Psychoneuroendocrinology 71 (2016) 180–188

Contents lists available at ScienceDirect

Psychoneuroendocrinology

journal homepage: www.elsevier.com/locate/psyneuen

A comparison of salivary testosterone measurement using immunoassays and tandem mass spectrometry

Keith M. Welker[a], Bethany Lassetter[b], Cassandra M. Brandes[c], Smrithi Prasad[c], Dennis R. Koop[d], Pranjal H. Mehta[c,*]

[a] University of Massachusetts Boston, Boston, MA 02125, United States
[b] University of Iowa, Iowa City, IA 52242, United States
[c] University of Oregon, Eugene, OR 97403, United States
[d] Oregon Health and Science University, Portland, OR 97239, United States

ARTICLE INFO

Article history:
Received 10 March 2016
Received in revised form 20 May 2016
Accepted 20 May 2016

Keywords:
Salivary testosterone
Immunoassays
Liquid chromatography
Mass spectrometry
Measurement

ABSTRACT

Enzyme immunoassays (EIAs) are widely used to measure salivary testosterone. However, little is known about how accurately different EIAs assess testosterone, partially because estimates across various EIAs differ considerably. We compared testosterone concentrations across EIAs of three commonly used manufacturers (DRG International, Salimetrics, and IBL International) to liquid chromatography tandem mass spectrometry (LC–MS/MS). Relative to EIAs from Salimetrics and IBL International, EIAs supplied by DRG International provided the closest approximation to LC–MS/MS testosterone concentrations, followed closely by EIAs from Salimetrics, and then IBL. Additionally, EIAs tended to inflate estimates of lower testosterone concentrations in women. Examining our results and comparing them to existing data revealed that testosterone EIAs had decreased linear correspondence with LC–MS/MS in comparison to cortisol EIAs. Overall, this paper provides researchers with information to better measure testosterone in their research and more accurately compare testosterone measurements across different methods.

1. Introduction

Testosterone—an androgenic steroid hormone and end product of the hypothalamic-pituitary-gonadal axis—has been increasingly studied by psychoneuroendocrinology researchers for its associations with behaviors and psychological processes implicated in social status and affiliation. These include dominance (e.g., Mazur and Booth, 1998), risk-taking (e.g., Apicella et al., 2014), romantic relationships (Edelstein et al., 2014), sexual behavior (e.g., Puts et al., 2015), aggression (e.g., Carré et al., 2014), and competition (e.g., Mehta et al., 2015), among others. To assess testosterone levels, researchers across disciplines have adopted the use of salivary hormone analysis. This technique has several advantages, including ease of collection, cost-effectiveness, non-invasiveness, and the ability to assess salivary testosterone changes over short intervals of time.

1.1. Assessment of salivary hormones through EIAs

Salivary hormone concentrations are frequently measured with enzyme immunoassays (EIAs). EIAs are cost-effective and convenient, making them attractive tools for measuring hormones like testosterone. Although many companies manufacture assay kits for commercial use, recent work suggests that these different kits vary in concentration predictions for the same hormones (e.g., Taieb et al., 2003; Baecher et al., 2013; Crewther et al., 2013). Some potential reasons for differences in the estimation of hormone levels across kits include differing levels of sensitivity and specificity across the range of hormonal concentrations, with the lowest and highest concentrations being most prone to quantification errors (Schultheiss and Stanton, 2009). This difference presents a potential obstacle to the accurate assessment of testosterone, especially in populations with lower levels of testosterone such as women and children (Rosner et al., 2007). Cross-reactivity acts as another potential source of inaccuracy in hormone assessment (i.e., inflation) with EIAs. Cross-reactivity occurs when chemically similar compounds are measured in addition to the hormone of interest because these compounds bind to the same receptors as the target analyte. Testosterone EIAs are found to cross-react with other steroids (Chattoraj, 1976) and biological agents (e.g., sex hormone-

* Corresponding author at: Department of Psychology, University of Oregon, United States.
E-mail address: mehta@uoregon.edu (P.H. Mehta).

http://dx.doi.org/10.1016/j.psyneuen.2016.05.022
0306-4530/

binding globulin; Pugeat et al., 1981). Consistent with this research on testosterone measurement, previous studies of cortisol have revealed differences in predicted salivary cortisol concentrations across assay kits (e.g., Baecher et al., 2013; Miller et al., 2013). These differences have been attributed to differing specifications of kits and cross-reactivity in EIAs.

Despite the limitations of EIA, the method's relative ease and cost-effectiveness has made it a popular technique for hormone measurement. Thus, it is critical to assess the validity of EIAs in the measurement of salivary testosterone. To do so, we propose to compare popular commercially available testosterone EIAs with testosterone measured using liquid chromatography tandem mass spectrometry (LC–MS/MS), an alternative measurement technique for salivary hormone analysis that is free from some of the limitations posed by EIAs.

1.2. Assessing the validity of EIAs with liquid chromatography tandem mass spectrometry

The use of LC–MS/MS as an analytical tool has become popular in a growing number of laboratories seeking to quantify hormone concentrations (Soldin and Soldin, 2009; Field, 2013; Keefe et al., 2014). The technique demonstrates impressive analytical specificity and is argued to be free of many of the limitations of immunoassays (e.g. matrix interference; Hoofnagle and Wener, 2009). Although the financial and logistical requirements of LC–MS/MS limit its widespread use, the technique is a sensitive reference measure, allowing for both the identification and quantification of compounds by combining the physical separation capacity of liquid chromatography with the mass analysis capability of mass spectrometry (Star-Weinstock et al., 2012; Turpeinen et al., 2012; Keevil, 2013). Miller et al. (2013) recently evaluated the validity and agreement between various immunoassays and LC–MS/MS assessing cortisol concentrations in adult saliva samples. Although one EIA (Demeditec) has been found to have a strong correlation with LC–MS/MS testosterone (Yasuda et al., 2008), no published research, to our knowledge, has directly compared how multiple EIA assessments of salivary testosterone are associated with those of LC–MS/MS.

1.3. Overview of the current research

In an effort to bridge this gap, we compared the accuracy of salivary testosterone measured via EIAs from Salimetrics, DRG International, and IBL International (hereby DRG and IBL, respectively) to LC–MS/MS in the present study. We also compared the shape, spread, and center of the concentration distributions from different EIAs with LC–MS/MS to examine how closely these commercially manufactured EIAs approximate LC–MS/MS concentrations of testosterone. As a secondary goal, we estimated salivary cortisol concentrations with LC–MS/MS to investigate if associations with testosterone and LC–MS/MS cortisol varied as a function of the testosterone EIA used.

2. Methods

2.1. Samples

One hundred saliva samples obtained via passive drool were drawn from a large pool previously assayed for salivary testosterone using Salimetrics EIAs (1-2402, Salimetrics, State College, PA, USA; *M* testosterone = 103.11 pg/mL; *SD* = 63.41 pg/mL) and cortisol using IBL EIAs (RE52631, IBL International, Toronto, ON, CA). All procedures were conducted with the adequate understanding and written consent of the participants in accordance with the Declaration of Helsinki. Samples from this large pool were collected over the course of the 2012 United States Presidential Election (Prasad et al., in prep; see Supplemental materials for methodological details). To ensure adequate volume for analysis across multiple methods, only samples with ≥1400 μL of saliva were selected. The measured testosterone concentrations in these samples represented a wide dispersion (Minimum = 13.79 pg/mL, 25th percentile = 56.29 pg/mL, Median = 83.52 pg/mL, 75th percentile = 145.43 pg/mL, Maximum = 281.07 pg/mL) and a relatively even gender split (58% female). An in-depth description of the aliquoting, storage, and shipping procedures is available in the Supplemental materials.

2.2. Analysis with EIAs

Saliva samples were assayed for testosterone in-house within the Social Psychoneuroendocrinology Laboratory at the University of Oregon (UO) using three commonly used commercially-available competitive EIAs (DRG, IBL, and Salimetrics) in accordance with protocols[1] and specifications provided by the manufacturers (DRG International, Inc., 2011; IBL International Corp., 2013; Salimetrics, LLC, 2014). Samples were assayed in duplicate, and those yielding coefficients of variation (CVs) between duplicate wells in the highest 10% of the range were re-assayed once to maximize measurement accuracy (Salimetrics: CVs > 17%; IBL: CVs > 11%; DRG: CVs > 18%). After re-assaying the samples with the highest CVs, the average intra-assay CVs for Salimetrics, IBL, and DRG were 6.97%, 5.33%, and 7.80%, respectively. The average inter-assay CVs were as follows: Salimetrics: 8.54%; IBL: 39.94%; DRG: 20.84%.[2] Further, different hormones can have similar chemical structures, which can reduce specificity. Manufacturers test and report the cross-reactivities of a variety of analytes for each of their assays. We present these cross-reactivities across EIAs in the Supplemental materials (see Supplemental Table S1).

As an additional step, the saliva samples were analyzed for cortisol using IBL EIA kits (average intra-assay CV = 7.41%, average inter-assay CVs were 13.35% and 8.35% for low and high controls, respectively). These cortisol data allowed us to conduct secondary analyses in which we compared the associations of testosterone EIAs and cortisol EIAs to LC–MS/MS testosterone concentrations in our own study. Finally, we compared the testosterone EIA and LC–MS/MS correlations in our research to the correlations between EIA cortisol and LC–MS/MS cortisol found by Miller et al. (2013). This allowed us to examine whether testosterone EIAs have similar correspondence with LC–MS/MS values as cortisol EIAs.

2.3. Analysis with liquid chromatography tandem mass spectrometry

Aliquots were analyzed for salivary testosterone and cortisol using LC–MS/MS at Oregon Health and Science University's Bioanalytical Shared Resource/Pharmacokinetics Core labs. LC–MS/MS analysis commenced approximately 5 months after concentrations were assessed with EIAs at UO. Cortisol was also assessed with LC–MS/MS with the secondary goal of providing converging findings to supplement the previously reported associated correlations

[1] At the request of a reviewer, we provide the analytical and functional sensitivities for DRG EIAs (1.9 pg/mL, 7.1 pg/mL, respectively) and IBL EIAs (2.0 pg/mL, 7.6 pg/mL, respectively). The analytical sensitivity of Salimetrics EIAs was <1.0 pg/mL, however Salimetrics did not provide data indicating functional sensitivity.

[2] Although the inter-assay CVs for IBL and DRG appear high, these inter-assay CVs are consistent with those we have found among other data with larger sample sizes for IBL (Mean inter-assay CV = 39.57%; across 5 assays) and DRG (Mean inter-assay CV = 19.42%; across 19 assays). Thus, these higher inter-assay CVs appear to be stable for IBL and DRG EIAs.

between cortisol derived from EIAs and LC–MS/MS (Miller et al., 2013).

Salivary testosterone and cortisol were determined by LC–MS/MS following extraction with ethyl acetate and derivatization with the novel quaternary aminooxy (QAO) mass tag reagent, Amplifex Keto Reagent®, as described by Star-Weinstock et al. (2012). Testosterone (Sigma-Aldrich, St. Louis, MO) concentrated standards were prepared in DMSO and were diluted on the day of analysis with mass spectral grade water:acetonitrile (1:1) with 0.1% formic acid. The internal standard, d_3-testosterone (Cerilliant reference standard, Sigma-Aldrich, St. Louis, MO), stock was prepared in acetonitrile and a working dilution prepared in methanol. Cortisol (Sigma-Aldrich, St. Louis, MO) stocks were prepared in 1:1 DMSO:methanol and working dilutions prepared in methanol. The internal standard, d_3-cortisol, (Cambridge Isotope Labs, Andover, MA) stock was prepared in methanol. The QAO working solution was prepared by adding equal volumes (0.75 mL) of the reagent and the diluent supplied in the Amplifex Keto reagent kit (AB Sciex, Redwood City, CA). The mixture was then further diluted with 4.5 mL of methanol containing 5% acetic acid to a working reagent solution of 2.5 mg/mL.

LC–MS/MS data were acquired and analyzed using Analyst 1.6.2 software. Sample values were calculated from standard curves generated from the peak area ratio of the analyte to internal standard versus the analyte concentration that was fit to a linear equation with 1/x weighting. R values of the regression were 0.999. The lower limit of quantification for testosterone was 1 pg/mL with an accuracy of 114% and precision (relative standard deviation) of 7.1% and the signal to noise (S/N) was 5:1. At a concentration of 2 pg/mL the accuracy was 109% and precision was 4.1% with a S/N of 10.5. The accuracy for cortisol at 50 pg/mL was 103% with a precision of 2.7% with a S/N of 30:1.

Standard curves were prepared by spiking 5 μL of stock solutions into 500 μL of phosphate buffered saline (PBS) to provide final concentrations of 1–500 pg/mL for testosterone and 50–5000 pg/mL for cortisol. The standards and 500 μL of saliva samples were spiked with 5 μL of an internal standard mixture containing d_3-testosterone (1 pg/μL) and d_3-cortisol (10 pg/μL). The samples and standards were vortexed briefly and then 5 mL of ethyl acetate was added. Samples and standards were vortexed for 60 s, centrifuged at 2,000xg and the organic phase removed to a clean glass tube. The ethyl acetate was removed under reduced pressure using a Savant speed vacuum system. The residue was dissolved in 50 μL QAO-reagent (2.5 mg/mL) and incubated at room temperature for 2 h. The solution was filtered with 0.22 μ spin filters (Millipore, Billerica, MA) and placed in sample vials for analysis by LC–MS/MS using an injection volume of 5 μL.[3]

2.4. Statistical analytic strategy

Our analyses used several different techniques to both compare the measurement of testosterone by different EIAs and LC–MS/MS, as well as examine agreement between our EIA measures and LC–MS/MS. In particular, we were interested in three facets of how EIAs assessments of testosterone correspond to those of LC–MS/MS: linear agreement, the presence of fixed and proportional bias, and systematic differences in the measurements of the methods. Pearson's correlations were used to examine linear correspondence between different kits and LC–MS/MS. Bland Altman plots (Bland and Altman, 1986) were used to evaluate the presence of fixed bias (e.g., whether one method systematically estimates higher concentrations compared to the other) and proportional bias (the methods do not agree equally across the range of measurement) in how EIAs assess testosterone with reference to LC–MS/MS testosterone. Typically, Bland Altman plots assess the correspondence between difference scores between two methods and the average of the two method measurements. However, because a highly accurate reference method was available (LC–MS/MS), we examined how differences between the concentrations assessed by two methods (an EIA and LC–MS/MS) were associated with LC–MS/MS testosterone (Krouwer, 2008). Finally, Deming regressions (Martin, 2000) were also used to compare the three EIAs to LC–MS/MS using the package mcr with the R statistical package (R Core Team, 2016). We regressed the EIA values on the LC–MS/MS reference values. This model was used to compare a regression equation between the EIAs and the reference method to an identity line, which assumes equality between the methods (intercept = 0, slope = 1).

3. Results

3.1. General analyses

Descriptive statistics and correlations for all testosterone immunoassays and LC–MS/MS, as well as intra-assay CVs, are presented in Table 1. Although all concentrations are presented in the metric of pg/mL, we provide a table of our descriptive statistics presented in Table 1 in the metric of molar units (See Supplemental Table 2). For interested readers, we also provide information on sex differences in testosterone as measured by the three EIAs and LC-MS/MS (See Supplemental Materials). The distributions of concentrations estimated from each of the four methods are presented in Fig. 1. One notable finding evident in the percentiles of Table 1 is that estimated concentrations of testosterone varied much more considerably in the upper ranges (75th percentiles range from 97.69 pg/mL [DRG] to 194.89 pg/mL [IBL]) compared to the lower ranges (25th percentiles range from 37.46 pg/mL [DRG] to 51.86 pg/mL [Salimetrics]).

Consistent with Table 1 and Fig. 1, IBL's testosterone concentrations showed the highest average concentrations and skewness, followed by Salimetrics, and then DRG. A repeated-measures ANOVA comparing testosterone concentrations revealed significant variability in the means of estimated testosterone concentrations among the four methods (Greenhouse-Geisser $F(1.48, 134.62) = 43.01$, $p < 0.001$, $\eta_p^2 = 0.32$). Furthermore, all possible pairwise comparisons between the concentrations of each kit were significant using a Bonferroni correction for multiple comparisons

[3] For interested readers, we also present the instrument specifications for LC–MS/MS: Derivatized extracts were analyzed using a 5500 QTRAP hybrid/triple quadrupole linear ion trap mass spectrometer (AB Sciex, Redwood City, CA) with electrospray ionization (ESI) in positive mode. The mass spectrometer was interfaced to a Shimadzu (Columbia, MD) SIL-20AC XR auto-sampler and 2 LC-20AD XR LC pumps. The instrument was operated with the following settings: source voltage 3000 kV, GS1 50, GS2 50, CUR 15, TEM 600, and CAD gas HIGH. QAO-testosterone and QAO-cortisol were separated on an Imtakt (Portland, OR) Cadenza CL-C18 3 μ (50 × 2 mm) column with matching guard column (2 × 10 mm) held at 35 °C in a Shimadzu CTO-20AC column oven using a gradient mobile phase delivered at a flow rate of 0.6 mL/min. The two solvents were, A: 0.1% formic acid in water and B: 0.1% formic acid in acetonitrile. Initial conditions were 20% B, increased to 60% B over 3 min, then increased to 95% B held at 95% B for 2 min, and then decreased to start conditions over 0.1 min, then held at start conditions of 20% B for 2 min. Each compound formed *E/Z* geometric isomers that were separated with the gradient (Star-Weinstock et al., 2012). Testosterone eluted with retention times of 2.14 and 2.29 min and cortisol with retention times of 1.38 and 1.48 min. The first peak was used for quantification for each compound. The multiple reaction monitoring (MRM) transitions were optimized by direct infusion of each QAO-derivative and were as follows: QAO-testosterone, m/z 403.2 → 164.2 and m/z 403.2 → 152.1; for QAO-d_3-testosterone, m/z 406.2 → 164.2 and m/z 406.2 → 152.1; for QAO-cortisol, m/z 477.3 → 388.3 and m/z 477.3 → 358.3; and QAO-d_3-cortisol, m/z 480.3 → 391.3 and m/z 480.3 → 361.3. Optimal intensities were obtained at a declustering potential of 76 V, collision energy of 59 eV, and an entrance potential of 10 V for the QAO-testosterone derivatives. For QAO-cortisol derivatives, the optimal intensities were obtained at a declustering potential of 81 V, collision energy of 55 eV, and an entrance potential of 10 V.

Table 1
Descriptive statistics and intercorrelations between EIAs and LC–MS/MS testosterone concentrations.

	LC–MS/MS	Enzyme Immunoassay Concentrations			Enzyme Immunoassay Intra-assay CVs		
		Salimetrics	DRG	IBL	Salimetrics	DRG	IBL
Moments							
Mean	48.58	99.56	71.97	141.10	6.97	7.80	5.33
SD	52.74	69.94	50.67	134.48	3.72	6.19	3.55
Skewness (SE)	1.34 (.24)	1.49 (.24)	1.42 (.24)	1.68 (.25)	0.08 (.24)	1.14 (.24)	1.10 (.25)
Kurtosis (SE)	1.48 (.48)	2.48 (.48)	1.47 (.48)	2.44 (.50)	−0.13 (.48)	1.20 (.48)	1.68 (0.50)
N	99	100	100	93			
Percentiles							
Minimum	1.10	4.94	13.22	8.88	0.03	0.04	0.22
1st Quartile	7.30	51.86	37.46	44.54	5.16	2.82	2.75
Median	19.90	78.54	53.08	87.12	7.21	6.01	4.95
3rd Quartile	76.20	127.80	97.69	194.89	9.04	11.59	7.35
Maximum	238.00	374.45	236.93	596.17	16.96	30.58	18.30
Pearson Correlations (All Samples)							
LC–MS/MS	–						
Salimetrics	0.55**	–			–		
DRG	0.57**	0.67**	–		−0.05	–	
IBL	0.47**	0.71**	0.67**	–	−0.16	−0.04	–
Pearson Correlations (Men)							
LC–MS/MS	–						
Salimetrics	0.17	–			–		
DRG	0.17	0.53**	–		0.39*	–	
IBL	0.10	0.62**	0.59**	–	−0.05	−0.07	–
Pearson Correlations (Women, Excluding an Outlier)							
LC–MS/MS	–						
Salimetrics	0.14	–			–		
DRG	0.22	0.30*	–		−0.19	–	
IBL	−0.17	0.43**	0.30*	–	−0.25†	−0.02	–

Note: Standard errors for skewness and kurtosis (SEs) are in parentheses. LC–MS/MS = Liquid chromatography tandem mass spectrometry, CV = coefficient of variation, SD = standard deviation, Levene's Test = Levene's test for the equality of variances between concentrations in the upper and lower 50% of the testosterone data. The metric of concentrations is pg/mL.

† $p < .10$.
* $p < .05$.
** $p < .001$.

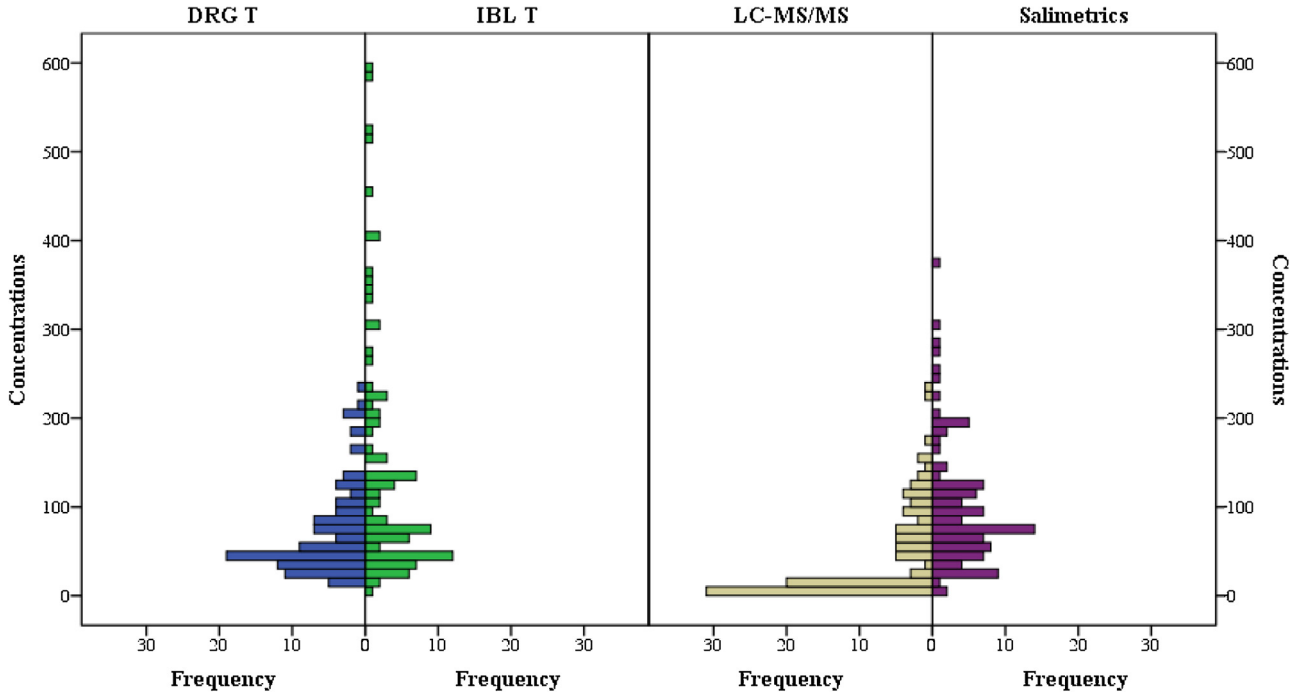

Fig. 1. Histograms of the distributions of testosterone concentrations (pg/mL) derived from each testosterone measurement method.

(all *ps* < 0.001). Examining the means and distributions of the EIAs and LC–MS/MS revealed that all EIAs measured inflated estimates of testosterone compared to LC–MS/MS. Despite having testosterone concentration estimates that were significantly higher than those of LC–MS/MS, the estimated concentrations from DRG EIAs were the most consistent with the concentrations from LC–MS/MS.

3.2. Correlational analyses

We then examined the linear correlations between EIAs and LC–MS/MS testosterone concentrations. DRG's concentrations were the most strongly correlated with the LC–MS/MS results ($r = 0.57$), followed closely by Salimetrics ($r = 0.55$), and then IBL ($r = 0.47$; all $ps < 0.001$). Although Fisher's r to Z transformations revealed no significant differences among these correlation coefficients (all $ps \geq 0.342$); the correlations among the EIAs were higher ($rs \geq 0.67$) than the correlations between LC–MS/MS testosterone and the three EIAs (rs 0.47–0.57). In contrast to previous research correlating cortisol EIAs to LC–MS/MS concentrations (rs from 0.90 to 0.97; Miller et al., 2013), our EIA and LC–MS/MS correlations for testosterone had considerably lower correlations between EIA Testosterone and LC–MS/MS Testosterone (rs 0.47–.57; according to r to Z transformations, $|Z|s \geq 6.57$, $ps \leq 0.001$). To corroborate correlations between testosterone EIAs and LC–MS/MS in this analysis, we also report similar testosterone cross-method immunoassay correlations measured in a preliminary validation study with a smaller sample size (N = 38) in our Supplemental materials.

Because we analyzed cortisol via LC–MS/MS and in previous EIA analysis of this data using IBL EIAs, we also calculated correlations and mean differences between these estimates to compare our results to Miller et al. (2013). IBL cortisol concentrations were strongly correlated with those of LC–MS/MS ($r = 0.80$, $p < 0.001$), but less so compared to the EIA and LC–MS/MS correlations reported by Miller et al. (2013) using a Fisher's r to Z transformation ($Zs \geq 3.01$, $ps < 0.003$), although Miller et al. (2013) did not assess cortisol with an IBL EIA.[4] Consistent with Miller et al. (2013), EIA cortisol concentrations were also considerably higher ($M = 6553.64$ pg/mL, $SD = 6203.27$ pg/mL) than those of LC–MS/MS ($M = 2773.80$ pg/mL, $SD = 2567.99$ pg/mL, $t(98) = -8.49$, $p < 0.001$). We also compared our EIA/reference method testosterone correlations to the EIA/reference method cortisol correlation in our own data using r to Z transformations. This revealed that the EIA to LC–MS/MS correlations for testosterone were all significantly lower than the correlation between IBL cortisol and LC–MS/MS cortisol ($Zs \geq 3.09$, $ps \leq 0.002$). Deming regressions and Bland Altman plots also revealed that the IBL cortisol EIA has significant positive fixed bias and proportional bias (See Supplemental materials). Collectively, our data and the results of Miller et al. (2013) suggest that testosterone EIAs have less linear correspondence with LC–MS/MS compared to cortisol EIAs.

Furthermore, we examined correlations between EIA testosterone concentrations with those of LC–MS/MS in the bottom 50% and top 50% of testosterone concentrations determined by LC–MS/MS. We did this for two primary reasons. First, many researchers are interested in understanding testosterone in populations with lower levels of testosterone (e.g., women, children). Second, researchers have identified that quantification errors in the lower ranges of testosterone may pose a limitation to analyzing testosterone in low-testosterone populations (Schultheiss and Stanton, 2009). In the bottom 50% of the testosterone data, IBL ($r = -0.07$) showed poor linear correspondence with LC–MS/MS testosterone concentrations, whereas DRG ($r = 0.24$) and Salimetrics ($r = 0.27$) were moderately associated with LC–MS/MS concentrations. Fisher's r to Z transformations indicated that these correlations did not significantly differ ($ps \geq 0.131$), with the exception of a marginally significant difference between how LC–MS/MS testosterone was associated with the concentrations of Salimetrics and IBL ($p = 0.097$). In the top 50% of the data, Salimetrics ($r = 0.31$), IBL ($r = 0.27$), and DRG ($r = 0.26$) had similar moderately-sized linear associations with LC–MS/MS testosterone concentrations. These correlations did not significantly differ from each other (Fisher's r to Z transformations, $ps \geq 0.795$). Although these values overall could be seen as suggesting that the three EIAs had only moderate correlations with those of LC–MS/MS, it is important to note that the range restriction introduced by median-splitting our data can restrict statistical power and effect sizes (Cohen, 1988). The directions and magnitude of these correlations suggest more accurate assessment of low-levels of testosterone in DRG and Salimetrics, compared to IBL (See Section 3.4 for analyses within men and women separately).

3.3. Bland Altman plots across all data

We used Bland Altman plots to assess the correspondence between EIA testosterone and LC–MS/MS testosterone. Fig. 2 depicts Bland Altman plots of associations between LC–MS/MS (abbreviated in the figure as MS to avoid appearing as a mathematical expression) and the three EIAs. Visual inspection of these plots indicated a degree of proportional bias for DRG and Salimetrics kits. DRG kits were more likely to estimate inflated values for lower testosterone concentrations, and so were Salimetrics kits, but to a lesser extent. Overall, all EIAs showed significant fixed bias in estimating LC–MS/MS testosterone (95% CIs did not include 0). IBL showed the highest degree of bias (Mean difference = 91.06, 95% CI: [67.11, 115.00]), followed by Salimetrics (Mean difference = 51.33, 95% CI: [39.29, 63.37]), then DRG, which showed the least bias (Mean difference = 23.77, 95% CI: [14.13, 33.42]).

To assess whether differences in assessments of measures changed as a function of average testosterone concentration (proportional bias), we examined associations between EIA and LC–MS/MS difference scores and the LC–MS/MS concentrations. There was a moderate significant association between the concentration differences of methods as testosterone increased for Salimetrics ($r = -0.25$, $p = 0.014$) and a large association for DRG ($r = -0.50$, $p < 0.001$), reflecting a tendency for lower concentrations of hormones to have inflated values from these EIAs. For IBL, there was no association between method differences and testosterone concentrations ($r = 0.05$, $p = 0.647$).

3.4. Sex differences in agreement trends

We also examined if the associations presented in the Bland Altman plots (Fig. 2) varied as a function of sex. For men, method difference scores were negatively associated with testosterone for Salimetrics ($r = -0.41$, $p = 0.007$) and DRG ($r = -0.58$, $p < 0.001$). Although the correlation was in a negative direction, men's IBL method difference scores were not significantly associated with testosterone ($r = -0.22$, $p = 0.196$). Within women, method differences were not associated with testosterone for Salimetrics ($r = -0.14$, $p = 0.287$) and DRG kits ($r = -0.17$, $p = 0.221$). However, IBL kits showed a negative association between difference scores and testosterone in women ($r = -0.31$, $p = 0.023$). Altogether, these analyses suggest that DRG and Salimetrics kits showed proportional bias in estimating men's testosterone, but IBL kits showed proportional bias in estimating women's testosterone.

We additionally conducted Bland Altman plots comparing percentage differences between EIAs and LC–MS/MS concentrations (See Fig. 3). Similar to the above analyses in Section 3.3., these analyses indicated that the EIAs inflated very low concentrations of testosterone (approximately <10 pg/mL) assessed by LC–MS/MS. This inflation of concentrations occurred in women, likely due to women having lower concentrations than men do. Visual

[4] In addition to measuring cortisol with LC–MS/MS, Miller et al. (2013) assessed cortisol with four EIAs (DRG, Salimetrics, DSL, and DELFIA) and an IBL chemiluminescence immunoassay. Our statistical comparisons between IBL EIAs and LC–MS/MS were to the correlations between the four EIAs and LC–MS/MS measurements provided by Miller et al. (2013).

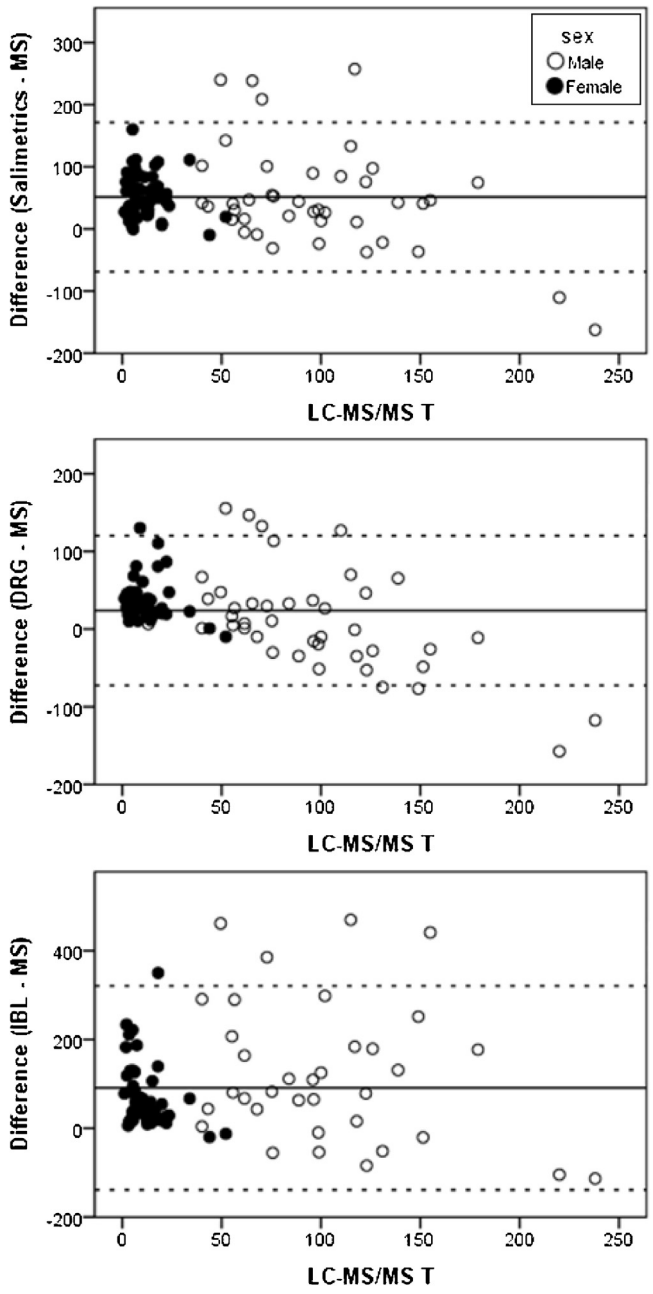

Fig. 2. Bland Altman Plots of Testosterone EIAs and the Mass Spectrometry Reference Method.
Note: The difference between EIA testosterone concentrations (Salimetrics, DRG, and IBL) and those of LC–MS/MS (indicated as MS on the y-axis of the figure to avoid being read as a mathematical expression) is drawn against the LC–MS/MS values. Solid lines indicate the mean difference in methods, whereas dashed lines represent the 95% limits of agreement (±2 SDs from the mean difference). Black dots represent female samples, whereas white dots represent male samples.

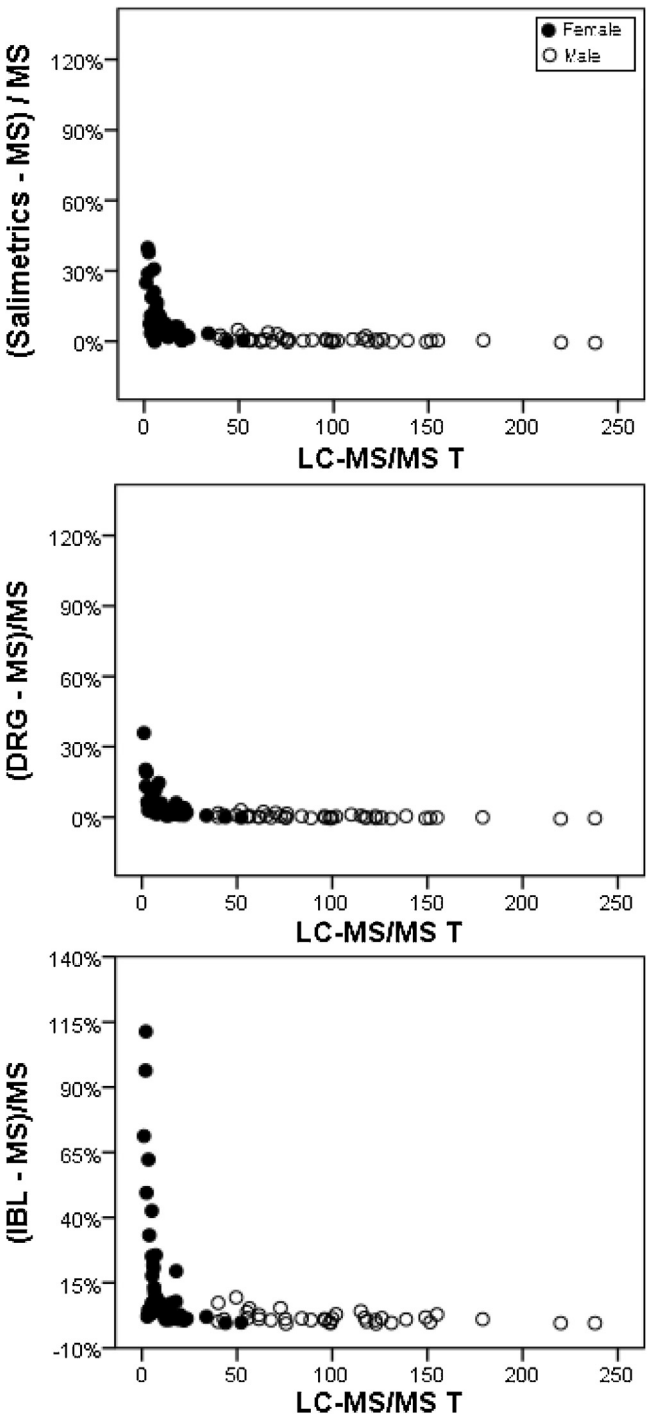

Fig. 3. Bland Altman Plots in the Metric of Percent Differences Between EIAs and the Mass Spectrometry Reference Method.
Note: The percent difference between EIA testosterone concentrations (Salimetrics, DRG, and IBL) and those of LC–MS/MS (indicated as MS on the y-axis of the figure to avoid being read as a mathematical expression) is drawn against the LC–MS/MS values.

inspection of this figure shows that the largest inflation of low concentrations occurred for IBL, whereas there was the least amount of inflation for DRG.

3.5. Deming regressions

Deming regressions were conducted to assess correspondence between EIAs and LC–MS/MS. Fig. 4 shows scatterplots for testosterone measured by the three EIAs compared to LC–MS/MS, as well as the line of best fit from Deming regressions, the identity line (intercept = 0, slope = 1), and the Deming regression equations. Of the three methods, DRG EIAs most closely approximated the line of identity (Intercept = 27.00 [95% CI: 7.71, 46.30], Slope = 0.93 [95% CI: 0.66, 1.20]), followed by Salimetrics (Intercept = 20.70 [95% CI: −14.94, 56.35], Slope = 1.64 [95% CI: 1.14, 2.14]), and last, IBL (Intercept = −78.95 [95% CI: −210.47, 52.57], Slope = 4.67 [95% CI: 2.83, 6.51]). Additionally, although the confidence interval for the slope of DRG contained the line of identity (i.e., included 1), the slopes of IBL and Salimetrics significantly differed from the line of identity

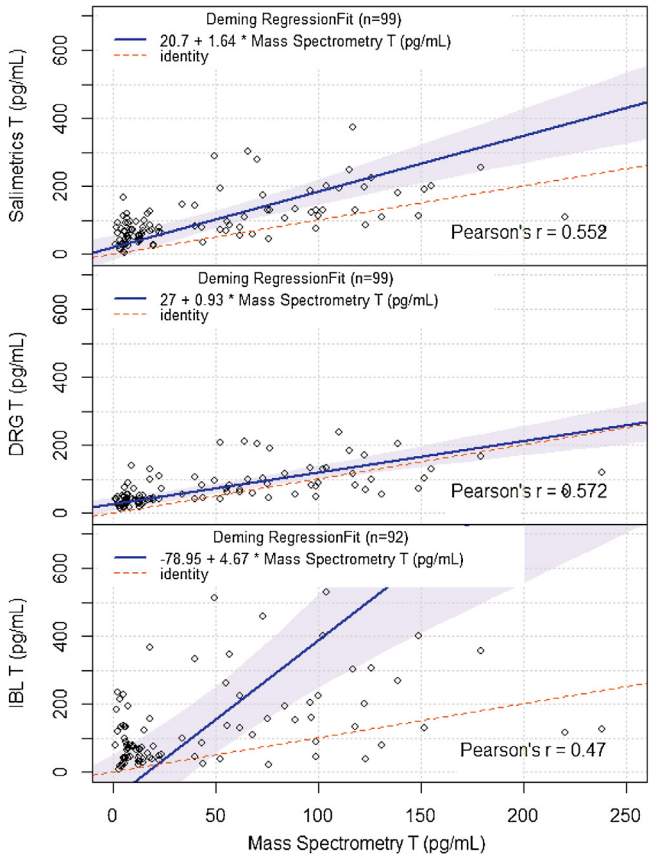

Fig. 4. Deming regressions between LC–MS/MS and the three EIA methods.
Note: The dashed line represents the line of identity (if DRG and LC–MS/MS concentrations were equivalent), whereas the blue line represents the Deming regression line. Shaded regions indicate 95% confidence regions for the Deming Regression line. The X-axis scaling is equivalent for all scatterplots and LC–MS/MS Testosterone is labeled as "Mass Spectrometry T" to avoid confusing LC–MS/MS with a mathematical expression when presented in equations. For interpretation of the references to color in this figure legend, please refer to the web version of the article.

(i.e., did not include 1). In our supplemental materials, we also conducted these Deming regressions in men and women separately, finding no statistically significant deviations from the line of identity.

3.6. Additional analyses

Given the growing interest in testing the interactive effects (Mehta and Josephs, 2010) and "coupling" (Shirtcliff et al., 2015) of the HPG and HPA axes, we investigated whether the correlations between testosterone and cortisol concentrations from LC–MS/MS in our data differed depending on the EIA (i.e., IBL, Salimetrics, or DRG) used to measure testosterone. Overall, testosterone and cortisol were not significantly correlated across all available measures ($|rs| \leq 0.14$) and correlations between the cortisol concentrations from LC–MS/MS concentrations and testosterone concentrations from available EIAs did not significantly differ across all of our data. Additionally, these associations were not statistically significant when examining men and women separately, although the correlations between DRG testosterone and LC–MS/MS were of moderate size (see Supplemental Materials). This suggests that the associations between testosterone and cortisol do not systematically differ depending on the method, be it EIAs or LC–MS/MS.

4. Discussion

This study compared salivary testosterone measured by frequently used EIAs to the values obtained from LC–MS/MS in a mixed-sex sample of healthy adults. Testosterone immunoassays from DRG provided assessments of salivary testosterone that were the most comparable to LC–MS/MS, followed closely by Salimetrics, and lastly by IBL. Consistent with previous literature, EIAs overestimated testosterone concentrations compared to LC–MS/MS (see Rosner et al., 2007 for a review). Among the three assay kits, DRG kits demonstrated the lowest estimates of testosterone concentrations, with Salimetrics showing higher values than DRG, and IBL showing the highest estimated concentrations of testosterone.

All EIAs in the current study tended to produce more inflated testosterone concentrations in samples assessed as having very low testosterone by LC–MS/MS (approximately <10 pg/mL). This was observed exclusively within female samples, who have lower concentrations of testosterone compared to men. This tendency to inflate very low concentrations of testosterone creates a substantial impediment to accurately assessing women's testosterone with EIAs. Researchers who are interested in studying testosterone in women may do so without this bias by using LC–MS/MS instead of EIAs. This systematic error in EIAs for assessing women's testosterone may inflate type 2 errors by obscuring behavioral and psychological effects of testosterone that could be assessed with relatively decreased measurement error in men. In retrospect, it is possible that psychological and behavioral effects and correlates of testosterone identified in men but not women by EIAs (e.g., Stanton et al., 2009; Carré et al., 2013) may have been influenced by this elevated systematic bias for testosterone EIAs in women.

Additionally, the analyses confirmed that cortisol measured by LC–MS/MS is similar (does not significantly differ) in its correlations with testosterone examined by the four methods included in this study. This finding is useful to future researchers interested in examining joint or concurrent effects of testosterone and cortisol (see Mehta and Prasad (2015) for a review; see also Shirtcliff et al. (2015)). It suggests that the method of testosterone measurement may not be a source of bias in investigating testosterone and cortisol as simultaneous predictors of psychological or behavioral outcomes. The increased accuracy of LC–MS/MS in assessing hormones compared to immunoassays, combined with the lack of method bias in immunoassays for testosterone, suggests that specific cortisol immunoassay kits would not influence associations between testosterone and cortisol.

This research is not without limitations. These findings would benefit from replication with larger sample sizes. For instance, we found gender differences in the measurement of testosterone and the extent to which EIAs overestimated testosterone. Future studies with better statistical power might be able to better disentangle these differences at extremely low testosterone concentrations. This work would help clarify whether certain EIAs are more beneficial for assessing testosterone within populations with very low testosterone (e.g., women, children). Larger sample sizes measuring testosterone via EIAs and LC–MS/MS are necessary to make stronger assessments of the measurement accuracy of EIAs. Another limitation of the current research is that the calibration consistency of LC–MS/MS could be verified more accurately. This research could have been improved if certified reference materials (CRM) were used for samples by adding CRM to the participants' saliva samples in addition to the internal standards. The standard concentration samples from each kit could also have been validated with LC–MS/MS to ensure differences in calibrator accuracy did not affect the comparison of the methods. Finally, this study did not assess all commercially available methods available for measuring testosterone (e.g., Demeditec EIAs, radioimmunoassays, luminescence immunoassays).

Altogether, our results suggested that DRG provided a closer approximation to LC–MS/MS testosterone data closely followed by Salimetrics, and then IBL. However, there is considerable room for improvement in assessing salivary testosterone through EIAs. The testosterone EIA and LC–MS/MS correlations in our study have considerably less linear correspondence compared to correlations between EIA-measured cortisol and cortisol assessed by LC–MS/MS (Miller et al., 2013).

5. Conclusion

Altogether, the assessment of testosterone by EIAs had only moderate correspondence with testosterone assessed by LC–MS/MS. Based on the increasing use of LC–MS/MS for measuring hormones (for reviews see Soldin and Soldin (2009) and Field (2013)), researchers may wish to adopt LC–MS/MS for more accurate testosterone measurement, when feasible. LC–MS/MS is more costly than conducting EIAs, but the advantages of LC–MS/MS to assess testosterone may outweigh the costs. In addition to having higher accuracy and sensitivity for hormone measurement compared to EIAs, LC–MS/MS can also allow for the assessment of several analytes from only one sample (e.g., Keefe et al., 2014). Based on the limitations of assessing testosterone through EIAs identified in this study, scientists may develop more accurate and affordable ways of assessing testosterone and other hormones. Improving the estimation of testosterone will increase precision and theoretical depth in the burgeoning research investigating testosterone's role in behaviors, psychological processes, and social functioning.

Conflict of interest

None of the authors report any conflict of interest.

Contributors

Keith M. Welker, Bethany Lassetter, Cassandra Brandes, Smrithi Prasad, Dennis R. Koop, Pranjal H. Mehta.

Funding

This research was supported by two National Science Foundation grants awarded to PHM (Award Numbers: 1451848 and 1303743). This funding sponsor was not involved in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Acknowledgement

We thank Jenny Luo (Oregon Health and Science University) for assisting with conducting liquid chromatography tandem mass spectrometry analyses.

Appendix A. Supplementary materials

Supplementary materials associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.psyneuen.2016.05.022.

References

Apicella, C.L., Dreber, A., Mollerstrom, J., 2014. Salivary testosterone change following monetary wins and losses predicts future financial risk-taking. Psychoneuroendocrinology 39, 58–64.
Baecher, S., Azad, S.C., Vogeser, M., 2013. Inter-method comparison of salivary cortisol measurement. J. Lab. Med. 37, 269–273.
Bland, J.M., Altman, D., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327, 307–310.
Carré, J.M., Campbell, J.A., Lozoya, E., Goetz, S.M., Welker, K.M., 2013. Changes in testosterone mediate the effect of winning on subsequent aggressive behaviour. Psychoneuroendocrinology 38, 2034–2041.
Carré, J.M., Iselin, A.M.R., Welker, K.M., Hariri, A.R., Dodge, K.A., 2014. Testosterone reactivity to provocation mediates the effect of early intervention on aggressive behavior. Psychol. Sci. 25, 1140–1146.
Chattoraj, S., 1976. Endocrine function. In: Tietz, N.W. (Ed.), Fundamentals of Clinical Chemistry. WB Saunders, Philadelphia, pp. 699–817.
Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, second ed. Hillsdale, New Jersey.
Crewther, B.T., Al-Dujaili, E., Smail, N.F., Anastasova, S., Kilduff, L.P., Cook, C.J., 2013. Monitoring salivary testosterone and cortisol concentrations across an international sports competition: data comparison using two enzyme immunoassays and two sample preparations. Clin. Biochem. 46, 354–358.
DRG International, Inc, 2011. DRG Salivary Testosterone. DRG International, Inc., USA, 1–13.
Edelstein, R.S., van Anders, S.M., Chopik, W.J., Goldey, K.L., Wardecker, B.M., 2014. Dyadic associations between testosterone and relationship quality in couples. Horm. Behav. 65, 401–407.
Field, H.P., 2013. Tandem mass spectrometry in hormone measurement. Methods Mol. Biol. 1065, 45–74.
Hoofnagle, A.N., Wener, M.H., 2009. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. J. Immun. 347, 3–11.
IBL International Corp., 2013. Testosterone Saliva EIA: Enzyme immunoassay for the in-vitro diagnostic quantitative determination of free testosterone in human saliva. IBL International, Hamburg, Germany. 1–6.
Keefe, C.C., Goldman, M.M., Zhang, K., Clarke, N., Reitz, R.E., Welt, C.K., 2014. Simultaneous measurement of thirteen steroid hormones in women with polycystic ovary syndrome and control women using liquid chromatography-tandem mass spectrometry. PLoS One 9, e93805, http://dx.doi.org/10.1371/journal.pone.0093805.
Keevil, B.G., 2013. Novel liquid chromatography tandem mass spectrometry (LC–MS/MS) methods for measuring steroids. Best Prac. Res. Clin. Endocrinol. Metab. 27, 663–674.
Krouwer, J.S., 2008. Why Bland-Altman plots should use X: not (Y+X)/2 when X is a reference method. Stat. Med. 27, 778–780.
Martin, R.F., 2000. General deming regression for estimating systematic bias and its confidence interval in method-comparison studies. Clin. Chem. 46, 100–104.
Mazur, A., Booth, A., 1998. Testosterone and dominance in men. Behav. Brain Sci. 21, 353–363.
Mehta, P.H., Josephs, R.A., 2010. Testosterone and cortisol jointly regulate dominance: evidence for a dual-hormone hypothesis. Horm. Behav. 58, 898–906.
Mehta, P.H., Prasad, S., 2015. The dual-hormone hypothesis: a brief review and future research agenda. Curr. Opin. Behav. Sci. 3, 163–168.
Mehta, P.H., Snyder, N.A., Knight, E.L., Lassetter, B., 2015. Close versus decisive victory moderates the effect of testosterone change on competitive decisions and task enjoyment. Adapt. Hum. Behav. Physiol. 1, 291–311.
Miller, R., Plessow, F., Rauh, M., Gröschl, M., Kirschbaum, C., 2013. Comparison of salivary cortisol as measured by different immunoassays and tandem mass spectrometry. Psychoneuroendocrinology 38, 50–57.
Pugeat, M.M., Dunn, J.F., Nisula, B.C., 1981. Transport of steroid hormones: interaction of 70 drugs with testosterone-binding globulin and corticosteroid-binding globulin in human plasma. J. Clin. Endocrinol. Metab. 53, 69–75.
Puts, D.A., Pope, L.E., Hill, A.K., Cárdenas, R.A., Welling, L.L., Wheatley, J.R., Breedlove, S.M., 2015. Fulfilling desire Evidence for negative feedback between men's testosterone, sociosexual psychology, and sexual partner number. Horm. Behav. 70, 14–21.
R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, Retrieved from http://www.R-project.org/.
Rosner, W., Auchus, R.J., Azziz, R., Sluss, P.M., Raff, H., 2007. Utility, limitations, and pitfalls in measuring testosterone: an endocrine society position statement. J. Clin. Endocrinol. Metab. 92, 405–413.
Salimetrics, LLC, 2014. Salivary Testosterone: Enzyme Immunoassay Kit Salimetrics. LLC., State College, PA, pp. 1–25.
Schultheiss, O.C., Stanton, S.J., 2009. Assessment of salivary hormones. In: Harmon-Jones, E., Beer, J.S. (Eds.), Methods in Social Neuroscience. Guilford Press, New York, pp. 17–44.
Shirtcliff, E.A., Dismukes, A.R., Marceau, K., Ruttle, P.L., Simmons, J.G., Han, G., 2015. A dual-axis approach to understanding neuroendocrine development. Dev. Psychobiol. 57, 643–653.
Soldin, S.J., Soldin, O.P., 2009. Steroid hormone analysis by tandem mass spectrometry. Clin. Chem. 55, 1061–1066, http://dx.doi.org/10.1373/clinchem.2007.100008.
Stanton, S.J., Wirth, M.M., Waugh, C.E., Schultheiss, O.C., 2009. Endogenous testosterone levels are associated with amygdala and ventromedial prefrontal cortex responses to anger faces in men but not women. Biol. Psychol. 81, 118–122.
Star-Weinstock, M., Williamson, B.L., Dey, S., Pillai, S., Purkayastha, S., 2012. LC–ESI–MS/MS analysis of testosterone at sub-picogram levels using a novel derivatization reagent. Anal. Chem. 84, 9310–9317.

Taieb, J., Mathian, B., Millot, F., Patricot, M.C., Mathieu, E., Queyrel, N., Lacroix, I., Somma-Delpero, S., Boudou, P., 2003. Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography–mass spectrometry in sera from 116 men women, and children. Clin. Chem. 49, 1381–1395.

Turpeinen, U., Hämäläinen, E., Haanpää, M., Dunkel, L., 2012. Determination of salivary testosterone and androstendione by liquid chromatography–tandem mass spectrometry. Clin. Chim. Acta 413, 594–599.

Yasuda, M., Honma, S., Furuya, K., Yoshii, T., Kamiyama, Y., Ide, H., Muto, S., Horie, S., 2008. Diagnostic significance of salivary testosterone measurement revisited: using liquid chromatography/mass spectrometry and enzyme-linked immunosorbent assay. J. Mens. Health 5, 56–63.