Title

Comparison of four clinically validated testosterone LC-MS/MS assays: Harmonization is an attainable goal

Permalink

https://escholarship.org/uc/item/25m1s34x

Authors

French, Deborah
Drees, Julia
Stone, Judith A
et al.

Publication Date

2019

DOI

10.1016/j.clinms.2018.11.005

Peer reviewed

Comparison of four clinically validated testosterone LC-MS/MS assays: Harmonization is an attainable goal

Deborah French[a,*], Julia Drees[b], Judith A. Stone[c], Daniel T. Holmes[d,e], J. Grace van der Gugten[d]

[a] Department of Laboratory Medicine, University of California San Francisco, San Francisco, CA 94107, United States
[b] Kaiser Permanente Northern California Regional Laboratory, Richmond, CA, United States
[c] University of California San Diego Health Center for Advanced Laboratory Medicine, San Diego, CA, United States
[d] Department of Pathology and Laboratory Medicine, St Paul's Hospital, Vancouver, British Columbia, Canada
[e] Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada

ARTICLE INFO

ABSTRACT

Introduction: Immunoassays and liquid chromatography-tandem mass spectrometry assays are commonly employed in clinical laboratories for measurement of total testosterone in serum. Results obtained from either of these methodologies compare poorly due to differences in calibration and/or inadvertent detection of interfering substances by the immunoassays. Standardization efforts are underway, but recent studies indicate that accuracy remains an issue.

Methods: This study compares the results from four independently developed and validated LC-MS/MS assays for total testosterone. The calibration for each assay was verified using National Institute of Standards and Technology Standard Reference Material 971.

Results: Initially, one of the four assays had a mean percent difference of +11.44%, compared to the All Method Mean, but following re-verification of all five non-zero calibrator concentrations with the NIST SRM 971, the mean percent difference decreased to −4.88%. Subsequently, the agreement between all four assays showed a mean bias of < 5% across the range of all testosterone concentrations (0.13–38.10 nmol/L; 3.7–1098 ng/dL), including at low concentrations of < 1 nmol/L (< 29 ng/dL).

Conclusions: Excellent agreement between four independently developed LC-MS/MS assays demonstrates that harmonization using standard reference material is attainable. However, as we found in this study, to ensure accurate calibration it is critical to validate the concentrations of new lots of calibrators.

1. Introduction

Total testosterone in serum has historically been measured with solvent extraction followed by radioimmunoassay (RIA) or gas chromatography-mass spectrometry (GC-MS) [1]. These methods are labor intensive and involved cumbersome manual extraction steps prior to the advent of automated liquid handling systems. Automated immunoassays for total testosterone were developed to be more amenable for use in a high-throughput, clinical laboratory environment [1].

However, post-implementation, immunoassays were found to be inaccurate at the low concentrations present in pediatric, female, and hypogonadal male patients; presumably due to endogenous and exogenous interferences in patient samples and issues with accurate calibration at the low analyte target concentrations [2,3,4].

In a 2003 editorial, the authors concluded that "guessing would be more accurate and additionally could provide cheaper and faster testosterone results for females – without even having to draw the patient's blood" in reference to the immunoassay methods for testosterone

quantitation in use at that time [5]. In the same journal issue, one group compared 10 testosterone immunoassays to a GC-MS assay, and demonstrated that immunoassays gave results that were up to five-fold higher in females (values that are clinically useless since they do not give an accurate picture of what is occurring in the patient and may lead to further clinical investigation) than the results from the same sample analyzed by GC-MS [2]. In 2007, the Endocrine Society proposed that "the best prospect for a gold standard (in testosterone testing) lies in extraction and chromatography followed by MS or MS/MS in which the chemical structure of the molecule measured is identified [1]."

To this end, the Centers for Disease Control and Prevention (CDC) launched the Hormone Standardization Program (HoSt) with the objective of ensuring that laboratory results are traceable to one accurate measurement and are comparable across methods, time and location [6]. To assess the extent of the lack of harmonization at the time, a study was performed using 30 human serum samples run on seven different liquid chromatography-tandem mass spectrometry (LC-MS/MS) assays and one gas chromatography-tandem mass spectrometry (GC-MS/MS) assay to quantify testosterone [7] and compared to the LC-MS/MS reference measurement procedure (RMP) at the National Institute of Standards and Technology (NIST) [8]. Female samples had concentrations between 0.17 and 2.98 nmol/L (4.9–85.9 ng/dL) and male samples had concentrations between 5.55 and 39.67 nmol/L (159.9–1143.23 ng/dL). The overall imprecision of the measurements compared to the RMP was < 15% at testosterone concentrations > 1.53 nmol/L (44.09 ng/dL) and < 34% at 0.3 nmol/L (8.65 ng/dL) with a mean percent difference of 11% [7]. The CDC developed a RMP using LC-MS/MS [9], which is currently used by the HoSt program, along with the NIST RMP, to determine the concentrations of testosterone in a set of serum samples. These samples can be purchased by laboratories or vendors and used to assess the accuracy of the calibration of their assays. In addition, laboratories or vendors can participate in the HoSt certification program where, following successful participation in two phases of testing, the assay will be given CDC certification [8,9,10]. However, participation in the HoSt program is cost prohibitive for many laboratories.

The NIST Standard Reference Material 971 (SRM 971) was developed to address the need for improved accuracy of routine clinical assays [11]. The NIST SRM 971 is value assigned using the NIST RMP and is much less costly than participation in the CDC HoSt program [12]. NIST SRM 971 consists of two human serum samples at low and high testosterone concentrations, and comes with a certificate of analysis (CofA) stating the target concentration for each sample with acceptable bias criteria [12]. While the CDC HoSt program requires submission of laboratory data that is analyzed by the CDC and, subsequently, returned to participating laboratories, there is no required submission of results to NIST for validation. It is, thus, the responsibility of the individual laboratory to ensure that their testosterone assay meets the NIST SRM 971 acceptance criteria as stated in the CofA.

While studies publishing comparisons between immunoassay(s) and LC-MS/MS assays for measuring testosterone are commonplace [2,4,13,14,15], those comparing LC-MS/MS assays with each other are rare. In one study, four LC-MS/MS assays were compared to a GC-MS RMP using 58 samples with concentrations ranging from 0.2 to 31.3 nmol/L (5.76–902.02 ng/dL; from 15 healthy males, eight hypogonadal males, 30 females and five pools of serum) [16]. The method comparison linear regression slopes ranged from 0.85 to 0.94 and the mean percent differences across all testosterone concentrations between the four LC-MS/MS assays and the RMP were −9.6%, 6.4%, 6.8% and 0.4%, respectively [16]. In another study, seven published LC-MS/MS assays for testosterone were compared using 55 patient samples with median testosterone concentrations ranging from 0.22 to 1.36 nmoL/L (6.34–39.19 ng/dL) for the female samples and 8.27 to 27.98 nmol/L (238.33–806.34 ng/dL) for the male samples [17]. The method comparison linear regression slopes between the reported

concentrations of the published assays and the median testosterone concentrations ranged from 0.92 to 1.04, with intercepts ranging from −0.07 to +0.21 [17]. The inter-method coefficient of variation (CV) for the female samples was 14% and for male samples was 8% [17]. The same group undertook a recent study comparing eight unpublished LC-MS/MS methods for measurement of testosterone in 60 serum samples from male and female volunteers with concentrations ranging from 6.15 to 24.44 nmol/L (177.23–704.32 ng/dL) and 0.05 to 1.26 nmol/L (1.44–36.31 ng/dL), respectively [18]. The linear regression slopes of the method comparisons ranged between 0.9 and 1.25 for the female samples and 0.87–1.24 for the male samples [18]. The inter-method CVs were 24% and 14% for the female and male samples, respectively [18]. These reports, which demonstrate significant variation between routine LC-MS/MS testosterone assays, indicate that there is room for improvement.

The goal of this study was to perform interlaboratory comparisons using human serum samples across a range of clinically relevant concentrations, using routine LC-MS/MS assays developed independently in four different laboratories each calibrated to the NIST SRM 971.

2. Materials and methods

2.1. Experimental information

2.1.1. St. Paul's Hospital (SPH)

Type II deionized water was from the in-house Barnstead water system (> 18.2 mOhms). Mass spectrometry grade methanol and acetonitrile and NUNC 2 mL 96-well plates were from Fisher Scientific (Ontario, Canada). Hexane and ethyl acetate (ACS grade) were from Sigma Aldrich (Ontario, Canada). NIST SRM 971 was obtained from NIST (Gaithersburg, MD). Double charcoal stripped human serum (MSG3000) was from Golden West Biologicals (Temecula, CA). Testosterone and d3-testosterone (≥98% purity; ≥99% isotope incorporation) were from Cerilliant Corporation (Round Rock, TX). Quality control samples used were Immunoassay Plus Lyphocheck levels 1, 2 and 3 plus a 1:10 diluted level 1 (BioRad Laboratories; Irvine, CA), plus a patient serum pool. The liquid chromatography column was a Luna® C18, 3 μm, 50 × 2 mm column from Phenomenex (Torrance, CA). Calibrators were prepared in-house by spiking known volumes of the Cerilliant testosterone standard solution into the double charcoal stripped serum with concentrations assigned using the NIST SRM 971. This was achieved by running the NIST SRM 971 in duplicate as calibrators and the in-house calibrators in duplicate as unknowns. The mean calculated concentrations of the in-house spiked calibrators were used as the assigned calibrator concentrations.

2.1.2. University of California San Francisco (UCSF)

Mass spectrometry grade water and solvent were from VWR International (Brisbane, CA). Human female serum samples spiked with testosterone at five different concentrations were obtained from UTAK Laboratories (Valencia, CA). NIST SRM 971 was obtained from NIST (Gaithersburg, MD). Double charcoal stripped human female serum (MSG3100) was from Golden West Biologicals (Temecula, CA) and d3-testosterone (≥98% purity; ≥99% isotope incorporation) was obtained from Cerilliant Corporation (Round Rock, TX). A low concentration quality control sample made from diluted female serum was from UTAK Laboratories (Valencia, CA) and Lyphocheck Immunoassay Controls levels 1 and 2 were from BioRad Laboratories (Irvine, CA). The Kinetex® C18, 2.6 μm, 100 × 3 mm liquid chromatography column was from Phenomenex (Torrance, CA). Calibrators were female serum spiked with testosterone from UTAK Laboratories with concentrations assigned using the NIST SRM 971 as calibrators across four different runs in duplicate and using the mean of the calculated concentrations obtained. The zero calibrator was double charcoal stripped female serum.

2.1.3. Kaiser Permanente Northern California Regional Laboratory (KP)

Mass spectrometry grade water and solvent were from Sigma-Aldrich Corporation (St Louis, MO). NIST SRM 971 was obtained from NIST (Gaithersburg, MD). Ninety six-well plates were obtained from Axygen Scientific Inc. (Union City, CA). Double charcoal stripped human serum (MSG4000) was from Golden West Biologicals (Temecula, CA). Testosterone and d3-testosterone ($\geq$ 98% purity; $\geq$ 99% isotope incorporation) were from Cerilliant Corporation (Round Rock, TX). Quality control samples were from UTAK Laboratories (Valencia, CA). The Kinetex® C18, 2.6 µm, 100 × 3 mm liquid chromatography column was from Phenomenex (Torrance, CA). Calibrators were prepared in-house by spiking the testosterone standard into the double charcoal stripped serum with concentrations assigned using the NIST SRM 971. Assignment was achieved by using the NIST SRM 971 run in duplicate as calibrators and back calculating the in-house spiked calibrators set as unknowns, also in duplicate. During method validation, KP purchased 40 samples from the CDC HoSt program to compare the calibration of this assay to the RMP. The comparison yielded a linear regression slope of 1.02 and a mean bias of −2.4% (data not shown).

2.1.4. University of California San Diego (UCSD)

Mass spectrometry grade solvents were obtained from Thermo Fisher Scientific (Fremont, CA), as were lithium chloride, ammonium hydroxide and ammonium acetate. Mass spectrometry grade deionized water was obtained from a Veolia Water Technologies ELGA purification system. NIST SRM 971 was obtained from NIST (Gaithersburg, MD). Tecan AC 96-well plates were obtained from Tecan (San Jose, CA). Double charcoal stripped human female serum (MSG3100) was from Golden West Biologicals (Temecula, CA). Testosterone was from Cerilliant Corporation (Round Rock, TX). ^{13}C-labeled testosterone ($\geq$ 98% purity; $\geq$ 99% isotope incorporation) was from Isosciences (King of Prussia, PA). Liquicheck Immunoassay Plus Quality Control samples were from BioRad Laboratories (Irvine, CA). The XSelect HSS C18 XP, 2.5 µm, 150 × 2.1 mm liquid chromatography column was from Waters Corporation (Milford, MA). Calibrators were prepared in-house using the testosterone standard spiked into the double charcoal stripped serum with concentrations assigned using the NIST SRM 971 as calibrators in four different runs in duplicate and using the mean of the calculated concentrations.

2.2. Patient samples

Institutional review board approval was determined not to be required for this study. One hundred two remnant patient samples that had been submitted for routine testing and had sufficient volume remaining were obtained from St Paul's Hospital (Vancouver, Canada). The testosterone concentrations in these samples spanned the analytical measurement range of the LC-MS/MS assays used in this study. The samples were anonymized, aliquoted, frozen at −70 °C and shipped overnight on dry ice to the other three laboratories with calibrators from the St Paul's Hospital testosterone LC-MS/MS assay. The patient samples and calibrators were thawed and run in each of the other laboratories using the routine testosterone LC-MS/MS procedures.

2.3. Sample preparation strategies

Sample preparation strategies are summarized in Table 1.

2.3.1. SPH

Liquid-liquid extraction was performed on 100 µL of serum and 40 µL of d3-testosterone (11.7 nmol/L; 337 ng/dL) using 750 µL of 90:10 (v/v) hexane: ethyl acetate in a 96-well plate using a Hamilton Microlab Starlet liquid handler. The samples were vortexed at high speed for three minutes followed by centrifugation at 3000 rpm (948g) for 10 min. Then, 500 µL of the organic layer was transferred to a new

Table 1
Summary of the sample preparation strategies used at each laboratory.

Parameter	SPH	UCSF	KP	UCSD
Sample volume	100 µL	200 µL	125 µL	100 µL
Type of extraction	Liquid-liquid extraction with 750 µL of 90:10 hexane:ethyl acetate	Liquid-liquid extraction with 1000 µL of 90:10 hexane:ethyl acetate	Liquid-liquid extraction with 725 µL of 90:10 hexane:ethyl acetate	TECAN AC Plate: Basic lithium chloride buffer mixed with sample. Plate washed. Elute with 100 µL 35:65 water:acetonitrile
Automated?	Semi: 96-well plates. Mixing, centrifuging and dry down step off-line from Hamilton Microlab Star^{LET} (Reno, NV)	No- manual in glass tubes	Semi: 96-well plates. Mixing, centrifuging and dry down step off-line from Hamilton Microlab Star^{LET} (Reno, NV)	Yes – all steps performed by TECAN SchweizG Freedom EVO® 100 ALH (San Jose, CA)
Reconstitution/final volume	200 µL of 75:25 (0.1% formic acid in water):(0.1% formic acid, 2 mM ammonium acetate in 70:30 methanol:acetonitrile	125 µL of 60:40 methanol:water	120 µL of 60:40 methanol:water	100 µL of 35:65 water:acetonitrile

deep well 2 mL 96-well plate, evaporated under air at 45 °C and reconstituted with 200 µL of 75:25 (0.1% formic acid in water):(0.1% formic acid, 2 mM ammonium acetate in 70:30 methanol:acetonitrile).

2.3.2. UCSF

As previously described [19], liquid–liquid extraction was performed on 200 µL of serum and 25 µL of d3-testosterone (3.47 nmol/L; 100 ng/dL in methanol) using 1 mL of 90:10 (v/v) hexane: ethyl acetate. The sample was vortexed and centrifuged at 3000 rpm (1200 g) for 10 min. The aqueous layer was frozen and the organic layer containing testosterone was poured off, dried under nitrogen at 45 °C and reconstituted with 125 µL of 60:40 (v/v) methanol: water.

2.3.3. KP

Briefly, liquid–liquid extraction was performed on 125 µL of serum and 40 µL of d3-testosterone (10.41 nmol/L; 300 ng/dL) using 725 µL of 90:10 (v/v) hexane: ethyl acetate in a 96-well plate using a Hamilton Microlab Starlet liquid handler. The sample was vortexed and centrifuged at 4700 rpm (2721 g) for 15 min. Five hundred and thirty µL of the organic layer containing testosterone was pipetted off, dried under nitrogen at 40 °C and reconstituted with 120 µL of 60:40 (v/v) methanol: water.

2.3.4. UCSD

As previously described [20], 100 µL of serum and 25 µL of C13-testosterone (1.72 nmol/L in 60:40 water:acetonitrile; 50 ng/dL) were added to the AC 96-well plate and shaken. Then, 175 µL of 60:24:6 (vol:vol) 0.33 mol/L lithium chloride/0.1% ammonium hydroxide:acetonitrile:water was added followed by 10 min of shaking. The extraction residue was removed and each well was washed with 0.2% ammonium hydroxide and shaken. The wash solution was discarded and the wash step was repeated. Testosterone was eluted from the plate with 100 µL of 35:65 (v:v) water: acetonitrile. All steps were performed using a TECAN SchweizG Freedom EVO® 100 ALH liquid handler.

2.4. LC-MS/MS assay information

Liquid chromatography conditions are summarized in Table 2 and the gradients are described in Table 3.

All laboratories used electrospray ionization (ESI) in positive mode. The selected reaction monitoring (SRM) transitions used for mass spectrometry are summarized in Table 4. In each laboratory, transition 1 was used to quantify the testosterone while transition 2 was used as the qualifier transition. Ion ratios were calculated for each calibrator, quality control and patient sample, using the peak area of one transition divided by the peak area of the other transition in order to increase the specificity of the methods.

Additional analytical parameters for each LC-MS/MS assay are shown in Table 5.

2.5. Method comparisons

UCSD and UCSF could not analyze one sample each due to insufficient sample volume or sample spillage, and, therefore, analyzed 101 samples using their routine LC-MS/MS assays. KP analyzed 102 patient samples using its routine testosterone LC-MS/MS assay. Each laboratory quantified testosterone values using their own calibrators, as well as calibrators provided by SPH. The results were tabulated and sent to SPH for data analysis. Regression and difference plots were prepared using cp-R [21].

3. Results

The patient samples had concentrations ranging from 0.13 to 38.10 nmol/L (3.7–1098 ng/dL) with a mean concentration of 8.33 nmol/L (241.9 ng/dL) and a median concentration of 5.87 nmol/L

Table 2
Summary of liquid chromatography conditions of the different laboratories.

Parameter	SPH	UCSF	KP	UCSD
Liquid chromatography system used	Shimadzu Prominence UFLC	Shimadzu Prominence XR UFLC	Shimadzu Prominence XR UFLC	Waters Acquity UPLC®
Mobile phase A	0.1% formic acid in water	0.1% formic acid in water	0.1% formic acid in water	0.1% formic acid in water with 2 mM ammonium acetate
Mobile phase B	0.1% formic acid in 70:30 methanol: acetonitrile with 2 mM ammonium acetate	0.1% formic acid in 70:30 methanol:acetonitrile	70:30 methanol:acetonitrile	0.1% formic acid in acetonitrile
Flow rate	0.5 mL/min	0.5 mL/min	0.5 mL/min	0.4 and 0.6 mL/min
Column	Luna® C18, 3 µm 50 × 2 mm	Kinetex® C18, 2.6 µm, 100 × 3 mm	Kinetex® C18, 2.6 µm, 100 × 3 mm	XSelect HSS C18 XP, 2.5 µm, 150 × 2.1 mm
Column Temperature	55 °C	40 °C	40 °C	45 °C
Injection volume	25 µL	50 µL	50 µL	10 µL
Run time	6 min	7 min	7 min	5.44 min

Table 3
Liquid chromatography gradients used in each laboratory.

SPH		UCSF		KP		UCSD	
Minutes	% MPB	Minutes	% MPB	Minutes	% MPB	Minutes	% MPB
0–0.25	25	0–1.5	10–65	0–5.2	10–93	0.01–2.8	10–40
0.25–3.25	25–95	1.5–2.5	65–70	5.2–5.9	hold at 93	2.8–3.8	40–90
3.25–3.95	hold at 95	2.5–4.3	hold at 70	5.9–5.95	93–10	3.8–4.18	hold at 90
4.0–6.0	hold at 25	4.3–5.5	70–95	5.95–7	hold at 10	4.19	90–95
		5.5–5.7	hold at 95			4.19–4.93	hold at 95
		5.9–7.0	hold at 10			4.94	95–10
						4.94–5.44	hold at 10

Table 4
Summary of mass spectrometry transitions used in the different laboratories.

Parameter	SPH	UCSF	KP	UCSD
Mass spectrometer used	SCIEX API5000	SCIEX 6500 QTRAP	SCIEX 5500	Waters XEVO TQS
Testosterone Transition 1	289/97	289/97	289/97	289/97
Testosterone Transition 2	289/109	289/109	289/109	289/109
Internal standard transition(s)	292/97	292/97 292/109	292/97	292/100

(169.2 ng/dL) ng/dL).

Each laboratory's testosterone results were compared to the All Method Mean (AMM) calculated from the testosterone results for each patient sample obtained from the four laboratories included in this study. Initially, the UCSF results showed the largest mean percent difference (+11.13%) from the AMM (Fig. 1A) versus a mean percent difference of −0.22% when testosterone was quantified by UCSF using the SPH calibrators (Fig. 1B).

To address the large mean difference for the AMM, UCSF adjusted the calibrator values based on the NIST SRM 971, and re-calculated the testosterone concentrations for the patient samples. After re-verification of the calibrator values, the mean percent difference between UCSF and the AMM was reduced to −4.88% (Fig. 1C).

Overall, there was excellent agreement between the testosterone concentrations reported by all 4 laboratories using their own calibrators with regression slopes ranging from 0.946 to 1.034, y-intercepts ranging from −0.121 to 0.078 and the coefficient of determination (R^2) ranging from 0.995 to 0.999 when comparing each laboratory's testosterone result to the AMM (Fig. 2A).

The mean percent bias from the AMM ranged from −4.88% to +3.68% across all concentrations (Fig. 2B).

Since the dispersion of the biases cannot be easily represented in Fig. 2B, they were calculated for data grouped by AMM concentrations as follows: ≤1 nmol/L (≤29 ng/dL) (n = 24, mean = 0.60 nmol/L (17 ng/dL)), 1 to ≤5 nmol/L (29 to ≤144 ng/dL) (n = 22,

mean = 2.28 nmol/L (66 ng/dL)), and >5 nmol/L (>144 ng/dL) (n = 56, 14.01 nmol/L (404 ng/dL)). The mean biases and standard deviation for each laboratory in the different AMM concentration groups are shown in Table 6.

At low concentrations of <1 nmol/L (<29 ng/dL), regression slopes ranged from 0.914 to 1.089, y-intercepts were between −0.040 and 0.044, R^2 ranged from 0.975 to 0.990 (Fig. 3A) and the mean percent bias from the AMM was between −4.47% to +3.63% (Fig. 3B).

4. Discussion

In clinical laboratories, total testosterone is most commonly measured by immunoassay or LC-MS/MS. Immunoassays are easier to implement, and are technically simple for clinical laboratory technicians to perform; however, they suffer from a lack of specificity, especially at low testosterone concentrations found in female and pediatric patients [1,22]. LC-MS/MS assays are more technically complex than immunoassays, but they offer increased specificity and sensitivity, provided they are developed correctly [1,22]. The potential for inaccurate results due to calibration errors, however, is shared by both platforms [22]. A number of papers have been published documenting the differences between immunoassays and GC-MS or LC-MS/MS assays measuring testosterone [2,4,13,14,15]. This has led to an increased awareness of the lack of standardization and harmonization for these assays and, subsequently, to the creation of the HoSt program by the CDC [6].

The broad range of testosterone results reported by proficiency testing programs for different immunoassays at low testosterone concentrations indicates the poor accuracy of these assays at low testosterone concentrations (for example, College of American Pathologists (CAP) and United Kingdom National External Quality Assessment Scheme (UK NEQAS)). One proficiency testing provider, recognizing this issue, has recently changed the target values for testosterone in serum, for both low and high concentration samples, to be the LC-MS/MS assay mean, rather than the AMM (UK NEQAS).

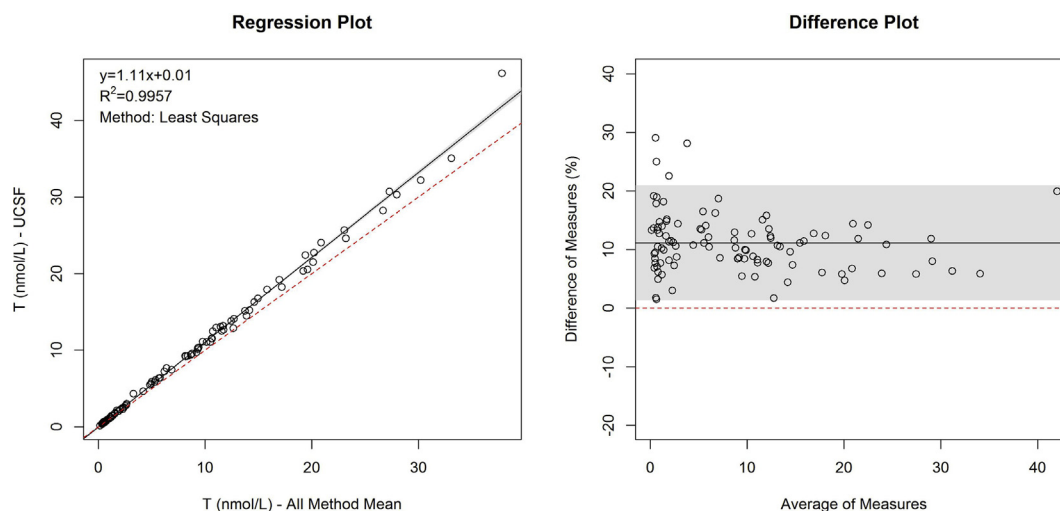The CDC HoSt program has been successful in improving

Table 5
Selected analytical parameters for the different LC-MS/MS assays.

Parameter	SPH	UCSF	KP	UCSD
Analytical measurement range	0.05–45 nmol/L (1.44–1296 ng/dL)	0.07–35.67 nmol/L (2 – 1028 ng/dL)	0.17–52.05 nmol/L (5–1500 ng/dL)	0.14–54.13 nmol/L (4–1560 ng/dL)
Reportable range	0.05–45 nmol/L (1.44–1296 ng/dL)	0.07–104.10 nmol/L (2 – 3000 ng/dL)	0.17–52.05 nmol/L (5–1500 ng/dL)	0.14–54.13 nmol/L (4–1560 ng/dL)
Concentration and imprecision of low QC sample	0.14 nmol/L (4.03 ng/dL) 6.6%	0.24 nmol/L (7 ng/dL) 7.5%	0.7 nmol/L (20.2 ng/dL) 3.4%	0.31 nmol/L (8.8 ng/dL) 5.2%
Concentration and imprecision of middle QC sample	3.80 nmol/L (109.4 ng/dL) 4.4%	6.07 nmol/L (175 ng/dL) 3.8%	6.66 nmol/L (191.9 ng/dL) 2.1%	3.09 nmol/L (89 ng/dL) 4.3%
Concentration and imprecision of high QC sample	21.75 nmol/L (626.7 ng/dL) 4.2%	17.11 nmol/L (493 ng/dL) 3.4%	20.26 nmol/L (584.0 ng/dL) 1.99%	17.72 nmol/L (511 ng/dL) 3.3%

Regression Plot

Difference Plot

Fig. 1A. Linear regression analysis and mean percent difference plot comparing testosterone results of the AMM versus results obtained at UCSF when calculated using UCSF calibrators.

measurement accuracy, as demonstrated by a 50% decrease in mean absolute bias between mass spectrometry assays compared to the CDC RMP between 2007 and 2011 [22]. With the same goal of traceability and accuracy to a RMP, the LC-MS/MS assays described in this manuscript were developed using the NIST SRM 971 to verify calibration. The assays were independently developed on different instrumentation, using different reagents and calibrators, and in some cases using different sample preparation procedures, with and without automation.

It should be noted that, although LC-MS/MS is considered a sensitive, specific, precise and accurate technique for the measurement of testosterone [1], itis also subject to errors that can result in incorrect measurement if methods are not properly developed, validated and implemented [23,24]. When developing a LC-MS/MS assay, it is critical to assess and mitigate sources of error that could contribute to over- or under-estimation of testosterone values [25,26]. Isobaric and isomeric steroids, such as dehydroepiandrosterone and epitestosterone, need to be chromatographically separated from testosterone, and care should be taken to minimize crosstalk from product ions of other steroid hormones common to testosterone [24,25,26]. Gel serum separator tubes have been demonstrated to be a source of interference that results in higher than actual testosterone concentrations reported by LC-MS/MS [19,27,28], and fluoride-containing tubes have been shown to result in

lower than actual testosterone concentrations reported by LC-MS/MS [28]. Matrix effects (commonly ion suppression or ion enhancement) can also be a cause of inaccuracy in LC-MS/MS assays due to differential ionization of sample components [24,29].

In this study, testosterone concentrations obtained at each laboratory were compared to the AMM, with testosterone values at UCSF showing the largest percent difference in values. To investigate this, the UCSF testosterone concentrations were re-calculated using the SPH calibrators, resulting in a decrease in the mean percent difference from the AMM indicating a bias in the initial calibration values. To address this, the UCSF testosterone assay calibration was adjusted using the NIST SRM 971 to re-verify the calibrator concentrations. The patient sample results were re-analyzed using the newly assigned calibrator concentrations and compared to the AMM with significant improvement. When UCSF had initially developed the testosterone LC-MS/MS assay, the NIST SRM 971 was used to verify the calibration of the assay [19]. Subsequently, when new lots of calibrators were put into use, the concentrations of the new lots were verified based upon comparisons with the current in-use lot of calibrators. UCSF was already reporting total testosterone by immunoassay when the LC-MS/MS assay was first implemented and so this method was used for CAP proficiency testing. Initially, the LC-MS/MS method was only compared with the
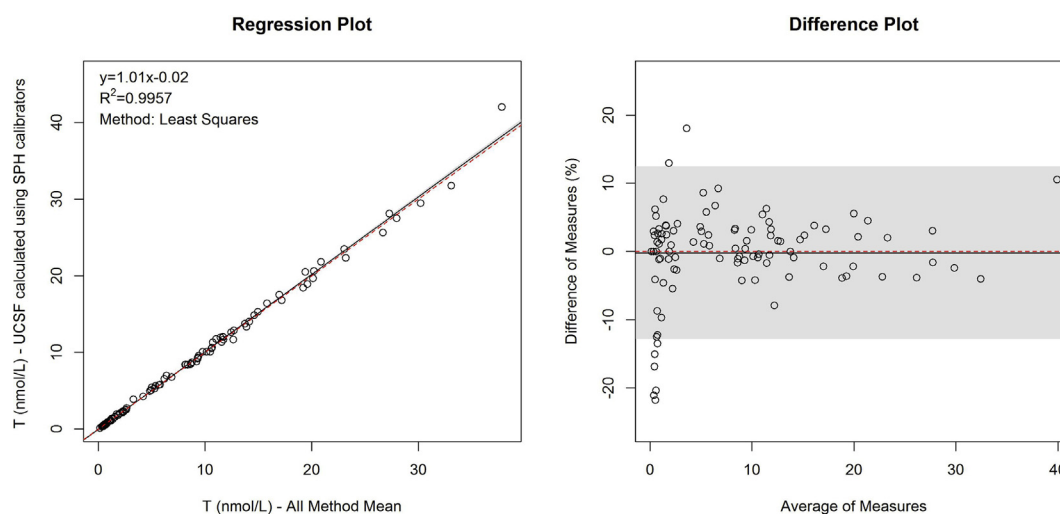
Regression Plot

Difference Plot

Fig. 1B. Linear regression analysis and mean percent difference plot comparing testosterone results of the AMM versus results obtained at UCSF when calculated using SPH calibrators.
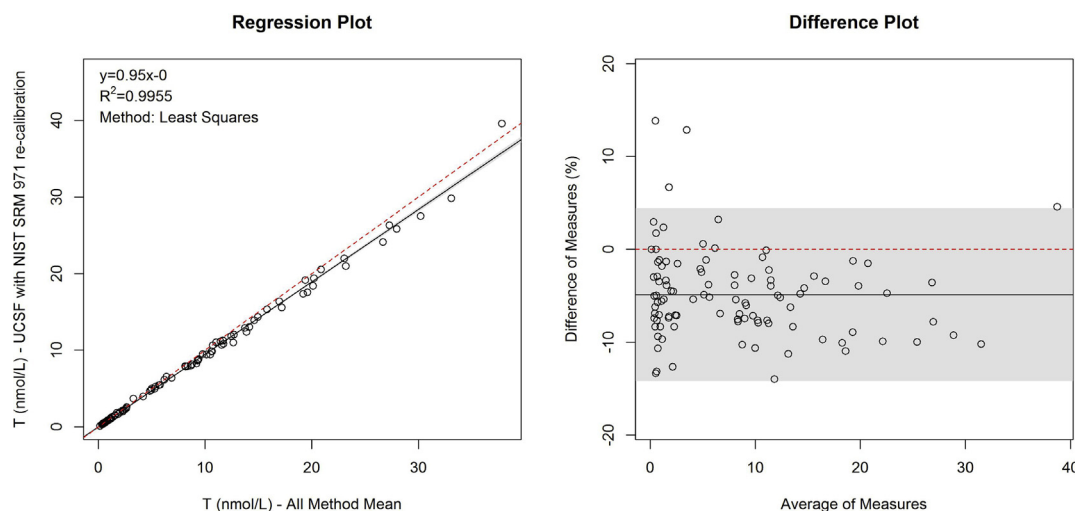
Regression Plot

Difference Plot

Fig. 1C. Linear regression analysis and mean percent difference plot comparing testosterone results of the AMM versus results obtained at UCSF when calculated using UCSF calibrators after re-verification of the concentrations using NIST SRM 971.
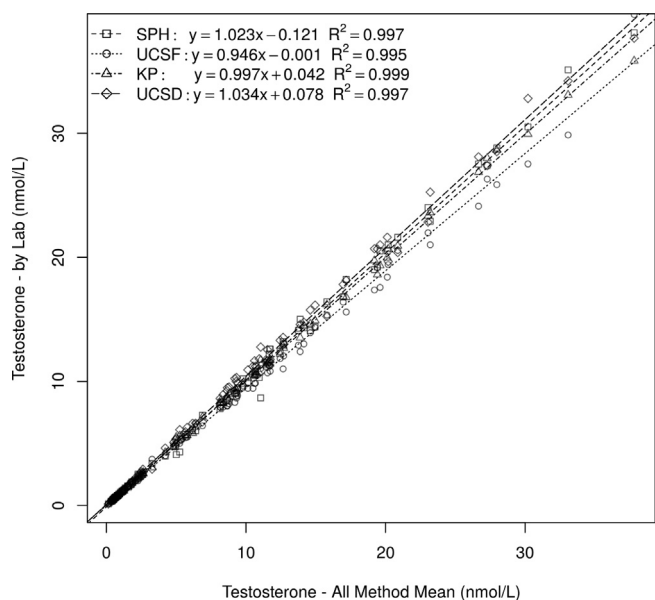

Fig. 2A. Linear regression analysis comparing testosterone results obtained by each laboratory.

Percent Bias

Fig. 2B. Mean percent difference plot comparing testosterone results obtained by each laboratory.

immunoassay every 6 months, per CAP guidelines. UCSF has since implemented a procedure to verify the concentration of new calibrator lots that includes using the NIST SRM 971, the CAP Y-ligand proficiency survey and the CAP ABS accuracy-based survey after the immunoassay method results have been reported to CAP and CAP has reported the results back to the laboratory allowing the LC-MS/MS results to be compared to the target concentration and/or the peer-group mean.

An alternative option for laboratories to reduce the cost associated with using the NIST SRM 971 material is to run accuracy-based proficiency samples or inter-method comparison samples to determine if any bias has been introduced by the new lot of calibrators. If a bias is observed, then the NIST SRM 971 could be used to verify the concentration of the new calibrators. If no bias is observed, it may not be necessary to use the NIST SRM 971 material.

Comparison of patient samples with concentrations < 1 nmol/L (< 29 ng/dL) is important as testosterone concentrations in this range are typical for pediatric and female patients. It has previously been shown that testosterone immunoassays are inaccurate in this concentration range [2]. Comparison between the four LC-MS/MS assays in
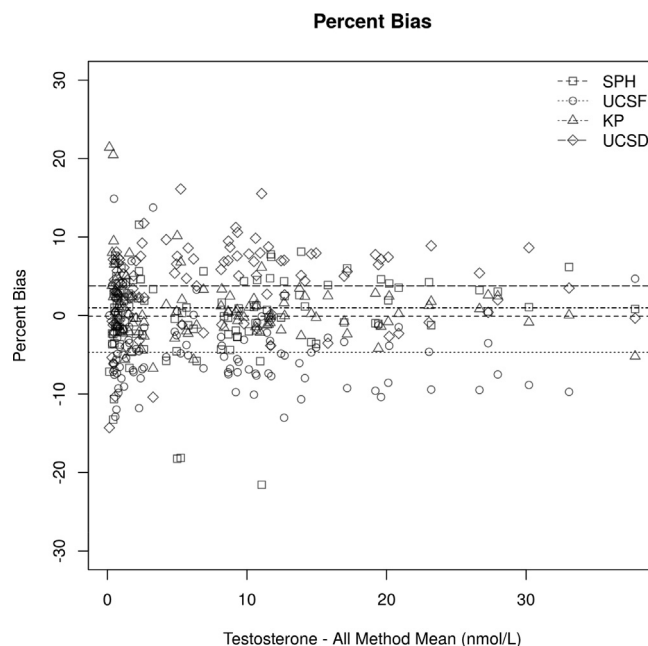
this study was excellent even at these low concentrations with a mean percent bias of −4.5 to +3.6%. This is well within the criteria set out by the CDC HoSt Program, namely +/− 6.4% mean bias to the CDC Testosterone RMP over the concentration range of 0.09–34.70 nmol/L (2.5–1000 ng/dL), which is derived from data on the published biological variation of testosterone [10]. However, since the measurement bias reported here was to an All Method Mean, it remains to be seen how these harmonized methods compare to a RMP.

While other publications have compared testosterone LC-MS/MS assays [16,17,18], this is the first study to compare assays that have each used the NIST SRM 971 to verify calibration. In one study comparing four LC-MS/MS assays to a RMP, one of the assays performed well with a +0.4% mean difference across the concentration range tested, but the others had mean differences equal to, or greater than, the CDC HoSt program criteria of +/− 6.4% [16]. Two other studies comparing routine LC-MS/MS methods found that the inter-method mean difference was 14% and 24% for female samples, respectively,

Table 6
Dispersion of biases in difference concentration ranges.

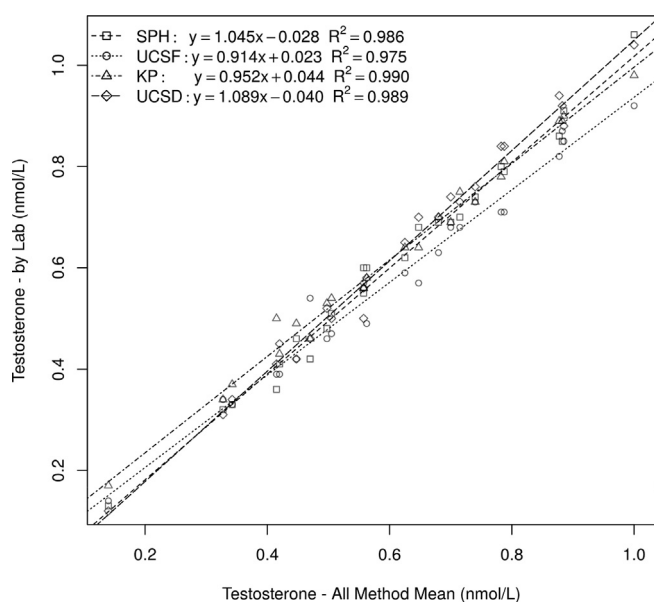| All method mean concentration group	Mean bias	Standard deviation
≤1 nmol/L (≤29 ng/dL)	SPH: 0.00 nmol/L (0 ng/dL) UCSF: −0.03 nmol/L (−0.86 ng/dL) KP: 0.02 nmol/L (0.58 ng/dL) UCSD: 0.01 nmol/L (0.29 ng/dL)	SPH: 0.03 nmol/L (0.86 ng/dL) UCSF: 0.04 nmol/L (1.15 ng/dL) KP: 0.02 nmol/L (0.58 ng/dL) UCSD: 0.03 nmol/L (0.86 ng/dL)
1 to ≤ 5 nmol/L (29 to ≤144 ng/dL)	SPH: 0.00 nmol/L (0 ng/dL) UCSF: −0.07 nmol/L (−2.02 ng/dL) KP: −0.02 nmol/L (−0.58 ng/dL) UCSD: 0.01 nmol/L (0.29 ng/dL)	SPH: 0.11 nmol/L (3.17 ng/dL) UCSF: 0.14 nmol/L (4.03 ng/dL) KP: 0.07 nmol/L (2.02 ng/dL) UCSD: 0.03 nmol/L (0.86 ng/dL)
> 5 nmol/L (144 ng/dL)	SPH: 0.13 nmol/L (3.75 ng/dL) UCSF: −0.79 nmol/L (−22.77 ng/dL) KP: 0.03 nmol/L (0.86 ng/dL) UCSD: 0.62 nmol/L (17.87 ng/dL)	SPH: 0.65 nmol/L (18.73 ng/dL) UCSF: 0.81 nmol/L (23.34 ng/dL) KP: 0.40 nmol/L (11.53 ng/dL) UCSD: 0.66 nmol/L (19.02 ng/dL)

SPH: $y = 1.045x − 0.028$ $R^2 = 0.986$
UCSF: $y = 0.914x + 0.023$ $R^2 = 0.975$
KP: $y = 0.952x + 0.044$ $R^2 = 0.990$
UCSD: $y = 1.089x − 0.040$ $R^2 = 0.989$

Fig. 3A. Linear regression analysis comparing low level testosterone results of < 1 nmol/L (< 29 ng/dL) obtained by each laboratory.

Fig. 3B. Mean percent difference plot comparing low level testosterone results of < 1 nmol/L (< 29 ng/dL) obtained by each laboratory.

and 8% and 14% for male samples, respectively [17,18]. Since testosterone concentrations in female samples are low, it follows that the percent differences are magnified in this range. In this study we have demonstrated excellent comparison at all concentrations with no significant increase in the percent mean difference between the individual methods and the AMM at low testosterone concentrations compared to the entire concentration range.

5. Conclusions

The results of this four-way comparison study demonstrate that independently developed LC-MS/MS testosterone assays can be harmonized using a standard reference material. Efforts to ensure accurate calibration should be taken not only when validating an assay, but also when new lots of calibrators are placed into use. Even though commercially supplied calibrators may have assigned values and certificates of analysis, it is suggested that the concentrations should still be verified with a standard reference material

Conflict of interest

Daniel T. Holmes reports the loan of an instrument from SCIEX. The rest of the authors have no conflicts of interest to disclose.
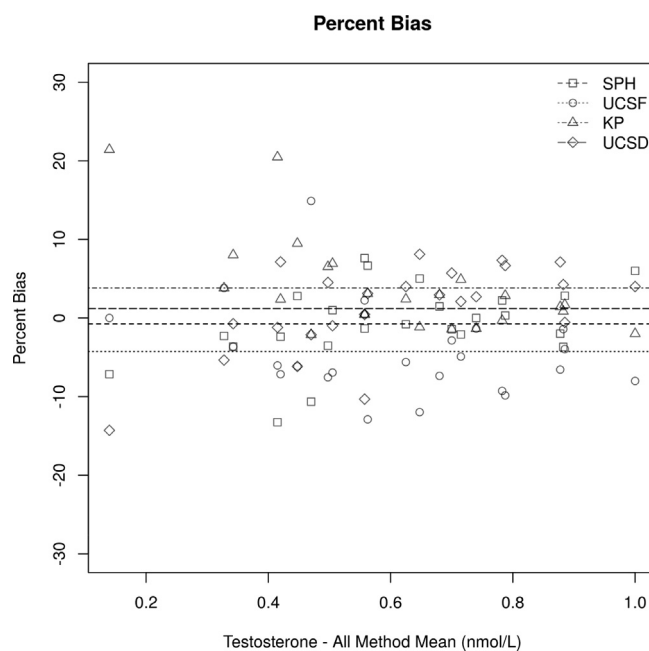
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clinms.2018.11.005.

References

[1] W. Rosner, R.J. Auchus, R. Azziz, P.M. Sluss, H. Raff, Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement, J. Clin. Endocrinol. Metab. 92 (2) (2007) 405–413, https://doi.org/10.1210/jc.2006-1864.
[2] J. Taieb, B. Mathian, F. Millot, M.C. Patricot, E. Mathieu, N. Queyrel, I. Lacroix, C. Somma-Delpero, P. Boudou, Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women and children, Clin. Chem. 49 (8) (2003) 1381–1395, https://doi.org/10.1373/49.8.1381.
[3] A.M. Matsumoto, W.J. Bremner, Serum testosterone assays – accuracy matters, J. Clin. Endocrinol. Metab. 89 (2) (2004) 520–524, https://doi.org/10.1210/jc.2003-032175.
[4] C. Wang, D.H. Catlin, L.H. Demers, B. Starcevic, R.S. Swerdloff, Measurement of total serum testosterone in adult men: comparison of current laboratory methods

versus liquid chromatography-tandem mass spectrometry, J. Clin. Endocrinol. Metab. 89 (2) (2004) 534–543, https://doi.org/10.1210/jc.2003-031287.

[5] D.A. Herold, R.L. Fitzgerald, Immunoassays for testosterone in women: better than a guess? Clin. Chem. 49 (8) (2003) 1250–1251, https://doi.org/10.1373/49.8.1250.

[6] https://www.cdc.gov/labstandards/hs.html, (Accessed 04 April 2018).

[7] H.W. Vesper, S. Bhasin, C. Wang, S.S. Tai, L.A. Dodge, R.J. Singh, J. Nelson, S. Ohorodnik, N.J. Clarke, W.A. Salameh, C.R. Parker Jr, R. Razdan, E.A. Monsell, G.L. Myers, Interlaboratory comparison study of serum total testosterone measurements performed by mass spectrometry methods, Steroids 74 (6) (2009) 498–503, https://doi.org/10.1016/j.steroids.2009.01.004.

[8] S.S.C. Tai, B. Xu, M.J. Welch, K.W. Phinney, Development and evaluation of a candidate reference measurement procedure for the determination of testosterone in human serum using isotope dilution liquid chromatography/tandem mass spectrometry, Anal. Bioanal. Chem. 388 (5–7) (2007) 1087–1094, https://doi.org/10.1007/s00216-007-1355-3.

[9] J.C. Botelho, C. Shacklady, H.C. Cooper, S.S. Tai, K. Van Uytfanghe, L.M. Thienpont, H.W. Vesper, Isotope-dilution liquid chromatography-tandem mass spectrometry candidate reference method for total testosterone in human serum, Clin. Chem. 59 (2) (2013) 372–380, https://doi.org/10.1373/clinchem.2012.190934.

[10] https://www.cdc.gov/labstandards/pdf/hs/CDC_Certified_Testosterone_Procedures.pdf, (Accessed 04 April 2018).

[11] https://www-s.nist.gov/srmors/view_detail.cfm?srm=971, (Accessed 26 April 2018).

[12] https://www-s.nist.gov/srmors/certificates/971.pdf. (Accessed April 4th 2018).

[13] W.E. Owen, M.L. Rawlins, W.L. Roberts, Selected performance characteristics of the Roche Elecsys testosterone II assay on the Modular analytics E170 analyzer, Clin. Chim. Acta 411 (15–16) (2010) 1073–1079, https://doi.org/10.1016/j.cca.2010.03.041.

[14] Y. Chen, M. Yazdanpanah, B.R. Hoffman, E.P. Diamandis, P.Y. Wong, Rapid determination of serum testosterone by liquid chromatography-isotope dilution tandem mass spectrometry and a split sample comparison with three automated immunoassays, Clin. Biochem. 42 (6) (2009) 484–490, https://doi.org/10.1016/j.clinbiochem.2008.11.009.

[15] V. Moal, E. Mathieu, P. Reynier, Y. Malthièry, Y. Gallois, Low serum testosterone assayed by liquid chromatography-tandem mass spectrometry. Comparison with five immunoassay techniques, Clin. Chim. Acta 386 (1–2) (2007) 12–19, https://doi.org/10.1016/j.cca.2007.07.013.

[16] L.M. Thienpont, K. Van Uytfanghe, S. Blincko, C.S. Ramsey, H. Xie, R.C. Doss, B.G. Keevil, L.J. Owen, A.L. Rockwood, M.M. Kushnir, K.Y. Chun, D.W. Chandler, H.P. Field, P.M. Sluss, State-of-the-art of serum testosterone measurement by isotope dilution-liquid chromatography-tandem mass spectrometry, Clin. Chem. 54 (8) (2008) 1290–1297, https://doi.org/10.1373/clinchem.2008.105841.

[17] R.M. Büttler, F. Martens, F. Fanelli, H.T. Pham, M.M. Kushnir, M.J. Janssen, L. Owen, A.E. Taylor, T. Soeborg, M.A. Blankenstein, A.C. Heijboer, Comparison of 7 published LC-MS/MS methods for the simultaneous measurement of testosterone, androstenedione and dehydroepiandrosterone in serum, Clin. Chem. 61 (12) (2015) 1475–1483, https://doi.org/10.1373/clinchem.2015.242859.

[18] R.M. Büttler, F. Martens, M.T. Ackermans, A.S. Davison, A.E. van Herwaarden, L. Kortz, J.G. Krabbe, E.G. Lentjes, C. Syme, R. Webster, M.A. Blankenstein, A.C. Heijboer, Comparison of eight routine unpublished LC-MS/MS methods for the simultaneous measurement of testosterone and androstenedione in serum, Clin. Chim. Acta 454 (2016) 112–118, https://doi.org/10.1016/j.cca.2016.01.002.

[19] D. French, Development and validation of a serum total testosterone liquid chromatography-tandem mass spectrometry (LC-MS/MS) assay calibrated to NIST SRM 971, Clin. Chim. Acta 415 (2013) 109–117, https://doi.org/10.1016/j.cca.2012.10.007.

[20] J.A. Stone, D.R. van Staveren, R.L. Fitzgerald, Automated sample preparation enables LC-MS/MS as a routine diagnostic analysis for serum testosterone, J. Appl. Lab. Med. 2 (1) (2017) 33–46, https://doi.org/10.1373/jalm.2016.022772.

[21] D.T. Holmes, cp-R, an interface the R programing language for clinical laboratory method comparisons, Clin. Biochem. 48 (3) (2015) 192–195, https://doi.org/10.1016/j.clinbiochem.2014.10.015.

[22] H.W. Vesper, J.C. Botelho, Y. Wang, Challenges and improvements in testosterone and estradiol testing, Asian J. Androl. 16 (2) (2014) 178–184, https://doi.org/10.4103/1008-682X.122338.

[23] C. Seger, M. Vogeser, Pitfalls of LC-MS/MS in the clinical laboratory, in: Q.A. Xu, T.L. Madden (Eds.), LC-MS in Drug Bioanalysis, first ed., Springer Science + Business Media, New York, 2012, pp. 109–126, , https://doi.org/10.1007/978-1-4614-3828-1_5.

[24] M. Vogeser, C. Seger, Pitfalls associated with use of liquid chromatography-tandem mass spectrometry in the clinical laboratory, Clin. Chem. 56 (8) (2010) 1234–1244, https://doi.org/10.1373/clinchem.2009.138602.

[25] CLSI. Mass Spectrometry for Androgen and Estrogen Measurements in Serum; Approved Guideline. CLSI document C57. Clinical and Laboratory Standards Institute, Wayne, PA, 2015.

[26] CLSI. Liquid Chromatography-Mass Spectrometry methods; Approved Guideline. CLSI document C62-A. Clinical and Laboratory Standards Institute, Wayne, PA, 2014.

[27] R.Z. Shi, H.H. van Rossum, R.A.R. Bowen, Serum testosterone quantitation by liquid chromatography-tandem mass spectrometry: interference from blood collection tubes, Clin. Biochem. 45 (18) (2012 Dec) 1706–1709, https://doi.org/10.1016/j.clinbiochem.2012.08.008.

[28] C. Wang, S. Shiraishi, A. Leung, S. Baravarian, L. Hull, V. Goh, P.W. Lee, R.S. Swerdloff, Validation of a testosterone and dihydrotestosterone liquid chromatography tandem mass spectrometry assay: interference and comparison with established methods, Steroids 73 (13) (2008) 1345–1352, https://doi.org/10.1016/j.steroids.2008.05.004.

[29] T.M. Annesley, Ion suppression in mass spectrometry, Clin. Chem. 49 (7) (2003) 1041–1044, https://doi.org/10.1373/49.7.1041.