992;201:762–767.

27. Korchia J, et al. Validation study of canine urine cortisol measurement with the Immulite 2000 Xpi cortisol immunoassay. J Vet Diagn Invest 2021;33(6).

28. Korchia J, et al. Total observable error, total allowable error, and QC rules for canine serum and urine cortisol achievable with the Immulite 2000 Xpi cortisol immunoassay. J Vet Diagn Invest. Submitted.

29. Le Boedec K. Sensitivity and specificity of normality tests and consequences on reference interval accuracy at small sample size: a computer-simulation study. Vet Clin Pathol 2016;45:648–656.

30. Mack RE, et al. Comparison of two low-dose dexamethasone suppression protocols as screening and discrimination tests in dogs with hyperadrenocorticism. J Am Vet Med Assoc 1990;197:1603–1606.

31. Mattingly D. A simple fluorimetric method for the estimation of free 11-hydroxycorticoids in human plasma. J Clin Pathol 1962;15:374–379.

32. Meijer JC, et al. Biochemical characterization of pituitary-dependent hyperadrenocorticism in the dog. J Endocrinol 1978;77:111–118.

33. Meijer JC, et al. Adrenocortical function tests in dogs with hyperfunctioning adrenocortical tumours. J Endocrinol 1979; 80:315–319.

34. Monroe WE, et al. Concentrations of noncortisol adrenal steroids in response to ACTH in dogs with adrenal-dependent hyperadrenocorticism, pituitary-dependent hyperadrenocorticism, and nonadrenal illness. J Vet Intern Med 2012;26: 945–952.

35. Nabity MB, et al. ASVCP guidelines: allowable total error hematology. Vet Clin Pathol 2018;47:9–21.

36. Nivy R, et al. The interpretive contribution of the baseline serum cortisol concentration of the ACTH stimulation test in the diagnosis of pituitary dependent hyperadrenocorticism in dogs. J Vet Intern Med 2018;32:1897–1902.

37. O'Neil MJ, et al. The Merk Index. 15th ed. RSPC Publishing, 2013:4824.

38. Peterson ME. Diagnosis of hyperadrenocorticism in dogs. Clin Tech Small Anim Pract 2007;22:2–11.

39. Peterson ME, et al. Plasma cortisol response to exogenous ACTH in 22 dogs with hyperadrenocorticism caused by adrenocortical neoplasia. J Am Vet Med Assoc 1982;180:542–544.

40. Rebane R. LC-MS method validation: repeatability, intermediate precision and reproducibility. LC-MS method validation, University of Tartu. [cited 2020 Aug 23]. https://sisu.ut.ee/lcms_method_validation/41-precision-trueness-accuracy

41. Reimers TJ, et al. Validation of radioimmunoassay for triiodothyronine, thyroxine, and hydrocortisone (cortisol) in canine, feline, and equine sera. Am J Vet Res 1981;42:2016–2021.

42. Reimers TJ, et al. Validation and application of solid-phase chemiluminescent immunoassays for diagnosis of endocrine diseases in animals. Comp Haematol Int 1996;6: 170–175.

43. Reusch CE, et al. Canine hyperadrenocorticism due to adrenocortical neoplasia. Pretreatment evaluation of 41 dogs. J Vet Intern Med 1991;5:3–10.

44. Rijnberk A, et al. Assessment of two tests for the diagnosis of canine hyperadrenocorticism. Vet Rec 1988;122:178–180.

45. Russell NJ, et al. Comparison of radioimmunoassay and chemiluminescent assay methods to estimate canine blood cortisol concentrations. Aust Vet J 2007;85:487–494.

46. Singh AK, et al. Validation of nonradioactive chemiluminescent immunoassay methods for the analysis of thyroxine and cortisol in blood samples obtained from dogs, cats, and horses. J Vet Diagn Invest 1997;9:261–268.

47. Smith MC, Feldman EC. Plasma endogenous ACTH concentrations and plasma cortisol responses to synthetic ACTH and dexamethasone sodium phosphate in healthy cats. Am J Vet Res 1987;48:1719–1724.

48. Theodorsson E, et al. Bias in clinical chemistry. Bioanalysis 2014;6:2855–2875.

49. Van Liew CH, et al. Comparison of results of adrenocorticotropic hormone stimulation and low-dose dexamethasone suppression tests with necropsy findings in dogs: 81 cases (1985–1995). J Am Vet Med Assoc 1997;211:322–325.

50. Vis JY, et al. Verification and quality control of routine hematology analyzers. Int J Lab Hematol 2016;38(Suppl 1): 100–109.

51. Watson AD, et al. Plasma cortisol concentrations in dogs given cortisone or placebo by mouth. Res Vet Sci 1993;55: 379–381.

52. Wenger-Riggenbach, et al. Salivary cortisol concentrations in healthy dogs and dogs with hypercortisolism. J Vet Intern Med 2010;24:551–556.

53. Westgard J. Rilibak—German guidelines for quality. Westgard QC, 2015. https://www.westgard.com/rilibak.htm

54. Westgard J. CLIA Requirements for analytical quality. [cited 2020 Aug 23]. https://www.westgard.com/clia.htm

55. Westgard S. Consolidated comparison of chemistry performance specifications. [cited 2021 Apr 21]. https://www.westgard.com/consolidated-goals-chemistry.htm

56. Zerbe CA, et al. Effect of nonadrenal illness on adrenal function in the cat. Am J Vet Res 1987;48:451–454.

57. Zerbe CA, et al. Adrenal function in 15 dogs with insulin-dependent diabetes mellitus. J Am Vet Med Assoc 1988;193:454–456.