Comprehensive Psychoneuroendocrinology 10 (2022) 100132

Contents lists available at ScienceDirect

Comprehensive Psychoneuroendocrinology

journal homepage: www.sciencedirect.com/journal/comprehensive-psychoneuroendocrinology

Chemiluminescent immunoassay overestimates hormone concentrations and obscures testosterone sex differences relative to LC-MS/MS in a field study of diverse adolescents

Julia E. Chafkin [a,*], Joseph M. O'Brien [a], Fortunato N. Medrano [a], Hae Yeon Lee [b], David S. Yeager [a], Robert A. Josephs [a]

[a] Department of Psychology, University of Texas at Austin, Austin, USA
[b] Department of Psychology, Yale-NUS, Singapore

ARTICLE INFO

Keywords:
Methods
Psychopathology
Immunoassay
LC-MS/MS
Adolescence

ABSTRACT

Background: Methodological comparisons of hormone quantification techniques have repeatedly demonstrated that, in adults, enzyme immunoassay (EIA) inflates steroid hormone concentrations relative to mass spectrometry. However, methodological comparisons in adolescent samples remain rare, and few studies have examined how chemiluminescent immunoassay (CLIA), another popular immunoassay, compares to mass spectrometry. Additionally, no studies have examined how differences in analytical techniques may be affecting relationships between steroid hormone levels and outcomes of interest, such as psychopathology. This pre-registered analysis of an existing dataset measured salivary cortisol and testosterone using both CLIA and liquid chromatography dual mass spectrometry (LC-MS/MS) in a repeated measures (516 samples) sample of 207 9th graders.
Methods: In aim 1, this study sought to expand on past findings by 1) measuring inflation of testosterone and cortisol by CLIA in a relatively large adolescent sample, and 2) showing that CLIA (like EIA) testosterone inflation was especially true in groups with low 'true' testosterone levels. In aim 2, this study sought to examine the impact of hormone quantification method on relationships between hormone levels and psychopathological measures (the Children's Depression Inventory, the Perceived Social Stress Scale, the UCLA Loneliness Scale, and the Anxious Avoidant and Negative Self Evaluation subscales of the Social Anxiety Scale for Adolescents).
Results: We found that CLIA, like EIA, inflated testosterone and cortisol levels and overestimated female testosterone resulting in suppressed sex differences in testosterone. We did not observe these same patterns when examining testosterone in individuals with differing levels of pubertal development. Results of psychopathology analyses demonstrated no significant method differences in hormone-psychopathology relationships.
Conclusions: Our findings show that CLIA introduces proportional bias in cortisol and testosterone in a manner that suppresses sex differences in testosterone. Steroid measurement method did not significantly moderate the relationship between hormones and psychopathology in our sample, though more work is needed to investigate this question in larger, clinical samples.

1. Introduction

A series of landmark methodological studies and reviews have demonstrated that steroid hormone concentrations [1–3] differ depending on the method used to quantify hormone levels. Comparisons of steroid hormone measurement techniques suggest that certain types of *immunoassays*, which are widely used and cost-effective, may inflate steroid hormone levels relative to *liquid chromatography tandem mass spectrometry* (LC-MS/MS), the gold standard in the field of clinical chemistry [4]. Despite the inflation introduced by some immunoassays, many researchers continue to employ immunoassay methods, making research that examines discrepancies between specific immunoassays and mass spectrometry techniques a valuable field resource.

Many questions concerning immunoassay bias remain unanswered. This study sought to address three main gaps in the literature. First, though enzyme immunoassay (EIA) techniques have been repeatedly

* Corresponding author.
E-mail address: Julia.Chafkin@utexas.edu (J.E. Chafkin).

https://doi.org/10.1016/j.cpnec.2022.100132
Received 22 March 2022; Accepted 22 March 2022
Available online 1 April 2022

found to introduce error into steroid hormone measurement relative to mass spectrometry, other immunoassays remain understudied. Does chemiluminescent immunoassay (CLIA), another popular immunoassay technique, distort cortisol and testosterone in a manner similar to EIA? Second, though many methodological comparisons have examined adult samples, studies in other age groups remain rare. How do CLIA and LC-MS/MS results differ in a relatively large sample of adolescents between age 13 and 16? And finally, how might hormone measurement method be impacting hormone-outcome relationships? Even if immunoassay produces distortion in hormone measurement relative to LC-MS/MS, are these distortions impacting the relationship between salivary hormone levels and outcomes of interest, such as measures of psychopathology? Using 516 salivary samples collected from 207 ninth grade students whose hormones were assayed using both CLIA and LC-MS/MS, we sought to answer these questions.

1.1. Immunoassay techniques and challenges

In order to understand some of the challenges faced by different hormone measurement techniques, it is helpful to briefly review the processes involved in immunoassay and mass spectrometry methods (for a more detailed discussion of hormone measurement techniques and their challenges, see Ref. [4]). Immunoassays generally rely on both the immune system's ability to develop antibodies that recognize specific antigens, and a labeling technique that can quantify the amount of antigen present in a sample. In radioimmunoassays, which have become less common over time due to their use of radioactive materials, radioisotopes emit radioactivity in inverse proportion to the amount of antigen. In EIA, antigens have been bound by an antibody-enzyme complex. This complex undergoes a chemical reaction producing a color change. The depth of color change indicates the amount of antigen in a sample. In CLIA, similar to EIA, antibody-bound antigen undergoes a chemical reaction, though in CLIA photons of light are emitted. CLIA is purported to be useful for measuring antigens present at very low volumes. Issues with EIA, and other immunoassays, include poor inter-lab reliability of hormone concentrations, variation in immunoassay sensitivity and specificity for certain hormones, cross-reactivity, or overestimation of antigen due to a sample containing molecules so similar in shape to the antigen that the immunoassay cannot distinguish between the antigen and the similar molecule, and matrix effects, or species-specific components of a sample that interfere with the function of the immunoassay [5,6]. As steroid hormones are all downstream products of the same molecule, cholesterol, and therefore share varying degrees of similarity in structure, cross-reactivity is of particular concern for the measurement of steroid hormones with immunoassay. Despite these concerns, immunoassays remain popular and widely used for hormone measurement in the field of psychology.

1.2. Mass spectrometry techniques and challenges

Mass spectrometry identifies compounds in an electromagnetic field by their mass-to-charge ratio, making it a highly specific method for measuring an analyte. However, mass spectrometry is vulnerable to a number of issues including difficulty in correctly identifying compounds with similar mass-to-charge ratios and ion suppression, which occurs when biological compounds in a substrate containing an analyte of interest are too similar to the analyte of interest or interfere with the ionization, or breakdown, process of the analyte. To combat these issues, mass spectrometry is coupled with chromatography (liquid or gas), a technique that separates chemical compounds by their chemical *and* physical properties. The combination of chromatography and mass spectrometry makes for a highly sensitive and specific analytical tool that is considered the gold standard in the field of clinical chemistry and is capable of providing an accurate standard against which to compare immunoassay results.

1.3. Methodological comparisons of immunoassay and LC-MS/MS in testosterone and cortisol

Systematic comparisons of EIA and LC-MS/MS for the measurement of testosterone and cortisol show that EIA, relative to LC-MS/MS, tends to inflate salivary testosterone and cortisol concentrations especially when these hormones are present at very low levels [1,3]. This inflation specifically effects demographic groups with low levels of testosterone, such as females and older males, resulting in suppression of testosterone differences between males and females and between younger and older males [3,7]. Other groups with low hormone levels, such as children and peripubertal teens, have not been examined for this type of EIA inflation, though it is reasonable to suggest that EIA's tendency to inflate low concentrations of testosterone and suppress group differences may behave similarly with pubertal development as it does with respect to sex and age. While these findings have not been replicated with CLIA, and chemiluminescent immunoassay techniques are often heralded for their ability to detect low volume substrates, CLIA, as an immunoassay, is vulnerable to many of the same challenges as EIA, meaning that CLIA results may produce the same biases as EIA when quantifying testosterone.

Past research has clearly demonstrated inflation of low steroid hormone levels by immunoassay techniques. Work examining higher concentrations of hormones has produced less uniform findings. In cortisol, some research has demonstrated that LC-MS/MS concentrations above 5 nmol/L show good agreement with cortisol measured with CLIA [1]. Other work has shown that cortisol concentrations measured by EIA are related to LC-MS/MS-measured cortisol in a nonlinear fashion [8] or that cortisol is overestimated by EIA and CLIA across the full range of true hormone concentrations [9]. In methodological comparisons of higher testosterone concentrations, some studies have demonstrated that testosterone concentrations measured with EIA show low linear correspondence with testosterone measured via LC-MS/MS, while other studies show good reliability between methods [10].

In addition to the lack of studies examining how CLIA hormone results may differ from LC-MS/MS, almost no methodological comparison study has tested how these methods compare to one another in non-adult samples. The one exception examined cortisol concentrations in children ages 8–14 [1], leaving a critical gap in the literature for adolescent individuals between 14 and 18 years of age. As puberty is precipitated by adrenal and gonadal steroid hormone surges, hormone levels are often incorporated into studies of adolescent development [11–14]. Therefore, comparisons of immunoassay and mass spectrometry techniques in this age group can provide information about technique-based errors in an important measure of adolescent development: hormones.

1.4. Associations with psychopathology

Immunoassay measurement methods are often employed in research seeking to elucidate relationships between steroid hormones and measures of psychopathology [15–23]. Results of this research comprise multiple fields of work that are beyond the scope of this study. However, given findings that many immunoassay techniques introduce error and suppress sex differences in testosterone, it is important to investigate whether hormone measurement method might, itself, be a moderator of hormone-psychopathology relationships. Though the participants in this study come from a non-clinical sample, a preliminary investigation of the moderating power of hormone measurement method on hormone-psychopathology correlations can provide a helpful framework for future research comparing methods in clinical samples.

1.5. Current study

The aims of the present study were two-fold. First, we sought to expand upon past findings comparing LC-MS/MS to EIA by measuring

testosterone and cortisol in an adolescent sample using LC-MS/MS and CLIA. Though one past study has compared CLIA and LC-MS/MS measures of cortisol [1], no study to date has compared CLIA and LC-MS/MS results of testosterone. Second, we sought to investigate the moderating effect of hormone measurement method on hormone-psychopathology associations by comparing the relationships between CLIA-or-LC-MS/MS-measured hormone concentrations and measures of concurrently measured, self-reported psychopathology.

Hypotheses for both aims of our study were pre-registered with the Open Science Framework. In the first aim, we hypothesized that we would see: 1) overestimation by CLIA (relative to LC-MS/MS) of both cortisol and testosterone concentrations, especially at low 'true values' ('true values' are defined as the LC-MS/MS values in line with past research) of both hormones, 2) CLIA underestimation of sex differences in testosterone concentrations owing to greater overestimation of female testosterone concentrations than male testosterone concentrations, and 3) greater CLIA overestimation of testosterone concentrations among low-pubertal status (less developed) individuals, with correspondingly weaker correlation between CLIA-measured testosterone and pubertal status than between LC-MS/MS-measured testosterone and pubertal status.

In our second aim, we sought to explore the relationship between CLIA-or-LC-MS/MS-measured hormones (cortisol and testosterone) and measures of psychopathology using regression models. We expected to see moderation of hormone-psychopathology relationships by hormone measurement method.

2. Methods

The present study 1) directly compared and examined systematic differences in hormone concentrations of cortisol and testosterone measured by CLIA and LC-MS/MS, and 2) sought to examine whether measurement method moderated hormone-psychopathology statistical associations. All statistical analyses were completed in R [24].

2.1. Sample

Analyses were conducted with a subsample[1] (n = 516 salivary samples in n = 207 individuals, 52.17% female) of individuals whose salivary samples were analyzed using both immunoassay and LC-MS/MS (n = 16 samples were excluded after analysis with CLIA, due to insufficient sample volume for subsequent analysis with LC-MS/MS) from the Texas Longitudinal Study of Adolescent Stress Resilience (TLSASR); a new public-use dataset funded by the NICHD.[2] Self-report measures were collected with the first day of salivary hormone samples, and the two remaining salivary samples were collected on consecutive days one week later, all in the fall semester. Parental consent, child assent, and saliva samples were provided for all individuals in this sample. Research protocols were approved by the institutional research review board at the authors' institution, by the research committee at the participating school district, and by the collaborating school principal. Demographics of the sample were self-reported by participants as follows: Asian 5.4%, Black 4.9%, Hispanic 25.4%, Native Hawaiian/Pacific Islander, American Indian/Alaskan Native, or reporting "Two or More Races" 5.4%, White 59%.

[1] Hormone data analyzed in this paper came from a pre-treatment day and two consecutive post-treatment days of a longitudinal study titled "Teaching teens that people can change." See supplemental materials for analyses showing no impact of intervention on hormone levels in this subsample.

[2] The TLSASR datasets are posted on the Inter-university Consortium for Political and Social Research (ICPSR) server. Pre-registrations can be found at https://osf.io/yg85b and https://osf.io/f7grk.

2.2. Procedures

Saliva samples were collected using 2.5 ml or 4.0 ml Salicap tubes (IBL International, Hamburg, Germany) in the early afternoon (1:30 p.m.–4:30 p.m.) to reduce variability due to diurnal rhythms in cortisol concentrations [25] (for more detail on passive drool procedures, see Ref. [26]. Time of sample collection was automatically recorded in an electronic daily intake questionnaire, and controlled for in analyses relating hormone concentrations to psychopathology measures. Students were asked to refrain from eating dairy products (e.g., yogurt, milk, cheese) as bovine hormones can cross-react with immunoassay antibodies [27], drinking caffeinated beverage (e.g., coffee, soda, tea, and energy drinks) as caffeine has been reported to increase cortisol and testosterone levels [28], taking nonprescribed medications which have been shown to have a range of effects on hormone levels, or engaging in strenuous physical exercise, which can increase testosterone and cortisol levels, at least 2 h prior to sample collection [29,30].

After collection, samples were transferred to a Yeti™ cooler (Austin, TX) at < 0 °C, before being moved to a −80 °C laboratory freezer on the UT Austin campus at the end of the same day. All samples were stored for 3–4 months in the same −80 °C freezer on the UT Austin campus (between September 2016 and late December 2016) before being shipped to the biological health psychology laboratory at Brandeis University, Waltham, MA (PIs, N. Rohleder and J. Wolf) for analysis using a chemiluminescence immunoassay (IBL International, Hamburg, Germany).

Samples were pipetted by a Hamilton Company liquid handling robot and measured in duplicate. Samples with a coefficient of variation (CV) > 10% underwent repeated analysis. Cortisol assay intra- and inter-assay CVs were 9.07% and 5.59%, respectively. Testosterone assay intra- and inter-assay CVs were 6.29% and 4.65% respectively.

For analysis using LC-MS/MS, samples were shipped to Dresden, Germany for analysis at Dresden Lab Services (PI, C. Kirschbaum). (Detailed methods can be found in Ref. [31]. Samples analyzed using LC-MS/MS underwent two thaw freeze cycles more than those analyzed at Brandeis University. For analyses of the potential impact of freeze-thaw cycles on hormone concentrations, which showed no significant effect of additional freeze-thaw cycles on hormone concentrations, see supplemental materials.

2.3. Measures

2.3.1. Depressive symptoms

Depression symptomatology (Female Mean: 0.51; SD: 0.36, Male Mean: 0.4; SD: 0.32) was measured concurrently with baseline saliva samples using the 27-item Children's Depression Inventory (CDI; [32], from which item 9, which assesses for suicidality, was removed due to concerns for student safety. A 2015 meta-analysis of the reliability of the English version of the CDI with item 9 removed resulted in a Cronbach's alpha of .841 (95% CI = 0.839-0.851) [33]. For comparison, our within-sample Cronbach's alpha was calculated to be 0.92 (95% CI = 0.88 0.94). Each of the CDI items asks participants to identify which of three concentrations of a symptom best describes how they feel (e.g. 0 = *I do most things O.K.;* 1 = *I do many things wrong;* 2 = *I do everything wrong*). Scores from each item were summed together and divided by the total number of items answered to compute an average item score (ranging between 0–2). This method was employed to assess average ratings of depression symptomatology, and to avoid issues with depression sum scores arising from omission of the suicidality item. For comparison, normative CDI scores in a population with a similar age range showed a mean of 0.3(SD: 0.26) in males and 0.34(SD: 0.28) in females [34].

2.3.2. Loneliness

Measures of perceived loneliness were assessed using the UCLA Loneliness Scale-short form (ULS-8; [19,35,36]): The ULS-8 is an 8-item

self-report scale used to measure loneliness that asks respondents to rate feelings of loneliness on a 4-point Likert scale ranging from 1 ("Never") to 4 ("Often"). Average scores (ranging from 1–4.5) were calculated to avoid issues due to item missingness (Mean = 2.54; SD = 0.84). For comparison, normative ULS-8 scores in a U.S. population showed a mean of 2.39(SD: 0.74) [37].

2.3.3. Social anxiety

Measures of adolescent social anxiety were assessed using the Social Anxiety Scale for Adolescents (SAS-A, [38]: Fear of Negative Evaluation (Mean = 2.45; SD = 1.08) and Social Avoidance (Mean = 2.69; SD = 1.02) subscales. The SAS-A is a self-report scale used to assess social anxiety in child and adolescent populations ranging in age from 13 to 18. Respondents are asked to choose items from a 5-point Likert scale with answers ranging from 1 ("Strongly disagree") to 5("Strongly Agree"). Subscales of Fear of Negative Evaluation and Social Avoidance (New) were used to measure adolescent social anxiety. Average scores (ranging from 1–5) were calculated in order to avoid any issues due to item missingness. For comparison, normative SAS-A Avoidance scores in a population with a similar age range showed a mean of 2.15(SD: 0.67) in males and 2.28(SD: 0.69) in females. Normative SAS-A Negative Evaluation scores showed a mean of 2.34(SD: 0.84) in males and 2.60 (SD: 0.86) in females [34].

2.3.4. Stress

The Perceived Stress Scale (PSS; [39]: The PSS, a global measure of perceived stress, is a 14-item self-report scale in which respondents report on the degree of perceived stress in their lives. Each item on the PSS is rated on a 5-point Likert scale ranging from 1 ("Never") to 5 ("Very Often"). Average scores (ranging from 1.8–4.6 in this sample) were calculated in order to avoid any issues due to item missingness (Mean = 3.04; SD = 0.56). For comparison, normative PSS scores in a population with a similar age range showed a mean of 2.66 (SD: 0.52) [40].

2.3.5. Pubertal development

The Pubertal Developmental Scale (PDS; [41] was administered concurrently with baseline saliva sample collection in the fall semester of 9th grade to assess adolescents' pubertal development stage (Mean = 3.06, SD = 0.58, range = 1.6–4). The PDS scale asks participants to rate progression of puberty-relevant physical changes, including breast development, presence of pimples, growth spurt, body hair, and presence or absence of menstruation, and has been shown to generally agree with clinician-rated Tanner stage [42]. It should be noted that the PDS scores in our sample were restricted, ranging from 1.6 to 4.0 instead of from 1.0 to 4. Though this restricted range was expected given the relatively late age of our participants, it is worth noting that we were unable to examine a sample representing the full range of PDS scores for this analysis. For more information about PDS scale scoring, see supplemental materials.

2.4. Analyses

Data were analyzed in r (Rstudio version 1.3.959). For aim one, all days of data were included. A behavioral intervention that was not hypothesized to have a main effect on hormone levels was administered after day one. To confirm no main effect of the intervention on hormones, multilevel analyses were completed (for results of this analysis, which revealed no impact of the behavioral intervention on cortisol or testosterone concentrations, see supplemental materials). Therefore, aim one used all three days of data. For aim two, inspecting hormone-psychopathology relationships, baseline data only was used as the intervention *was* intended to impact psychopathology measures.

2.4.1. Aim 1 analyses

Pearson correlations were used to confirm general linear agreement between immunoassay and LC-MS/MS methods for the measurement of testosterone and cortisol (overall, as well as by sex and PDS). Additionally, we used Bland Altman plots to visually inspect differences between CLIA and LC-MS/MS methods. Bland Altman plots compare the mean of two measurements taken of a single object or substance against the difference between those measurements to allow for a visual comparison of the two methods [43–46]. The traditional Bland Altman method assumes independence between observations and a normal distribution of differences between methods being compared [45,47]. Our sample included up to three repeated measures collected from individuals. Inspection of qqplots and histograms showing the distribution of differences between methods showed right skew in all hormone results (see supplement for qqplots and histograms). To handle issues related to the use of repeated measures in our sample, we used a version of the Bland Altman plot that controls for multiple measures within individuals (https://stat.ethz.ch/pipermail/r-help/2008-July/166921.html). To handle issues related to the non-normality of the distribution of differences between methods, hormone data was log-transformed in line with recommendations for handling the issue of non-normality in Bland Altman plots [45,47]. Subsequent inspection of qqplots and histograms indicated amelioration of skew. Results of Bland Altman plots before transformation can be found in the supplemental materials. Additional Bland Altman plots with LC-MS/MS results on the x axis can also be found in the supplement.

Passing-Bablok regression was added to the analysis plan after preregistration in order to have a quantifiable measure of method bias in addition to the visual information provided by Bland Altman plots. While other techniques, such as Deming regression, can be used to compare method differences, Deming regression assumes that error is normally distributed, whereas Passing-Bablok makes no assumption about sample or error distributions, requiring only that the two methods have a linear correlation. Originally developed for use in chemistry, Passing-Bablok regression allows for an estimation of the extent of agreement between methods of measurement without assuming prior knowledge of the direction of imprecision of one vs another method (although it is not without its critics [46,48]), making it a useful tool for method comparisons [49–52]. Traditionally, Passing-Bablok plots are used to determine whether methods being compared are sufficiently similar to one another to merit use of either method for the measurement of the substrate in question. Recent research has suggested, however, that Passing-Bablok plots may not be sufficient to determine the answer to this question [46]. For the purposes of this analysis, Passing-Bablok plots have been used to inspect trends in the *differences*, as opposed to the similarities, between CLIA and LC-MS/MS for the measurement of testosterone and cortisol, a task for which Passing-Bablok plots are appropriate. Additionally, as the Passing-Bablok approach to analyses of method equality and inequality was originally developed for use in chemistry, it is particularly well-suited to examinations of bias between salivary hormone quantification techniques. A Passing-Bablok slope is equal to the median of all slopes of the straight lines between any two points on the Passing-Bablok plot (excluding slopes equal to 0 or -1). As the lack of independence of these slopes leads to estimation bias, the median (Passing-Bablok slope) is shifted by a factor, K, equivalent to the number of slopes less than −1. A Passing-Bablok slope confidence interval that does not include one ("1") denotes the presence of proportional bias, meaning that the methods do not agree equally across the range of data. A slope greater than 1 indicates that the method being compared to the standard (CLIA in this study) proportionally overestimates the substrate being measured. A Passing-Bablok intercept is equal to the median of all intercepts created by each slope used to calculate the overall slope. A Passing-Bablok intercept with a confidence interval that does not include zero denotes the presence of fixed bias, meaning that one method differs systematically from the other. Normally a Passing Bablok intercept above 0 indicates positive systematic bias in the method being compared to the standard. It should be noted that in cases of high positive proportional bias, Passing-Bablok

intercepts can be negative, despite an overall trend of overestimation rather than underestimation [53]. In these cases, an intercept confidence interval that does not include 0 represents the presence, and not the directionality, of systematic bias. Confidence intervals for Passing-Bablok regression were produced with bootstrapping.

Passing-Bablok regression was completed using raw values of hormone data. We completed Passing-Bablok regressions to examine systematic and proportional bias between CLIA and LC-MS-MS testosterone and cortisol data in 1) all three days of data, 2) male and female participants separately, and 3) "low" and "high" PDS individuals in male and female participants separately. As Passing-Bablok plots are recommended to have no fewer than 30 samples for comparison, analyses of pubertal development that contain fewer than 30 samples should be considered highly prone to error [52]. In these cases, visual inspections of Bland Altman plots should be used instead of Passing-Bablok results.

2.4.2. Aim 2 analyses

To test the hypothesis that hormone analysis method (CLIA or LC-MS/MS) moderates the association between hormones and psychopathology, hormone data was first processed using log-transformation to reduce right skew and normalize residuals [54]. Once log-transformed, hormone data were winsorized and z-scored within sex (In females, one CLIA-assessed cortisol value (0.38%), three LC-MS/MS-assessed cortisol values (1.15%), two CLIA-assessed testosterone values (0.77%), and two LC-MS/MS-assessed testosterone values (0.82%) were winsorized. In males, three LC-MS/MS-assessed cortisol values (1.2%), two CLIA-assessed testosterone values (0.81%), and two LC-MS/MS assessed testosterone values (0.82%) were winsorized).

Cortisol and testosterone values measured with CLIA and LC-MS/MS were used to predict scores on the Children's Depression Inventory, Perceived Stress Scale, UCLA Loneliness Scale, and subscales of the Social Anxiety Scale for Adolescents. GAMs were pre-registered, and used in order to guide the choice of parametric test for all analyses in aim two, as this portion of the analysis was pre-registered as exploratory in nature to investigate hormone-psychopathology relationships when hormones are assessed using CLIA and LC-MS/MS. A GAM is a non-parametric regression technique that allows for the use of smooth terms to describe the relationship between predictor and outcome variables [55]. GAMs supported the use of simple linear regressions in all but two cases, but after correcting for multiple comparisons, GAMs supported linear regressions in all cases. For all models, linear regressions were performed within sex and, as reported above, with the baseline day only of participant data (plots of linear models can be found in the supplement). Analyses examined the association between hormone concentrations and psychopathological outcomes with and without holding time constant. Because holding time constant did not affect the results, results presented here hold time constant, and results found in the supplemental materials show results without time held constant. Relationships between hormones measured using either CLIA or LC-MS/MS and psychopathological symptoms were compared to hormone-psychopathology relationships in which hormones were measured using the alternate method.

Two physiological outliers were observed in this data set. One, measured in cortisol (CLIA result = 178.36 nmol/L, LC-MS/MS result = 155.51 nmol/L) and one measured in testosterone (CLIA result = 1687.92 pg/ml, LC-MS/MS result = 155.51 pg/ml). These values were observed in two separate participants, and were excluded from all analyses as they exceeded upper limits of CLIA reportable ranges of 82.8 nmol/L for cortisol and 500 pg/ml for testosterone (IBL International, Hamburg Germany).

3. Results

3.1. Aim 1 results

Direct comparisons of CLIA and LC-MS/MS values showed that in our adolescent sample, hormone quantification with CLIA resulted in *proportional* inflation bias of testosterone and cortisol concentrations, relative to LC-MS/MS-quantified testosterone and cortisol (see Tables 1 and 2 for descriptive statistics and results of Passing-Bablok regressions and Fig. 1 for Passing-Bablok charts). *Systematic* bias between methods was observed in testosterone but not in cortisol. Additionally, male and female testosterone distributions overlapped to a greater extent in CLIA testosterone Bland Altman plots than LC-MS/MS-assessed testosterone plots (see supplement), suggesting reduced discrimination between male and female testosterone values when testosterone was quantified with CLIA.

3.1.1. Sex differences

As hypothesized, CLIA produced significantly more inflation in female testosterone than male testosterone values (Table 2: slope [Confidence Interval] = 3.3[2.77, 3.86] in males and 17.94[13.1, 26.66] in females). Additionally, method agreement was significantly lower (bias was significantly higher) in female, compared to male, testosterone (Table 2). Visual depiction of the proportional overestimation of female testosterone values by CLIA can be clearly seen in Fig. 2, which shows female CLIA-measured testosterone values tending to be *lower* (more left on the x axis) and *more overestimated* (higher on the y axis) than male CLIA-measured testosterone values. This pattern, of significant overestimation by sex, was not observed in cortisol comparisons (Table 2: slope [Confidence Interval] = 2.64[2.3, 3.01] in males and 2.58[2.27, 3.07] in females). In both males and females, however, cortisol was proportionally overestimated by CLIA relative to LC-MS/MS (Fig. 2).

3.1.2. Differences by pubertal status

No group differences in CLIA-LC-MS/MS agreement by PDS were observed in our sample (Table 2). This result was evident from overlapping slopes and confidence intervals (Table 2). In females, a small number of participants (n = 14) reported low PDS, rendering results of Passing-Bablok equations in low PDS females highly prone to error and resulting in large confidence intervals that are difficult to interpret.

3.2. Bland Altman plots

3.2.1. Sex differences

As referenced above, Bland Altman plots of CLIA-LC-MS/MS comparisons provided clear visual depictions of patterns of CLIA overestimation by sex in testosterone but not in cortisol (Fig. 2). Testosterone plots support previous method comparison studies showing that overestimation of female testosterone levels by CLIA is a consequence of *low values tending to be female,* rather than an issue of female values tending to be error-prone when measured by CLIA. That low male testosterone values, when they occur, also appear to be more overestimated (higher on the y axis), just like female values, and that high female testosterone values, when they occur, tend to be less overestimated (lower on the y axis) just like male values, provides support for this conclusion. Cortisol Bland Altman plots reflect the proportional bias found in Passing-Bablok equations, showing that method differences in cortisol quantification increased over the range of mean values of CLIA and LC-MS/MS cortisol (b = 0.19, p = 5.56e-07, se = 0.04). Replicating prior research, the extent of CLIA and LC-MS/MS cortisol agreement did not differ based on biological sex. Instead, male and female cortisol value distributions appeared to overlap across the range of data.

3.2.2. Differences by pubertal status

Prior to visual inspection with Bland Altman plots, initial analyses were completed to assess for the presence of significant linear hormone-pubertal development relationships. Results of these analyses can be found in the supplement. PDS was positively correlated with CLIA and LC-MS/MS testosterone in males but not in females. Agreement between methods did not vary systematically at different PDS levels in testosterone or cortisol (see supplement for figures).

Table 1
Descriptive statistics of hormone concentrations.

sample size, mean, standard deviation, and variable distribution				
	Testosterone (CLIA)	Testosterone (LC-MS/MS)	Cortisol (CLIA)	Cortisol (LC-MS/MS)
Mean	46.6	12.38	3.8	1.47
SD	38.81	14.09	3.76	1.36
%Female	52.17%	51.94%	52.17%	52.17%
Skewness	2.35	1.94	8.11	3.37
Kurtosis	7.46	4.40	114.17	19.22
N(samples)	515	509	514	515
N(participants)	207	206	207	207
Minimum	1.2	0.18	0.12	0.03
1st Quartile	21.86	2.97	1.97	0.67
Median	35.99	6.04	2.92	1.09
3rd Quartile	58.49	17.77	4.55	1.81
Maximum	268.27	93.6	62.42	14.25
Pearson Correlations				
Whole Group				
T LC-MS/MS	T CLIA	.61**		
C LC-MS/MS	C CLIA	.42**		
Female				
T LC-MS/MS	T CLIA	.06		
C LC-MS/MS	C CLIA	.33**		
Male				
T LC-MS/MS	T CLIA	.60**		
C LC-MS/MS	C CLIA	.64**		

Note: Descriptive statistics (mean, variance, range) and Pearson correlations are based on raw values of variables. Significant correlations ($p < 0.001$) are indicated with **. One outlier in female cortisol (CLIA cortisol, 62.42) was detected but retained in data. When outlier was removed, pearson correlations of CLIA and LC-MS/MS cortisol were 0.61** in the full dataset and 0.58** in females only.

Table 2
Passing-Bablok regression results.

	Passing-Bablok (CLIA vs LCMS) Raw Values			
	N	Proportional Bias	Systematic Bias	Pearson's r
		Slope (LCI-UCI)	Intercept (LCI-UCI)	
T all	509	**3.95 [3.43, 4.61]**	**4.69 [1.9, 7.33]**	.61
T F	258	**17.94 [13.1, 26.66]**	**−31.62 [-57.83, -16.92]**	.06
T M	251	**3.3 [2.77, 3.86]**	−2.21[-8.88, 2.98]	.60
T F low PDS	14†	−90.56[-178.71, 1407.33]†	99.59[-2749.23, 385.1]†	−0.44†
T F high PDS	203	**16.54 [12.02, 24.41]**	**−27.72 [-53.82, -14.95]**	.17
T M low PDS	66	**5.85 [3.81, 8.93]**	−10.41[-28.70, 1.04]	.57
T M high PDS	142	**3.43 [2.7, 4.22]**	−9.53[-21.12, 2.2]	.64
C All	514	**2.62 [2.38, 2.92]**	-.005[-.25, .19]	.42
C F	263	**2.58 [2.27, 3.07]**	.03[−.41, .34]	.33
C M	251	**2.64 [2.3, 3.01]**	-.04[-.33, 0.24]	.64
C F low PDS	14†	**2.87 [1.46, 4.76]** †	-.11[-2.48, 1.34]†	0.77†
C F high PDS	208	**2.4 [2.09, 2.84]**	.14[-.25, .49]	.56
C M low PDS	67	**2.5 [2.03, 3.25]**	.05[-.48, .48	.81
C M high PDS	141	**3 [2.36, 3.79]**	-.24[-.79, .31	.60

Note: C = cortisol, T = tesosterone, M = Male, F = Female, PDS = Pubertal Development Scale. LCI = Lower Confidence Interval and UCI = Upper Confidence Interval. LCI and UCI based on bootstrapped (n = 1000) Passing-Bablok regression. Slope CI not including 1 indicates at least proportional bias. Intercept CI not including 0 indicates at least systematic bias. Systematic and proportional bias indicated with bold text. † indicates sample size <30 and unreliable Passing-Bablok results.

3.3. Aim 2 results

3.3.1. Models of testosterone and cortisol predicting psychopathology

Results examining the moderating effect of hormone quantification method on hormone-psychopathology associations showed no statistically significant moderation in the whole group or in males or females separately (Table 3).

4. Discussion

This study sought to add to the existing literature comparing analysis methods for the measurement of salivary hormone concentrations. We 1) expanded on previous method comparison studies by comparing CLIA and LC-MS/MS results of testosterone and cortisol in a relatively large adolescent sample, and 2) tested for the moderating impact of hormone analysis method on hormone-psychopathology relationships.

Our findings in aim one support findings by Bae and colleagues showing proportional differences between CLIA and LC-MS/MS-assessed cortisol measurements in 8–14 year-olds [1]. This study replicates those same findings in cortisol in our sample of 13–16 year-old high school students, and shows that CLIA, like EIA, inflates testosterone results relative to LC-MS/MS. Observing these systematic differences in pubertal teens suggests that, despite the surges in steroid hormones, especially testosterone, that define and precipitate puberty, differences in CLIA or LC-MS/MS-assessed hormone concentrations persist, with particularly large method discrepancies occurring when testosterone values are low. Indeed, female testosterone values, which tended to be lower, were significantly more proportionally biased by CLIA measurement than male testosterone values, suggesting that CLIA measurement of low testosterone may be producing more error than signal.

Our results showed no differences in CLIA and LC-MS/MS agreement at different levels of PDS (see Table 2 and Bland Altman plots) and a significant relationship between testosterone and PDS in males but not females. These results are interesting to consider in light of the restricted age and PDS range of our present sample. In our sample of 13–16 year-olds, PDS ranged from 1.6 to 4 in males, and 1.8–4 in females (out of a possible 1–4). We additionally observed left skew of PDS concentrations in females, suggesting that a substantial proportion of females reported high levels of pubertal development. Studies of longitudinal adolescent hormone changes over the course of development have shown that pubertal hormones begin to increase starting at around ages 6–8.

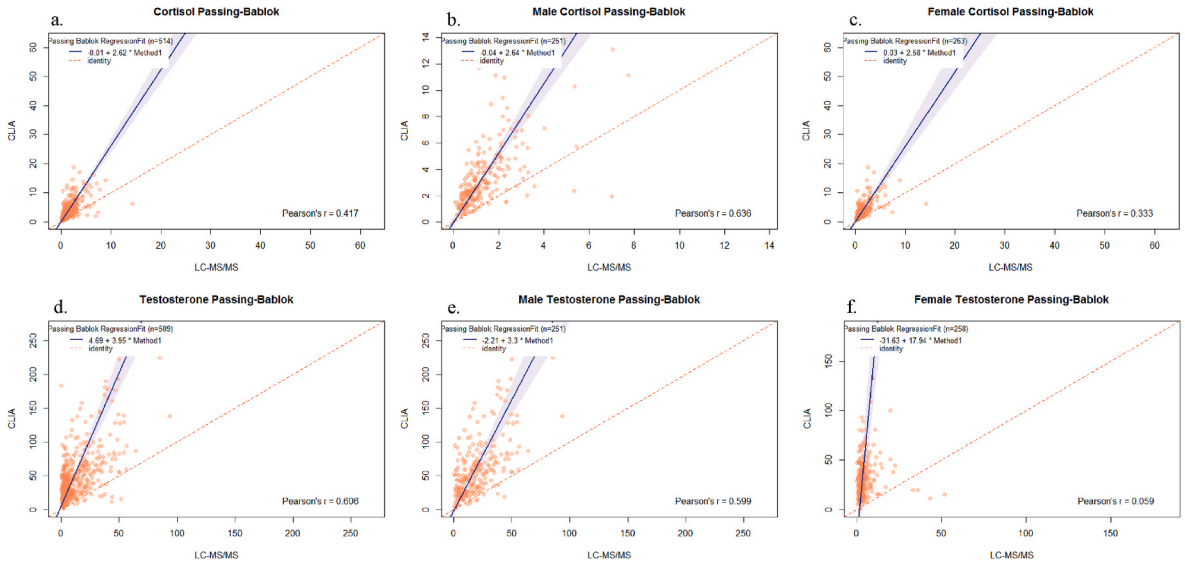

Fig. 1. CLIA and LC-MS/MS are differentially correlated by sex in testosterone, not cortisol. Note: Passing-Bablok regressions identify proportional (slope confidence intervals that do not include 1) and systematic (intercept confidence intervals that do not include 0) bias. Orange dotted line represents exact agreement between methods. Results show that CLIA (y-axis) proportionally biases testosterone and cortisol relative to LC-MS/MS (x-axis) such that CLIA overestimates testosterone and cortisol in the full group (a. and d.) and in males (b. and e.) and females (c. and f.) separately. Testosterone slopes in males (e.) significantly differ (do not overlap with) slopes in females (f.). These sex differences were not observed in cortisol (b. and c.). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

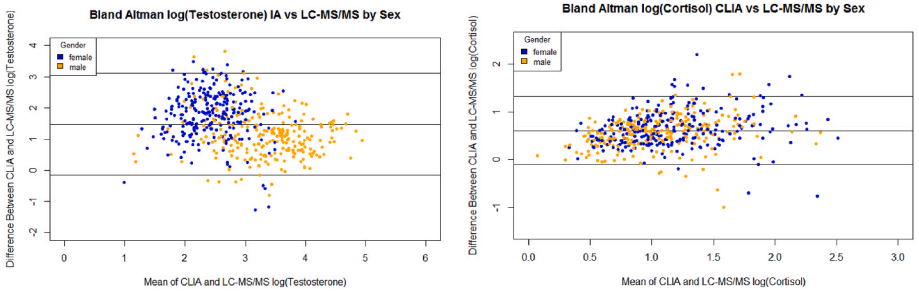

Fig. 2. Low values of testosterone, not cortisol, tend to be over-estimated by CLIA relative to LCMS/MS. Note: Bland Altman Plot depicts difference between log(testosterone) CLIA and log(testosterone) LC-MS/MS values (y-axis) along the range of average testosterone and cortisol values (from both methods; x-axis). Blue datapoints represent females and orange datapoints represent males. Plots show that CLIA particularly overestimates low values of testosterone, where a high percentage of female values are present. No sex differences were observed in the extent to which CLIA overestimated cortisol levels relative to LC-MS/MS (orange and blue dots overlap over the span of the x-axis values). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Therefore, it may be that our sample had too small of an age range and too restricted a PDS range to see any differences in method agreement by pubertal development. Indeed, by age 14, group differences in hormone concentrations by pubertal status are less pronounced than they are in samples with larger age ranges [56]. It is our recommendation that a replication of the analyses presented here in a sample with a wider age and developmental range that includes individuals at both low and high extremes of pubertal development would allow for a more thorough examination of our questions surrounding method agreement and pubertal development. Such a study would enable the field to address the issue of accuracy of testosterone measurement in participants who are just starting the transition to puberty. As more research focuses on this critical developmental milestone, and as more findings highlight the pubertal tempos, trajectories, and hormonal profiles associated with psychopathological risk, the importance of accuracy of hormone measurement cannot be overstated.

Results from our second aim found no significant impact of hormone measurement method on the linear relationships between hormone levels and measures of psychopathology in our non-clinical sample. Given the proportional bias we observed in CLIA-assessed cortisol and testosterone, however, we suggest that it may be important to look at this question in a larger clinical participant pool with expanded measures of psychopathology. It may be the case that, for instance, the relationship between testosterone and female externalizing behaviors is suppressed when testosterone is measured with CLIA, due to overestimation of female testosterone values. Additionally, it may be the case that CLIA bias introduces measurement error in assessments of cortisol release in response to a stressor, which may suppress our ability to identify patterns of stress-linked cortisol release. Given these possibilities, it is our recommendation that testosterone values and cortisol values should be measured with LC-MS/MS whenever possible.

4.1. *Conclusions and limitations*

Overall, the findings presented here expanded on past research by examining cortisol and testosterone bias in CLIA relative to LC-MS/MS in a sample of adolescents, a novel age group in which to investigate method differences in the quantification of both HPA and HPG axis hormones. Our findings showed that, in our sample of 207 males and females, CLIA measurement introduced proportional bias in cortisol and testosterone measurement, and suppressed sex differences in testosterone. We did not observe significant differences in method agreement at different levels of pubertal development and did not observe significant moderation of hormone-psychopathology relationships by method. The proportional bias introduced by CLIA measurement, however, suggests that researchers measuring testosterone and cortisol may

Table 3
Linear regression results.

Outcome	Moderators	hormone	*b*	se	*p*	r
CDI	Method	testosterone	-.01	.04	.85	.00
PSS	Method	testosterone	.03	.06	.59	.00
SAS-Avoid	Method	testosterone	-.06	.11	.57	.01
SAS-Neg Eval	Method	testosterone	.06	.12	.63	.00
Loneliness	Method	testosterone	-.08	.09	.37	.00
CDI	Method, Females	testosterone	-.01	.05	.92	.00
PSS	Method, Females	testosterone	-.01	.09	.91	.00
SAS-Avoid	Method, Females	testosterone	-.13	.15	.41	.01
SAS-Neg Eval	Method, Females	testosterone	-.02	.18	.91	.01
Loneliness	Method, Females	testosterone	-.03	.13	.83	.01
CDI	Method, Males	testosterone	-.02	.05	.63	.00
PSS	Method, Males	testosterone	.03	.08	.71	.00
SAS-Avoid	Method, Males	testosterone	-.06	.16	.71	.02
SAS-Neg Eval	Method, Males	testosterone	.06	.15	.67	.00
Loneliness	Method, Males	testosterone	-.16	.14	.25	.01
CDI	Method	cortisol	-.02	.04	.61	.00
PSS	Method	cortisol	-.04	.06	.53	.01
SAS-Avoid	Method	cortisol	-.11	.11	.34	.00
SAS-Neg Eval	Method	cortisol	-.11	.12	.35	.00
Loneliness	Method	cortisol	-.01	.10	.93	.00
CDI	Method, Females	cortisol	-.02	.05	.71	.00
PSS	Method, Females	cortisol	-.04	.09	.67	.00
SAS-Avoid	Method, Females	cortisol	.01	.15	.96	.00
SAS-Neg Eval	Method, Females	cortisol	-.01	.17	.94	.00
Loneliness	Method, Females	cortisol	.02	.12	.84	.00
CDI	Method, Males	cortisol	-.01	.05	.86	.01
PSS	Method, Males	cortisol	-.02	.09	.83	.02
SAS-Avoid	Method, Males	cortisol	-.23	.18	.20	.01
SAS-Neg Eval	Method, Males	cortisol	-.21	.17	.20	.01
Loneliness	Method, Males	cortisol	-.04	.15	.78	.00

Note: Table presents results of linear regressions examining impact of method (CLIA or LC-MS/MS) on hormone-psychopathology relationships. CDI = Children's Depression Inventory, PSS = Perceived Stress Scale, SAS-Avoid = Social Anxiety Scale for Adolescents, Avoidance subscale, SAS-Neg Evaluation = Social Anxiety Scale for Adolescents, Fear of Negative Evaluation subscale, Loneliness = UCLA Loneliness Scale 8. Results revealed no significant method differences predicting psychopathology outcomes.

benefit from use of mass spectrometry methods for analysis of salivary steroid hormone samples.

While the findings presented in this manuscript provide useful information for researchers seeking to incorporate hormone analysis into studies, they should be evaluated conservatively for a number of reasons. First, samples collected and quantified for this study were all first quantified with CLIA, underwent two additional freeze-thaw cycles, and then were quantified with LC-MS/MS. Findings in the literature are mixed as to the effect of multiple freeze-thaw cycles on steroid hormones, which overall tend to be more robust than other hormones and other analytes [57–59] in cattle; [60]. Additionally, studies have shown that multiple freeze-thaw cycles are more likely to impact immunoassay-quantification than LC-MS/MS due to the higher sensitivity of immunoassay methods to matrix interferences such as bacterial contamination [61]. This suggests that multiple freeze-thaw cycles may have less impact on LC-MS/MS methods, and indeed in our sample, time-and-participant-matched samples measured only with LC-MS/MS that had the same number of freeze-thaw cycles as CLIA data were not significantly different (p values ranged between 0.1665 and 0.778687) from the LC-MS/MS data presented here. While these analyses provide preliminary support for the validity of our multiple freeze-thaw LC-MS/MS results, our findings may be impacted in ways we cannot assess.

In addition to the issue of number of freeze-thaw cycles that samples in this study underwent, it is also important to note that samples were analyzed by two different methods in two different labs. While each participant's samples traveled to each lab, previous work examining the impact of laboratory on concentrations of hormones (round robin studies) has shown that differences often arise as a result of analysis in different laboratories [62,63]. Additionally, as we only examined LC-MS/MS and CLIA we are unable to draw conclusions about other immunoassays, such as radioimmunoassay, that may produce results more similar to LC-MS/MS. These impacts are important to consider when evaluating our findings.

Finally, aim two examined hormone-psychopathology relationships in a field study and in a healthy sample of adolescents. Variability in the time of day of collection was present, though limited, as was variability in exposure to environmental stimuli, as subjects were not brought to a lab. Additionally, aim two analyses included only one afternoon sample from each participant. For the purposes of our first aim, a direct comparison of methods, these facts of our sample did not produce substantial issues. Analyses of hormone-psychopathology relationships, however, may be suppressed or inflated due to these characteristics of our sample. Analyses of findings before and after setting time to a fixed constant did not produce different findings, however there may be effects of time of day that were not measured by our study. Additionally, it may be the case that a sample over-selected for participants who meet clinical criteria for psychopathological diagnoses may show different trends in hormone concentrations that may be more or less impacted by issues related to methodology. For instance, if individuals with a diagnosis of depression tend to have low testosterone concentrations, it may be the case that measurement of testosterone in these individuals, regardless of biological sex, are more likely to be universally inflated, leading to erroneous conclusions, or possibly obscuring true testosterone-depression relationships in this population. All of these limitations are important to consider when evaluating the findings presented in this study.

4.2. Recommendations and future directions

The two aims of this study were to 1) expand on previous methodological comparisons by comparing cortisol and testosterone concentrations measured with CLIA to those measured with LC-MS/MS in a relatively large sample of adolescents, and 2) examine the impact of hormone analysis method on hormone-psychopathology relationships. Based on our results, we make the following recommendations to future researchers planning to measure testosterone or cortisol in participant samples with this age range:

1) Testosterone should be quantified with mass spectrometry methods to avoid the proportional bias introduced by CLIA. This is especially vital if researchers intend to examine a population with known low testosterone concentrations such as are present in female populations.
2) Cortisol should, where possible, be quantified with mass spectrometry methods to avoid proportional bias introduced by CLIA.

It is our hope that the work presented here may prove useful for researchers seeking to plan and design new studies that involve salivary hormone collection and measurement. The addition of hormone measures to studies seeking to better understand the biological fluctuations that influence and respond to environments and behavior can be made more accurate by starting from a place of accurate measurement.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Acknowledgement

Research reported in this publication was supported by the National Institutes of Health under award number R01HD084772. This research was also supported by grant, P2CHD042849, Population Research

Center, awarded to the Population Research Center at The University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpnec.2022.100132.

References

[1] Y.J. Bae, A. Gaudl, S. Jaeger, S. Stadelmann, A. Hiemisch, W. Kiess, A. Willenberg, M. Schaab, K. von Klitzing, J. Thiery, U. Ceglarek, M. Döhnert, J. Kratzsch, Immunoassay or LC-MS/MS for the measurement of salivary cortisol in children? Clin. Chem. Lab. Med. 54 (5) (2016) 811–822, https://doi.org/10.1515/cclm-2015-0412.

[2] O.C. Schultheiss, G. Dlugash, P.H. Mehta, Hormone measurement in social neuroendocrinology: a comparison of immunoassay and mass spectrometry methods, in: Routledge International Handbook of Social Neuroendocrinology, Routledge, 2018.

[3] K.M. Welker, B. Lassetter, C.M. Brandes, S. Prasad, D.R. Koop, P.H. Mehta, A comparison of salivary testosterone measurement using immunoassays and tandem mass spectrometry, Psychoneuroendocrinology 71 (2016) 180–188, https://doi.org/10.1016/j.psyneuen.2016.05.022.

[4] Schultheiss, O. C., & Stanton, S. J. (n.d.). Assessment of Salivary Hormones. vol. 28.

[5] M. Mezzullo, F. Fanelli, A. Fazzini, A. Gambineri, V. Vicennati, G. Di Dalmazi, C. Pelusi, R. Mazza, U. Pagotto, R. Pasquali, Validation of an LC-MS/MS salivary assay for glucocorticoid status assessment: evaluation of the diurnal fluctuation of cortisol and cortisone and of their association within and between serum and saliva, J. Steroid Biochem. Mol. Biol. 163 (2016) 103–112, https://doi.org/10.1016/j.jsbmb.2016.04.012.

[6] S.A. Wudy, G. Schuler, A. Sánchez-Guijo, M.F. Hartmann, The art of measuring steroids: principles and practice of current hormonal steroid analysis, J. Steroid Biochem. Mol. Biol. 179 (2018) 88–103, https://doi.org/10.1016/j.jsbmb.2017.09.003.

[7] A. Mazur, S. Clifton, Enzyme immunoassay may be inadequate for measuring salivary testosterone in older men, in: The Aging Male, vols. 1–9, 2018, https://doi.org/10.1080/13685538.2018.1509206.

[8] R. Miller, F. Plessow, M. Rauh, M. Gröschl, C. Kirschbaum, Comparison of salivary cortisol as measured by different immunoassays and tandem mass spectrometry, Psychoneuroendocrinology 38 (1) (2013) 50–57, https://doi.org/10.1016/j.psyneuen.2012.04.019.

[9] S. Baecher, S.C. Azad, M. Vogeser, Inter-method comparison of salivary cortisol measurement, LaboratoriumsMedizin 37 (5) (2013) 269–273, https://doi.org/10.1515/labmed-2013-0008.

[10] M. Yasuda, S. Honma, K. Furuya, T. Yoshii, Y. Kamiyama, H. Ide, S. Muto, S. Horie, Diagnostic significance of salivary testosterone measurement revisited: using liquid chromatography/mass spectrometry and enzyme-linked immunosorbent assay, J. Mens Health 5 (1) (2008) 56–63, https://doi.org/10.1016/j.jomh.2007.12.004.

[11] A. Angold, E.J. Costello, A. Erkanli, C.M. Worthman, Pubertal changes in hormone levels and depression in girls, Psychol. Med. 29 (5) (1999) 1043–1053, https://doi.org/10.1017/S0033291799008946.

[12] W.E. Copeland, C. Worthman, L. Shanahan, E.J. Costello, A. Angold, Early pubertal timing and testosterone associated with higher levels of adolescent depression in girls, J. Am. Acad. Child Adolesc. Psychiatry 58 (12) (2019) 1197–1206, https://doi.org/10.1016/j.jaac.2019.02.007.

[13] D.A. Granger, E.A. Shirtcliff, C. Zahn–Waxler, B. Usher, B. Klimes–Dougan, P. Hastings, Salivary testosterone diurnal variation and psychopathology in adolescent males and females: individual differences and developmental effects, Dev. Psychopathol. 15 (2) (2003), https://doi.org/10.1017/S0954579403000233.

[14] H.S. Kamin, D.A. Kertes, Cortisol and DHEA in development and psychopathology, Horm. Behav. 89 (2017) 69–85, https://doi.org/10.1016/j.yhbeh.2016.11.018.

[15] N.L. Colich, K. Kircanski, L.C. Foland-Ross, I.H. Gotlib, HPA-axis reactivity interacts with stage of pubertal development to predict the onset of depression, Psychoneuroendocrinology 55 (2015) 94–101, https://doi.org/10.1016/j.psyneuen.2015.02.004.

[16] J. LeMoult, S.J. Ordaz, K. Kircanski, M.K. Singh, I.H. Gotlib, Predicting first onset of depression in young girls: interaction of diurnal cortisol and negative life events, J. Abnorm. Psychol. 124 (4) (2015) 850–859, https://doi.org/10.1037/abn0000087.

[17] N.L. Lopez-Duran, M. Kovacs, C.J. George, Hypothalamic–pituitary–adrenal axis dysregulation in depressed children and adolescents: a meta-analysis, Psychoneuroendocrinology 34 (9) (2009) 1272–1283, https://doi.org/10.1016/j.psyneuen.2009.03.016.

[18] M.C. Morris, U. Rao, J. Garber, Cortisol responses to psychosocial stress predict depression trajectories: social-evaluative threat and prior depressive episodes as moderators, J. Affect. Disord. 143 (1–3) (2012) 223–230, https://doi.org/10.1016/j.jad.2012.05.059.

[19] E. Iob, J.R. Baldwin, R. Plomin, A. Steptoe, Adverse childhood experiences, daytime salivary cortisol, and depressive symptoms in early adulthood: A longitudinal genetically informed twin study, Transl. Psychiatry 11 (1) (2021) 1–10, https://doi.org/10.1038/s41398-021-01538-w.

[20] A.E. Johnson, N.B. Perry, C.E. Hostinar, M.R. Gunnar, Cognitive–affective strategies and cortisol stress reactivity in children and adolescents: Normative development and effects of early life stress, Dev. Psychobiol. 61 (7) (2019) 999–1013, https://doi.org/10.1002/dev.21849.

[21] K.E. Speer, S. Semple, N. Naumovski, N.M. D'Cunha, A.J. McKune, HPA axis function and diurnal cortisol in post-traumatic stress disorder: a systematic review, Neurobiology of Stress 11 (2019) 100180, https://doi.org/10.1016/j.ynstr.2019.100180.

[22] G.E. Tafet, R. Bernardini, Psychoneuroendocrinological links between chronic stress and depression, Prog. Neuro Psychopharmacol. Biol. Psychiatr. 27 (6) (2003) 893–903, https://doi.org/10.1016/S0278-5846(03)00162-3.

[23] M. Zitzmann, Testosterone, mood, behaviour and quality of life, Andrology 8 (6) (2020) 1598–1605, https://doi.org/10.1111/andr.12867.

[24] R Studio Team, *RStudio: Integrated Development For R* (1.2.1335, RStudio, Inc, 2018 [R Studio] http://www.rstudio.com/.

[25] R.M. Rose, L.E. Kreuz, J.W. Holaday, K.J. Sulak, C.E. Johnson, Diurnal variation of plasma testosterone and cortisol, J. Endocrinol. (1972) 177–178.

[26] D.S. Yeager, H.Y. Lee, J.P. Jamieson, How to improve adolescent stress responses: insights from integrating implicit theories of personality and biopsychosocial models, Psychol. Sci. 27 (8) (2016) 1078–1091, https://doi.org/10.1177/0956797616649604.

[27] E.A. Shirtcliff, D.A. Granger, E. Schwartz, M.J. Curran, Use of salivary biomarkers in biobehavioral research: cotton-based sample collection methods can interfere with salivary immunoassay results, Psychoneuroendocrinology 26 (2) (2001) 165–173, https://doi.org/10.1016/S0306-4530(00)00042-1.

[28] C.M. Beaven, W.G. Hopkins, K.T. Hansen, M.R. Wood, J.B. Cronin, T.E. Lowe, Dose effect of caffeine on testosterone and cortisol responses to resistance exercise, Int. J. Sport Nutr. Exerc. Metabol. 18 (2) (2008) 131–141, https://doi.org/10.1123/ijsnem.18.2.131.

[29] E.K. Adam, M. Kumari, Assessing salivary cortisol in large-scale, epidemiological research, Psychoneuroendocrinology 34 (10) (2009) 1423–1436, https://doi.org/10.1016/j.psyneuen.2009.06.011.

[30] D.V. Leal, L. Taylor, J. Hough, Exercise-induced salivary hormone responses to high-intensity, self-paced running, Int. J. Sports Physiol. Perform. 1 (aop) (2021) 1–9, https://doi.org/10.1123/ijspp.2020-0541.

[31] W. Gao, T. Stalder, C. Kirschbaum, Quantitative analysis of estradiol and six other steroid hormones in human saliva using a high throughput liquid chromatography–tandem mass spectrometry assay, Talanta 143 (2015) 353–358, https://doi.org/10.1016/j.talanta.2015.05.004.

[32] M. Kovacs, A. Beck, An empirical-clinical approach toward a definition of childhood depression, Depression in Childhood: Diagnosis, Treatment, and Conceptual Models (1977) 1–25.

[33] S. Sun, S. Wang, The children's depression inventory in worldwide child development research: a reliability generalization study, J. Child Fam. Stud. 24 (8) (2015) 2352–2363, https://doi.org/10.1007/s10826-014-0038-x.

[34] H.M. Inderbitzen-Nolan, K.S. Walters, Social anxiety scale for adolescents: normative data and further evidence of construct validity, J. Clin. Child Psychol. 29 (3) (2000) 360–371, https://doi.org/10.1207/S15374424JCCP2903_7.

[35] D.W. Russell, UCLA loneliness scale (version 3): reliability, validity, and factor structure, J. Pers. Assess. 66 (1) (1996) 20–40, https://doi.org/10.1207/s15327752jpa6601_2.

[36] R. Hays, M. DiMatteo, A Short-Form Measure of Loneliness, J. Pers. Assess. 51 (1987) 69–81, https://doi.org/10.1207/s15327752jpa5101_6.

[37] J. Hudiyana, T.M. Lincoln, S. Hartanto, M.A. Shadiqi, M.N. Milla, H. Muluk, E. S. Jaya, How universal is a construct of loneliness? Measurement invariance of the UCLA loneliness scale in Indonesia, Germany, and the United States, Assessment (2021), https://doi.org/10.1177/10731911211034564, 10731911211034564.

[38] A.M. La Greca, N. Lopez, Social anxiety among adolescents: linkages with peer relations and friendships, J. Abnorm. Child Psychol. 26 (2) (1998) 83–94, https://doi.org/10.1023/A:1022684520514.

[39] Cohen, S. (n.d.-a). PERCEIVED STRESS SCALE. vol. 5.

[40] Cohen, S. (n.d.-b). PERCEIVED STRESS SCALE. vol. 5.

[41] A.C. Petersen, L. Crockett, M. Richards, A. Boxer, A self-report measure of pubertal status: reliability, validity, and initial norms, J. Youth Adolesc. 17 (2) (1988) 117–133, https://doi.org/10.1007/BF01537962.

[42] M.E. Koopman-Verhoeff, C. Gredvig-Ardito, D.H. Barker, J.M. Saletin, M. A. Carskadon, Classifying pubertal development using child and parent report: comparing the pubertal development scales to tanner staging, J. Adolesc. Health 66 (5) (2020) 597–602, https://doi.org/10.1016/j.jadohealth.2019.11.308.

[43] J. Martin Bland, DouglasG. Altman, Statistical methods for assessing agreement between two methods OF clinical measurement, Lancet 327 (8476) (1986) 307–310, https://doi.org/10.1016/S0140-6736(86)90837-8.

[44] J.M. Bland, D.G. Altman, Measuring agreement in method comparison studies, Stat. Methods Med. Res. 8 (2) (1999) 135–160, https://doi.org/10.1177/096228029900800204.

[45] N.Ö. Doğan, Bland-Altman analysis: A paradigm to understand correlation and agreement, Turkish J. Emerg. Med. 18 (4) (2018) 139–141, https://doi.org/10.1016/j.tjem.2018.09.001.

[46] B. Mayer, W. Gaus, U. Braisch, The fallacy of the Passing-Bablok-regression, Jokull 66 (6) (2016) 95–106.

[47] D. Giavarina, Understanding Bland Altman analysis, Biochem. Med. 25 (2) (2015) 141–151, https://doi.org/10.11613/BM.2015.015.

[48] I. Marín-Franch, Passing–Bablok regression is inappropriate for assessing association between structure and function in glaucoma, Investig. Ophthalmol. Vis. Sci. 54 (8) (2013) 5848–5849, https://doi.org/10.1167/iovs.13-12372.
[49] R.B. Payne, Method comparison: Evaluation of least squares, deming and passing/Bablok regression procedures using computer simulation, Ann. Clin. Biochem. 34 (3) (1997) 319–320, https://doi.org/10.1177/000456329703400317.
[50] H. Passing, W. Bablok, A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, part I, Clin. Chem. Lab. Med. 21 (11) (1983), https://doi.org/10.1515/cclm.1983.21.11.709.
[51] W. Bablok, H. Passing, Application of statistical procedures in analytical instrument testing, J. Automat. Chem. 7 (2) (1985) 74–79, https://doi.org/10.1155/S1463924685000177.
[52] H. Passing, W. Bablok, Comparison of several regression procedures for method comparison studies and determination of sample sizes application of linear regression procedures for method comparison studies in clinical chemistry, part II, J. Clin. Chem. Clin. Biochem. 22 (6) (1984) 431–445, https://doi.org/10.1515/cclm.1984.22.6.431.
[53] L. Bilić-Zulle, Comparison of methods: passing and Bablok regression, Biochem. Med. 21 (1) (2011) 49–52, https://doi.org/10.11613/BM.2011.010.
[54] R. Miller, F. Plessow, Transformation techniques for cross-sectional and longitudinal endocrine data: application to salivary cortisol concentrations, Psychoneuroendocrinology 38 (6) (2013) 941–946, https://doi.org/10.1016/j.psyneuen.2012.09.013.
[55] T. Hastie, R. Tibshirani, Generalized additive models for medical research, Stat. Methods Med. Res. 4 (3) (1995) 187–196, https://doi.org/10.1177/096228029500400302.
[56] A. Khairullah, L.C. Klein, S.M. Ingle, M.T. May, C.A. Whetzel, E.J. Susman, T. Paus, Testosterone trajectories and reference ranges in a large longitudinal sample of male adolescents, PLoS One 9 (9) (2014), e108838, https://doi.org/10.1371/journal.pone.0108838.
[57] G.W. Comstock, A.E. Burke, E.P. Norkus, G.B. Gordon, S.C. Hoffman, K. J. Helzlsouer, Effects of repeated freeze-thaw cycles on concentrations of cholesterol, micronutrients, and hormones in human plasma and serum, Clin. Chem. 47 (1) (2001) 139–142, https://doi.org/10.1093/clinchem/47.1.139.
[58] G. Gholib, S. Wahyuni, M. Akmal, M. Hasan, M. Agil, B. Purwantara, The validation of a commercial enzyme-linked immunosorbent assay and the effect of freeze-thaw cycles of serum on the stability of cortisol and testosterone concentrations in Aceh cattle, *F1000Research*, https://doi.org/10.12688/f1000research.19804.3, 2020, 8, 1220.
[59] A.H. Garde, Å.M. Hansen, Long-term stability of salivary cortisol, Scand. J. Clin. Lab. Investig. 65 (5) (2005) 433–436, https://doi.org/10.1080/00365510510025773.
[60] R.E. Gislefoss, M. Lauritzen, H. Langseth, L. Mørkrid, Effect of multiple freeze-thaw cycles on selected biochemical serum components, Clin. Chem. Lab. Med. 55 (7) (2017) 967–973, https://doi.org/10.1515/cclm-2016-0892.
[61] S. Prasad, B. Lassetter, K.M. Welker, P.H. Mehta, Unstable correspondence between salivary testosterone measured with enzyme immunoassays and tandem mass spectrometry, Psychoneuroendocrinology 109 (2019), https://doi.org/10.1016/j.psyneuen.2019.104373.
[62] J.M. Dabbs Jr., B.C. Campbell, B.A. Gladue, A.R. Midgley, M.A. Navarro, G.F. Read, E.J. Susman, L.M. Swinkels, C.M. Worthman, Reliability of salivary testosterone measurements: a multicenter evaluation, Clin. Chem. 41 (11) (1995) 1581–1584, https://doi.org/10.1093/clinchem/41.11.1581.
[63] Å.M. Hansen, A.H. Garde, R. Persson, Sources of biological and methodological variation in salivary cortisol and their impact on measurement among healthy adults: a review, Scand. J. Clin. Lab. Investig. 68 (6) (2008) 448–458, https://doi.org/10.1080/00365510701819127.

Further reading

[46] J.M. Bland, D.G. Altman, Comparing methods of measurement: why plotting difference against standard method is misleading, Lancet 346 (8982) (1995) 1085–1087, https://doi.org/10.1016/S0140-6736(95)91748-9.