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

Report from the HarmoSter study: different LC-MS/MS androstenedione, DHEAS and testosterone methods compare well; however, unifying calibration is a double-edged sword

https://doi.org/10.1515/cclm-2023-1138
Received October 12, 2023; accepted December 26, 2023; published online January 12, 2024

Abstract

Objectives: Current liquid chromatography-tandem mass spectrometry (LC-MS/MS) applications for circulating androgen measurements are technically diverse. Previously, variable results have been reported for testosterone. Data are scarce for androstenedione and absent for dehydroepiandrosterone sulfate (DHEAS). We assessed the agreement of androstenedione, DHEAS and testosterone LC-MS/MS measurements among nine European centers and explored benefits of calibration system unification.

Methods: Androgens were measured twice by laboratory-specific procedures in 78 patient samples and in EQA materials. Results were obtained by *in-house* and external calibration. Intra- and inter-laboratory performances were valued.

Results: Intra-laboratory CVs ranged between 4.2–13.2 % for androstenedione, 1.6–10.8 % for DHEAS, and 4.3–8.7 % and 2.6–7.1 % for female and male testosterone, respectively. Bias and trueness in EQA materials were within ±20 %. Median inter-laboratory CV with *in-house* vs. external calibration were 12.0 vs. 9.6 % for androstenedione ($p<0.001$), 7.2 vs. 4.9 % for DHEAS ($p<0.001$), 6.4 vs. 7.6 % for female testosterone ($p<0.001$) and 6.8 and 7.4 % for male testosterone ($p=0.111$). Median bias vs. all laboratory median with *in-house* and external calibration were −13.3 to 20.5 % and −4.9 to 18.7 % for androstenedione, −10.9 to 4.8 % and −3.4 to 3.5 % for DHEAS, −2.7 to 6.5 % and −11.3 to 6.6 % for testosterone in females, and −7.0 to 8.5 % and −7.5 to 11.8 % for testosterone in males, respectively.

***Corresponding author: Flaminia Fanelli**, Senior Assistant Professor, Department of Medical and Surgical Sciences, Endocrinology Research Group, Center for Applied Biomedical Research, Alma Mater Studiorum University of Bologna, Via Massarenti 9, 40138 Bologna, Italy, Phone: +39 051 2143902, Fax: +39 051 2143902, E-mail: flaminia.fanelli2@unibo.it. https://orcid.org/0000-0003-2601-2694

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)

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)

Marco Cantù, Laboratory of Clinical Biochemistry and Pharmacology, Institute of Laboratory Medicine EOLAB, Ente Ospedaliero Cantonale, Bellinzona, Switzerland

Anastasia Temchenko and Marco Mezzullo, Department of Medical and Surgical Sciences, Endocrinology Research Group, Center for Applied Biomedical Research, Alma Mater Studiorum University of Bologna, 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

James M. Hawley and Brian G. Keevil, Department of Clinical Biochemistry, University Hospital South Manchester, Manchester NHS Foundation Trust, Manchester, UK

Mariette T. Ackermans, Faculty of Science, Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, Netherlands

Jody Van den Ouweland, Department of Clinical Chemistry, Canisius-Wilhelmina Hospital, Nijmegen, Netherlands

Daniel Koeppl and Manfred Rauh, Department of Pediatrics and Adolescent Medicine, University Hospital, Erlangen, Germany

Elena Nardi, Department of Medical and Surgical Sciences, Alma Mater Studiorum University of Bologna, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy

Finlay MacKenzie, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

Annemieke C. Heijboer, Endocrine Laboratory, Department of Clinical Chemistry, Amsterdam UMC, University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Uberto Pagotto, Department of Medical and Surgical Sciences, Endocrinology Research Group, Center for Applied Biomedical Research, Alma Mater Studiorum University of Bologna, Bologna, Italy; and Unit of Endocrinology and Prevention and Care of Diabetes, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy

Conclusions: Methods showed high intra-laboratory precision but variable bias and trueness. Inter-laboratory agreement was remarkably good. Calibration system unification improved agreement in androstenedione and DHEAS, but not in testosterone measurements. Multiple components, such as commutability of calibrators and EQA materials and internal standard choices, likely contribute to inter-laboratory variability.

Keywords: liquid chromatography-tandem mass spectrometry; harmonization; androstenedione; dehydroepiandrosterone sulfate; testosterone; inter-laboratory performance

Introduction

Circulating testosterone, androstenedione and dehydroepiandrosterone sulfate (DHEAS) are routinely evaluated to characterize the extent and the nature of androgen imbalance in diseases that include gonadal or adrenal insufficiency, tumors, genetic enzyme deficits, Cushing syndrome and polycystic ovary syndrome in women [1, 2]. Clinically relevant levels of these hormones span both well above and below the normal range. Moreover, therapies that impact androgen availability require monitoring for changes over time, even within the normal range. Consequently, androgen measurement requires optimal precision and accuracy over a wide interval of concentrations. Another requirement is to effectively distinguish the analyte from other steroidal species sharing similar structure and properties [3].

By virtue of intrinsic high specificity and sensitivity, liquid chromatography-tandem mass spectrometry (LC-MS/MS) is today implemented in routine and research laboratories for steroid measurements via laboratory developed tests (LDT) or commercial kits [3, 4]. Participation of LC-MS/MS users in External Quality Assessment (EQA) programs is increasing [4, 5]. Nevertheless, reproducibility among LC-MS/MS measurements in complex patient samples remains unclear [4, 6]. So far, only two studies analyzed the comparability of LC-MS/MS methods for androstenedione [7, 8], while none have investigated DHEAS. A few studies on testosterone demonstrated that LC-MS/MS measurements exhibit superior consistency compared to immunoassays [7–12]. Nonetheless, these studies reported diverse degrees of agreement between LC-MS/MS methods. Unfortunately, no single cause has yet been identified for persistent variability [11, 13].

The HarmoSter consortium was created with the primary objective to investigate the harmonization status of LC-MS/MS measurements of 10 circulating steroids by nine European centers (Supplementary Table 1). Methods were tested on different sample matrices and calibration systems. Results for cortisol, 17OH-progesterone, aldosterone, corticosterone, 11-deoxycortisol and cortisone were recently published [14, 15], showing acceptable consistency among methods. Furthermore, we highlighted that the strategy of unifying calibration may sometimes improve overall inter-laboratory agreement, but not always [14, 15].

In this new study, we focus on LC-MS/MS measurements of androstenedione, DHEAS and testosterone in plasma and serum samples from patients and in EQA materials to assess the intra- and inter-laboratory performance, as well as any impact of unification of calibration systems.

Materials and methods

Consortium and methods

The HarmoSter study was approved by Bologna Ethics Committee (no. 141/2017/U/Tess) [14]. Nine laboratories with 10 LDTs overall (Laboratories B to I) and the MassChrom® kit (Chromsystems; Munich, Germany; https://chromsystems.com) (Laboratory L) participated. All measured testosterone; eight measured androstenedione and seven measured DHEAS (Table 1 and Supplementary Table 1). 6PLUS1® Multilevel Serum Calibrator set (Chromsystems) was used for *in-house* calibration by Laboratories D, E and F. Technical details and *in-house* measurement ranges are shown in Table 1 and Supplementary Table 1.

Study samples

As previously detailed [14], the sample set included 78 serum and plasma samples from 26 volunteers (women/men: 13/13; age: 20–69 years) via gel-separator, bead clot activator and lithium-heparin containing tubes. The 6PLUS1® Multilevel Serum Calibrator set (Chromsystems; lot.5016, different from lots used for *in-house* calibration by Laboratories D, E, F and L) was measured within the study batch and used as the external calibration set (androstenedione: 0.64–50.6 nmol/L; DHEAS: 280–15,460 nmol/L; and testosterone: 0.173–39.9 nmol/L). Chromsystems' calibrators were traceable to methanol certified reference materials for androstenedione and DHEAS, and to NIST SRM 972 for testosterone. EQA materials evaluated in this study included the Reference Institute for Bioanalytics (RfB; Bonn, Germany; www.rfb.bio) HM40121, HM40122, HM40123 and HM40124 (lyophilized human recalcified plasma spiked with steroids, no preservatives), assigned with target values determined by reference measurement procedure (RMP) for testosterone and by mean of all MS-methods for DHEAS; the United Kingdom National External Quality Assessment Service (UKNEQAS; Birmingham, UK; https://ukneqas.org.uk) materials (liquid off-the-clot minimally manipulated human serum) A382 (spiked with 10 nmol/L androstenedione standard) with androstenedione target assigned as mean of all MS-methods; the Instand e.V. (Düsseldorf, Germany; https://www.instand-ev.de) material pair N°302 (liquid human serum spiked with steroids, no additives) assigned with RMP-determined testosterone target value.

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

Lab/ method	Ref	*In-house* calibration	Sample volume, µL	Sample preparation	Instrument	Run time, min	Purification column	LC column; temperature	Mobile phases (A; B)	Source
B	[20]	Cerilliant CRM[a] in 4 % BSA	600[a]	PP: $ZnSO_4$ 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	[21]	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	[22]	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	[23]	6PLUS1®[a,b], Chromsystems	100	PP: $ZnSO_4$ in MeOH	1,260 Infinity, Agilent; API6500 Qtrap, Sciex[a]	7.0	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]
F	[24]	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.1	[26]	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 H_2O; 0.1 % FA in ACN	ESI
G.2	[25]	Gravimetric in ethanol	25	PP: ACN	Acquity-Xevo TQ-S Micro, Waters	4.7		Kinetex C18 75 × 2.1 mm, 2.6 µm	2.4 mM NH_4HCO_2 in H_2O; ACN	ESI
H	n.a.	Cerilliant CRM in 10 % MeOH	1,000	PP: $ZnSO_4$ 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.1	[27, 28]	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
I.2	[27, 28]	Gravimetric in charcoal stripped serum	100	PP: H_3PO_4; SPE: Oasis MCX	I-class Acquity; TQS, Waters[a]	5.0		C18 Zorbax eclipse plus 2.1 × 50 mm, 1.8 µm[a]	H_2O; 0.0125 % NH_4OH in ACN	
L	n.a., [c]	6PLUS1®[b], Chromsystems	500	Chromsystems' intellectual property	1,290 series HPLC; 6,490, Agilent	11.5		Chromsystems' proprietary column	Chromsystems' intellectual property	ESI

Lab/ method	Androstenedione			DHEAS			Testosterone		
	Analyte (ion mode) mass transition; retention time, min	IS	IS (ion mode) mass transition; retention time, min	Analyte (ion mode) mass transition; retention time, min	IS	IS (ion mode) mass transition; retention time, min	Analyte (ion mode) mass transition; retention time, min	IS	IS (ion mode) mass transition; retention time, min
B	(+) 287/97; 11.40	Androstenedione-[2,2,4,6,6-D5][a]	(+) 292/100; 11.06				(+) 289/97; 11.73	Testosterone-[3,4-^{13}C2]	(+) 291/111; 11.73
C				(−) 367/97; 1.68	DHEAS-[2,2,3,4,4-D5]	(−) 372/98; 1.66	(+) 289/97; 4.80	Testosterone-[2,3,4-^{13}C3]	(−) 292/97; 4.80
D	(+) 287/97; 6.90	Androstenedione-[2,2,4,6,6,16,16-D7]	(+) 294/100; 6.85	(+) 271/255; 5.45	DHEAS-[16,16-D2]	(+) 273/255; 5.43	(+) 289/97; 7.35	Testosterone-[2,2,4,6,6-D5]	(+) 294/100; 7.32
E	(+) 287/97; 4.62	17-OH-progesterone [2,3,4-^{13}C3]	(+) 334/112; 4.96	(+) 271/213; 4.07	DHEAS-[2,2,3,4,4-D5]	(+) 276/218; 4.04	(+) 289/97; 4.87	17-OH-progesterone [2,3,4-^{13}C3]	(+) 334/112; 4.96
F	(+) 287/97; 2.30	Androstenedione-[2,3,4-^{13}C3]	(+) 290/100; 2.30				(+) 289/97; 2.00	Testosterone-[2,3,4-^{13}C3]	(+) 292/100; 2.00
G.1	(+) 287/97; 7.17	Androstenedione-[2,2,4,6,6,16,16-D7]	(+) 294/100; 7.11				(+) 289/97; 6.56	Testosterone-[16,16,17-D3]	(+) 292/97; 6.53
G.2				(−) 367/97; 1.82	DHEAS-[2,3,4-^{13}C3][a]	(−) 373/98; 1.82			
H	(+) 287/97; 6.53	Testosterone-[16,16,17-D3]	(+) 292/97, 6.95	(+) 271/213; 7.32	DHEAS-[2,2,3,4,4-D5]	(+) 276/218; 7.29	(+) 289/97; 6.97	Testosterone-[16,16,17-D3]	(+) 292/97; 6.95
I.1	(+) 287/97; 5.70	Androstenedione-[2,2,4,6,6-D5]	(+) 292/100; 5.60				(+) 289/97; 6.30	Testosterone-[16,16,17-D3]	(+) 292/97; 6.20
I.2				(−) 367/97; 2.00	DHEAS-[2,2,3,4,4-D5]	(−) 372/98; 2.00			
L	(+) 287/109; 7.49	Chromsystems' intellectual property	(+) 290/112; 7.48	(+) 369/99; 3.00	Chromsystems' intellectual property	(+) 375/100; 2.93	(+) 289/97; 5.95	Chromsystems' intellectual property	(+) 292/97; 5.89

[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. IS, internal standard; CRM, certified reference material; PP, protein precipitation; $ZnSO_4$, zinc sulfate; MeOH, methanol; SPE, solid phase extraction; RP, reversed phase; ZnTFA, zinc trifluoroacetate; HLB, hydrophilic-lipophilic balance; SLE, supported liquid extraction; FA, formic acid; ACN, acetonitrile; HSS, high strength silica; BEH, ethylene bridged hybrid; H_3PO_4, phosphoric acid; MCX, mixed-mode strong cation-exchange; n.a., not applicable.

Running scheme and quantitation

Patient and EQA samples and external calibrators were measured in duplicate in two independent runs according to the protocols of each laboratory (Table 1). Each batch was quantified both by using *in-house* and external calibration sets. All calibration curves displayed R^2>0.97.

Data analysis and statistics

Androstenedione, DHEAS and testosterone values are reported in nmol/L; to convert to ng/mL, multiply by 0.286, 0.368 and 0.288, respectively. Results were excluded if below the lower limit of quantification (LLOQ) or above the upper LOQ (ULOQ) of the *in-house* measurement range. Data were also excluded from the external calibration dataset when outside of the measurement range. Means and CVs of duplicate measurements were calculated (Supplementary Tables 3–6). Within – (CVi) and between – (CVg) subjects biological variabilities [16, 17] were used to assess the maximum allowable imprecision (MAI) and bias (MAB) and total allowable error (TAE) via the following equations [18]:

$$\mathrm{MAI} = 0.5 \times \mathrm{CVi};$$

$$\mathrm{MAB} = 0.25 \times \left(\mathrm{CVi}^2 + \mathrm{Cvg}^2\right)^{0.5};$$

$$\mathrm{TAE} = 0.25 \times \left(\mathrm{CVi}^2 + \mathrm{CVg}^2\right)^{0.5} + 1.65 \times (0.5 \times \mathrm{CVi});$$

(androstenedione: 7.9, 10.5 and 23.5 %; DHEAS: 6.7, 8.1 and 19.0 %; testosterone: 6.3, 6.2 and 16.5 %; respectively) [16, 17].

Intra-laboratory performance: The overall intra-laboratory imprecision was determined as the duplicate measurement CV, as follows: $\mathrm{CV} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} D_i^2} \times 100$, where $D_i = \frac{(x_{i1}-x_{i2})/\sqrt{2}}{\overline{X}_i}$, x_{i1} and x_{i2} are duplicate measures of each sample, $\overline{X}_i$ is the arithmetic mean of the duplicates and n is the total number of duplicates [19]. Intra-laboratory CVs from each laboratory were then compared with the MAI. The impact of calibration within each method was evaluated using the Friedman test. Trueness and bias were estimated in EQA materials as % difference from target values (RMP-determined or mean of all MS methods in the survey, respectively) and compared with the MAB. In addition, the bias of laboratory measures of EQA materials and patient samples vs. all laboratory median were determined. For each laboratory, the Bland-Altman 95 % limits of agreement were calculated for the patient sample pool. Commutability of EQA materials for each method was considered acceptable when the % bias in EQA material measures was within the above agreement interval.

Inter-laboratory performance: Between-method reproducibility, valued by the inter-laboratory CV, was compared with the MAI. Passing-Bablok and Bland-Altman analyses were performed. The % bias was compared with the TAE. Wilcoxon and F tests were used to compare inter-laboratory CV, median bias and bias variance by *in-house* vs. external calibration. The Friedman test was used to compare the measurements between laboratories. Statistics were performed by SPSS (v.20, IBM Co., Somers, NY) and MedCalc (v.18.2.1; Mariakerke, Belgium).

Results

Androstenedione

Patient values ranged from 0.88 to 5.74 nmol/L (all laboratory median by *in-house* calibration) (Supplementary Table 3). The intra-laboratory CVs ranged from 4.2 to 13.2 %, which included six laboratories within the MAI (Supplementary Table 2). Cases with duplicate CVs exceeding the MAI were up to 65.4 % for Laboratory H (Supplementary Table 3). Replacing *in-house* with external calibration determined modest deviations of calculated concentrations, except for Laboratory D (–9.7 %) and H (–21.7 %) (Supplementary Table 2). Laboratories D, H and E showed a positive bias in the A382 material (target value: 10.35 nmol/L), which was corrected by using external calibration in Laboratories D and H but not in E (Figure 1). Median inter-laboratory CV was 12.0 % with *in-house*, and lowered to 9.6 % with external calibration (p<0.001). Most of the cases displayed an inter-laboratory CV above the MAI with either calibrations (Figure 2 and Supplementary Table 7). By Passing-Bablok analysis of individual laboratory measures vs. all laboratory median (Supplementary Figure 1 and Table 2), seven laboratories showed a significant slope coefficient (range [90 CI]: 0.851 [0.811–0.891] to 1.186 [1.133–1.252]). Using external calibration improved the slope in all laboratories, except E. By Bland-Altman analysis (Figure 3, Supplementary Table 8 and Supplementary Figure 2), median bias vs. all laboratory median ranged from –13.3 to 20.5 % using *in-house* calibration. Laboratories E and H showed the largest variance of bias (95.3 and 123.2 %, respectively), with several cases exceeding the TAE. When using external calibration, median bias reduced to within ±5 % in all except Laboratory E, whose median bias almost doubled to 18.7 % (p<0.001) (Figure 3, Supplementary Table 8 and Supplementary Figure 2). Supplementary Table 9 shows that androstenedione measurements by *in-house* calibration were not comparable between any laboratories, while, with external calibration, some laboratories showed non different measures.

DHEAS

Patient samples ranged from 541 to 14,372 nmol/L (all laboratory median by *in-house* calibration) (Supplementary Table 4). With *in-house* calibration, values were above the ULOQ of Laboratory D and C in one and two samples, respectively, and below the LLOQ of Laboratory E in nine samples. With external calibration, three samples were below the measurement range in Laboratory G. The intra-laboratory CV ranged from 1.6 to 10.8 % (Supplementary Table 2), within the MAI in all except Laboratory L, showing up to 73.1 % cases with duplicate CV above the threshold (Supplementary Table 4). Changing *in-house* with external calibration determined only modest deviations in measures within each laboratory, except for Laboratory H (18.9 %) (Supplementary Table 2). HM40122 target value (14,216 nmol/L)

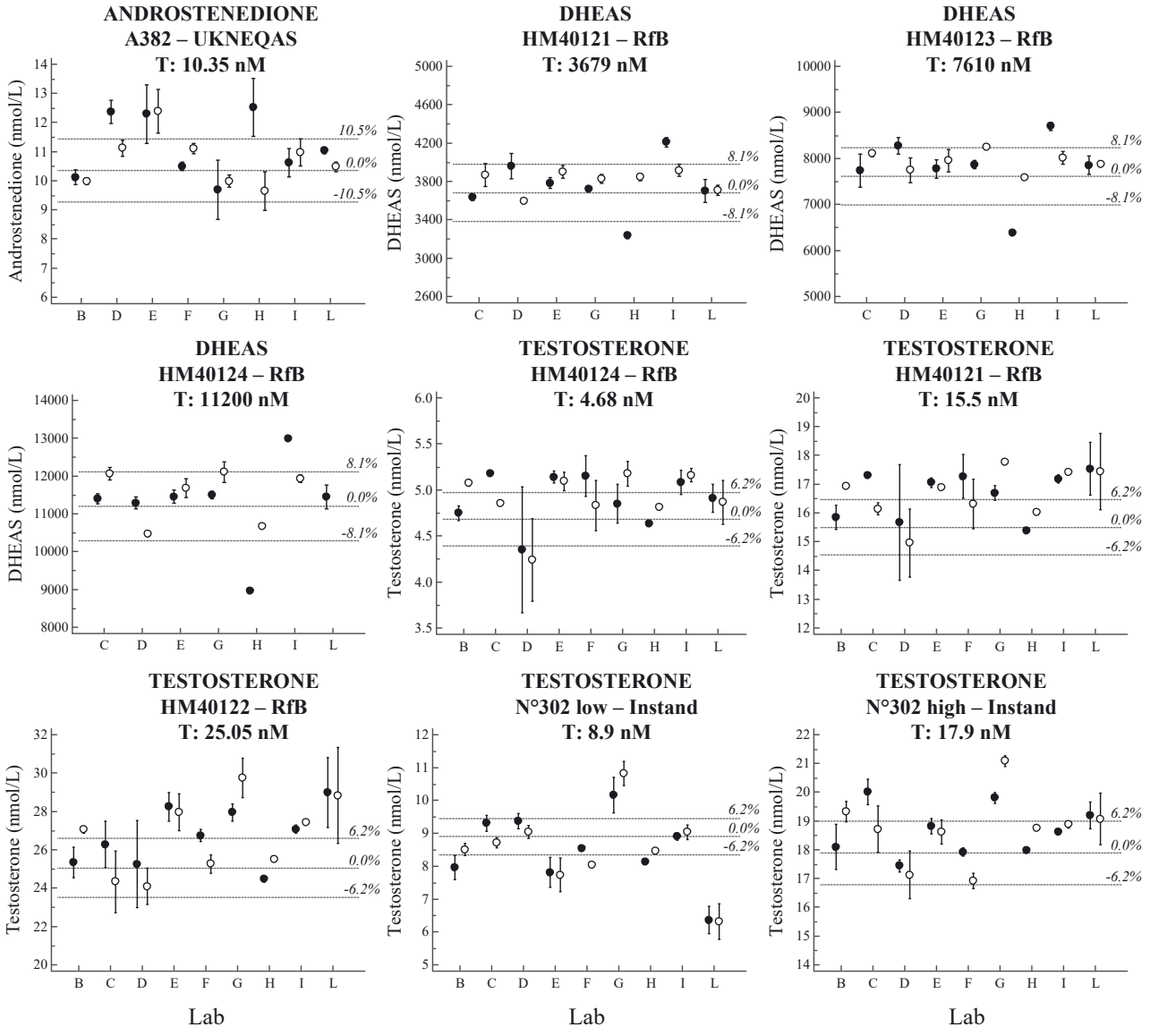

Figure 1: Laboratory measurements in external quality assessment (EQA) materials as function of the calibration system. Black dots: *in-house* calibration; white dots: external calibration. Bars: range of the duplicate measurements. Dashed lines: zero ± maximum allowable bias (androstenedione: 10.5 %; DHEAS: 8.1 %; testosterone: 6.2 %). T, target value. Target values of A382 for androstenedione, and of HM40121, HM40123 and HM40124 for DHEAS were determined as mean/median of all MS methods participating in the EQA survey. Target values of HM40121, HM40122, HM40124 and N°302 pair for testosterone were determined by reference measurement procedure.

was out of range for most of the laboratories and was therefore not evaluated. Targets of other HMs ranged from 3,679 to 11,200 nmol/L. Bias were within the MAB in all except Laboratories H and I, showing respective under- and over-estimation which was corrected when applying external calibration (Figure 1). The median inter-laboratory CV was 7.2 % with *in-house*, and lowered to 4.9 % with external calibration ($p<0.001$). Similarly, the prevalence of cases with inter-laboratory CV above MAI reduced from 55.1 % with *in-house* to 15.4 % with external calibration; with the latter, all samples above 8,000 nmol/L were within the threshold (Figure 2 and Supplementary Table 7). At Passing-Bablok analysis, when using *in-house* calibration, laboratories showed modest slope deviations (from 0.979 [0.964–0.997] to 1.087 [1.071–1.105]), except for Laboratory H (0.825 [0.803–0.847]). External calibration improved the overall slope range (from 0.929 [0.895–0.950] to 1.056 [1.049–1.064]). However, large intercepts were noted for some methods with both calibration systems (Supplementary Figure 1 and Table 2). Median bias vs. all laboratory median was within ±4.8 % for all except Laboratory H (–10.9 %) (Figure 3, Supplementary Table 8 and Figure 3). Laboratory D and H

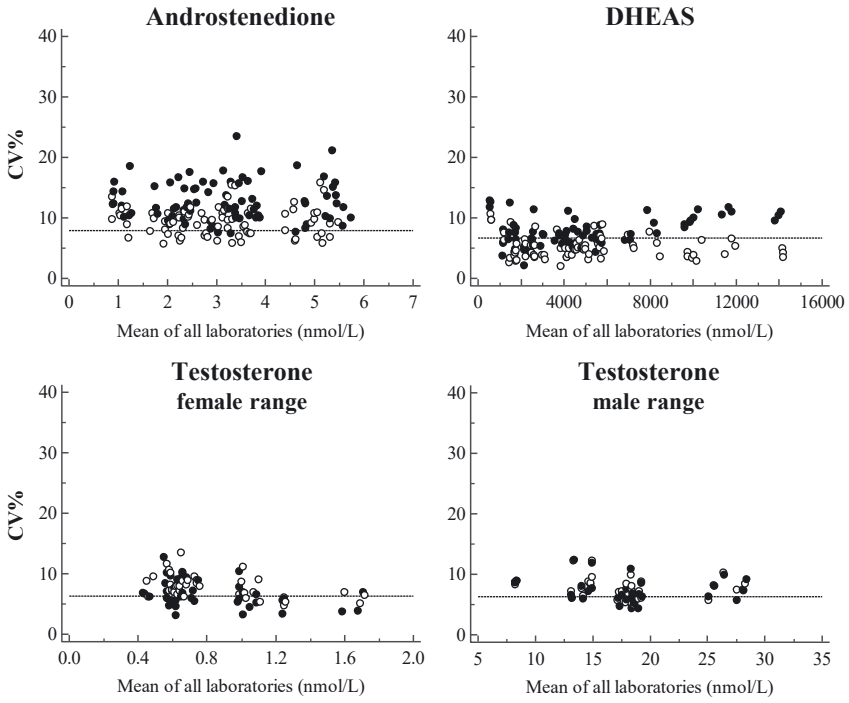

Figure 2: Inter-laboratory coefficient of variations according to the calibration system. Black dots: *in-house* calibration; white dots: external calibration. Horizontal lines: maximal allowable imprecision (androstenedione: 7.9 %; DHEAS: 6.7 %; testosterone: 6.3 %).

exhibited large variance of bias and a few cases exceeding the TAE. Using external calibration overall reduced the median bias (within ±3.5 %) and cases exceeding the TAE (Figure 3, Supplementary Table 8 and Figure 3). Supplementary Table 9 shows that DHEAS measures by *in-house* calibration were mostly non comparable between laboratories, whereas some similarities were obtained by external calibration.

Testosterone

Results were subdivided in female and male ranges. The former included 42 samples from 13 women and 1 severely hypogonadal male, with values between 0.43 and 1.71 nmol/L. The latter included 36 samples from 12 men, with values between 8.11 and 28.45 nmol/L (all laboratory median by *in-house* calibration) (Supplementary Tables 5 and 6). The intra-laboratory CV ranged 4.3–8.7 % for female and 2.6–7.1 % for male levels, which included four and eight laboratories within the MAI, respectively (Supplementary Table 2). Cases with duplicate CVs above the MAI were up to 65.4 % for Laboratory D at female levels (Supplementary Tables 5 and 6). Applying external instead of *in-house* calibration demonstrated modest deviation of measurements overall, except for Laboratories D (13.5 %), F (−9.0 %) and G (9.2 %) in the female range (Supplementary Table 2). Material HM40123 (target: 40.8 nmol/L) was above most of the measurement ranges, and was therefore not evaluated. Targets in other materials ranged from 4.68 to 25.05 nmol/L (Figure 1). When using *in-house* calibration, trueness ranged from −7.1 to 15.7 % in HM and in N°302 high materials, with six laboratories outside of the MAB in two materials or more. Employing external calibration determined worse (Laboratories B and G), similar (Laboratories D, E, H, I and L) or better (Laboratories C and F) trueness depending on the laboratory. Interestingly, some methods performed markedly differently with material N°302 low (overall trueness: −28.6 to 14.2 %) compared to other materials, such as Laboratories B, E and L (Figure 1). N°302 low also failed the commutability test with results either above (Laboratories G and I) and below (Laboratory L) the agreement interval (Supplementary Figure 4). Small deviations from commutability were also noted for materials HM40121 (Supplementary Figure 4). The median inter-laboratory CV was similar between female and male ranges when using *in-house* calibration (6.4 and 6.8 %, respectively). However, when applied to the female range, the external calibration significantly increased the inter-laboratory CV to 7.6 % ($p<0.001$) and cases above MAI from 50 to 81 % (Figure 2 and Supplementary Table 7). With Passing-Bablok analysis (Supplementary Figure 1 and Table 2), when using *in-house* calibration, methods' slopes performed better in the female (from 0.979 [0.938–1.000] to 1.100 [1.030–1.177]) than in the male range (from 0.928 [0.877–0.979] to 1.180 [1.097–1.316]), and were mostly not significant. Using external calibration resulted in similar results for male levels, but poorer for female levels (from 0.913 [0.875–0.944] to 1.090 [1.019–1.143]). Large intercepts were also noted for Laboratories I and L at male levels,

Table 2: Passing-Bablok analyses of steroid measurements from each laboratory vs. the medians of all laboratories as a function of the calibration system.

Analyte	Lab	*In-house* calibration				External calibration			
		n	R (95 CI)	Slope (95 CI)	Intercept (95 CI)	n	R (95 CI)	Slope (95 CI)	Intercept (95 CI)
Androstenedione	B	78	0.988 (0.981–0.992)	0.962 (0.938–0.988)	0.043 (−0.020 to 0.104)	78	0.988 (0.980–0.992)	0.994 (0.980–1.011)	0.043 (0.009–0.081)
	D	78	0.988 (0.981–0.992)	1.074 (1.039–1.108)	−0.053 (−0.136 to 0.036)	78	0.993 (0.988–0.995)	1.024 (1.000–1.046)	−0.048 (−0.113 to 0.000)
	E	78	0.973 (0.958–0.983)	1.104 (1.058–1.167)	−0.019 (−0.192 to 0.096)	78	0.976 (0.962–0.985)	1.194 (1.138–1.263)	−0.017 (−0.208 to 0.090)
	F	78	0.995 (0.992–0.997)	0.978 (0.956–0.996)	−0.025 (−0.082 to 0.019)	78	0.994 (0.991–0.996)	1.011 (0.987–1.032)	−0.097 (−0.152 to −0.039)
	G	78	0.972 (0.957–0.982)	0.851 (0.811–0.891)	0.058 (−0.063 to 0.150)	78	0.973 (0.958–0.983)	0.938 (0.897–0.985)	0.025 (−0.088 to 0.116)
	H	78	0.974 (0.959–0.983)	1.186 (1.133–1.252)	0.038 (−0.148 to 0.158)	78	0.980 (0.969–0.987)	0.964 (0.926–1.009)	0.052 (−0.048 to 0.141)
	I	75	0.986 (0.978–0.991)	0.941 (0.915–0.969)	−0.070 (−0.142 to 0.004)	75	0.985 (0.976–0.990)	1.016 (0.993–1.040)	−0.079 (−0.146 to −0.014)
	L	78	0.992 (0.987–0.995)	1.016 (0.997–1.041)	−0.022 (−0.089 to 0.038)	78	0.987 (0.980–0.992)	1.013 (0.989–1.040)	−0.025 (−0.098 to 0.040)
DHEAS	C	76	0.996 (0.994–0.998)	0.979 (0.964–0.997)	−39.6 (−99.0 to 18.4)	76	0.996 (0.994–0.998)	1.004 (0.996–1.017)	−8.7 (−63.4 to 14.5)
	D	77	0.991 (0.987–0.995)	1.030 (0.994–1.053)	76.6 (−20.1 to 142.0)	77	0.992 (0.988–0.995)	0.929 (0.895–0.950)	120.1 (62.8–183.6)
	E	69	0.995 (0.993–0.997)	0.998 (0.983–1.003)	−49.4 (−63.6 to 12.5)	69	0.996 (0.994–0.998)	1.000 (0.998–1.010)	0.0 (−54.2 to 4.3)
	G	75	0.997 (0.996–0.998)	1.000 (1.000–1.006)	−0.0 (−15.3 to 0.0)	75	0.996 (0.993–0.997)	1.056 (1.049–1.064)	−138.7 (−174.4 to −109.3)
	H	78	0.998 (0.996–0.998)	0.825 (0.803–0.847)	201.2 (131.8–287.9)	78	0.998 (0.997–0.999)	0.972 (0.954–0.991)	192.0 (118.3–273.6)
	I	75	0.998 (0.996–0.999)	1.087 (1.072–1.105)	−108.1 (−173.4 to −69.7)	75	0.997 (0.996–0.998)	0.994 (0.980–1.006)	−78.2 (−128.1 to −38.9)
	L	78	0.997 (0.995–0.998)	1.005 (0.998–1.015)	47.1 (0.2–78.5)	78	0.996 (0.994–0.998)	1.009 (1.000–1.023)	−22.2 (−58.0 to 0.0)
Testosterone – female range	B	42	0.988 (0.977–0.993)	1.024 (0.980–1.077)	−0.023 (−0.058 to 0.009)	42	0.985 (0.971–0.992)	1.081 (1.019–1.140)	−0.037 (−0.075 to 0.010)
	C	42	0.967 (0.939–0.982)	1.100 (1.030–1.177)	−0.025 (−0.075 to 0.026)	42	0.962 (0.930–0.979)	1.000 (0.976–1.062)	0.000 (−0.043 to 0.025)
	D	42	0.957 (0.921–0.977)	1.054 (1.000–1.125)	−0.057 (−0.101 to 0.015)	42	0.951 (0.909–0.973)	1.000 (0.938–1.077)	0.040 (−0.006 to 0.093)
	E	42	0.948 (0.905–0.972)	1.026 (0.991–1.094)	−0.013 (−0.060 to 0.019)	42	0.965 (0.936–0.981)	1.019 (0.991–1.068)	−0.021 (−0.060 to 0.005)
	F	42	0.973 (0.949–0.985)	1.000 (0.973–1.000)	0.000 (0.000–0.021)	42	0.963 (0.932–0.980)	0.913 (0.875–0.944)	−0.015 (−0.040 to 0.011)
	G	42	0.931 (0.874–0.962)	1.025 (0.985–1.082)	−0.022 (−0.064 to 0.012)	42	0.928 (0.869–0.961)	1.090 (1.019–1.143)	−0.024 (−0.067 to 0.031)
	H	42	0.966 (0.937–0.982)	0.979 (0.938–1.000)	0.016 (0.000–0.048)	42	0.961 (0.928–0.979)	1.000 (0.980–1.050)	0.000 (−0.038 to 0.015)
	I	42	0.954 (0.915–0.975)	1.000 (0.972–1.022)	−0.020 (−0.041 to 0.003)	42	0.958 (0.922–0.977)	1.000 (0.989–1.030)	−0.010 (−0.040 to 0.004)
	L	42	0.938 (0.887–0.967)	1.020 (0.973–1.107)	−0.012 (−0.071 to 0.027)	42	0.934 (0.880–0.964)	1.016 (0.957–1.096)	−0.043 (−0.109 to 0.009)
Testosterone – male range	B	36	0.943 (0.890–0.971)	0.928 (0.877–0.979)	0.052 (−0.819 to 0.954)	36	0.971 (0.943–0.985)	0.994 (0.957–1.020)	0.105 (−0.366 to 0.723)
	C	36	0.937 (0.880–0.968)	1.074 (1.031–1.140)	−0.328 (−1.354 to 0.356)	36	0.966 (0.933–0.983)	1.000 (0.979–1.041)	0.000 (−0.761 to 0.386)
	D	36	0.910 (0.829–0.953)	0.998 (0.927–1.046)	0.048 (−0.612 to 1.296)	36	0.905 (0.820–0.951)	0.943 (0.869–1.011)	0.606 (−0.581 to 1.832)
	E	36	0.947 (0.897–0.973)	1.016 (0.942–1.095)	0.044 (−1.450 to 1.247)	36	0.929 (0.865–0.964)	0.996 (0.916–1.082)	0.364 (−1.172 to 1.716)
	F	36	0.982 (0.966–0.991)	1.000 (0.998–1.028)	−0.095 (−0.570 to −0.049)	36	0.983 (0.968–0.992)	0.940 (0.917–0.966)	−0.189 (−0.637 to 0.148)
	G	36	0.941 (0.886–0.969)	1.012 (0.971–1.060)	−0.361 (−0.453 to 1.040)	36	0.927 (0.860–0.962)	1.089 (1.043–1.145)	0.466 (−0.454 to 1.201)
	H	36	0.966 (0.934–0.983)	0.969 (0.917–1.022)	−0.307 (−1.225 to 0.467)	36	0.984 (0.968–0.992)	1.000 (0.949–1.040)	−0.015 (−0.741 to 0.721)
	I	33	0.980 (0.959–0.990)	1.019 (0.978–1.070)	−0.994 (−1.817 to 0.347)	33	0.973 (0.945–0.987)	1.017 (0.985–1.074)	−0.469 (−1.467 to −0.030)
	L	36	0.924 (0.855–0.961)	1.180 (1.097–1.316)	−1.668 (−4.047 to 0.022)	36	0.930 (0.866–0.964)	1.161 (1.086–1.269)	−1.464 (−3.299 to −0.265)

n, number of samples; CI, confidence interval.

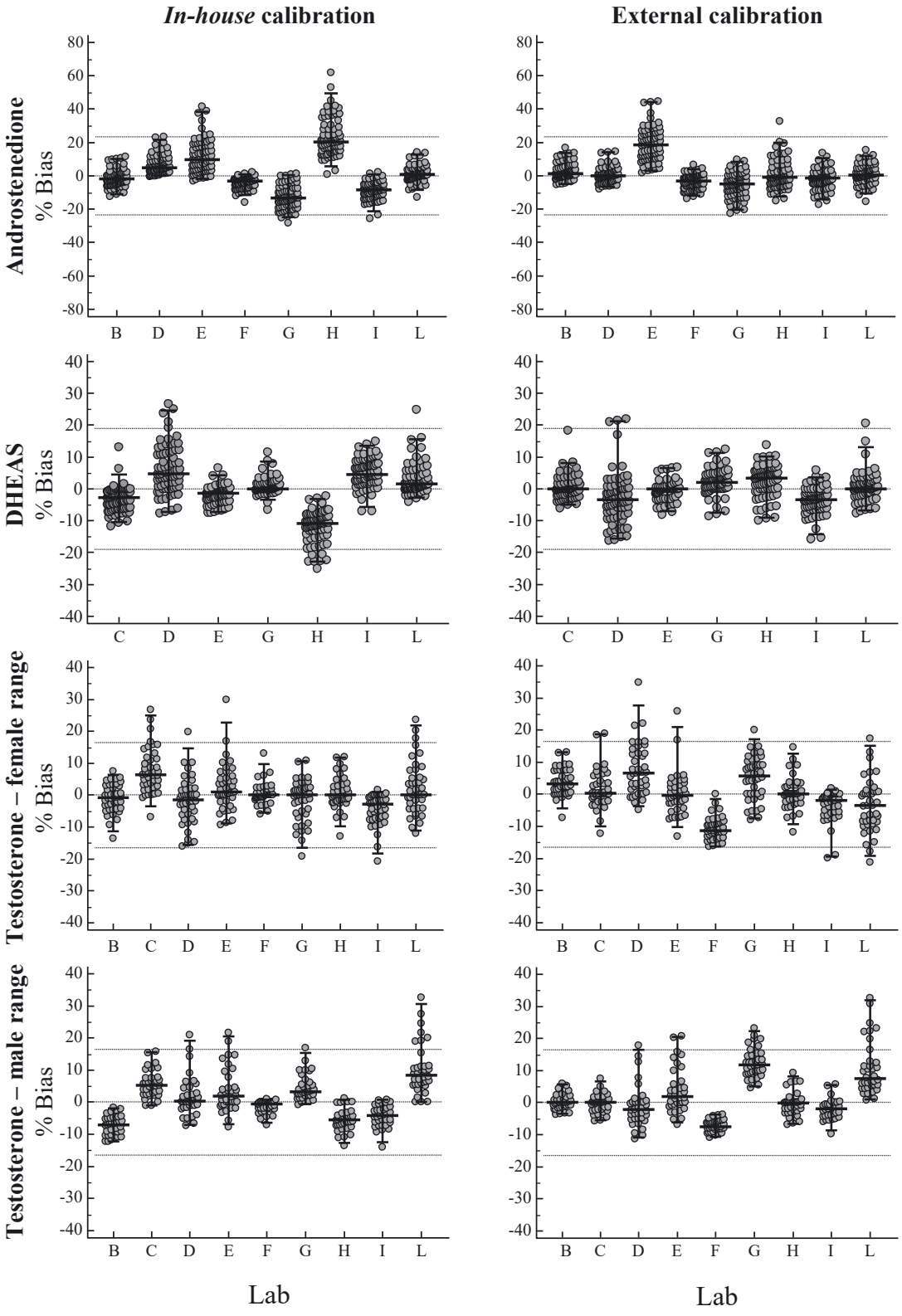

Figure 3: Laboratories % bias vs. median of all laboratories as function of the calibration system. *% bias = ([laboratory value – median of all laboratories]/ median of all laboratories) × 100*. Segments: median; bars: 2.5 and 97.5 centiles; horizontal lines: zero ± total allowable error (androstenedione: 23.5 %; DHEAS: 19.0 %; testosterone: 16.5 %).

which were not corrected by external calibration. With *in-house* calibration, the median bias vs. all laboratory median ranged from −2.7 to 6.5 in the female, and −7.0 to 8.5 % in the male range (Figure 3, Supplementary Table 8, Supplementary Figures 5 and 6). External calibration overall enlarged the bias to −11.3 to 6.6 % in the female and −7.5 to 11.8 % in the male range. Deviations were mostly within the TAE for both levels and calibration systems. The largest bias variance, up to 67.6 %, was noted for Laboratory L in all tested conditions (Figure 3, Supplementary Table 8, Supplementary Figures 5 and 6). Supplementary Table 9 shows that testosterone measures in the female range were non different between various laboratories when using *in-house* calibration. However, comparability was observed only in a few cases when using external calibration and when moving to the male range.

Discussion

Our study contributes to harmonization/standardization knowledge and status of androgen measurement by LC-MS/MS. While some previous studies have reported on testosterone and androstenedione, none focused on DHEAS [7–12]. We interpreted our results on precision, trueness, reproducibility and agreement by use of maximum allowable performances calculated from recently updated data on biological variability [16, 17]. The methods here compared were validated according to recommended guidelines [20–28]. They encompass different procedures for sample preparation, LC and calibration. Methods that included measurements of DHEAS adopted either positive or negative ionization and four different ion transitions. In contrast, the same ion transition was used by all methods for testosterone, and all but one for androstenedione. At least four different IS were used for DHEAS and five for testosterone and androstenedione quantitation.

According to our data, all except two methods demonstrated good precision, within or slightly above the MAI. Less than optimal precision for measurements of androstenedione by Laboratory H may be attributable to use of a non-matching isotopic IS (D3-testosterone). In contrast, suboptimal precision for measurements of DHEAS by Laboratory L may have reflected initial inexperience with maintenance of the MS [14, 15]. In EQA materials, deviations of androstenedione and DHEAS were within ±20 % for all methods, while trueness for testosterone was within ±15 %. Of note, some methods exceeded the MAB thresholds.

Although the inter-laboratory performance was generally acceptable for all three hormones, agreement was less satisfactory for androstenedione than the other two androgens, both in absolute terms and according to allowable thresholds. Although the inter-laboratory CVs for DHEAS and testosterone were within or slightly above the MAI, those for androstenedione were mostly above the threshold. All laboratories displayed median bias largely within the TAE. Unexpectedly, inter-laboratory agreement was similar or better for testosterone in females than males. This contrasts with earlier studies [7, 8, 10, 11], but not with a more recent one [12].

Some methods displayed relevant random errors as indicated by bias variance. The largest were reported for androstenedione by Laboratory E and H, which may have reflected use of non-matching isotopic IS ^{13}C3-17OH-progesterone and D3-testosterone in those respective methods. Laboratory E used the same ^{13}C3-17OH-progesterone for testosterone measurement, and showed a relatively large bias variance also for this analyte. These data support the importance of a suitable IS for the reliability of quantitation. Non-matching isotopic IS may not efficiently cope for variability in pre-analytics and analytics [29] and should be avoided. In a study that focused on testosterone, it was even reported that different forms of isotopically-labeled testosterone may influence quantitation up to 15 % [13]. Differences in the retention times between the analyte and respective IS, though small in nature, have been suggested to play a role. Within the current study, methods using the same IS did not show closer inter-laboratory agreement compared to others (data not shown). This suggests that the random variability among LC-MS/MS methods, i.e., variability not explained by calibration bias, results from a complex interaction between individual matrix, extraction procedure, chromatographic separation, MS detection and the ability of the IS to correct for all confounders. Furthermore, several isobaric compounds exist that may impact the accuracy of LC-MS/MS quantitation of the three androgens [3]. Notably, only part of the methods verified the resolution between testosterone and epitestosterone.

Due to different designs among previous studies on androgen harmonization, direct comparison of those and our results is not possible. Most studies employed limited numbers of samples [7–11]. Some, similar to our study, were ring trials [7, 8, 11, 12], while others included a reference technique [9, 10]. Overall, our results compare favorably with most previous studies on testosterone [7, 8, 10, 11] and androstenedione [7, 8]. We can also confirm previous observations of higher inter-laboratory agreement for testosterone than androstenedione [7, 8]. In addition, our results compare similarly to a study measuring testosterone in 58 samples comparing four LC-MS/MS to one GC-MS method [9]. Recently, French et al. [12] compared testosterone measurements of 102 samples among four LC-MS/MS methods; they reported a mean bias between −4.9 and +3.7 %. Such agreement may reflect technical similarities among investigated methods. Indeed, all were calibrated

using the NIST SRM 971; four used a C18 stationary phase and acetonitrile-based mobile phases; three used D3--testosterone as the IS; and three used the same sample extraction procedure [12].

In agreement with our previous reports, the strategy to unify calibration resulted in different findings depending on the analyte [14, 15]. Here, this strategy was beneficial for androstenedione and DHEAS comparability, as it reduced the CV for both hormones by about a third, and also enhanced correlation and agreement. Reasons are yet to be found for the correction of large calibration bias of Laboratories D and H for androstenedione, and of Laboratories H and I for DHEAS. Notably, improper calibration was not the reason for proportional overestimation demonstrated by Laboratory E for androstenedione. Once again, this may depend on the use of non-matched isotopic IS. As for testosterone, in line with previous observations [13], calibration was not a relevant source of error. Indeed, external calibration resulted in contrasting effects on testosterone trueness, and it worsened the inter-laboratory performance in the female range. Such findings are possibly due to a suboptimal commutability of the serum-derived matrix of the commercial set.

Our study highlighted a remarkable status of harmonization of androgen measurement by LC-MS/MS methods. Unfortunately, we were not able to include DHEA in this report, as only three laboratories measured this analyte, and levels were below the LOQ in several cases. One limitation is that no reference is available to assess true methods' deviations in patient samples. In principle, our present and previous studies support the unification of the calibration material as a strategy to improve harmonization, provided that calibrators are traceable to higher order materials or RMP [14, 15]. However, we repeatedly observed issues with commutability of calibration matrix that need to be addressed before such a strategy can be introduced on a large scale [14, 15]. As reported here for N°302 and previously [14, 15], EQA materials are also not free from matrix effects. It is important that commutability is not to be taken for granted with LC-MS/MS, or the result of these precious surveys may be counterproductive [30]. Sources of variability other than calibration should be investigated. In these regards, our study design opens the discussion for the type of vacuum tube used to derive serum or plasma samples for steroid investigations.

Satisfactory results were observed for DHEAS and testosterone. Reasons are possibly found in the wider availability of EQA surveys for both hormones. Most probably, testosterone performance benefits from the availability of the RMP for assigning target values [31, 32]. Conversely, no RMP is available for DHEAS [32]. The less than ideal performance noted for androstenedione may relate to the scarce availability of surveys for this hormone. Recently, a RMP has been listed in the Joint Committee for Traceability in Laboratory Medicine [32, 33]. We hope that such an addition will soon be translated into benefits for EQA surveys. That underlines the importance of a reference measurement system which generally entails a RMP and higher order reference materials to improve measurement accuracy and reduce variability between laboratories. Another important limitation of EQA materials can be their usual high target values, sometimes even exceeding the calibration range. This implies that laboratories can only verify their trueness in severe hyperandrogenic conditions or, for testosterone, in the normal male range. Contrarily, to ensure effective disease comprehension and treatment, modern clinical practice and research need to rely on methods that perform optimally for both sexes and in physiologic and pathologic ranges.

In conclusion, our study highlights improvements in the inter-laboratory comparability of androgen measurements made possible by LC-MS/MS and supports the feasibility of achieving a successful harmonization and standardization. In these regards, we shed light on important critical points, such as commutability of calibrator and EQA materials and avoiding non-matching isotopic IS, that need to be addressed to achieve the desirable goal.

Acknowledgments: We thank Dr. Oliver Midasch from Chromsystems for donating calibration materials used in the study.

Research ethics: The study received approval by the local Ethics Committee (no. 141/2017/U/Tess).

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

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: The authors state no conflict of interest.

Research funding: This study was supported by the Emilia-Romagna Region, Alessandro Liberati Young Researcher Grants (project number PRUA 1-2012-004, granted to FF) and

by Deutsche Forschungsgemeinschaft (DFG), German Research Foundation (project number 314061271-CRC/TRR 205, granted to MP and GE).

References

1. Pugeat M, Plotton I, de la Perrière AB, Raverot G, Déchaud H, Raverot V. Management of endocrine disease hyperandrogenic states in women: pitfalls in laboratory diagnosis. Eur J Endocrinol 2018;178:R141–54.
2. Guzelce EC, Galbiati F, Goldman AL, Gattu AK, Basaria S, Bhasin S. Accurate measurement of total and free testosterone levels for the diagnosis of androgen disorders. Best Pract Res Clin Endocrinol Metab 2022;36:101683.
3. Wudy SA, Schuler G, Sánchez-Guijo A, Hartmann MF. The art of measuring steroids: principles and practice of current hormonal steroid analysis. J Steroid Biochem Mol Biol 2018;179:88–103.
4. French D. Clinical utility of laboratory developed mass spectrometry assays for steroid hormone testing. J Mass Spectrom Adv Clin Lab 2023; 28:13–19.
5. Lentjes EGWM, Bui HN, Ruhaak LR, Kema IP, Coene KLM, van den Ouweland JMW. LC-MS/MS in clinical chemistry: did it live up to its promise? Consideration from the Dutch EQAS organization. Clin Chim Acta 2023;546:117391.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. French D, Drees J, Stone JA, Holmes DT, van der Gugten JG. Comparison of four clinically validated testosterone LC-MS/MS assays: harmonization is an attainable goal. Clin Mass Spectrom 2018;11:12–20.
13. 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.
14. Fanelli F, Cantù M, Temchenko A, Mezzullo M, Lindner JM, Peitzsch M, et al. Report from the HarmoSter study: impact of calibration on comparability of LC-MS/MS measurement of circulating cortisol, 17OH-progesterone and aldosterone. Clin Chem Lab Med 2022;60:726–39.
15. Fanelli F, Bruce S, Cantù M, Temchenko A, Mezzullo M, Lindner JM, et al. Report from the HarmoSter study: inter-laboratory comparison of LC-MS/MS measurements of corticosterone, 11-deoxycortisol and cortisone. Clin Chem Lab Med 2022;61:67–77.
16. Westgard QC. Desirable biological variation database specifications. https://www.westgard.com/biodatabase1.htm [Accessed 28 Aug 2023].
17. The European federation of clinical chemistry and laboratory medicine database. https://biologicalvariation.eu/meta_calculations [Accessed 28 Aug 2023].
18. 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.
19. Hyslop NP, White WH. Estimating precision using duplicate measurements. J Air Waste Manag Assoc 2009;59:1032–9.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. Büttler RM, Kruit A, Blankenstein MA, Heijboer AC. Measurement of dehydroepiandrosterone sulphate (DHEAS): a comparison of isotope-dilution liquid chromatography tandem mass spectrometry (ID-LC-MS/MS) and seven currently available immunoassays. Clin Chim Acta 2013;424:22–6.
26. Büttler RM, Martens F, Kushnir MM, Ackermans MT, Blankenstein MA, Heijboer AC. Simultaneous measurement of testosterone, androstenedione and dehydroepiandrosterone (DHEA) in serum and plasma using isotope-dilution 2-dimension ultra high performance liquid-chromatography tandem mass spectrometry (ID-LC-MS/MS). Clin Chim Acta 2015;438:157–9.
27. 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.
28. 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.
29. Vogeser M, Seger C. Pitfalls associated with the use of liquid chromatography-tandem mass spectrometry in the clinical laboratory. Clin Chem 2010;56:1234–44.
30. Miller WG, Myers GL, Rej R. Why commutability matters. Clin Chem 2006;52:553–4.

31. Greaves RF. The central role of external quality assurance in harmonisation and standardisation for laboratory medicine. Clin Chem Lab Med 2017;55:471–3.
32. 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 28 Aug, 2023].
33. Gradl K, Taibon J, Singh N, Albrecht E, Geistanger A, Pongratz S, et al. An isotope dilution LC-MS/MS-based candidate reference method for the quantification of androstenedione in human serum and plasma. Clin Mass Spectrom 2020;23:1–10.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/cclm-2023-1138).