

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

Jeremie Korchia,¹ Kathleen P. Freeman²

Abstract. We report here validation of the Immulite 2000 Xpi cortisol immunoassay (Siemens; with kit lot numbers <550) for measurement of urine cortisol in dogs, with characterization of the precision (CV), accuracy (spiking-recovery [SR] bias), and observed total error ($TE_o = \text{bias} + 2CV$) across the reportable range. Linearity assessed by simple linear regression was excellent. Imprecision, SR bias, and TE_o increased markedly with decreasing urine cortisol concentration. Interlaboratory comparison studies determined range-based (RB) bias and average bias (AB). The 3 biases (SR, RB, and AB) and resulting TE_o differed markedly. At 38.6 and 552 nmol/L (1.4 and 20 $\mu\text{g/dL}$), between-run CVs were 10% and 4.5%, respectively, and $TE_{o_{RB}}$ were ~30% and 20%, respectively, similar to observations in serum in another validation study. These analytical performance parameters should be considered for urine cortisol:creatinine ratio (UCCR) result interpretation, given that, for any hypothetical errorless urine creatinine measurement, the error % on UCCR mirrors the error % on urine cortisol. Importantly, there is no commonly used interpretation threshold for UCCR, given that UCCR varies greatly depending on measurement methods and threshold computation. To date, there is no manufacturer-provided quality control material (QCM) with target values for urine cortisol with an Immulite; for Liquicheck QCM (Bio-Rad), between-run imprecision was ~5% for both QCM levels. Acceptable QC rules are heavily dependent on the desired total allowable error (TEa) for the QCM system, itself limited by the desired clinical TEa.

Keywords: bias; cortisol; dogs; endocrinology; hyperadrenocorticism; total error; urine; urine cortisol:creatinine ratio.

The urine cortisol:creatinine ratio (UCCR) is reputed to have high sensitivity for hyperadrenocorticism (Cushing disease) in dogs and is commonly used to rule out hyperadrenocorticism when the UCCR is below a certain interpretation threshold (IT). The UCCR IT varies mostly with the method used to measure urine cortisol because measurement of cortisol requires an immunoassay, and immunoassays are characterized by significantly lower precision and accuracy than biochemical assays. Further minor variation may be introduced, depending on the method of creatinine measurement. One of the earliest methods used to measure canine urine cortisol¹⁹ was a non-immunologic fluorometric assay (method of Mattingly) with only minimal validation in urine.¹⁴ In the 1990s, a radioimmunoassay (RIA) method (Gammacoat RIA clinical assays kit; Travenol-Genentech Diagnostics) was characterized for canine urine cortisol in studies relying on dilutional linearity, spiking-recovery, and intra-assay precision.^{4,9} One study showed linearity with 1:2 and 1:4 dilution in phosphate-buffered saline, and intra-assay precision of 4.3% at 80 nmol/L (2.9 $\mu\text{g/dL}$) and 3.0% at 180 nmol/L (6.5 $\mu\text{g/dL}$).⁹ The second study showed linearity with 1:2 and 1:4 dilution from 2 canine urine pools, and intra-assay precision of 1.9% at 28 nmol/L (1 $\mu\text{g/dL}$) and 5.2% at 70 nmol/L (2.5 $\mu\text{g/dL}$).⁴ Both studies stated recovery of 90–105% at 3 different concentrations (without providing

the volume used to calculate the concentration of the spiked aliquots).^{4,9} Both studies found inter-assay imprecision lower than the intra-assay imprecision, suggesting a questionable design for the inter-assay investigation.^{4,9} In 1996, a study focused on serum cortisol also validated canine urine cortisol with the Immulite 1, the first version of the Immulite created by Diagnostic Products Corporation (DPC) Cirrus before it was purchased by Siemens Healthineers¹⁵; it demonstrated dilutional parallelism from 1:1 to 1:16, as well as satisfactory comparison with a previously validated RIA (simple linear regression: $y = 1.55x + 0.43$ in nmol/L).

Because UCCR is a ratio, there is no IT for canine urine cortisol alone. Thus, a somewhat low (L4) and a somewhat high (L8) urine cortisol concentration have been elected to investigate between-run imprecision and resulting observed total error (TE_o); we chose them to correspond with the 2 serum levels of interest to allow for comparison of matrix

¹Texas A&M Veterinary Medical Diagnostic Laboratory, Texas A&M University, College Station, TX, USA (Korchia); SYNLAB-VPG/Exeter, Exeter, United Kingdom (Freeman).

²Corresponding author: Jeremie 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

properties. Importantly, percentages of variation related to urine cortisol CV, bias, and TEo can still be applied to the UCCR, provided that urine creatinine remains constant; in other words, if we assume the approximation that urine creatinine is virtually errorless, then the % of error in the urine cortisol corresponds with the % of error of the UCCR. UCCR is a common test in small animal internal medicine, mostly in dogs, used to exclude hyperadrenocorticism when below a certain IT. The computation formula, as well as unit translation, are as follow¹⁸:

Formula:

$$\text{UCCR} = U_{\text{cort (nmol/L)}} / U_{\text{creat (nmol/L)}}$$

$$\text{UCCR} = U_{\text{cort (nmol/L)}} / (U_{\text{creat (}\mu\text{mol/dL)}} / 1,000)$$

$$\text{UCCR} = U_{\text{cort (}\mu\text{g/dL)}} \times 27.6 / [(U_{\text{creat (mg/dL)}} / 1,000) \times 88.4]$$

Unit translation:

$$U_{\text{creat (nmol/L)}} = U_{\text{creat (}\mu\text{mol/dL)}} / 1,000$$

$$U_{\text{cort (nmol/L)}} = U_{\text{cort (}\mu\text{g/dL)}} \times 27.6$$

$$U_{\text{creat (}\mu\text{mol/dL)}} = U_{\text{creat (mg/dL)}} \times 88.4$$

UCCR being (nmol/L)/(μmol/L)/1,000, has no units; however, it is expressed with a $10^{-9}/10^{-6}/10^3 = a 10^{-6}$ factor (which makes sense as there is much more creatinine than cortisol in urine).

Urine cortisol was measured for many years at the Texas Veterinary Medical Diagnostic Laboratory (TVMDL; College Station, TX, USA) with the RIA ImmuChem *for in vitro diagnostic use* (MP Biomedical Diagnostic), which includes a trichloromethane extraction step. This assay was discontinued in late 2018 and replaced by another version labeled *for research use only* (with identical package inserts). CortiCote (MP Biomedical Diagnostic) is a different RIA from the same manufacturer, with claimed validation in urine in addition to serum and plasma (in humans), that does not require a trichloromethane extraction step (making it less laborious to perform) but uses a different capture antibody for cortisol with more extensive cross-reaction with prednisone.

The high output, automation, and cost-effective properties of the Immulite 2000 Xpi support its validation for canine urine cortisol measurement. Assuming virtually errorless creatinine measurement, the % of variation in urine cortisol determines the % of variation in UCCR. To date, there are no consensus-based recommendations for the total allowable error (TEa) for UCCR in veterinary medicine. In such a situation, knowing the TEo is very useful in establishing a reasonable consensus for TEa, which in turn will impact interpretation and clinical decisions. As demonstrated in serum,¹¹ the CV, bias, and resulting TEo vary with the cortisol concentration. Our objectives were: 1) to validate urine cortisol measurement in dogs with the Immulite 2000 Xpi, 2) to characterize test performances (CV, bias, TEo) across the reportable concentration range, 3) to compare performance

in urine with that in serum, and 4) to validate quality control (QC) rules for QC materials (QCMs) to monitor adequately the system performance of the Immulite 2000 Xpi.

Materials and methods

Study overview

Our study was designed to validate canine urine cortisol measurement with the Immulite 2000 Xpi (Siemens) cortisol immunoassay (intended for serum), following the immunoassay validation protocol recommended by the Quality Assurance and Laboratory Standards committee of the American Society for Veterinary Clinical Pathology.¹ The dilutions used for the spiking-recovery study (from which are assessed linearity, precision, recovery bias, and detection limit) were similar to our previous study¹¹ of canine serum cortisol (except that endogenous cortisol in the urine matrix was so low that it was neglected for computations: see urine matrix constitution). In our current study, the immunoassay validation protocol was as follows: 1) linearity study (reportable range determination), 2) within-run replication study, 3) between-run replication study, 4) recovery study, 5) detection limit study, 6) comparison of methods, and 7) QCM rule validation.

We did not perform interference studies, and we did not determine reference intervals. We expanded the final step of QC rule determination by investigating performance not only on the 2 commercial QCM levels, but also in spiked urine samples at the 2 relevant serum ITs to allow comparison of both matrices.

We conducted 3 complementary studies: 1) a spiking-recovery study in cortisol-free canine urine from a dog with hypoadrenocorticism (Addison disease), 2) an interlaboratory comparison study with the Animal Health Diagnostic Center of Cornell University (AHDC-CU; Ithaca, NY, USA), and 3) a QC study with QCM data from TVMDL over one month (April 2019) and with data from the spiking-recovery phase at both IT levels over one week (Fig. 1). Utilization of excess samples was undertaken with owners' permission, as accepted when submitting samples to the TVMDL. Permission for utilization of the urine from the dog with hypoadrenocorticism for the spiking-recovery phase was obtained from the owner through the submitting veterinarian.

Immunoassays

The Immulite 2000 cortisol immunoassay (Siemens) is a competitive, heterogeneous-phase chemiluminoassay for cortisol that uses a rabbit surface-bound capture anti-cortisol polyclonal antibody and cortisol-alkaline phosphatase as tracer. This assay was validated by the manufacturer for human serum cortisol. It cross-reacts with prednisone (and the prednisolone metabolite of prednisone). We used this assay to measure urine cortisol in dogs in our study at TVMDL.

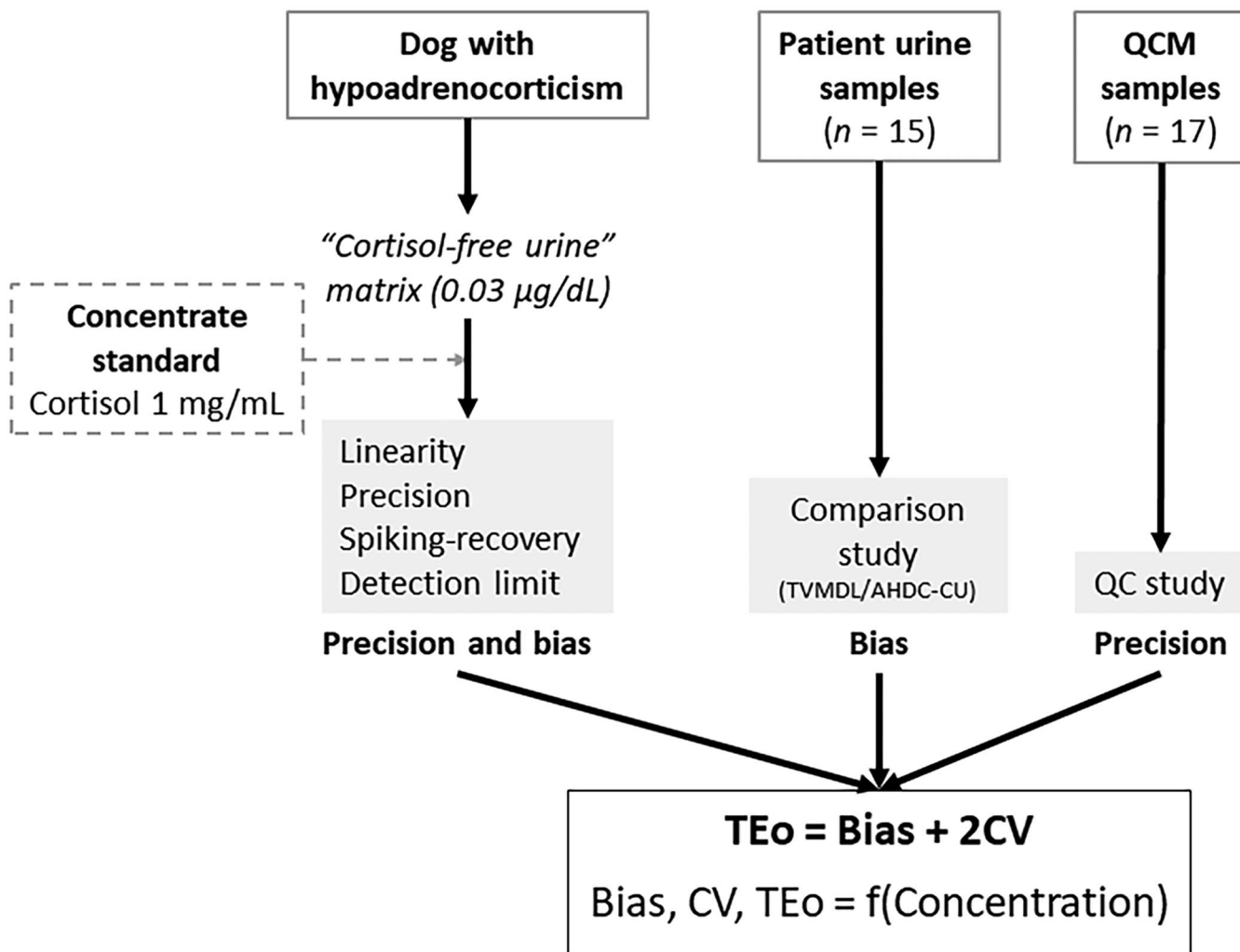


Figure 1. Canine urine cortisol validation study consisted of a spiking-recovery phase, an interlaboratory comparison phase, and a quality control material (QCM) phase. AHDC-CU=Animal Health Diagnostic Center of Cornell University; LDD=low-dose dexamethasone suppression test; TEo=observed total error; TVMDL=Texas A&M Veterinary Medical Diagnostic Laboratory.

The Immulite 1000 cortisol immunoassay (Siemens) is, according to the manufacturer's technical information, similar to the Immulite 2000 cortisol immunoassay, except that the kits cannot be used interchangeably. This is the assay used at the reference laboratory, AHDC-CU, to measure canine urine cortisol.

The ImmuChem RIA kit (for *in vitro* diagnostic use, referenced: 0722110, discontinued; for research use only, referenced: 07221102R; MP Biomedicals) is a competitive, heterogeneous-phase RIA for cortisol, using a rabbit surface-bound capture anti-cortisol antibody and cortisol- I^{125} as tracer. This assay includes a trichloromethane extraction step. This assay was validated by the manufacturer for human serum cortisol and is used for urine cortisol, as well. It cross-reacts with prednisone (and the prednisolone metabolite of prednisone). The *in vitro diagnostic use* version of this assay was used for years at

TVMDL to measure urine cortisol, and we used it for the intralaboratory comparison study.

Measurement of cortisol concentrations outside the reportable range: calibration verifier mode and dose-response calculation

Immulite analyzers are configured by the manufacturer to report numerical results only within the predefined reportable range of 28–1,380 nmol/L (1–50 µg/dL) during routine use (values below or above these limits are reported as < or > these limits). However, values outside these ranges can be reported in the analyzer's "Calibration Verifier Mode" (CVM). The CVM was used to measure concentrations <28 nmol/L (1 µg/dL) for the urine matrix constitution as well as for the detection limit study, and concentrations potentially >1,380 nmol/L (50 µg/dL) for the spiked L10.

The CVM was used for days 4 and 5 (of 5) of the urine spiking-recovery study. The CVM was not activated during the first 3 d of the urine study. Thus, concentration values were retrieved through a dose-response calculation. The latter allows analyte concentrations to be retrieved from the CPS (counts per second, corresponding with the measurement of chemiluminescence), the slope and the intercept defined by the adjustors, as well as the calibration curve parameter values. Generated urine cortisol values were judged straightforwardly acceptable given the homogeneity within the generated values (days 1–3), the homogeneity between generated values (days 1–3) and measured values (days 4, 5), and the homogeneity between the generated values and expectations.

Urine matrix constitution

An adrenocorticotropic hormone (ACTH) stimulation test was submitted for a castrated male, 2-y-old Labrador dog with strong clinical and biochemical signs suspicious for hypoadrenocorticism. Serum cortisol was undetectable (<28 nmol/L, <1 µg/dL) both pre- and 1 h post-cosyntropin (Cortrosyn; Amphastar) stimulation, measured with the Immulite 2000 Xpi.¹¹ Given the absence of interfering external factors that commonly suppress the hypothalamo-hypophysio-adrenal axis (no exogenous steroids, no fluid therapy, etc.), hypoadrenocorticism was diagnosed confidently. A free-catch urine sample was collected from the dog by the submitting veterinarian, sent overnight to TVMDL, tested in duplicate to confirm undetectable cortisol, and immediately frozen at -74°C until thawed for the spiking-recovery study 10 d later.

On the day of thawing, the cortisol concentration of the urine matrix was immediately remeasured in quadruplicate to confirm undetectable cortisol before being used for the spiking-recovery study, and a non-thawed smaller tube was sent out to the AHDC-CU for confirmation in duplicate (both measurements were <5.5 nmol/L, or 0.2 µg/dL). The actual cortisol concentration of the urine matrix was also determined by using the CVM of the Immulite 2000 Xpi. The urine matrix was measured in quadruplicate immediately (day 1) with a mean of 0.9 nmol/L (0.0325 µg/dL), and in quadruplicate daily for the 5 consecutive days of the study (days 1–5, $n=20$) with a mean of 1.1 nmol/L (0.0397 µg/dL). It was judged negligible, and consequently, the endogenous urine cortisol was not subtracted from the spiking-recovery computations.

Cortisol spiking

The “cortisol-free” canine urine was spiked with a standard cortisol concentrate (Cerilliant; MilliporeSigma), a certified reference material (1 mg/mL, 100,000 µg/dL, 2.8×10^6 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 one week, with samples stored at 4°C.

Choice of level concentration. The cortisol spiking protocol for canine urine was identical to the one elected for canine serum.¹¹ The spiking concentrations were chosen based on the relevant values for serum cortisol testing, providing coverage of the reportable range for cortisol measurement by the Immulite 2000 Xpi. Because the upper limit of the reportable range stated within the package insert of the Siemens cortisol assay for 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. To allow comparison with the serum cortisol study,¹¹ 552 nmol/L (20 µg/dL) and 38.6 nmol/L (1.4 µg/dL) were also included into the spiking strategy. Thus, the spiking scheme included the levels: 1,380, 1,035, 552, 345, 172, 69, 38.6, 13.8, and 6.9 nmol/L (50, 37.5, 20, 12.5, 6.25, 2.5, 1.4, 0.5, and 0.25 µg/dL, respectively); the last 2 concentrations were 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 2.8×10^3 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 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 by mixing 100 µL of L10 at 1,380 nmol/L (50 µg/dL) with 900 µL of the matrix (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 which ordinary least squares simple 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, which provided an estimate of the within-run precision across the intended reportable testing range. The CVs were plotted as a function of the cortisol concentration, and trendlines were generated (Excel 2016). The comprehensive within-run precision ($n=20$, intra-run) also was assessed at concentrations of 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 L8 at 552 nmol/L (20 µg/dL) and L4 at 38.6 nmol/L (1.4 µg/dL) from 20 replicates (4 replicates each day for 5 consecutive days), and for 2 commercial QCM levels based on the QCM data from 1 mo

Table 1. Urine 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.

(April 2019; single reagent and QCM lots). The commercial QCM was the Liquicheck urine chemistry (Bio-Rad): for this QCM, there are no target values available for urine cortisol measured with Immulite analyzers. The means of the daily results were 185 nmol/L (6.7 µg/dL) for QCM1, and 648 nmol/L (23.5 µg/dL) for QCM2. QCM1 and QCM2 were measured once daily for 17 d over 1 mo with a single QCM lot.

Recovery study

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

$$\text{recovery (\%)} = \left[\frac{L_{x_{\text{recovered}}} - L_{x_{\text{spiked}}}}{L_{x_{\text{spiked}}}} \right] \times 100\%.$$

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

Detection limit study

Exploration of the detection limits was performed by measuring the blank (L0), 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 available from the between-run replication study.

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

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

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

$$\text{LOD} = \text{mean}_{\text{low urine cortisol}} + 1.65 \text{ SD.}$$

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

$$\text{LOQ} = \text{mean}_{\text{low urine cortisol}} + 2 \text{ SD.}$$

Interlaboratory comparison study

The interlaboratory comparison study was performed by selecting excess patient samples over 2 wk, already tested at TVMDL for urine cortisol, and stored at -74°C . Samples were sent out (overnight and refrigerated) in one batch to the AHDC-CU for canine urine cortisol measurement (Immulite 1000, Siemens; assay initially intended for serum, and internally characterized to be reliable in urine). Over 2 wk, 15 urine samples for which urine cortisol had been measured were available. Four urine samples were excluded because of a urine cortisol concentration beyond the reportable range ($>1,380$ nmol/L, >50 µg/dL), precluding comparison between laboratories. Average bias (AB) between institutions was calculated. Estimation of range-based (RB) bias was performed by grouping various numbers of samples in slots (134–243, 320–342, and 623–924 nmol/L, or 4.84–8.82, 11.6–12.4, and 22.6–33.5 µg/dL, respectively), roughly matching the slots assessed in our study on serum,¹¹ when available.

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,² was performed (www.acomed-statistik.de/en-gb). Simple linear regression was also performed with Excel 2016 for comparison.

A Bland–Altman comparison plot was manually performed on Excel 2016 (y-axis: difference of our result from the reference laboratory result; x-axis: mean of our result and the reference laboratory result). First, the line of the mean of the differences (M) was traced, surrounded by its 95% confidence intervals (95% CI) calculated as^{3,6}:

$$M \pm t \times \text{SE},$$

where t was the t value, taken from the t -distribution table, SE was the standard error, computed as $\sqrt{(\text{SD}^2/n)}$, for which

Table 2. Types of calculated 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 calculations	Spiking-recovery bias	Average bias	Range-based bias*	SR, AB, RB†
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*, L8)	NA
Between-run CV (used for QC rule validation)	TEo _{SR} (L4, L8)	TEo _{AB} (L4, L8)	TEo _{RB} (equivalent L4*, L8)	TEo _{QCM,SR}} TEo _{QCM,AB}} TEo _{QCM,RB}} (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.

* RB bias from groups of the comparison study of the closest ranges from those spiked levels. The RB bias around L4 was approximated by the RB bias observed on samples around L6 (closest available).

† There is no target value for urine cortisol QCM with the Immulite 2000 Xpi cortisol immunoassay; thus, to calculate TEo, we attempted the use of the 3 other types of biases: SR and RB at a roughly similar concentration as QCM1 and QCM2, and AB.

SD was the standard deviation of the differences. The *t* value for urine (95%, *n*-1 = 10 degrees of freedom) was 1.81.

Then, the normality of the differences was investigated with a D'Agostino–Pearson normality test, with significance set at a *p* value threshold of 0.3.¹³ Because normality of the differences could not be demonstrated for urine (*p*=0.004), agreement limits could not be computed.

Intralaboratory comparison study for derivation of IT

In December 2018, 19 canine frozen urine samples for which the cortisol concentration had been previously measured at TVMDL with the ImmuChem RIA kit *for in vitro diagnostic use* were thawed, and urine cortisol was measured on the Immulite 2000 (not Xpi) on a single batch. A simple linear regression was performed to assess the constant and proportional biases between methods to approximate the IT of UCCR when cortisol is measured with the Immulite 2000 Xpi cortisol immunoassay.

Observed total error calculation

TEo was calculated according to the formula: 2CV+absolute bias (%). Four types of TEo were calculated depending on the methods used to determine 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 then plotted on a graph. TEo_{SR} was also calculated with a longer-term precision (*n*=20, between-run: 4 samples/d for 5 d) for L4 and L8 only; for the latter computations, 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) in order to not add some

error from imprecision, generated by sample conservation, into the bias computation.

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

TEo_{RB} was calculated from either the within-run (L6–L8) or the between-run (L4 and L8) precision, and from the concentration targeted bias of the comparison study. The RB bias consisted of the bias observed between TVMDL and the reference institution (AHDC-CU), in a subset of samples grouped by similar concentration ranges. Because there were no urine samples available in the concentration around L4, the RB bias for L4 was approximated by using the RB bias of L6.

TEo_{QCM} was calculated at both QCM levels (QCM1 and QCM2) from the between-run CV observed on the data over one month (April 2019, *n*=22, one QCM lot). There is no manufacturer-provided commercially available QCM intended for urine cortisol on this analyzer. Because there are no target values for urine cortisol with the QCM we used (Liquicheck urine chemistry), the bias could not be determined. To allow for further computations (TEo and QC rule validation), we used instead the SR bias (at a similar concentration range), the RB bias (at a similar concentration range), and the AB bias, generating respectively TEo_{QCM,SR}}, TEo_{QCM,RB}}, and TEo_{QCM,AB}}. For the RB bias, QCM1 having a mean of 185 nmol/L (6.7 µg/dL), the bias from the 134–243 nmol/L (4.84–8.82 µg/dL) range was used; similarly, QCM2 having a mean of 648 nmol/L (23.5 µg/dL), the bias from the 623–924 nmol/L (22.6–33.5 µg/dL) range was used.

Quality control rule validation study

Usable QC rules were explored for both commercial QCM levels (QCM1 and QCM2), as well as for 2 concentration levels (L4 and L8). We determined the acceptable QC rules manually, from normalized operational process specifications

(OPSpec) charts, plotting each operating point determined from the bias (as a % of TEa) on the y-axis, and the CV (as a % of TEa) on the x-axis, as explained previously in a study of canine serum cortisol.¹¹ We investigated the sigma metric and the acceptable QC rules:

- at low TEa (20%, or TEa=TEo when TEo > 20%) and at an arbitrarily chosen high TEa (50%, or TEa=TEo when TEo > 50%),
- at high probability of error detection (P_{ed} ; 90%) and arbitrarily chosen low P_{ed} (50%),
- at $N=2$ (2 levels analyzed once; results provided: all of the results of our 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. Because of the manual use of OPSpec charts, the probability of false rejection (P_{fr}) was fixed per QC rule in all QC combinations:

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} :	$P_{fr}=0.00$
$1_{3,5S}$:	$P_{fr}=0.00$

Results

Reportable range study

Linearity appeared excellent based on visual examination of the graph (Fig. 2) and on the linear regression characteristics. Indeed, linear regression yielded a slope of 1.025 (close to 1), an intercept of 0.706 (close to 0), and a coefficient of determination R^2 of 0.998 (close to 1). Linearity was thus confirmed between 6.9 and 1,380 nmol/L (0.25 and 50 $\mu\text{g/dL}$, respectively).

Within-run replication study

CV increased markedly (trendline: power function) with decreasing urine cortisol concentration, remaining <3% beyond 69 nmol/L (2.5 $\mu\text{g/dL}$), ~7% between 38.6 and 6.9 nmol/L (1.4 and 0.25 $\mu\text{g/dL}$, respectively), and increasing to 30% and 85% closer to zero (Table 3; Fig. 3A). The shift of the curve happened at ~38.6 nmol/L (1.4 $\mu\text{g/dL}$), with a CV of 6.29%.

Between-run replication study

Between-run CV was barely <10% for L4, barely <5% for L8, and ~5% for QCM1 and QCM2 (Table 4). For spiked samples, CV respected the following expected relationships:

- Between-run precision CV > within-run precision CV
- CV from L8 < CV from L4

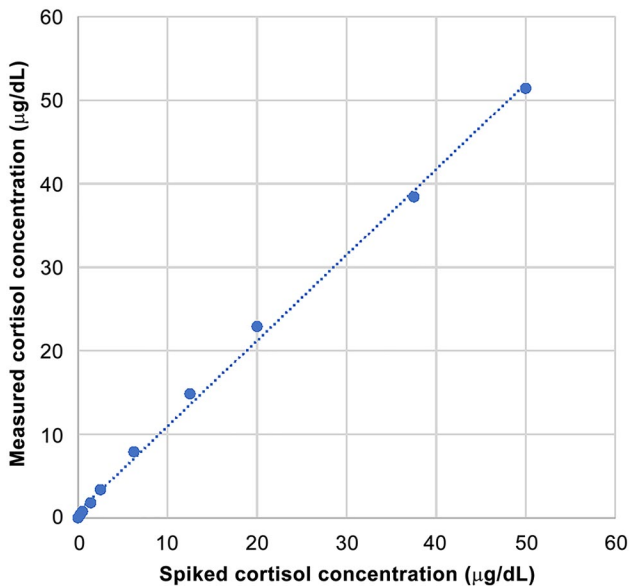


Figure 2. Reportable range study, with simple linear regression for within-run spiking-recovery canine urine cortisol. The simple linear regression slope=1.025 (close to 1), intercept=0.706 (close to 0). The coefficient of determination R^2 of 0.998 (close to 1); >0.975 indicates that the assessed range is sufficient.¹ Linearity was confirmed between 0.25 and 50 $\mu\text{g/dL}$.

Recovery study

The SR bias was positive (overestimation) and increased markedly with decreasing urine cortisol concentration; yet the consequences on the absolute concentration values were minor because the increasing percentages were offset by the decreasing concentrations to which they were applied (Table 3; Fig. 3B).

Detection limit study

The LOB on saline was 1.38 nmol/L (0.05 $\mu\text{g/dL}$). Because the LOB was higher than the cortisol concentration of the urine matrix (L1), determined to be 1.1 nmol/L (0.0397 $\mu\text{g/dL}$) from the study ($n=20$ over 5 d), it confirmed that the endogenous cortisol concentration of L1 could be neglected. When determined on L1 (accounting for matrix properties) instead of saline, the LOB was 2.21 nmol/L (0.08 $\mu\text{g/dL}$). The LOD was determined on L2 and was 15.7 nmol/L (0.57 $\mu\text{g/dL}$). The LOQ was determined on L2 and was 16.6 nmol/L (0.60 $\mu\text{g/dL}$), with corresponding CV of 16.5% and TEo of 113%.

Interlaboratory comparison study

The AB was lower than any RB bias. The Passing–Bablok regression was good, with a slope of 0.863 (95% CI: 0.773–1.089, including 1) and an intercept of 1.626 (95% CI: –0.089

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

Level	Spiked cortisol concentration, nmol/L ($\mu\text{g/dL}$)	Urine		
		Mean measured concentration, nmol/L ($\mu\text{g/dL}$)	Within-run CV (%)	Recovery (%)
L10	1,380 (50)	1,424 (51.6)	2.18	103
L9	1,035 (37.5)	1,098 (39.8)	6.01	106
L8	552 (20)	662 (24.0)	1.69	120
L7	345 (12.5)	439 (15.9)	2.95	127
L6	172 (6.25)	234 (8.49)	2.56	136
L5	69 (2.5)	102 (3.68)	2.38	147
L4	38.6 (1.4)	55 (1.99)	6.29	142
L3	13.8 (0.5)	23.5 (0.853)	7.25	171
L2	6.9 (0.25)	12.8 (0.465)	7.55	186
L1 (matrix)	0	0.9 (0.0325*)	29.5	NA
L0 (blank)	0	0.69 (0.025)	84.9	NA

NA=not applicable.

* The intrinsic cortisol concentration of the urine matrix (0.0325 $\mu\text{g/dL}$) was neglected for the recovery and then for the bias computations.

to 2.498, including 0) (Table 5; Fig. 4). The simple linear regression obtained very similar results, with a slope of 0.829 (close to 1) and an intercept of 2.045 (somewhat close to 0); coefficient of determination $R^2=0.991$ (close to 1). In the Bland–Altman comparison graph (Fig. 5), the mean of differences with its 95% CI was -3.86 [-32.6 to 25.1] nmol/L (-0.14 [-1.18 to 0.91] $\mu\text{g/dL}$). The normality of the differences could not be verified according to the D’Agostino–Pearson normality test ($p=0.004$; <0.05 and 0.3 taken as ITs), hence the agreement limits and their 95% CIs could not be calculated.

Intralaboratory comparison study for derivation of IT

Linear regression on 19 canine urine samples between the previous RIA results and the new Immulite 2000 results yielded the equation: Immulite 2000 = $2.59 \times \text{RIA} + 1.15$ ($R^2=0.909$), resulting in a shift of IT from 10×10^{-6} to 27×10^{-6} (Fig. 6). The Immulite 2000 results were systematically higher than those obtained from the RIA.

Observed total error calculations

The within-run CV (Fig. 3A), the SR bias (Fig. 3B), and the TEo_{SR} (Fig. 3C) increased with decreasing serum cortisol concentration (Table 6). The AB bias was $<$ RB bias, and the RB bias was $<$ SR bias regardless of the concentration level. TEo (with within-run data) verified $\text{TEo}_{\text{AB}} < \text{TEo}_{\text{RB}} < \text{TEo}_{\text{SR}}$ (Table 6).

The between-run CVs for L4 & L8 and for QCM 1 & QCM 2 were $>$ within-run CVs (Table 7). The corresponding TEo for L4 & L8 (using between-run CV but keeping

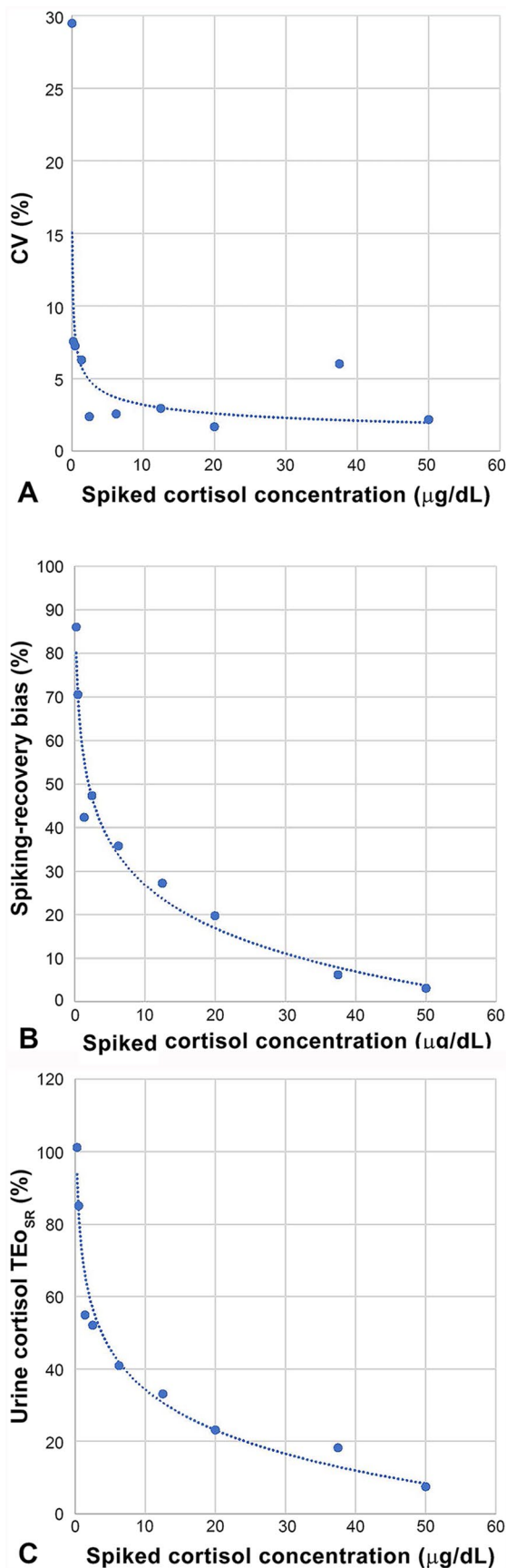
within-run bias) were highly discrepant, verifying $\text{TEo}_{\text{AB}} < \text{TEo}_{\text{RB}} < \text{TEo}_{\text{SR}}$ (Table 7), unlike what was observed previously in serum.¹¹ Similarly, TEo_{QCM} verified $\text{TEo}_{\text{QCM,AB}} < \text{TEo}_{\text{QCM,RB}} < \text{TEo}_{\text{QCM,SR}}$.

Quality control rule validation study

Almost no QC rules were acceptable at high P_{ed} (90%) and low TEa (Table 8), regardless of the considered level ($1_{2\text{S}}$ was acceptable for $L8_{(\text{AB})}$ and $\text{QCM}2_{(\text{AB})}$ but $1_{2\text{S}}$ should be avoided given an excessively high P_{fr} of 0.09). For all levels, decreasing P_{ed} from 90% to 50% resulted in an increased proportion of acceptable QC rules; however, this proportion of newly acceptable QC rules remained very limited. On the other hand, for all levels, increasing TEa to 50% resulted in a markedly 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} .

Discussion

Similar to our study on canine serum cortisol with the Immulite 2000 Xpi,¹¹ we validated the Immulite 2000 Xpi for measurement of urine cortisol in dogs, and also characterized the CV, bias, and TEo function of concentrations spanning the reportable range, for kit lot numbers less than 550. For lots 550 and greater, the anti-cortisol capture antibody has been recently modified by Siemens and has shown significant negative biases in serum (average -23.1%) and urine (average -72.7%) compared to kit lots <550 ,⁷ whereas not affecting the measurements in human samples. This observation (November 2020) led to the recent commercialization (January 2021) by Siemens of a “veterinary cortisol kit” containing the new antibody but also integrating a



(continued)

Figure 3. **A.** Evolution of within-run ($n=4$) CV (%) across urine cortisol concentrations. Spiked cortisol levels for the linearity study were run in quadruplicate: CV remained low (1.5–2.5%) between 1,380 and 69 nmol/L (50 and 2.5 $\mu\text{g/dL}$, respectively), except for an outlier at 1,035 nmol/L (37.5 $\mu\text{g/dL}$), before increasing markedly at lower concentrations. **B.** Evolution of the spiking-recovery bias (%) across urine cortisol concentrations. 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 increased markedly from the high end to the low end of the reportable range. **C.** Evolution of observed total error (TEo; %) across urine cortisol concentrations. Within-run CV (Fig. 3A) and spiking-recovery bias (Fig. 3B) were combined to calculate TEo across serum cortisol concentrations.

correction formula (historical value = $(1.1 \times \text{new antibody kit value}) + 4.14 \text{ nmol/L}$), acceptably correcting serum cortisol (AB of -8.6%), but unsuccessful in correcting urine cortisol (AB of -61.4%). Repeating our study with the new “veterinary cortisol kit” to investigate linearity, precision, and SR bias is warranted; independent satisfactory analytical performance may open the door to new determination of relevant IT for this method.

Based on the simple linear regression on the spiking-recovery study, urine cortisol exhibited excellent linearity with the Immulite 2000 Xpi cortisol immunoassay between 6.9 and 1,380 nmol/L (0.25 and 50 $\mu\text{g/dL}$, respectively). An upper limit of the reportable range $>1,380 \text{ nmol/L}$ was deemed unnecessary, given that the UCCR would need an unrealistically high urine creatinine to drop below the IT; for example, with a urine cortisol of 1,380 nmol/L, a urine creatinine of at least 51.1 $\mu\text{mol/L}$ (578 mg/dL) would be needed to drop UCCR below the IT of 27×10^{-6} mentioned below. This observation, combined with the detection limit study results showing a LOQ of 16.6 nmol/L (0.60 $\mu\text{g/dL}$), supports the reportable range of 28–1,380 nmol/L (1–50 $\mu\text{g/dL}$) from the manufacturer for serum cortisol as also acceptable for urine cortisol. This LOQ is associated with a between-run CV of 16.5%, which is slightly higher (less precise) than the CV observed in serum¹¹ (13%), but remains satisfactorily low for such a low cortisol concentration. On the other hand, the TEo of this LOQ was much higher in urine (113%) than in serum¹¹ (43%) because of a much higher SR bias in urine (+86%) than in serum (-6.4%) at this concentration. Yet, this high bias compared to spiking has limited impact for clinical decisions given the low concentration to which it is applied (6.9 nmol/L [0.25 $\mu\text{g/dL}$]). Admittedly this could be of concern provided that the IT of UCCR is a lower limit; however, this bias should be compensated by the determination of the IT itself, given that the IT determination is performed based on clinical studies, not based on spiking-recovery.

Despite an outlier (high CV of L9; Table 3, Fig. 3A), the within-run CV, SR bias, and TEo across the reportable concentration range reflected similar trends, increasing with

Table 4. Within-run and between-run precision for 2 spiked canine urine cortisol levels and 2 levels of quality control material.

Precision	<i>n</i>	Days	Level: nmol/L (µg/dL)	CV (%)
Within-run (spiked)	20	1	L4: 38.6 (1.4)	6.4
	20	1	L8: 552 (20)	2.5
Between-run (spiked)	20	5	L4: 38.6 (1.4)	9.6
	20	5	L8: 552 (20)	4.6
Between-run (QCM)	17	17	QCM1: 185 (6.7)*	5.4
	17	17	QCM2: 648 (23.5)*	4.5

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

* Concentrations for QCM are not target values; the manufacturer does not provide target values for urine cortisol measured with an Immulite. Instead, means are provided as an estimate.

Table 5. Interlaboratory comparison study results for canine urine cortisol range-based bias and average bias, based on 11 urine samples.

Increasing concentration groups	Serum cortisol: nmol/L (µg/dL)		Range-based bias
	TVMDL	AHDC-CU	
Group C	924 (33.5)	1,036 (38.5)	-10.7%
	624 (22.6)	670 (24.3)	
Group B	342 (12.4)	309 (11.2)	5.3%
	328 (11.9)	284 (10.3)	
	328 (11.9)	367 (13.3)	
	320 (11.6)	292 (10.6)	
Group A	243 (8.82)	226 (8.18)	8.7%
	243 (8.80)	240 (8.69)	
	187 (6.77)	164 (5.96)	
	152 (5.52)	142 (5.15)	
	134 (4.84)	110 (4.00)	
Average bias	-1.1%		

AHDC-CU = Animal Health Diagnostic Center of Cornell University; TVMDL = Texas Veterinary Medical Diagnostic Laboratory.

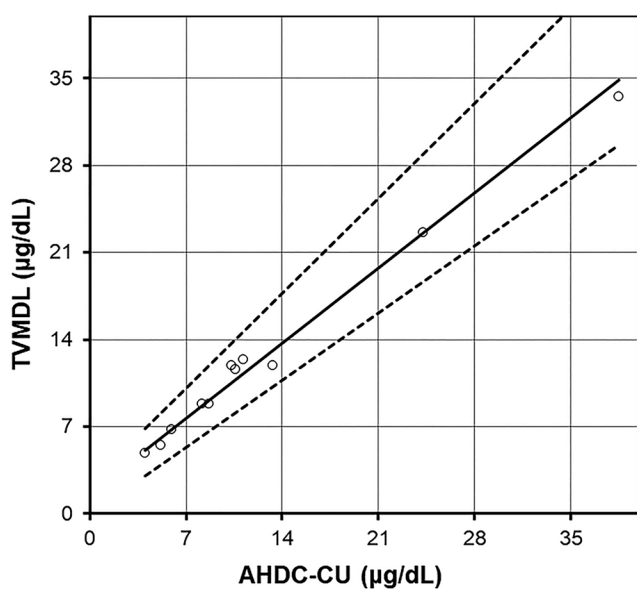


Figure 4. Interlaboratory comparison study for canine serum cortisol (*n*=11), with Passing–Bablok regression (plain line) and its 95% CI (dotted lines). AHDC-CU=Animal Health Diagnostic Center of Cornell University; TVMDL=Texas A&M Veterinary Medical Diagnostic Laboratory.

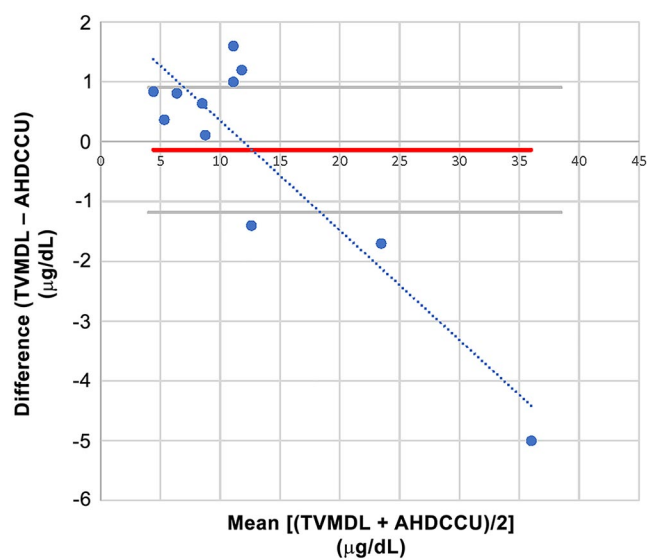


Figure 5. Bland–Altman comparison for the interlaboratory canine urine cortisol comparison study (*n*=11), with simple linear regression of the differences between laboratories. TVMDL used the Immulite 2000 Xpi cortisol immunoassay, and AHDC-CU used the Immulite 1000 cortisol immunoassay. There was one freeze–thaw
(continued)

Figure 5. (continued)

cycle for AHDC-CU; there was none for TVMDL. The simple linear regression of the difference followed the equation: $y = -0.1834x + 2.1856$. Thus, and in light of the regression curve, there was a small positive constant bias most visible at low concentration, and a small negative proportional bias at higher concentration; none would significantly impact clinical interpretation. AHDC-CU = Animal Health Diagnostic Center of Cornell University; TVMDL = Texas A&M Veterinary Medical Diagnostic Laboratory.

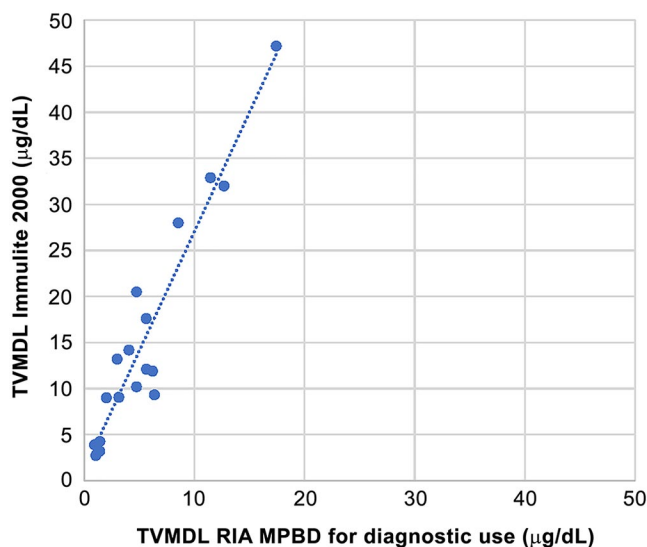


Figure 6. Intralaboratory comparison by simple linear regression of the Immulite 2000 for *in vitro* diagnostic use radioimmunoassay (RIA; MPBD, MP Biomedicals) and the Immulite 2000 chemiluminescence immunoassay (Siemens Healthineers) for canine urine cortisol. TVMDL = Texas A&M Veterinary Medical Diagnostic Laboratory.

decreasing urine cortisol concentration. Those trends forming defined patterns, as opposed to random scatter plots, support a true effect of urine cortisol concentration on CV and bias, and then on TEo. Of note, between-run CV of L4 and L8 in urine were \leq between-run CV determined in serum¹¹ (9.6% for urine L4 was interpreted as not significantly different from 9.5% for serum L4).

In our intralaboratory comparison of methods, the number of compared samples was limited ($n = 19$); however, the equation for simple linear regression of $y = 2.589x + 1.157$ is enough to highlight the difference of methods, and to emphasize that UCCR IT should be determined for each method individually. After months of using the estimated UCCR IT of 27×10^{-6} with the Immulite 2000 Xpi, it seems satisfactory in a vast majority of cases, although sensitivity and specificity remain to be determined in a dedicated study. Because of the UCCR IT variation, the goal of our study was not to assess the relevance of a defined UCCR IT, but rather to assess the TEo one could consider in interpreting UCCR results.

In the literature, most of the validated assays for measuring cortisol in canine urine are RIAs.^{4,9} The IT may not be comparable from one method to another (Table 9), as illustrated by a study using an ELISA⁸ and finding a much higher IT (60×10^{-6}) by ROC curve than those described for the RIA^{4,10,16,17} (10×10^{-6} to 30×10^{-6}). The difference in ITs depends on the method not only because of the type of signal (radioactivity, chemiluminescence, enzymology), but also because of the specificity of the anti-cortisol antibody for cortisol.²¹ The cortisol molecule is too small to elicit an immune response and needs to be bound to a macromolecule, for example, bovine serum albumin (BSA).²¹ BSA masks some haptens on the cortisol molecule at the site of binding, and the produced antibodies are never entirely specific for pure cortisol. Depending on the site of binding of BSA to cortisol, on C3 or C21, the specificity varies greatly; anti-cortisol antibodies produced from cortisol bound to BSA on C3 are much less specific than those on C21 because many conjugations of cortisol metabolites happen on C3. Cortisol metabolites being abundant in urine (as opposed to serum), this analytical specificity of anti-cortisol antibodies is much more problematic in urine samples. For this reason, it is more correct to talk about urinary corticoids than urinary cortisol. Of note, excreted urinary corticoids may vary across animal species. Consequently, it is observed and expected that urine cortisol and the resulting UCCR vary between antibodies, those created with BSA binding in C3 (more frequent) being less specific and yielding significantly higher values than those created with BSA binding in C21. The type of anti-cortisol antibody in the Siemens immunoassay is not disclosed and could also have evolved within the new kits (lot numbers >550). Our extrapolation of IT at 27×10^{-6} , resulting from linear regression with our former RIA for which the IT was 10×10^{-6} , closely matches the findings from another study¹⁹ using the Immulite 1000, which determined the optimal exclusion IT by ROC curve analysis between dogs with confirmed hyperadrenocorticism and dogs suspect of hyperadrenocorticism at 26.5×10^{-6} . This could suggest anti-cortisol antibody created by binding in C3 (higher measured values given lower analytical specificity). On the other hand, another study⁵ comparing 5 immunoassays including the Immulite 2000 and the in-house RIA from the University of Utrecht, known to use antibodies produced by cortisol bound in C21 (lower values as a result of higher analytical specificity), did not show overtly different ranges of results between the 2 methods. Finally, the presence or absence of an organic solvent extraction initial step is also critical in the method²¹ because it removes most (not all) polar molecules (conjugated metabolites) while leaving the hydrophobic cortisol in the sample. A study showed that measured urinary corticoid concentrations were significantly lower post-extraction for both types of antibodies (anti-C3 bound or anti-C21 bound).²¹ Our previous RIA did include an initial trichloromethane extraction step, whereas the Immulite 2000 assay

Table 6. Coefficient of variation (within-run), bias, and observed total error results across the canine urine 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)	2.18	3.11	NA	-1.09	7.48	NA	5.5
L9	1,035 (37.5)	6.01	6.20	NA		18.2	NA	13.1
L8	552 (20)	1.69	19.8	-10.7		23.1	14.1	4.5
L7	345 (12.5)	2.95	27.2	5.3		33.1	11.2	7.0
L6	172 (6.25)	2.56	35.8	8.7		40.9	13.8	6.2
L5	69 (2.5)	2.38	47.3	NA		52.1	NA	5.9
L4	38.6 (1.4)	6.29	42.3	8.7*		54.9	21.3*	13.7
L3	13.8 (0.5)	7.25	70.5	NA		85.0	NA	15.6
L2	6.9 (0.25)	7.55	86.0	NA		101.1	NA	16.2

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.

* Using the closest available bias.

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

Spiked urine samples

Cortisol level	Target values, nmol/L (µg/dL)	Precision (%)	Bias (%)			TEo (%)		
		Between-run CV	SR	RB	AB	TEo _{SR}	TEo _{RB}	TEo _{AB}
L4	38.6 (1.4)	9.55	42.3	8.7*	-1.09	61.4	27.8	20.2
L8	552 (20)	4.60	19.8	-10.7		28.9	19.9	10.3

QCM

Cortisol level	Target values,* nmol/L (µg/dL)	Precision (%)	Bias (%)†			TEo _{QCM} (%)		
		Between-run CV	SR	RB	AB	TEo _{QCM,SR}	TEo _{QCM,RB}	TEo _{QCM,AB}
QCM1	193 (7.0)	5.42	35.8	8.7*	-1.09	46.6	19.5	11.9
QCM2	389 (14.1)	4.51	19.8	-10.7		28.7	19.7	10.1

AB=average bias; CV=coefficient of variation; Lx=level x (see dilution Table 1); RB=range-based bias; SR=spiking-recovery bias; TEo=observed total error.

* Closest available bias.

† There were no target values provided by the manufacturer of the QCM for canine urine cortisol. Means are provided to give an idea of the tested level. Because of the absence of target values, the QCM bias could not be assessed. Instead, TEo_{QCM} was calculated using the biases from the spiking recovery study (SR) and the comparison study (RB and AB) at similar urine cortisol concentrations.

does not, which certainly accounts for at least part of the measurement differences between those 2 techniques. Thus, it is not good practice to accept for one method¹⁶ an IT determined with another method¹⁹ (Table 9). Especially for UCCR, the IT is highly method-dependent.

In addition to the method, the IT is also influenced by the elected method for IT determination (mean + xSD, range comparison between populations, ROC curve) and the involved populations (healthy dogs, dogs with hyperadrenocorticism [HAC], dogs suspected to have but not having hyperadrenocorticism [HAC-S], dogs with non-adrenal illness [NAI], etc.). With variation on all those parameters, the UCCR IT is reported here from 10×10^{-6} to 60×10^{-6} (Table 9), so it is expected to be even more laboratory-dependent than the IT for serum cortisol post-

ACTH stimulation or post-dexamethasone suppression. Ideally, the UCCR IT should be determined by ROC curve analysis, between HAC and HAC-S populations, with the method of use in the laboratory. A study assessing the UCCR diagnostic performance with the Immulite 1000 further illustrates the variation of the IT depending on the method of calculation and on the goal; the upper limit of the reference interval was 30.81×10^{-6} , and the ROC curve-optimized IT for a test of exclusion was 26.5×10^{-6} , whereas the ROC curve-optimized IT for a confirmatory test was 161.2×10^{-6} .²⁰

The Bland-Altman plot used for the between-laboratory comparison study suggested a small positive bias at low urine cortisol concentrations and a small negative bias at high urine cortisol concentrations. The small number of

Table 8. Quality control (QC) rule validation for 2 cortisol concentrations (L4=38.6 nmol/L=1.4 µg/dL; L8=552 nmol/L=20 µg/dL) and both QC material levels in canine urine.

Urine 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)	61.4	61.4	2	None	L4 _(SR)	61.4	61.4	2	None
	L4 _(RB)	27.8	33	2.5	None	L4 _(RB)	27.8	33	2.5	None
	L4 _(AB)	20.2	33	3.3	None	L4 _(AB)	20.2	33	3.3	1_{2s}
	L8 _(SR)	28.9	28.9	2	None	L8 _(SR)	28.9	28.9	2	None
	L8 _(RB)	19.9	20	2.0	None	L8 _(RB)	19.9	20	2.0	None
	L8 _(AB)	10.3	20	4.1	1 _{2s}	L8 _(AB)	10.3	20	4.1	1 _{2s} ; 1_{2.5s} ; 1_{3s}/2_{2s}/R_{4s} ; 1_{3s}
	QCM1 _(SR)	46.6	46.6	2	None	QCM1 _(SR)	46.6	46.6	2	None
	QCM1 _(RB)	19.5	20	2.1	None	QCM1 _(RB)	19.5	20	2.1	None
	QCM1 _(AB)	11.9	20	3.5	None	QCM1 _(AB)	11.9	20	3.5	1_{2s} ; 1_{2.5s}
	QCM2 _(SR)	28.7	28.7	2	None	QCM2 _(SR)	28.7	28.7	2	None
	QCM2 _(RB)	19.7	20	2.1	None	QCM2 _(RB)	19.7	20	2.1	None
	QCM2 _(AB)	10.1	20	4.2	1 _{2s}	QCM2 _(AB)	10.1	20	4.2	1 _{2s} ; 1_{2.5s} ; 1_{3s}/2_{2s}/R_{4s} ; 1_{3s}
High TEa	L4 _(SR)	61.4	61.4	2	None	L4 _(SR)	61.4	61.4	2	None
	L4 _(RB)	27.8	50	4.3	1_{2s}	L4 _(RB)	27.8	50	4.3	1 _{2s} ; 1_{2.5s} ; 1_{3s}/2_{2s}/R_{4s} ; 1_{3s}
	L4 _(AB)	20.2	50	5.1	1_{2s} ; 1_{2.5s} ; 1_{3s}/2_{2s}/R_{4s}	L4 _(AB)	20.2	50	5.1	All
	L8 _(SR)	28.9	50	6.6	All	L8 _(SR)	28.9	50	6.6	All
	L8 _(RB)	19.9	50	8.5	All	L8 _(RB)	19.9	50	8.5	All
	L8 _(AB)	10.3	50	10.6	All	L8 _(AB)	10.3	50	10.6	All
	QCM1 _(SR)	46.6	50	2.6	None	QCM1 _(SR)	46.6	50	2.6	None
	QCM1 _(RB)	19.5	50	7.6	All	QCM1 _(RB)	19.5	50	7.6	All
	QCM1 _(AB)	11.9	50	9.1	All	QCM1 _(AB)	11.9	50	9.1	All
	QCM2 _(SR)	28.7	50	6.7	All	QCM2 _(SR)	28.7	50	6.7	All
	QCM2 _(RB)	19.7	50	8.7	All	QCM2 _(RB)	19.7	50	8.7	All
	QCM2 _(AB)	10.1	50	10.9	All	QCM2 _(AB)	10.1	50	10.9	All

Bold rules are those added when switching from [Low TEa and Ped90%] to [Low TEa and Ped50%], from [Low TEa and Ped90%] to [High TEa and Ped90%], and from [High TEa and Ped90%] to [High TEa and Ped50%]. AB=average bias; All=includes all 4 QC rules (1_{2s}; 1_{2.5s}; 1_{3s}; 1_{3s}/2_{2s}/R_{4s}); Lx=level x (see dilution Table 1); P_{ed}=probability of error detection (by the QCM); QCM=quality control material; RB=range-based bias; SR=spiking-recovery bias; TEa=total allowable error; TEo=observed total error.

samples for the comparison ($n=11$) is a major limitation, and a more extensive study would be needed to confirm those preliminary results. Because of the small number of samples, the normality of the differences between methods could not be demonstrated and agreement limits could not be determined. We elected a threshold for the p value of the D'Agostino–Pearson normality test of 0.3 rather than 0.05, in order to decrease type I error (distribution falsely determined to be Gaussian) in a clinical setting at low n .¹³ Because the p value of the differences between methods was 0.004, normality would not have been demonstrated even with the classic threshold of 0.05.

The CV for both commercial QCM levels over one month was ~5%. The Immulite analyzers had not been validated for urine cortisol previously, hence there is no Siemens commercial QCM available for urine cortisol. Thus, we used the Liquicheck urine QCM (Bio-Rad); it provides target values for (human) urine cortisol with several different methods (Table 10) but does not provide target values for the Immulite analyzers. Had the target values been somewhat similar across the different methods and analyzers within the same level, we could have attempted to use these approximate

target values to assess the QCM bias. Yet, target values were very diverse across the different methods and analyzers, being scattered within 81.4–220 nmol/L (2.95–7.99 µg/dL) for QCM1 and within 422–745 nmol/L (15.3–27.0 µg/dL) for QCM2. In light of the wide range of acceptable values across the available methods, there was no rationale to derive appropriate target values for the Immulite 2000.

To overcome the challenge of the QCM bias, we attempted to use the SR, RB, and AB biases determined previously in urine for levels best matching the QCM level concentrations. Because SR and RB were much higher than AB, in turn itself 2–3 times higher than the QCM bias observed in serum,¹¹ we believe that the best approximations (bias and resulting TEo) for urine QCM are the AB. Thus, AB was elected for determination of QCM acceptable QC rules. Given QCM performance of the K9CON (Immolute Systems; Siemens) for serum cortisol used in our previous study,¹¹ if some new urine cortisol commercial QCM is developed, or if some target values for urine cortisol using the Immulite are determined, then the QCM bias might be as good as the AB we used. We used 2 spiked samples (L4 and L8) to model patient samples at 2 different cortisol concentrations (chosen similar

Table 9. Chronologic and technical evolution of the urine cortisol:creatinine ratio interpretation threshold to exclude the differential of hyperadrenocorticism in dogs.

Ref.	Cortisol assay	UCCR	Population with hyperadrenocorticism	Compared population	IT	Source of IT	Se (%)	Sp (%)
Stolp et al. ¹⁹	FMM (non-immune)	Morning urine	27 HAC	28 adult healthy pet dogs	Not provided Deductible IT from provided data: $8-10 \times 10^{-6}$ • UCCR range normal dogs: $1.2-6.9 \times 10^{-6}$ • UCCR range HAC: $11-23.5 \times 10^{-6}$	19	100‡	‡
Rijnberk et al. ¹⁶	RIA	Mean of 2 consecutive mornings	93 HAC	57 HAC-S	$10 \cdot 10^{-6}$	Quoted study ¹⁹	99	77§
Feldman et al. ⁴	RIA‡‡	Morning urine	40 HAC	20 healthy dogs 23 PUPD non-HAC (including 11 DM)	13.5×10^{-6} ‡ • UCCR normal dogs: mean + 2SD = 13.5×10^{-6} range $0.5-17.7 \times 10^{-6}$ • UCCR PUPD non-HAC dogs: mean + 2SD = 61.4×10^{-6} range: $8.0-144 \times 10^{-6}$ • UCCR HAC dogs: mean ± SD = $337.6 \pm 72.0 \times 10^{-6}$ range: $20-2,100 \times 10^{-6}$		100	22#
Smiley et al. ¹⁷	RIA‡‡	Morning urine	25 HAC	31 normal dogs 21 HAC-S 28 severe NAI	30×10^{-6} Remark from provided data: • UCCR normal dogs: mean + 2SD = 27.1×10^{-6} range: $0.1-31.2 \times 10^{-6}$ • UCCR HAC-S dogs: mean + 2SD = 30.3×10^{-6} range: $6.9-36.1 \times 10^{-6}$ • UCCR NAI dogs: mean + 2SD = 30.1×10^{-6} range: $2.7-524 \times 10^{-6}$ • UCCR HAC dogs: range: $21.3-432 \times 10^{-6}$	No source	9	Sp _{normal} 97** Sp _{HAC-S} 95** Sp _{NAI} 21**
Kaplan et al. ¹⁰	RIA§§	Morning urine	20 HAC PDH	59 NAI	About 20×10^{-6} ††		75	24
Jensen et al. ⁸	ELISA§§	Morning urine; 1 or mean of 2 consecutive mornings	18 HAC Mean = 142.4×10^{-6}	20 dogs without HAC (no other selection criteria) Mean = 34.2×10^{-6}	60×10^{-6} • ROC curve optimal threshold: 62.5×10^{-6}	8	100	85§
Zaugswetter et al. ²⁰	Immunitite 1000	Morning urine collected at home	66 HAC	50 healthy dogs 87 HAC-S	30.81×10^{-6} • Upper limit of the RI on 50 healthy dogs of normal distribution as mean + 2SD 26.5×10^{-6} • ROC curve HAC vs. HAC-S optimal threshold for a test of exclusion 161.2×10^{-6} • ROC curve HAC vs. HAC-S optimal threshold for a confirmatory test	20	97	65 100 54.2 98.8

DM = diabetes mellitus; FMM = fluorometric method of Matingly; HAC = dogs with hyperadrenocorticism; HAC-S = dogs suspected of hyperadrenocorticism confirmed to not have hyperadrenocorticism; IT = interpretation threshold; LDDST = low-dose dexamethasone suppression test; NAI = dogs with non-adrenal illness; PDH = dogs with pituitary-dependent hyperadrenocorticism; RI = reference interval; RIA = radioimmunoassay; Se = sensitivity; Sp = specificity; UCCR = urine cortisol:creatinine ratio.

* The article also included investigation with a LDDST.

† The article also included investigation with an adrenocorticotrophic hormone stimulation test.

‡ The sensitivity is not mentioned in the article but can be retrieved from the data. Here the threshold is low enough to allow a sensitivity of 100%. The specificity has never been quoted as "100%" from this article, even if this is what data support, as the comparison group is healthy and has an especially low range of urinary cortisol, not reflecting the situation in real life.

§ Surprising high specificity, mismatching with other studies, and not reflecting the performances of the UCCR when applied in a dog reasonably suspected of HAC.

¶ Authors chose to define the IT as (mean + 2SD) of the healthy population (13.5×10^{-6}). Yet, because the lower limit of the range for HAC dogs is substantially higher (20×10^{-6}), they could have set up the IT at 19×10^{-6} , which would likely have (slightly) improved the low specificity when compared to PUPD non-HAC dogs.

This is the specificity according to the chosen IT applied to the PUPD non-HAC dog population. The specificity within the healthy population is useless and not reported.

¶ The IT has been chosen as (mean + 2SD) of the HAC-S or NAI populations, which may be better than based on the healthy population, when the sensitivity does not drop too much with this approach.

** It is surprising to find a specificity that high in the HAC-S population, very close from the specificity in the normal population, and very far from the low specificity in the NAI population. Explanations may be that HAC-S suspicions were too readily applied, and also that NAI were severe. Yet, the reality of the specificity of UCCR for HAC is closer from what is observed in the NAI population.

†† 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.

‡‡ Article including some intrinsic validation of the cortisol assay in canine urine.

§§ Article quoting a source for validation of the cortisol assay in canine urine.

Table 10. Target values for urinary cortisol for the Liquicheck QCM (lot 68530), reproduced from the one provided by the manufacturer (Bio-Rad) for different methods and analyzers. This example illustrates how one quality control material (QCM) can have widely different acceptable ranges for a given level of a measurand depending on the considered methods. The acceptable range for the Immulite 2000 Xpi is not provided by the manufacturer, preventing determination of the QCM bias.

Method	Cortisol ($\mu\text{g/dL}$)			
	Level 1: lot 68531		Level 2: lot 68532	
	Mean	Range	Mean	Range
Abbott Architect iSystems	2.95	2.09–3.80	15.4	12.3–18.5
Beckman Coulter Access / 2 / 2i	6.63	2.54–10.7	24.1	17.8–30.3
Beckman Coulter Access UniCel DxI	6.28	2.42–10.1	22.9	14.3–31.6
Ortho Vitros MicroWell Series (7)	3.49	2.41–4.57	15.3	11.7–18.9
Roche Elecsys/E170/Cobas eSystems	4.69	3.10–6.29	21.4	15.4–27.4
Siemens Advia Centaur CP (COR)	7.99	5.23–10.7	27.0	19.6–34.3
Siemens Advia Centaur XO Systems (COR)	6.66	3.08–10.2	17.9	12.2–23.7

to relevant serum concentration to allow for direct performance comparison between urine and serum) and considered 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 TE_o. As opposed to serum, the 3 investigated types of bias (SR, AB, RB) are markedly different in urine (AB < RB < SR) resulting in TE_o verifying TE_{oAB} < TE_{oRB} < TE_{oSR}. Yet, similarly to serum, 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}), whereas the chosen TE_a goal has a large effect on the set of acceptable QC rules. This emphasizes the major influence of the chosen TE_a on the resulting acceptable QC rules, further illustrating the direct relationship between the usable QC rules and the quality goal quantified by TE_a. On the other hand, similarly to what was already demonstrated in serum,¹¹ our study highlights the limited influence of P_{ed} on acceptable QC rules; P_{ed} needs to be maintained >90%.¹¹

For L4_{AB}, more QC rules were acceptable given a low AB achieved by averaging multiple RB biases and thus providing an illusion of a better-performing test than that which occurs when various concentrations of cortisol are considered separately. Especially in endocrinology, 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 most relevant type of bias to consider is certainly worth further discussion; however, it is important to realize that the considered type of bias has less impact on the acceptable QC rules (homogeneity of QC rules within one level regardless of the considered type of bias) than the considered level of TE_a (heterogeneity of acceptable QC rules within one level depending on the amount of TE_a). The bias of most relevance is the SR bias. It shows a marked positive bias with the Immulite 2000 Xpi cortisol immunoassay, especially at lower urine cortisol concentration. If the corresponding TE_o

(61.4% at 38.6 nmol/L, and 28.9% at 552 nmol/L) cannot verify TE_o < TE_a, this method may need to be rejected. However, first, a consensus recommendation for TE_a currently does not exist for urine cortisol in veterinary medicine. Second, provided that the assay has demonstrated satisfactory linearity in urine across the reportable range and that the precision in urine is roughly similar to that in serum, the interpretation of the test can consider the identified positive bias by an adapted IT. In our study, comparison of the Immulite 2000 with the RIA that we used previously for urine cortisol reflected that adaptation by increasing the IT of UCCR from 10×10^{-6} to 27×10^{-6} . The fact that the pre-spiking urine cortisol concentration was nearly 0 confirms that the positive bias identified after pure cortisol spiking cannot be attributed to measurement of other urinary corticoids, but is indeed related to the urine matrix. On the other hand, the negative bias observed with the new Siemens cortisol assay (lot numbers >550) in veterinary urine cortisol values in dogs (and cats) compared to values obtained with the assay used in our study (lot numbers <550)⁷ may result from higher specificity of the new antibody for cortisol versus urinary corticoids. Third, the SR bias is not always available, given that many hormones from veterinary species are not available for spiking, and it is not always possible to obtain a matrix depleted of the hormone of interest as we achieved here. Additionally, the TE concept is most commonly used to compare a result to a given IT for clinical interpretation, and this IT has been determined by studies using previous methods. One could then argue that the comparison with the previous method provides a more relevant bias, rather than a SR bias, specifically for interpretation purposes, at least until new studies determine the best IT with the newer method.

In light of these observations, we cannot yet provide recommendations for the use of specific QC rules, as those rules depend mostly on the TE_a, for which there is currently no consensus in veterinary medicine. However, a “reverse approach” to determine the level of total error that could be controlled with a simple QC rule, N=2 QCM levels,

$P_{ed} > 0.90$, and $P_{fr} < 0.05$, is conceivable.¹² These specifications for QC are considered reasonable for most veterinary laboratories based on QC expense, time involved in technical training for QC, and time spent on QC analysis, documentation, and troubleshooting.

We took extensive precautions to minimize limitations in our study; however, some limitations could not be overcome and require acknowledgment. The SR bias was calculated from a limited number of samples ($n=4$); $TE_{O_{SR}}$ results are approximate. Similarly, because the RB bias was generated from limited numbers of samples, $TE_{O_{RB}}$ results are approximate and would definitely benefit from further investigation with larger numbers of samples approximating those concentration levels tested in 2 different laboratories. Our study does not focus on a defined IT for UCCR interpretation when urine cortisol is measured with the Immulite 2000 Xpi cortisol immunoassay, but rather on the TEO one should be aware of for the measurement of canine urine cortisol used in the calculation of UCCR. The characterized TEO may also be of benefit in the determination of a future consensus about UCCR TEa in dogs. In addition, knowledge of biologic variation of UCCR is not currently available but would be a valuable contribution to our understanding of this parameter. It is certainly a limitation to use an older version of the method as a reference method for assay validation. On the other hand, the Immulite 1000 assay has been used satisfactorily for a long time to measure urine cortisol and to calculate the UCCR in a clinical setting. Furthermore, despite the marked negative bias identified in new kits from Siemens for “veterinary cortisol” (lot numbers ≥ 550), it would be worth investigating the linearity of the new kit to determine if it could qualify for clinical studies aimed at determining the most relevant IT from scratch. Finally, the absence of target values for the commercial QCM prevented the computation of the QCM bias and thus of $TE_{O_{QCM}}$; QCM bias was approximated primarily by the AB from the comparison study, which is certainly debatable, but might consist of a realistic approximation if the QCM bias of a future urine cortisol commercial QCM turns out to be as small as the one observed in serum.¹¹ Investigation to determine optimal simple QC rules function of desired $TE_{a_{QCM}}$ is also warranted.

Acknowledgments

We thank Mindy Borst (TVMDL) for performing the dilutions, Amy Siller (TVMDL) for performing and reporting the measurements, and Dr. Kelly M. Deewall for obtaining owner’s authorization and providing urine of the dog with hypoadrenocorticism as the matrix for the spiking-recovery study. Study results are part of a presentation given at the Annual Meeting of the American Association of Veterinary Laboratory Diagnosticians, Oct 5–21, 2020.

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

Our project was funded internally by the Texas A&M Veterinary Medical Diagnostic Laboratory.

ORCID iDs

Jeremie Korchia  <https://orcid.org/0000-0002-9344-6639>

Kathleen P. Freeman  <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. Bilić-Zulle L. Comparison of methods: Passing and Bablok regression. *Biochem Med (Zagreb)* 2011;21:49–52.
3. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.
4. Feldman EC, et al. Urine cortisol:creatinine ratio as a screening test for hyperadrenocorticism in dogs. *J Am Vet Med Assoc* 1992;200:1637–1641.
5. Galeandro L, et al. Urinary corticoid concentrations measured by 5 different immunoassays and gas chromatography-mass spectrometry in healthy dogs and dogs with hypercortisolism at home and in the hospital. *J Vet Intern Med* 2014;28:1433–1441.
6. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)* 2015;25:141–151.
7. 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/2020/20201109cortisolmeasurement_PreliminaryReport_Immulite2000Impact.pdf
8. Jensen AL, et al. Evaluation of the urinary cortisol:creatinine ratio in the diagnosis of hyperadrenocorticism in dogs. *J Small Anim Pract* 1997;38:99–102.
9. Jones CA, et al. Changes in adrenal cortisol secretion as reflected in the urinary cortisol/creatinine ratio in dogs. *Domest Anim Endocrinol* 1990;7:559–572.
10. 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.
11. Korchia J, Freeman KP. Validation study of canine serum cortisol measurement with the Immulite 2000 Xpi cortisol immunoassay. *J Vet Diagn Invest* 2021;33.
12. Korchia J, Freeman KP. Total observable error, total allowable error, and QC rules for canine serum and urine cortisol achievable with the Immulite 2000 Xpi. *J Vet Diagn Invest*. Submitted.
13. 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.
14. Mattingly D. A simple fluorimetric method for the estimation of free 11-hydroxycorticoids in human plasma. *J Clin Pathol* 1962;15:374–379.
15. 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.

16. Rijnberk A, et al. Assessment of two tests for the diagnosis of canine hyperadrenocorticism. *Vet Rec* 1988;122:178–180.
17. Smiley LE, et al. Evaluation of a urine cortisol:creatinine ratio as a screening test for hyperadrenocorticism in dogs. *J Vet Intern Med* 1993;7:163–168.
18. Stockham SL, et al. Urinary system. In: Stockham SL, ed. *Fundamentals of Veterinary Clinical Pathology*. 2nd ed. Wiley-Blackwell, 2008:415–494.
19. Stolp R, et al. Urinary corticoids in the diagnosis of canine hyperadrenocorticism. *Res Vet Sci* 1983;34:141–144.
20. Zeugswetter F, et al. Tailored reference limits for urine corticoid:creatinine ratio in dogs to answer distinct clinical questions. *Vet Rec* 2010;167:997–1001.
21. Zeugswetter FK, et al. Configuration of antibodies for assay of urinary cortisol in dogs influences analytic specificity. *Domest Anim Endocrinol* 2013;45:98–104.