Flaminia Fanelli*, Marco Cantù, Anastasia Temchenko, Marco Mezzullo, Johanna M. Lindner, Mirko Peitzsch, James M. Hawley, Stephen Bruce, Pierre-Alain Binz, Mariette T. Ackermans, Annemieke C. Heijboer, Jody Van den Ouweland, Daniel Koeppl, Elena Nardi, Finlay MacKenzie, Manfred Rauh, Graeme Eisenhofer, Brian G. Keevil, Michael Vogeser and Uberto Pagotto

Report from the HarmoSter study: impact of calibration on comparability of LC-MS/MS measurement of circulating cortisol, 17OH-progesterone and aldosterone

Abstract

Objectives: Liquid chromatography-tandem mass spectrometry (LC-MS/MS) is recommended for measuring circulating steroids. However, assays display technical heterogeneity. So far, reproducibility of corticosteroid LC-MS/MS measurements has received scant attention. The aim of the study was to compare LC-MS/MS measurements of cortisol, 17OH-progesterone and aldosterone from nine European centers and assess performance according to external quality assessment (EQA) materials and calibration.

Methods: Seventy-eight patient samples, EQA materials and two commercial calibration sets were measured twice by laboratory-specific procedures. Results were obtained by *in-house* (CAL1) and external calibrations (CAL2 and CAL3). We evaluated intra and inter-laboratory imprecision, correlation and agreement in patient samples, and trueness, bias and commutability in EQA materials.

Results: Using CAL1, intra-laboratory CVs ranged between 2.8–7.4%, 4.4–18.0% and 5.2–22.2%, for cortisol, 17OH-progesterone and aldosterone, respectively. Trueness and bias in EQA materials were mostly acceptable, however, inappropriate commutability and target value assignment were highlighted in some cases. CAL2 showed suboptimal accuracy. Median inter-laboratory CVs for cortisol, 17OH-progesterone and aldosterone were 4.9, 11.8

*Corresponding author: Flaminia Fanelli, Junior Assistant Professor, Department of Medical and Surgical Sciences, Unit of Endocrinology and Prevention and Care of Diabetes, Center for Applied Biomedical Research, University of Bologna, S. Orsola Policlinic, Via Massarenti 9, Bologna 40138, Italy, Phone and Fax: +39 051 2143902, E-mail: flaminia.fanelli2@unibo.it. https://orcid.org/0000-0003-2601-2694
Marco Cantù, Laboratory of Clinical Biochemistry and Pharmacology, Institute of Laboratory Medicine EOLAB, Ente Ospedaliero Cantonale, Bellinzona, Switzerland
Anastasia Temchenko, Marco Mezzullo and Uberto Pagotto, Department of Medical and Surgical Sciences, Unit of Endocrinology and Prevention and Care of Diabetes, Center for Applied Biomedical Research, University of Bologna, S. Orsola Policlinic, Bologna, Italy. https://orcid.org/0000-0001-6748-3041 (M. Mezzullo)
Johanna M. Lindner and Michael Vogeser, Institute of Laboratory Medicine, Hospital of the University of Munich (LMU), Munich, Germany
Mirko Peitzsch and Graeme Eisenhofer, Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. https://orcid.org/0000-0002-8601-9903 (G. Eisenhofer)

James M. Hawley and Brian G. Keevil, Department of Clinical Biochemistry, University Hospital South Manchester, Manchester NHS Foundation Trust, Manchester, UK
Stephen Bruce and Pierre-Alain Binz, Clinical Chemistry Laboratory, University Hospital of Lausanne (CHUV), Lausanne, Switzerland. https://orcid.org/0000-0002-0045-7698 (P.-A. Binz)
Mariette T. Ackermans, Department of Clinical Chemistry, Endocrine Laboratory, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
Annemieke C. Heijboer, Department of Clinical Chemistry, Endocrine Laboratory, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; and Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Jody Van den Ouweland, Department of Clinical Chemistry, Canisius-Wilhelmina Hospital, Nijmegen, The Netherlands
Daniel Koeppl and Manfred Rauh, Department of Pediatrics and Adolescent Medicine, University Hospital, Erlangen, Germany
Elena Nardi, Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy
Finlay MacKenzie, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

and 13.8% with CAL1 and 3.6, 10.3 and 8.6% with CAL3 (all p<0.001), respectively. Using CAL1, median bias vs. all laboratory-medians ranged from –6.6 to 6.9%, –17.2 to 7.8% and –12.0 to 16.8% for cortisol, 17OH-progesterone and aldosterone, respectively. Regression lines significantly deviated from the best fit for most laboratories. Using CAL3 improved cortisol and 17OH-progesterone between-method bias and correlation.

Conclusions: Intra-laboratory imprecision and performance with EQA materials were variable. Inter-laboratory performance was mostly within specifications. Although residual variability persists, adopting common traceable calibrators and RMP-determined EQA materials is beneficial for standardization of LC-MS/MS steroid measurements.

Keywords: 17OH-progesterone; aldosterone; calibration; cortisol; external quality control; harmonization; inter-laboratory performance; liquid chromatography-tandem mass spectrometry; method comparison; steroid hormones.

Introduction

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) is recommended for steroid measurement in high-throughput settings thanks to its analytical specificity and large dynamic range [1–5]. However, LC-MS/MS combines different pre-analytical, analytical and post-analytical strategies, resulting in a heterogenous spectrum of methods. The harmonization of LC-MS/MS methods has previously been investigated for testosterone, a few other sex steroids and 25OH-vitamin D [6–13]. Cortisol, 17OH-progesterone and aldosterone are routinely measured in the work-up of hypercortisolism, endocrine hypertension, congenital steroidogenesis defects, adrenal insufficiency and female hyperandrogenism; however, little is known concerning LC-MS/MS reproducibility for these measurements [14, 15].

Due to the abundance of isobars within the steroid family, chromatographic separation is critical for LC-MS/MS specificity. Using stable isotope-labeled analytes as internal standards (IS) minimizes procedural variability and matrix interference. However, IS with different isotopes and substitutions may influence measurement accuracy [15–17]. Therefore, LC-MS/MS methods may display different susceptibilities to isobaric and matrix-related interferences. Whether performance of LC-MS/MS methods differs according to serum or plasma sample matrix and

associated anticoagulants or coagulation supports has received limited examination [18].

Calibration is a major determinant of accuracy. Solvent-based certified reference materials (CRMs) are currently available for most steroids. However, to prepare calibrators, CRMs need to be diluted in solvents or surrogate matrices, introducing procedural and matrix variability. Adopting common calibrators was found to improve inter-laboratory performance in some studies [7, 13] but not in others [9].

External quality assessment (EQA) programs are crucial tools for harmonization [19]. EQA materials differ in the assignment of target values, including reference measurement procedures (RMPs) or mean/median of survey results, and experimental evidence on their commutability by LC-MS/MS remains scarce [20].

The extent to which the aforementioned factors impact LC-MS/MS results and reproducibility can be significant. International initiatives promoting harmonization and traceability, and databases of RMPs and CRMs are now available [21, 22]. Based on these, standardization of clinically relevant steroids by LC-MS/MS appears to be a realistic goal.

The high quality of the assays is a prerequisite for achieving harmonization of laboratory tests [23]. However, even when methods are validated by recommended guidelines, actual performance is hardly inferable from publications. Consequently, there is increasing attention toward applying strict validation requirements and performance reporting protocols [24, 25]. Moreover, aiming at easing the harmonization process, the incoming EU *in vitro* diagnostic regulation (IVDR) [26] states that commercial tests with appropriate analytical and clinical performance should be preferred to equivalent laboratory developed tests (LDTs). However, comparing LDTs and commercial kits based on the reported performance can be cumbersome [25].

The HarmoSter initiative aims to evaluate the harmonization status of LC-MS/MS measurement of ten circulating steroids (cortisol, 17OH-progesterone, aldosterone, dehydroepiandrosterone, dehydroepiandrosterone-sulfate, androstenedione, testosterone, corticosterone, 11-deoxycortisol and cortisone) by nine European centers (Supplemental Table 1). Authentic samples collected by three different vacuum tubes, EQA materials and commercial calibrators were tested. The present work focuses on the impact of calibration on intra- and inter-laboratory variability for three of the ten steroids of the HarmoSter initiative: cortisol, 17OH-progesterone and aldosterone.

Materials and methods

Consortium and methods

The Bologna ethics committee approved the HarmoSter study (n° 141/2017/U/Tess). Laboratory A coordinated the study, recruited patients and collected samples. Laboratories B to L were measuring centers: all measured cortisol and 17OH-progesterone; five also measured aldosterone (Supplemental Table 1 and Table 1). Eleven LDTs (Laboratories B to I) and two panel MassChrom® kit (Chromsystems; Munich, Germany; https://chromsystems.com) (Laboratory L) were involved. Among LDTs, 6PLUS1® Multilevel Serum Calibrator set (Chromsystems) was used for *in-house* calibration for cortisol and aldosterone by Laboratories D and E, and for 17OH-progesterone by Laboratories D, E and F. Technical details and *in-house* measurement ranges are shown in Table 1 and Table 2.

Study samples

The sample set included authentic samples, EQA materials and calibrators. We recruited 26 volunteers (women/men: 13/13; age: 20–69 years) following informed consent. Nine subjects were healthy and medication-free. Six women had hyperandrogenism and one had an estroprogestogen patch; two men had hypogonadism and one inhaled budesonide; others were taking one or more medications (cholecalciferol, thyroxine, folic acid, insulin sensitizers, cholesterol-lowering, anti-hypertensive, expectorants and antibiotics). Blood was taken at 7:30–8:30 am after overnight fasting. For each subject, three types of tubes were randomly alternated, containing gel separator (Vacutainer™ SST™ II Advance, cat.366468, Becton Dickinson, Milan, Italy), lithium-heparin (Vacutainer™, 102 IU, cat.368886, Becton Dickinson) or beads clot activator (Vacutest, cat.10636, KIMA S.R.L., Azergrande, Italy). After 30 min settling, samples were centrifuged (2000 rcf, 10 min, room temperature). Supernatants of each specimen were mixed, aliquoted and stored at −80 °C. Seventy-eight samples were obtained.

EQA materials were sent to Laboratory A and stored according to manufacturers' specifications. The Reference Institute for Bioanalytics (RfB; Bonn, Germany; www.rfb.bio) donated four materials (HM40121, HM40122, HM40123 and HM40124; lyophilized human recalcified plasma spiked with steroids and no preservatives), and assigned with cortisol, 17OH-progesterone and aldosterone RMP target values. The United Kingdom National External Quality Assessment Service (UKNEQAS; Birmingham, UK; https://ukneqas.org.uk) donated eight liquid materials (off-the-clot minimally manipulated human serum), four with cortisol (C568 and C532), 17OH-progesterone (H408, spiked with 17OH-progesterone) and aldosterone (L125, spiked with aldosterone) target values determined as means of all MS-based methods in the survey. Instand e.V (Düsseldorf, Germany; https://www.instand-ev.de). donated three low/high concentration paired materials (human serum spiked with steroids and no additives), two with target values assigned by RMP for cortisol (N°302, liquid), and as means of all methods in the survey for aldosterone (N°304, lyophilic). None of the EQA materials were directly tested for commutability.

The Biological Sales Network (BSN Srl, Castelleone, Italy; https://www.bsn-srl.it) donated seven level serum-based liquid calibrators for a ten-steroid panel (EUM01041, lot.M01411808VEQ) traceable to the Royal College of Pathologists of Australasia Quality Assurance Programs (RCPAQAP). Target values of RCPAQAP materials were determined by LC-MS/MS by the National Measurement Institute Australia (traceable to CRM-6007a) for cortisol, and by all laboratory medians for 17OH-progesterone and aldosterone. Freshly prepared calibrators were delivered to Laboratory A, immediately aliquoted and stored at −80 °C. Chromsystems donated serum-based lyophilic calibrators for a fifteen-steroid panel (6PLUS1® Multilevel Serum Calibrator set, lot.5016, different from lots used for *in-house* calibration by Laboratories D, E, F and L). Cortisol was traceable to NIST SRM-971 – frozen human serum; 17OH-progesterone and aldosterone were traceable to CRMs – methanol – from ISO 17025 and 17034 certified supplier. At Laboratory A, calibrators of each level were reconstituted according to manufacturer's instructions, mixed together, aliquoted and stored at −80 °C. BSN and Chromsystems measurement ranges were 11.92–3, 182 and 25.6–806 nmol/L for cortisol, 0.61–166.2 and 0.27–43.9 nmol/L for 17OH-progesterone, and 0.11–25.68 and 0.075–12.5 nmol/L for aldosterone, respectively.

Running scheme and quantitation

Two aliquots from 110 samples were shipped to measuring centers on the same day and stored at −80 °C. EQA materials were stored and handled according to manufacturers' indications. All were measured within 4 months by two identical runs, each including 110 singlets and an independent *in-house* calibration set, according to protocols ordinarily used by each laboratory. Results by *in-house* calibration (CAL1) were sent to Laboratory A before measuring centers received external calibrators' nominal values. They were then asked to use BSN (CAL2) and Chromsystems (CAL3) sets for re-quantification of results. External calibrators were included in the curve calculation if their nominal concentration was within each method's range. Calibration curves displayed $R^2 > 0.98$.

Data analysis and statistics

Hormone values are reported in nmol/L. To convert cortisol, 17OH-progesterone and aldosterone to ng/mL, multiply by 0.362, 0.33 and 0.36, respectively. Results were excluded if below the lower limit of quantification (LLOQ) or above the upper LOQ (ULOQ) of CAL1. Moreover, results were excluded from CAL2 or CAL3 datasets when out of the respective measurement range. Duplicate means and CVs were calculated (Supplemental Tables 2–4). Within-subject (CVi) and between-subjects (CVg) biological variabilities [27] were used to assess the maximum allowable imprecision (MAI; $0.5 \times CVi$) and bias (MAB; $0.25 \times (CVi^2 + CVg^2)^{0.5}$) and total allowable error (TAE; $0.25 \times (CVi^2 + CVg^2)^{0.5} + 1.65 \times (0.5 \times CVi))$ (cortisol: 11.6, 13.5 and 32.5%; 17OH-progesterone: 14.2, 12.0 and 35.3%; aldosterone 18.3, 12.6 and 42.8%; respectively) [28].

Intra-laboratory performance: Within-method imprecision was calculated in authentic samples according to the formula intra-laboratory CV : $\sqrt{\left\{ \left[\sum (a - b)^2 \right] / 2N \right\} \left(N / \sum \bar{x} \right)}$ (Σ: sum; a and b: duplicate measures of each sample; N: total number of duplicates; $\bar{x}$:

Table 1: Assays' pre-analytical, liquid chromatography and mass spectrometry features.

Lab	Ref	*In-house* calibration	Sample volume, µL	Sample preparation	Instrument	Run time, min	Purification column	LC column; temperature	Mobile phases (A; B)	Source
B	[2]	Cerilliant CRM[a] in 4% BSA	600[a]	PP: ZnSO₄ in MeOH; SPE: C18	Series 200, Perkin Elmer; API4000 QTrap, Sciex	21	POROS R1/20	Luna RP-C8 100 × 4.6 mm, 5 µm; 40 °C	20%MeOH in H_2O; MeOH	APCI
C	[35]	Gravimetric in charcoal stripped serum	50	PP: ZnTFA in MeOH	Acquity UPLC; Xevo TQ-S, Waters	10		Kinetex Biphenyl 150 × 2.1 mm, 1.7 µm	0.2 mM NH_4F in H_2O; MeOH	ESI
D	[31]	6PLUS1[®a,b], Chromsystems	500	SPE: Oasis HLB	Acquity UPLC, Waters; API5500 Qtrap, Sciex	14		Kinetex C18 100 × 2.1 mm, 2.5 µm	5 mM NH_4F in H_2O; MeOH	ESI[a]
E	[32]	6PLUS1[®a,b], Chromsystems	100	PP: ZnSO₄ in MeOH	1260 Infinity, Agilent; API6500 Qtrap, Sciex[a]	7	Chromolith SpeedRod C18 4.6 × 50 mm	Chromolith performance C18, 100 × 2.1 mm, 30 °C	5 mM NH_4HCO_2 in H_2O; 5 mM NH_4HCO_2 in MeOH[a]	ESI[a]
	n.a.	6PLUS1[®a,b], Chromsystems	500	SPE HR-X, 50 mg	1260 Infinity, Agilent; API6500 Qtrap, Sciex	7	Chromolith SpeedRod, C18 4.6 × 50 mm	Chromolith performance C18, 100 × 2.1 mm, 30 °C	0.2 mM NH_4F in H_2O; ACN	ESI
F	[34]	Cerilliant CRM[a] in PBS 0.1% BSA	20	PP: ZnSO₄ in MeOH	Acquity UPLC; TQD, Waters	2.6		Kinetex C8, 50 × 2.1 mm 40 °C	2 mM $C_2H_7NO_2$ + 0.1% FA in H_2O; 2 mM $C_2H_7NO_2$ + 0.1% FA in MeOH	ESI
	[36]	6PLUS1[®a,b], Chromsystems	50	SLE: Isolute 200	Acquity UPLC; TQS, Waters	4.0		Acquity T3 C18 50 × 2.1 mm, 50 °C	2 mM $C_2H_7NO_2$ + 0.1% FA in H_2O; ACN	ESI
G	[12]	Cerilliant CRM in steroid free plasma	25	PP: ACN	Acquity-Xevo TQ-S, Waters	9.9		HSS T3 2.1 × 100 mm, 1.8 µm	0.1% FA in H2O; 0.1% FA in ACN	ESI
	[33]	Cerilliant CRM in ACN	200	SLE: DCM	Acquity-Xevo TQ-S, Waters	6.1	Protein BEH C4 300 Å, 2.1 × 50 mm, 1.7 µm	BEH phenyl 2.1 × 50 mm, 1.7 µm	H_2O; MeOH	ESI
H	n.a.	Cerilliant CRM in 10% MeOH	1,000	PP: ZnSO₄ in MeOH; SPE: C18	Acquity-Xevo TQ-S, Waters	12		BEH C18 2.1 × 100 mm, 1.7 µm	0.05% FA in H_2O; 0.05% FA in MeOH	ESI
I	30, 37	Gravimetric in charcoal stripped serum	100	PP: H_3PO_4; SPE: Oasis MCX	I-class Acquity; TQS, Waters[a]	15[a]		C18 Zorbax Eclipse plus 2.1 × 50 mm, 1.8 µm[a]	H_2O; MeOH	ESI
L	[c]	6PLUS1[®b], Chromsystems	500	Patented	1290 series HPLC 6490, Agilent	11.5		Patented	Patented	ESI
	[c]	6PLUS1[®b], Chromsystems	500	Patented	1290 series HPLC 6490, Agilent	11.7		Patented	Patented	ESI

duplicate (a and b) mean), and compared with the MAI. Within-method impact of calibration was evaluated by the Friedman test and Passing-Bablok regression. Within-method trueness and bias were estimated in EQA materials as % difference of RMP or mean/median of the surveys as target values, respectively, and were compared with the MAB. Least-squares regression lines were calculated in authentic sample CAL1 results from each laboratory vs. all laboratories-medians; 95% prediction intervals were used to test EQA materials commutability [29].

Inter-laboratory performance: Analyses were performed in authentic samples measured with duplicate-CV <30% and within both CAL1 and CAL3 ranges (Supplemental Tables 2–4). Between-method reproducibility, valued by the inter-laboratory CV, was compared with the MAI. Between-method regression was assessed by Passing-Bablok analysis. Between-method agreement was valued by %-bias vs. all methods median and Bland-Altman; results were compared with the TAE. Wilcoxon and F tests were used to compare CAL1 and CAL3 inter-laboratory CV and bias. Statistics were performed by SPSS (v.20, IBM Co., Somers, NY) and MedCalc (v.18.2.1; Mariakerke, Belgium).

Results

Cortisol

Authentic sample results ranged from 221 to 994 nmol/L (median of all laboratories by CAL1). Three to nine samples were above CAL1 ULOQ in three laboratories. The highest CAL2 calibrator was above the *in-house* measurement range of all laboratories. Conversely, three samples were above the CAL3 measurement range (Supplemental Table 2). The intra-laboratory CVs ranged from 2.8 to 7.4%, all below the MAI (Table 2). Calibration influenced the measures within all laboratories (p<0.001). Laboratory D results were largely lowered when using CAL2 (–40.2%) and CAL3 (–31.8%) compared to CAL1. For other laboratories, when compared with CAL1, CAL2 determined –18.8 to –9.7% deviation, while CAL3 yielded modest deviations

Table 1: (continued)

Lab	Ref	Cortisol			17OH-Progesterone			Aldosterone		
		Analyte (ion mode) mass transition	IS	IS (ion mode) mass transition	Analyte (ion mode) mass transition	IS	IS (ion mode) mass transition	Analyte (ion mode) mass transition	IS	IS (ion mode) mass transition
B	[2]	(+) 363/267[a]	cortisol-[9,11,12,12-D4]	(+) 367/97	(+) 331/97	17OH-progesterone-[2,2,4,6,6,21,21,21-D8]	(+) 339/100			
C	[35]	(+) 363/121	cortisol-[9, 12, 12-D3]	(+) 366/121	(+) 331/97	17OH-progesterone-[2,2,4,6,6,21,21,21-D8]	(+) 339/100			
D	[31]	(+) 363/121	cortisol-[1,2-D2]	(+) 365/347	(+) 331/109	17OH-progesterone-[2,2,4,6,6,21,21,21-D8]	(+) 339/100	(+) 361/343	Aldosterone-[2,2,4,6,6,21,21-D7]	(+) 368/350
E	[32]	(+) 363/121	cortisol-[9,11,12,12-D4]	(+) 367/121	(+) 331/109	17-OH-progesterone [2,3,4-13C3]	(+) 334/112			
	n.a.							(−) 359/189	Aldosterone-[2,2,4,6,6,21,21-D7]	(−) 366/338
F	[34]	(+) 363/97	D4-cortisol [9,11,12,12-D4]	(+) 367/97						
	[36]				(+) 331/97	17-OH-progesterone [2,3,4-13C3]	(+) 334/100			
G	[12]	(+) 363/121	cortisol-[9,11,12,12-D4]	(+) 367/121	(+) 331/97	17OH-progesterone-[2,2,4,6,6,21,21,21-D8]	(+) 339/100			
	[33]							(−) 359/189	Aldosterone-[9,11,12,12-D4]	(−) 363/190
H	n.a.	(+) 363/121	cortisol-[9,11,12,12-D4]	(+) 367/121	(+) 331/97	17OH-progesterone-[2,2,4,6,6,21,21,21-D8]	(+) 339/100	(+) 361/343	Aldosterone-[2,2,4,6,6,21,21-D7]	(+) 368/350
I	30, 37	(+) 363/121	cortisol-[9,11,12,12-D4]	(+) 367/121	(+) 331/97	17OH-progesterone-[2,2,4,6,6,21,21,21-D8]	(+) 339/100			
L	[c]	(+) 363/97	Patented	(+) 367/97				(−) 359/331	Patented	(−) 363/335
	[c]				(+) 331/109	Patented	(+) 334/112			

[a]Modified from original publication. [b]Lots different from the one distributed in the study. [c]https://chromsystems.com/en/products/steroids/masschromr-steroids-in-serum-plasma-with-96-spe-well-plate-lc-ms-ms-72072-96%20&%2072072-480.html. CRM, certified reference material; IS, internal standard; PP, protein precipitation; ZnSO4, zinc sulfate; MeOH, methanol; SPE, solid phase extraction; RP, reversed phase; ZnTFA, zinc trifluoroacetate; HLB, hydrophilic-lipophilic balance; FA, formic acid; ACN, acetonitrile; PBS, phosphate buffered saline; HSS, high strength silica; SB, selectivity for base; SLE, supported liquid extraction; DCM, dichloromethane; BEH, ethylene bridged hybrid; H_3PO_4, phosphoric acid; MCX, mixed-mode strong cation-exchange.

(Supplemental Table 5). Comparing CAL1 vs. CAL3 by Passing-Bablok within laboratories using Chromsystems' as *in-house* calibration, confirmed the large proportional overestimation of the former in Laboratory D and the consistency of results in Laboratories E and L (Supplemental Figure 1)

EQA material target values ranged from 265.0 to 948.5 nmol/L; two were above CAL1 and CAL3 measurement ranges in Laboratories D, E and L (Supplemental Table 2). Trueness and bias by CAL1 and CAL3 were mostly within MAB, except Laboratory D, displaying a large positive bias with CAL1. CAL2 determined a negative deviation, with several cases exceeding the MAB, especially for Laboratories D and F (Figure 1). Commutability was demonstrated except for some high level EQA materials slightly outside the interval for three laboratories (Supplemental Figure 2).

Given the large deviation shown by Laboratory D, inter-laboratory analyses were performed both with and without data from that laboratory, the latter reported as follows. Median inter-laboratory CV by CAL1 was 4.9%, and it was significantly reduced to 3.6% with CAL3 (p<0.001). No cases were detected with CV >MAI (Figure 2 and Supplemental Table 6).

At Passing-Bablok analysis, slopes were similar to 1 in two laboratories with CAL1 and in six with CAL3 (Table 3).

Median bias ranged from −6.6 to 6.9% for CAL1 and −2.3 to 5.5% for CAL3, with all results within the TAE (Figure 3). Compared to CAL1, CAL3 significantly reduced the bias median within six and variance within three laboratories (Supplemental Table 7). Bland-Altman analyses per laboratory are shown in Supplemental Figures 3 and 4.

Table 2: Intra-laboratory measurement range, measures and imprecision.

Analyte	Lab	LLOQ-ULOQ, nmol/L	n	Mean (min-max), nmol/L	Intra-lab CV% (min-max)
Cortisol	B	0.3–1,380	78	393 (204–981)	3.8 (0.1–9.7)
	C	3–1,480	78	436 (238–1,135)	5.3 (0.0–19.7)
	D	7–830[a]	69	496 (303–784)	5.1 (0.3–15.8)
	E	0.3–830[a]	75	377 (210–623)	3.0 (0.1–6.1)
	F	12.5–2000	78	380 (207–922)	2.8 (0.0–6.8)
	G	2–1,000	78	404 (224–943)	4.2 (0.1–21.2)
	H	3.3–1,660	78	415 (222–1,013)	2.8 (0.0–8.2)
	I	5–1,400	75	422 (188–1,005)	4.3 (0.0–25.3)
	L	25–800[a]	75	387 (221–658)	7.4 (0.0–19.1)
17OH-Progesterone	B	0.03–150	78	1.98 (0.35–6.91)	5.3 (0.2–16.3)
	C	0.24–110	78	1.75 (0.28–5.96)	6.3 (0.0–20.3)
	D	0.19–80[a]	74	1.80 (0.22–6.63)	8.9 (0.0–25.3)
	E	0.15–45[a]	78	1.93 (0.34–7.18)	6.3 (0.2–27.3)
	F	0.25–80.3[a]	78	1.90 (0.34–6.65)	4.4 (0.0–25.3)
	G	1–400	39	2.49 (1.05–5.63)	10.7 (0.7–23.0)
	H	0.5–60	64	1.95 (0.65–5.44)	18.0 (0.8–42.5)
	I	0.2–80	75	1.82 (0.30–5.99)	6.4 (0.0–33.1)
	L	0.29–45[a]	78	1.90 (0.33–7.43)	15.0 (0.0–41.1)
Aldosterone	D	0.04–14[a]	77	0.26 (0.07–1.42)	17.2 (0.0–54.4)
	E	0.01–7.1[a]	77	0.24 (0.08–1.16)	8.6 (0.4–26.4)
	G	0.03–5	78	0.20 (0.07–1.00)	5.2 (0.0–17.7)
	H	0.05–4	78	0.22 (0.06–1.23)	5.4 (0.0–16.3)
	L	0.07–12[a]	43	0.33 (0.10–1.30)	22.2 (0.0–70.5)

Data from *in-house* calibration. Intra-laboratory imprecision was calculated as duplicate measurement CV. [a]*In-house* calibration material from Chromsystems. n, number of samples.

17OH-Progesterone

Authentic sample results ranged from 0.34 to 6.63 nmol/L (median of all laboratories by CAL1). Values were below the CAL2 measurement range in 33 authentic samples (Supplemental Table 3). The intra-laboratory CV ranged from 4.4 to 18.0%, exceeding the MAI in Laboratory H and L (Table 2). Duplicate-CV increased at values <1.0 nmol/L. Calibration influenced results within all laboratories (p<0.001). Compared to CAL1, CAL2 determined –20.3 to –5.7% lower values, while CAL3 determined modest variations (Supplemental Table 5). CAL1 vs. CAL3 Passing-Bablok analysis within laboratories using Chromsystems' as *in-house* calibration detected small deviations in Laboratory D and E, but substantial consistency in Laboratories F and L (Supplemental Figure 1).

EQA materials target values ranged from 2.41 to 12.73 nmol/L. Trueness and bias by CAL1 and CAL3 were mostly within the MAB, except for Laboratory G (–22.2 to –12.3%), and Laboratory H, showing higher values at lower levels (–25.8 to 109.1%). CAL2 increased the negative biases, with several cases exceeding the MAB (Figure 1). A slight deviation from the commutability interval was observed in high level EQA materials in two laboratories. Moreover, HM materials lacked commutability in Laboratory H (Supplemental Figure 2).

The 11.8% median inter-laboratory CVs observed with CAL1 reduced to 10.3% with CAL3 (p<0.001), while cases with inter-laboratory CV>MAI were 33.3 and 7.7%, respectively. Notably, CAL1 CVs increased at levels <1 nmol/L (Figure 2 and Supplemental Table 6).

Passing-Bablok slopes were similar to 1 in two laboratories with both CAL1 and CAL3. Laboratory G and H displayed the lowest slope coefficients (95CI) when using CAL1 (0.844 (0.789–0.896) and 0.777 (0.738–0.809), respectively). Laboratory H also showed a large intercept (95CI) (0.318 (0.274–0.363)). CAL3 improved Laboratory H (slope: 0.922 (0.868–0.954); intercept: 0.066 (0.014–0.112)) but not Laboratory G regression (Table 3).

Median bias vs. median of all laboratories ranged from –17.2 to 7.8% for CAL1 and -20.9 to 10.3% for CAL3 (Figure 3). With both calibrations, Laboratory G showed the largest negative median and Laboratory H the largest variance of bias. Compared to CAL1, CAL3 significantly reduced the bias median within four and variance within three laboratories (Supplemental Table 7). Moreover, CAL3
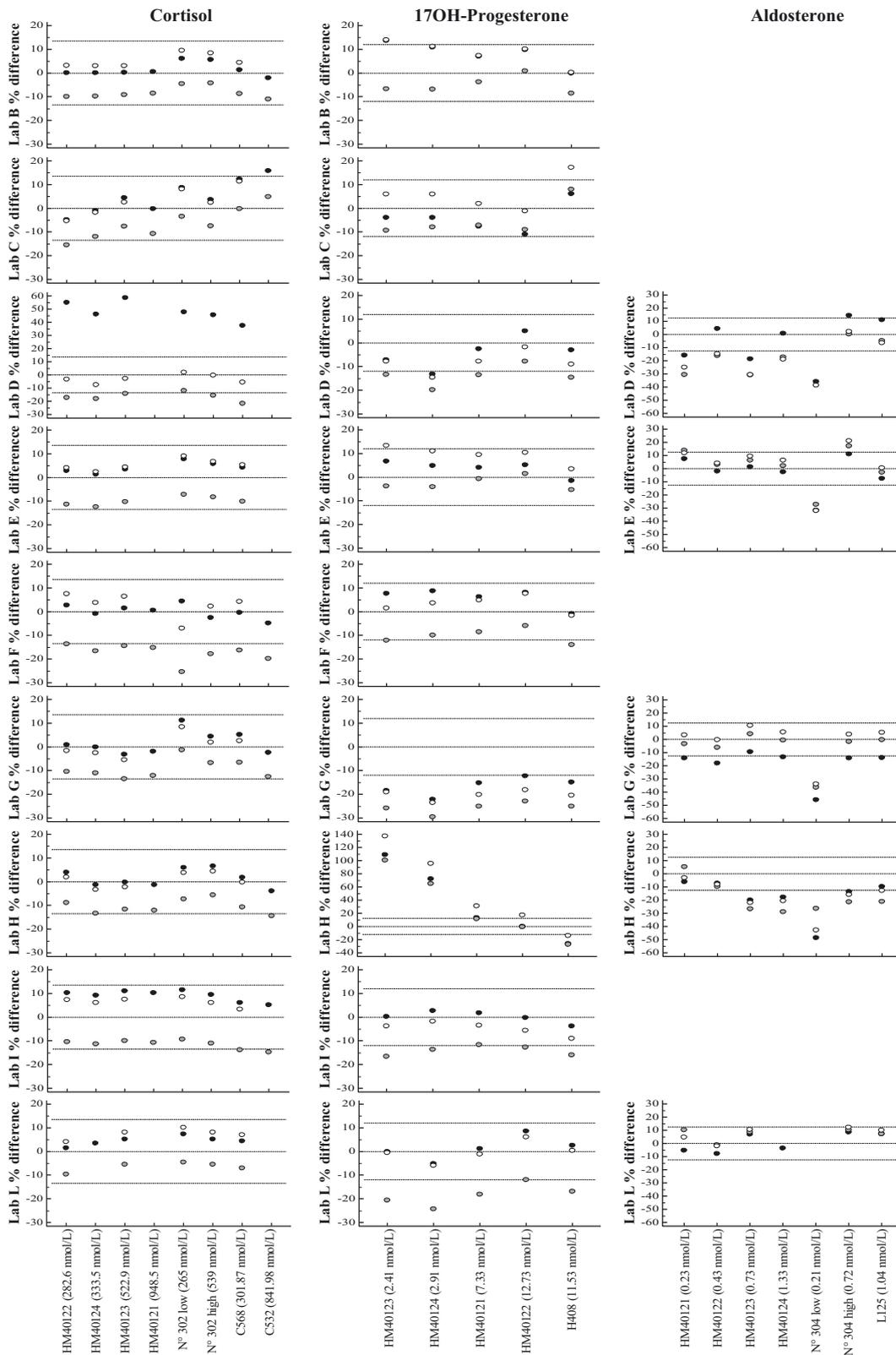
Figure 1: Trueness and bias of steroid measurement in external quality assessment materials as function of the calibration system.
% Difference = ((laboratory value – target value)/target value) × 100. Trueness vs. target values determined by reference measurement procedure was evaluated in materials HM40121, HM40122, HM40123, HM40124 for all analytes and in material N°302 for cortisol. Bias vs. target values determined as mean/median of the EQA survey was evaluated in materials C568 and C532 for cortisol, H408 for 17OH-progesterone, N°304 and L125 for aldosterone. Lines: zero ± maximum allowable bias (cortisol: 13.5%, 17OH-progesterone: 12.0% and aldosterone: 12.6%). Black dots: *in-house* calibration (CAL1); gray dots: BSN calibration (CAL2); white dots: Chromsystems calibration (CAL3).
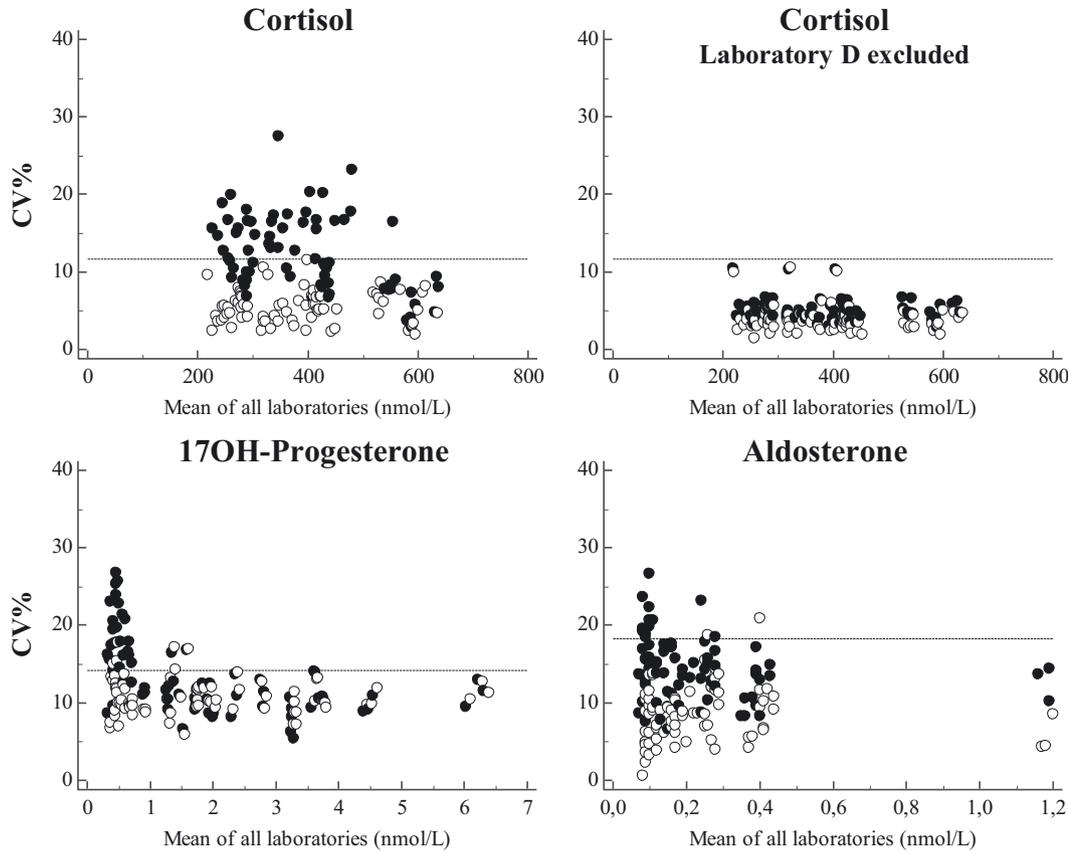
Figure 2: Inter-laboratory coefficient of variation according to the calibration system.
Black dots: *in-house* calibration (CAL1); white dots: Chromsystems calibration (CAL3). Horizontal lines: maximal allowable imprecision (cortisol: 11.6%; 17OH-progesterone: 14.2%; aldosterone: 18.3%).

reduced the opposite systematic bias shown by Laboratory D and H at lowering concentrations with CAL1 (Supplemental Figure 5), almost eliminating cases exceeding the TAE (Figure 3).

Aldosterone

Authentic sample results ranged from 0.07 to 1.22 nmol/L (median of all laboratories by CAL1). Values were below the CAL2 measurement range in 32 authentic samples for most of the laboratories (Supplemental Table 4).

The intra-laboratory CV ranged from 5.2 to 22.2%, with Laboratory L exceeding the MAI (Table 2). Duplicate-CV increased at values <0.2 nmol/L. Calibration influenced results within all laboratories (p<0.001). Compared to CAL1 values, the deviation was −13.1 to 25.2% for CAL2 and −10.0 to 22.1% for CAL3 (Supplemental Table 5). CAL1 vs. CAL3 Passing-Bablok analysis within laboratories using Chromsystems' as *in-house* calibrators detected deviations

in Laboratory D and E, but consistent results in Laboratory L (Supplemental Figure 1).

EQA material target values ranged from 0.21 to 1.33 nmol/L. A large negative bias was shown for N°304 low (0.21 nmol/L) by all laboratories and calibration sets (−48.6 to −31.6%, or below measurement range). Trueness and bias by CAL1 mostly exceeded the MAB in Laboratory D, G and H. External calibrations improved trueness and bias in Laboratory G only (Figure 1). A slight deviation from the 95% commutability interval was noted in high level EQA materials in three laboratories (Supplemental Figure 2).

Median inter-laboratory CVs with CAL1 was 13.8% and reduced to 8.6% with CAL3 (p<0.001), while cases with inter-laboratory CVs>MAI were 15.4 and 2.6%, respectively (Figure 2 and Supplemental Table 6).

Passing-Bablok slopes in CAL1 and CAL3 were similar to 1 in three and two laboratories, respectively. Compared to CAL1, CAL3 improved the slope (95CI) in Laboratory G (CAL1: 0.846 (0.821–0.870); CAL3: 1.000 (1.000–1.000))

Table 3: Passing-Bablok analysis of steroid measures from each laboratory vs. the median of all laboratories as function of the calibration system.

Analyte	Lab		CAL1				CAL3		
		n	r (95CI)	Slope (95CI)	Intercept (95CI)	n	r (95CI)	Slope (95CI)	Intercept (95CI)
Cortisol	B	75	0.995 (0.992–0.997)	1.007 (0.990–1.024)	−18.2 (−24.6−−11.8)	75	0.996 (0.994–0.998)	1.008 (1.000–1.020)	−4.7 (−8.9−−1.3)
	C	75	0.982 (0.971–0.988)	1.083 (1.055–1.111)	−8.7 (−18.4–1.4)	75	0.984 (0.975–0.990)	1.057 (1.029–1.085)	−0.3 (−9.4–10.4)
	D	69	0.910 (0.859–0.944)	1.268 (1.122–1.496)	21.1 (−46.2 – 71.6)	69	0.912 (0.861–0.945)	0.813 (0.739–0.923)	30.9 (−6.4–58.6)
	E	75	0.995 (0.991–0.997)	1.024 (1.008–1.040)	−19.3 (−25.0−−13.0)	75	0.997 (0.995–0.998)	1.013 (1.000–1.025)	−7.7 (−12.1−−3.7)
	F	75	0.993 (0.988–0.995)	0.946 (0.927–0.963)	−6.2 (−12.3–0.4)	75	0.996 (0.993–0.997)	1.006 (0.995–1.019)	0.6 (−3.9–4.4)
	G	75	0.994 (0.991–0.996)	0.974 (0.952–0.990)	7.7 (2.8–13.6)	75	0.993 (0.989–0.996)	0.938 (0.918–0.954)	14.7 (9.3–21.4)
	H	75	0.992 (0.987–0.995)	1.031 (1.008–1.055)	−6.7 (−14.8–3.7)	75	0.989 (0.983–0.993)	1.000 (0.978–1.022)	0.0 (−7.5–8.6)
	I	72	0.992 (0.988–0.995)	1.068 (1.040–1.099)	−14.4 (−26.5−−5.9)	72	0.994 (0.990–0.996)	1.004 (0.991–1.028)	−1.0 (−8.9–3.8)
	L	75	0.983 (0.974–0.990)	1.000 (0.969–1.009)	0.0 (−3.0 – 12.4)	75	0.980 (0.968–0.987)	1.002 (0.973–1.036)	10.3 (−0.3–20.8)
Cortisol Laboratory D excluded	B	75	0.995 (0.992–0.997)	1.004 (0.988–1.022)	−14.0 (−20.5−−7.7)	75	0.996 (0.994–0.998)	1.006 (0.995–1.019)	−5.2 (−9.4−−1.5)
	C	75	0.983 (0.973–0.989)	1.083 (1.054–1.112)	−5.8 (−17.0−−5.9)	75	0.983 (0.974–0.990)	1.053 (1.024–1.076)	1.1 (−7.5–11.5)
	E	75	0.995 (0.991–0.997)	1.022 (1.006–1.038)	−14.5 (−21.1−−9.2)	75	0.997 (0.995–0.998)	1.006 (0.995–1.018)	−7.1 (−11.3−−3.5)
	F	75	0.992 (0.987–0.995)	0.945 (0.929–0.959)	−3.5 (−9.3–2.0)	75	0.996 (0.994–0.998)	1.005 (0.993–1.016)	−0.1 (−3.9–4.9)
	G	75	0.994 (0.990–0.996)	0.969 (0.948–0.984)	10.8 (5.3–18.1)	75	0.993 (0.990–0.996)	0.933 (0.917–0.947)	15.6 (10.8–21.5)
	H	75	0.992 (0.987–0.990)	1.033 (1.007–1.059)	−4.7 (−14.4–5.5)	75	0.988 (0.982–0.993)	0.995 (0.971–1.020)	1.1 (−7.6–10.2)
	I	72	0.992 (0.988–0.995)	1.067 (1.040–1.097)	−12.2 (−21.6−−3.3)	72	0.995 (0.991–0.997)	1.004 (0.988–1.021)	−0.9 (−7.0–4.9)
	L	75	0.984 (0.975–0.990)	0.992 (0.962–1.018)	5.0 (−4.3 – 16.7)	75	0.979 (0.968–0.987)	0.998 (0.967–1.032)	11.2 (−0.1 – 21.9)
17OH-Progesterone	B	78	0.992 (0.988–0.995)	1.071 (1.047–1.091)	0.007 (−0.019–0.033)	78	0.993 (0.989–0.996)	1.072 (1.050–1.088)	−0.021 (−0.036–0.006)
	C	78	0.995 (0.993–0.997)	0.974 (0.954–0.994)	−0.033 (−0.047−0.009)	78	0.995 (0.992–0.997)	1.042 (1.016–1.062)	−0.009 (−0.023–0.018)
	D	74	0.996 (0.993–0.997)	1.039 (1.025–1.054)	−0.193 (−0.210−−0.170)	74	0.994 (0.990–0.996)	0.952 (0.938–0.972)	−0.029 (−0.060−−0.020)
	E	78	0.998 (0.996–0.999)	1.052 (1.037–1.063)	−0.012 (−0.025–0.004)	78	0.995 (0.993–0.997)	1.090 (1.077–1.107)	0.012 (−0.005–0.037)
	F	76	0.994 (0.991–0.996)	1.009 (1.000–1.023)	0.026 (0.007–0.035)	76	0.994 (0.991–0.996)	1.033 (1.021–1.045)	−0.084 (−0.098−−0.070)
	G	39	0.957 (0.920–0.978)	0.844 (0.789–0.896)	−0.030 (−0.225–0.115)	39	0.967 (0.938–0.983)	0.773 (0.727–0.811)	0.065 (−0.061–0.195)
	H	64	0.978 (0.963–0.986)	0.777 (0.738–0.809)	0.318 (0.274–0.363)	64	0.981 (0.968–0.988)	0.922 (0.868–0.954)	0.066 (0.014–0.112)
	I	75	0.995 (0.992–0.997)	1.056 (1.036–1.084)	−0.057 (−0.080−−0.038)	75	0.995 (0.993–0.997)	1.007 (0.988–1.028)	−0.033 (−0.055−−0.016)
	L	78	0.993 (0.990–0.996)	1.026 (1.000–1.063)	−0.024 (−0.054–0.000)	78	0.993 (0.990–0.996)	1.000 (0.981–1.022)	0.000 (−0.021–0.024)

Table 3: (continued)

Analyte	Lab	CAL1				CAL3			
		n	r (95CI)	Slope (95CI)	Intercept (95CI)	n	r (95CI)	Slope (95CI)	Intercept (95CI)
Aldosterone	D	77	0.981 (0.970–0.988)	1.118 (1.052–1.172)	0.007 (−0.004–0.015)	77	0.987 (0.980–0.992)	0.936 (0.909–0.969)	0.011 (0.005–0.018)
	E	73	0.977 (0.964–0.986)	1.000 (1.000–1.000)	0.010 (0.010–0.010)	73	0.980 (0.968–0.987)	1.077 (1.039–1.125)	−0.010 (−0.019–−0.004)
	G	77	0.986 (0.978–0.991)	0.846 (0.821–0.870)	0.005 (0.003–0.010)	77	0.986 (0.979–0.991)	1.000 (1.000–1.000)	0.000 (0.000–0.000)
	H	78	0.979 (0.968–0.987)	0.969 (0.909–1.000)	−0.006 (−0.010–0.005)	77	0.980 (0.969–0.987)	0.875 (0.829–0.933)	0.011 (0.001–0.017)
	L	43	0.976 (0.955–0.987)	1.010 (1.000–1.115)	−0.003 (−0.030–0.000)	43	0.975 (0.954–0.986)	1.080 (1.000–1.143)	−0.013 (−0.029–0.010)

CAL1, *in-house* calibration; CAL3, Chromsystems calibration.

but worsened it in Laboratory H (CAL1: 0.969 (0.909–1.000); CAL3: 0.875 (0.829–0.933)) (Table 3).

Using CAL1, median bias vs. median of all methods ranged from −12.0 to 16.8%, with almost all cases within the TAE. CAL3 significantly reduced bias median, ranging from −8.8 to 3.4%, in three laboratories and variance in one (Figure 3, Supplemental Table 7 and Supplemental Figure 6).

Discussion

This is the first study examining the variability among LC-MS/MS measurements for circulating cortisol, 17OH-progesterone and aldosterone. Methods were validated according to recommended guidelines, most were published [2, 12, 30–37], and exhibited considerable technical heterogeneity. Within- and between-laboratory performances were interpreted by means of allowable imprecision, bias and total error, which, however, are still derived from immunoassay-based biological variability studies [27]. Intra-method imprecision for cortisol was within MAI for all, but unsatisfactory for 17OH-progesterone and aldosterone in some laboratories, typically worsening at lower concentrations. Trueness and bias in EQA materials for aldosterone were unsatisfactory in three laboratories, but mostly acceptable for cortisol and 17OH-progesterone. Exceptions were Laboratory D for cortisol, due to a miscalibration, and, for 17OH-progesterone, Laboratory G, showing a constant bias, and Laboratory H, showing a severe commutability problem in HM materials. Hence, the study allowed the discovery and correction of some laboratory-specific defects, such as in Laboratory D, where the incorrect dilution

of newly utilized commercial calibrators in place of previously used *in-house* calibrators caused large cortisol overestimation. These data underline the importance of promoting uniform protocols for method validation, performance reporting and monitoring beyond the validation stage [24, 25, 38]. Findings also implies that LC-MS/MS is not immune from commutability issues. Testing EQA materials from different providers could assist identification of potential commutability problems. Moreover, our data about material N°304 low suggests that assigning target values as all methods means/median may confound method evaluation. Therefore, application of RMPs should be encouraged among EQA providers [22].

Current guidelines for clinical LC-MS/MS assays recommend calibrators being prepared in the same matrix as samples to be tested [39]. Some laboratories used steroid-depleted serum/plasma processed by charcoal-stripping. Others used surrogate matrices (bovine serum albumin or phosphate buffer solutions) or solvents. Although the adequacy of these alternative matrices in comparison with the native matrix has been tested by each of the participant, we cannot exclude that *in-house* calibrator commutability is contributing to the overall inter-laboratory variability.

Despite the aforementioned heterogeneities, between-method performance was substantially within specifications, indicating impressive consistency of the LC-MS/MS methods under investigation. When evaluating the commercial calibrators, we found that the BSN highest cortisol calibrator was above all the *in-house* measurement ranges, while Chromsystems' was too low, preventing the measurement in one woman taking estroprogestin and in two EQA materials. Moreover, BSN calibration did not cover the
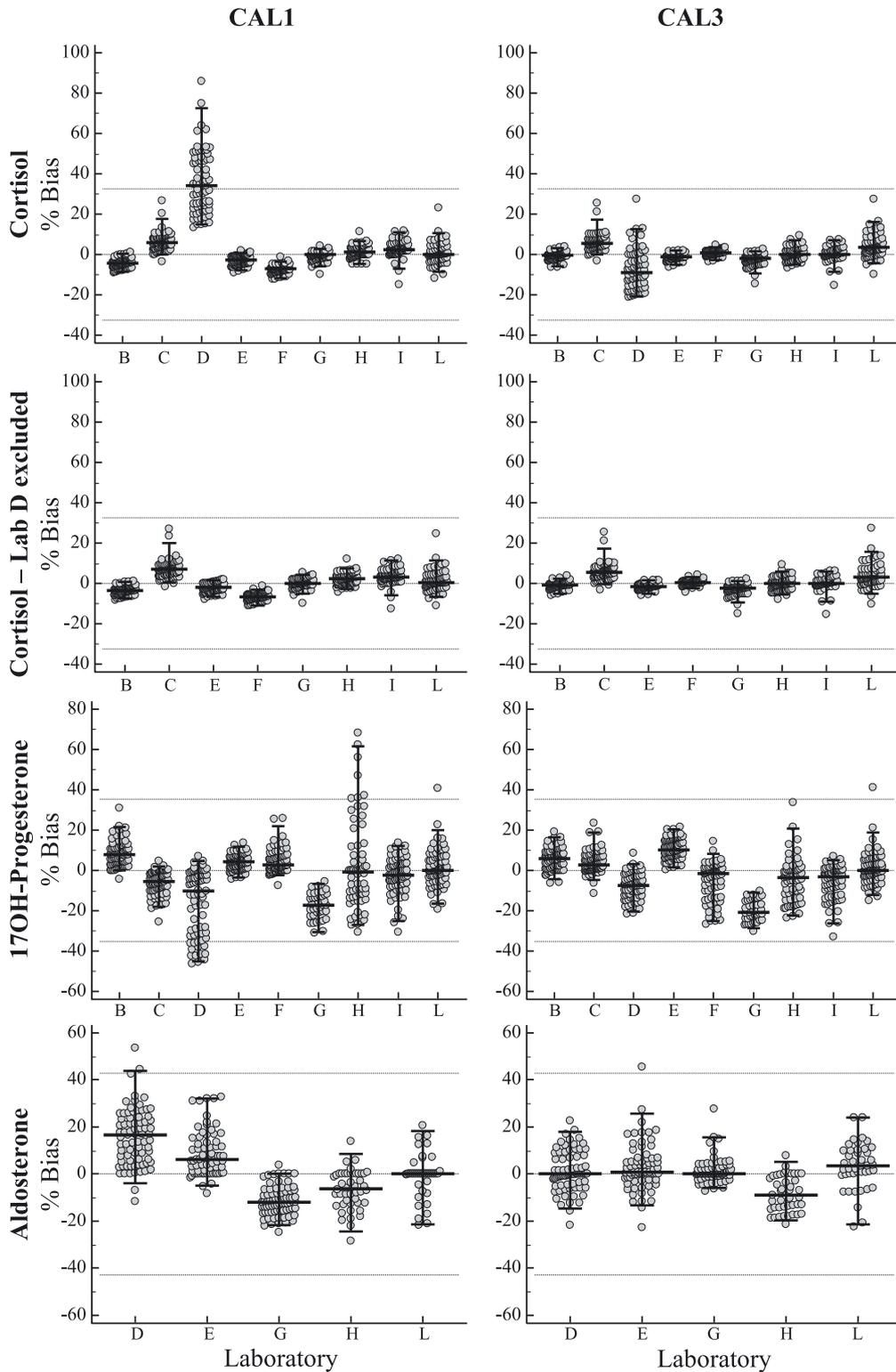
Figure 3: Laboratories % bias vs. median of all laboratories as function of the calibration system.
% Bias = ((laboratory value − median of all laboratories)/median of all laboratories) × 100. Segments, median; error bars, 2.5 and 97.5 centiles; horizontal lines, zero ± total allowable error (cortisol: 32.5%; 17OH-progesterone: 35.3%; aldosterone: 42.8%). CAL1, *in-house* calibration; CAL3, Chromsystems calibration.

low-physiologic concentrations of 17OH-progesterone and aldosterone. Therefore, commercial calibrators need to be adapted to assay features and to reporting purposes of laboratories. Compared to *in-house*, BSN calibration seemed to worse methods' trueness for cortisol and 17OH-progesterone. This may derive from BSN calibrators being traceable to EQA programs, such as RCPAQAP, and not to reference materials. At variance, Chromsystems calibrators are traceable to NIST and CRMs, and showed a good overlap with most of the *in-house* calibrations. Unifying calibration using Chromsystems' set improved between-laboratory performance, particularly reducing the systematic bias at low 17OH-progesterone levels.

Laboratory L, the only laboratory using a commercial kit, exhibited high intra-laboratory imprecision. The possible reason involved the need for more frequent cleaning of the ionization source. Therefore, laboratory experience in instrument maintenance and method monitoring is recommended even when using commercial kits. Four laboratories used Chromsystems' as *in-house* calibrators. A part from the procedural problem of Laboratory D, modest deviations between *in-house* and external calibrations were found in some cases, which may be procedural or due to lot-to-lot variability.

Our study supports the use of common calibrators for improving the harmonization of measurements. However, we showed that commercial calibrators or kits are not necessarily superior to *in-house* procedures and do not guarantee adequate performance *per se*. Conversely, setting more stringent requirements and standardized reporting protocols for analytical performance should be encouraged for improving research quality and clinical effectiveness. This is particularly relevant in view of the EU-IVDR, effective in May 2022, promoting commercial devices over equivalent LDTs [26].

Non-negligible residual between-laboratory variability was observed when unifying calibration, implying that other components have to be considered. Monitoring the qualifier ions is often ineffective in detecting interferences among isobars. Therefore, LC resolution is critical in ensuring specificity. Various 17OH-progesterone isobars circulate at relevant levels [14]. All methods in this study achieved baseline separation of 17OH-progesterone from 11-deoxycorticosterone. 17OH-progesterone separation from 16OH-progesterone has been verified by Laboratories B, C and L, and from 11OH-progesterone by Laboratories B, I and L. Interference from these or other isobars may explain the overestimation in HM materials by Laboratory H.

Although multiple isotopes were used as IS, laboratories using the same IS did not show a better between-laboratory performance. Nonetheless, LC and IS contributions need to

be directly addressed in purposely designed studies. Notably, Loh et al. recently reported that deuterium and ^{13}C-based IS provided similar 17OH-progesterone results by three LC-MS/MS methods [15].

Our design included many laboratories and duplicate measurements of individual samples. Moreover, three different tubes were used, whose impact on measurements will be investigated in a dedicated study. While these aspects reinforced the robustness of findings, they required large blood volumes, which are not easily obtained from patients. Consequently, we recruited a suboptimal number of 26 volunteers [40] mostly showing normal steroid concentrations. Further studies are needed to assess harmonization at high or low concentrations typical of hypercortisolism, hyperaldosteronism, adrenal insufficiency or functional tests. Nevertheless, reproducible measurement of physiological levels is relevant for therapeutic monitoring and for emerging multi-analyte approaches to diagnosis, subtyping and risk stratification of complex conditions, such as female hyperandrogenism, endocrine tumors and non-communicable diseases [41]. Finally, our study was designed as a ring trial. Studies including target values achieved by RMPs or GC-MS methods are needed to address standardization, which is mandatory for establishing laboratory-independent reference intervals and decision limits.

Our study described pitfalls in method performance and EQA materials, as well as advantages and limitations of calibration materials. Unifying the calibration could significantly reduce between-laboratory variability, while adopting traceable calibrators and RMP-determined EQA materials could ease standardization. Residual disagreement requires investigation. Evidence generated by our study supports LC-MS/MS utility for achieving steroid harmonization.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission. FF conceived, designed and coordinated the study, performed the statistical analysis and wrote the manuscript; MC, BGK, MV and UP

contributed to the study design; MC, MM, JML, MP, JMH, SB, MTA, JVDO and DK carried out sample measurement and data exports; AT coordinated subject recruitment and sample management; EN performed the statistical analysis; UP conceived the study; MC, MP, JMH, SB, PAB, MTA, ACH, JVDO, FM, MR, GE, BGK, MV and UP contributed in result interpretation and in writing the manuscript.

Competing interests: Authors state no conflict of interest.

Informed consent: Informed consent was obtained from all individuals included in this study.

Ethical approval: The local Institutional Review Board deemed the study exempt from review. The Bologna Ethics Committee approved the study n° 141/2017/U/Tess. All volunteers participating in the study signed their informed consent.

References

1. Olesti E, Boccard J, Visconti G, González-Ruiz V, Rudaz S. From a single steroid to the steroidome: trends and analytical challenges. J Steroid Biochem Mol Biol 2021;206:105797.

2. Fanelli F, Belluomo I, Di Lallo VD, Cuomo G, De Iasio R, Baccini M, et al. Serum steroid profiling by isotopic dilution-liquid chromatography-mass spectrometry: comparison with current immunoassays and reference intervals in healthy adults. Steroids 2011;76:244–53.

3. Ray JA, Kushnir MM, Palmer J, Sadjadi S, Rockwood AL, Meikle AW. Enhancement of specificity of aldosterone measurement in human serum and plasma using 2D-LC-MS/MS and comparison with commercial immunoassays. J Chromatogr B 2014;970: 102–7.

4. Hawley JM, Owen LJ, Lockhart SJ, Monaghan PJ, Armston A, Chadwick CA, et al. Serum cortisol: an up-to-date assessment of routine assay performance. Clin Chem 2016;62:1220–9.

5. Hamer HM, MJJ Finken, van Herwaarden AE, du Toit T, Swart AC, Heijboer AC. Falsely elevated plasma testosterone concentrations in neonates: importance of LC-MS/MS measurements. Clin Chem Lab Med 2018;56:e141–e3.

6. Thienpont LM, Van Uytfanghe K, Blincko S, Ramsay CS, Xie H, Doss RC, et al. State-of-the-art of serum testosterone measurement by isotope dilution-liquid chromatography-tandem mass spectrometry. Clin Chem 2008;54:1290–7.

7. Yates AM, Bowron A, Calton L, Heynes J, Field H, Rainbow S, et al. Interlaboratory variation in 25-hydroxyvitamin D2 and 25-hydroxyvitamin D3 is significantly improved if common calibration material is used. Clin Chem 2008;54:2082–4.

8. Vesper HW, Bhasin S, Wang C, Tai SS, Dodge LA, Singh RJ, et al. Interlaboratory comparison study of serum total testosterone [corrected] measurements performed by mass spectrometry methods. Steroids 2009;74:498–503.

9. Owen LJ, MacDonald PR, Keevil BG. Is calibration the cause of variation in liquid chromatography tandem mass spectrometry testosterone measurement? Ann Clin Biochem 2013;50: 368–70.

10. Vesper HW, Botelho JC, Vidal ML, Rahmani Y, Thienpont LM, Caudill SP. High variability in serum estradiol measurements in men and women. Steroids 2014;82:7–13.

11. Büttler RM, Martens F, Fanelli F, Pham HT, Kushnir MM, Janssen MJ, et al. Comparison of 7 published LC-MS/MS methods for the simultaneous measurement of testosterone, androstenedione, and dehydroepiandrosterone in serum. Clin Chem 2015;61: 1475–83.

12. Büttler RM, Martens F, Ackermans MT, Davison AS, van Herwaarden AE, Kortz L, et al. Comparison of eight routine unpublished LC-MS/MS methods for the simultaneous measurement of testosterone and androstenedione in serum. Clin Chim Acta 2016;454:112–8.

13. Dirks NF, Vesper HW, van Herwaarden AE, van den Ouweland JM, Kema IP, Krabbe JG, et al. Various calibration procedures result in optimal standardization of routinely used 25(OH)D ID-LC-MS/MS methods. Clin Chim Acta 2016;462:49–54.

14. Greaves RF, Ho CS, Loh TP, Chai JH, Jolly L, Graham P, et al. Working group 3 "harmonisation of laboratory assessment" European Cooperation in Science and Technology (COST) action BM1303 "DSDnet". Current state and recommendations for harmonization of serum/plasma 17-hydroxyprogesterone mass spectrometry methods. Clin Chem Lab Med 2018;56: 1685–97.

15. Loh TP, Ho CS, Hartmann MF, Zakaria R, Lo CWS, van den Berg S, et al. Influence of isotopically labeled internal standards on quantification of serum/plasma 17alpha-hydroxyprogesterone (17OHP) by liquid chromatography mass spectrometry. Clin Chem Lab Med 2020;58:1731–9.

16. Owen LJ, Keevil BG. Testosterone measurement by liquid chromatography tandem mass spectrometry: the importance of internal standard choice. Ann Clin Biochem 2012;49:600–2.

17. Huang M, Cadwallader AB, Heltsley R. Mechanism of error caused by isotope-labeled internal standard: accurate method for simultaneous measurement of vitamin D and pre-vitamin D by liquid chromatography/tandem mass spectrometry. Rapid Commun Mass Spectrom 2014;28:2101–10.

18. Hepburn S, Wright MJP, Boyder C, Sahertian RC, Lu B, Zhang R, et al. Sex steroid hormone stability in serum tubes with and without separator gels. Clin Chem Lab Med 2016;54:1451–9.

19. Greaves RF. The central role of external quality assurance in harmonisation and standardisation for laboratory medicine. Clin Chem Lab Med 2017;55:471–3.

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

21. International Consortium for Harmonization of Clinical Laboratory Results. https://www.harmonization.net [Accessed Feb 2021].

22. Joint Committee for Traceability in Laboratory Medicine (JCTLM). Database of higher-order reference materials, measurement methods/procedures and services. https://www.bipm.org/jctlm/ [Accessed Feb 2021].

23. Miller GW, Myers GL, Lou Gantzer M, Kahn SE, Schönbrunner ER, Thienpont LM, et al. Roadmap for harmonization of clinical laboratory measurement procedures. Clin Chem 2011;57: 1108–17.

24. Vogeser M, Schuster C, Rockwood AL. A proposal to standardize the description of LC–MS-based measurement methods in laboratory medicine [Editorial]. Clin Mass Spectrom 2019;13: 36–8.

25. Dirks NF, Ackermans MT, Martens F, Cobbaert CM, de Jonge R, Heijboer AC. We need to talk about the analytical performance of our laboratory developed clinical LC-MS/MS tests, and start separating the wheat from the chaff. Clin Chim Acta 2021;514: 80–3.

26. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. Official Journal of the European Union.

27. The European Federation of Clinical Chemistry and Laboratory Medicine database. https://biologicalvariation.eu/meta_calculations [Accessed Feb 2021].

28. Oosterhuis WP, Bayat H, Armbruster D, Coskun A, Freeman KP, Kallner A, et al. The use of error and uncertainty methods in the medical laboratory. Clin Chem Lab Med 2018;56:209–19.

29. Phinney KW, Sempos CT, Tai SS, Camara JE, Wise SA, Eckfeldt JH, et al. Baseline assessment of 25-hydroxyvitamin D reference material and proficiency testing/external quality assurance material commutability: a vitamin D standardization program study. J AOAC Int 2017;100:1288–93.

30. Bruce SJ, Rey F, Béguin A, Berthod C, Werner D, Henry H. Discrepancy between radioimmunoassay and high performance liquid chromatography tandem-mass spectrometry for the analysis of androstenedione. Anal Biochem 2014;455:20–5.

31. Peitzsch M, Dekkers T, Haase M, Sweep FC, Quack I, Antoch G, et al. An LC-MS/MS method for steroid profiling during adrenal venous sampling for investigation of primary aldosteronism. J Steroid Biochem Mol Biol 2015;145:75–84.

32. Fahlbusch FB, Heussner K, Schmid M, Schild R, Ruebner M, Huebner H, et al. Measurement of amniotic fluid steroids of midgestation via LC-MS/MS. J Steroid Biochem Mol Biol 2015; 152:155–60.

33. Bekkach Y, Heijboer AC, Endert E, Ackermans MT. Determination of urinary aldosterone using a plasma aldosterone 2D ID LC-MS/MS method. Bioanalysis 2016;8:1765–75.

34. Owen LJ, Adaway JE, Davies S, Neale S, El-Farhan N, Ducroq D, et al. Development of a rapid assay for the analysis of serum cortisol and its implementation into a routine Service laboratory. Ann Clin Biochem 2013;50:345–52.

35. Lindner JM, Vogeser M, Grimm SH. Biphenyl based stationary phases for improved selectivity in complex steroid assays. J Pharm Biomed Anal 2017;142:66–73.

36. Hawley JM, Adaway JE, Owen LJ, Keevil BG. Development of a total serum testosterone, androstenedione, 17-hydroxyprogesterone, 11β-hydroxyandrostenedione and 11-ketotestosterone LC-MS/MS assay and its application to evaluate pre-analytical sample stability. Clin Chem Lab Med 2020;58:741–52.

37. Laszlo CF, Montoya JP, Shamseddin M, De Martino F, Beguin A, Nellen R, et al. A high resolution LC–MS targeted method for the concomitant analysis of 11 contraceptive progestins and 4 steroids. J Pharm Biomed Anal 2019;175:112756.

38. Vogeser M, Stone JA. A suggested standard for validation of LC-MS/MS based analytical series in diagnostic laboratories. Clin Mass Spectr 2020;16:25–32.

39. Liquid Chromatorgraphy-Mass Spectrometry Methods; Approved Guideline. CLSI C62 Revision A. Wayne, PA USA: Clinical and Laboratory Standard Institute; 2014.

40. International Consortium for Harmonization of Clinical Laboratory Results. Toolbox of technical procedures to be considered when developing a process to achieve harmonization for a measurand; 2013. www.harmonization.net.

41. Eisenhofer G, Masjkur J, Peitzsch M, Di Dalmazi G, Bidlingmaier M, Grüber M, et al. Plasma steroid metabolome profiling for diagnosis and subtyping patients with cushing syndrome. Clin Chem 2018;64:586–96.