



# Stability and validity of steroid hormones in hair and saliva across two ovulatory cycles

Julia Stern<sup>a,b,\*</sup>, Ruben C. Arslan<sup>c,d</sup>, Lars Penke<sup>b,e</sup>

<sup>a</sup> Department of Psychology, University of Bremen, Grazer Strasse 2c, 28359, Bremen, Germany

<sup>b</sup> Department of Psychology, University of Goettingen, Gosslerstrasse 14, 37073, Goettingen, Germany

<sup>c</sup> Department of Psychology, University of Leipzig, Neumarkt 9-19, 04109, Leipzig, Germany

<sup>d</sup> Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195, Berlin, Germany

<sup>e</sup> Leibniz ScienceCampus Primate Cognition, Germany

## ARTICLE INFO

### Keywords:

LC-MS/MS

Hair hormones

Saliva hormones

Ovulatory cycle

Steroid hormones

## ABSTRACT

Steroid hormones are often assessed via saliva samples, as they are noninvasive and easy to collect. However, hormone levels in saliva can fluctuate from moment-to-moment, are influenced by factors such as momentary emotional states and food intake, and some vary strongly across women's ovulatory cycle. In contrast, hormone levels in hair seem to be more robust against these influences and were previously suggested to be a good alternative to obtain women's baseline hormone levels. In the current study, we investigated whether hormone levels are stable across multiple assays and whether hormone levels from saliva and hair samples correlate. We collected saliva and hair samples from  $N = 155$  naturally cycling women across two ovulatory cycles. All samples were analyzed for progesterone, testosterone and cortisol levels via mass spectrometry (LC-MS/MS). Results showed that both averaged saliva and hair hormone levels were moderately stable across cycles. Hair progesterone levels showed higher stability than the respective levels from saliva. Saliva and hair levels for progesterone and testosterone were moderately correlated, whereas cortisol levels from saliva and hair were only weakly correlated. Results suggest that the type of sample from which baseline hormone levels are assessed and the cycle phase in which saliva samples are collected may have a high impact on the obtained results. Implications for future studies are suggested.

## 1. Introduction

Steroid hormone assays of saliva samples have often been used to study effects of hormonal changes across women's ovulatory cycle on diverse psychological outcomes, such as sexual desire, mate preferences, or stress (e.g. Refs. [1–11]). Saliva assays capture moment-to-moment fluctuations in steroid hormones [12,13], are non-invasive and easy to assess (as opposed to, for example, blood samples; [14]). As recent studies suggest that some of the previously reported associations between hormones and women's mating psychology are between-women, rather than within-woman hormonal effects [15,16], there is a need to reliably obtain baseline hormone levels. For this purpose, most previous studies simply averaged multiple salivary hormone measures. However, saliva assays also come with disadvantages. They have to be stored in a freezer immediately after collection and assays are somewhat error-prone because they can be influenced by gum bleeding, smoking,

food, coffee, or alcohol intake and change rapidly in response to emotional states, exercise, competition or stress [17]. Thus, Wang and colleagues (2019) investigated whether hair samples might be an attractive alternative for measuring basal hormone levels across the ovulatory cycle. Whereas salivary hormones fluctuate momentarily, diurnally and across women's ovulatory cycles, hormone levels in hair reflect an average hormone level for a particular time span (e.g. one month in a 1 cm hair segment). Indeed, Wang and colleagues (2019) report a high correspondence between DHEA and testosterone in one hair sample and four averaged saliva sample concentrations in  $N = 10$  participants assessed across two ovulatory cycles ( $r = 0.65$  for DHEA,  $r = 0.67$  for testosterone). Hormone concentrations were stable across hair and saliva samples from both cycles. The authors concluded that hair hormone assays can provide reliable estimates of long-term average hormones and reduce the day-to-day variability of DHEA and testosterone levels.

\* Corresponding author. Department of Psychology, University of Bremen, Grazer Strasse 2c, 28359, Bremen, Germany.

E-mail address: [jstern@uni-bremen.de](mailto:jstern@uni-bremen.de) (J. Stern).

<https://doi.org/10.1016/j.cpnec.2022.100114>

Received 31 October 2021; Received in revised form 17 December 2021; Accepted 9 January 2022

Available online 11 January 2022

2666-4976/© 2022 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

A steroid hormone of particular interest in ovulatory cycle research, is progesterone. Progesterone levels fluctuate strongly across the ovulatory cycle (e.g. Ref. [7]). Whereas levels are rather low in the follicular phase (usually <100 pmol/l), levels rise strongly in the luteal phase with a mid-luteal peak (usually >300 pmol/l; [18]). Thus, on which particular days across the cycle saliva hormone assays are taken and aggregated to get an indicator of between-women progesterone levels might have a strong impact on the results (especially if only two samples are collected, as often done in ovulatory cycle research, e.g. in Refs. [1,19–25]). In recent years, interest in assessing between-women associations of progesterone and individual differences (e.g. in mating psychology) increased, as studies reported that progesterone might rather be associated with between-women than within-woman differences in mate preferences [15,16]. Further, assessing progesterone seems to be particularly important when investigating effects between different ovulatory cycles, as progesterone levels were reported to be higher in more fertile cycles (for a review see Ref. [26]). However, so far, the question whether progesterone in hair samples might be a more stable indicator of between-women differences in hormone levels than progesterone across different saliva samples remains unanswered.

In ovulatory cycle research, testosterone levels are often also assessed as a variable of interest. Testosterone fluctuates across the ovulatory cycle and seems to be higher in the follicular phase as compared to the luteal phase (e.g. Ref. [7]). Importantly, rather between-women (averaged values of two assays) than within-woman testosterone levels have been linked to individual differences in women's sexual motivation [24], showing the importance of a reliable indicator of between-women testosterone levels. As described above, one previous study [27] reported that salivary testosterone levels obtained across women's ovulatory cycle are highly stable (ICC = 0.91) and substantially correlated with hair hormones ( $r = 0.67$ ). Wang and colleagues (2019) assessed four saliva samples across the ovulatory cycle, but cycle studies often only assess two samples per cycle and woman (usually one fertile, one luteal phase sample). The open question whether testosterone levels are already considerably stable across two assays per cycle and how an average of two samples (fertile and luteal phase) is associated with hair testosterone remains.

In other previous studies, hair samples have also been used to investigate long-term effects of cortisol as a biological marker of chronic stress and its effects on health (e.g. Ref. [28]). A number of different studies have already investigated whether long-term cortisol levels in hair are linked to averaged short-term cortisol levels in saliva: In a meta-analysis, Stalder and colleagues (2017) reported that hair and salivary cortisol are significantly but only weakly correlated ( $r$ s between 0.13 and 0.19). As previous studies had only small sample sizes, with a mean  $N = 53$  in studies with single and  $N = 31$  with multiple salivary cortisol assays [29], it seems doubtful that the point estimates are precise and generalizable. Nevertheless, cortisol levels seem to be considerably stable in saliva ( $r = 0.78$  two weeks apart [30]), and in hair ( $r$ s between 0.68 and 0.79 across one year, [31]). While fluctuations in cortisol do not necessarily occur systematically across women's ovulatory cycle [30], it is still important to study cortisol in ovulatory cycle studies, as cortisol might interfere with cycle effects. For example, it has been reported that cortisol might suppress women's preferences for masculine faces across the cycle [1]. Further, high levels of cortisol could lead to decreases in other hormone levels, such as testosterone [32] and the dual-hormone hypothesis suggests that cortisol and testosterone jointly regulate social behavior (e.g. status-relevant behavior; [33]). Consequently, when studying cycle effects, or within-woman or between-women effects of testosterone, it seems crucial to also reliably assess cortisol levels.

### 1.1. Methodological issues in hormone assessments

Besides the disadvantages that come with saliva samples stated above, there is another important issue that arises for both saliva and

hair assays: There are different methods of hormone measurement that differ in validity. More precisely, analyzing hormone assays via liquid chromatography mass spectrometry (LC-MS/MS) is currently seen as the gold standard [34]. While LC-MS/MS is frequently used to analyze hair hormone levels, most previous studies used either radioimmunoassays or enzymatic immunoassays for analyzing salivary hormone assays (see Ref. [34] for a detailed discussion of the differences between immunoassays). Immunoassays have especially been criticized for suffering from cross-reactivities to other substances and for overestimating hormone levels, especially in small ranges [34]. Importantly, estradiol and testosterone levels in women's saliva, as well as cortisol levels in men and women are usually in such a small range and, thus, likely overestimated [34–37]. This may lead to differences in results, depending on the analysis method, which likely impacts replicability. Indeed, estradiol levels in saliva analyzed via immunoassays and LC-MS/MS were reported to be uncorrelated ( $r = 0.06$ ) in a study focussing on changes in women's mate attraction and preferences across the ovulatory cycle [38]. Nevertheless, in comparison to LC-MS/MS, immunoassays are more often used to analyze salivary hormone levels, because they are cheaper and analysts need less training [27,34]. As a consequence, the literature lacks validation studies using saliva and hair LC-MS/MS analyses.

### 1.2. Current study

This study's aim is to investigate the correspondence and stability of progesterone, testosterone and cortisol levels in saliva and hair samples, analyzed via LC-MS/MS. For this purpose, we examine the relationship between hair steroid hormones and the average values of these hormones in saliva samples collected in a comparatively large sample of  $N = 155$  naturally cycling women, with multiple assays across two ovulatory cycles per participant.

## 2. Methods

### 2.1. Participants

A total of 157 participants took part in this study. One participant dropped out because her hair was too short to collect hair samples, one participant did not want to provide hair samples and two other participants only provided one hair sample, resulting in a final sample of 155 participants (with 308 observations). Participants were female, between 18 and 35 years old ( $M = 23.3$ ,  $SD = 3.4$ ), naturally cycling (no hormonal contraception for at least three months, no switch to hormonal contraception during the study, no current pregnancy or breastfeeding, no birth or breastfeeding in the previous three months, not taking hormone-based medication or anti-depressants). Additionally, they had to report that their ovulatory cycles had a regular length between 25 and 35 days during the last 3 months. All participants signed a written consent form and the local ethics committee approved the study protocol (no. 144). Upon completion of all sessions (see below), participants received a payment of 80€ or course credit.

### 2.2. Procedure

All participants took part in five individually scheduled sessions. In the first introductory session, participants received detailed information about the general procedure, duration of the study, and compensation. Furthermore, the experimenter checked the inclusion criteria (e.g. whether participants were naturally cycling). To plan the dates of the next four sessions, participant's average cycle length as well as the dates of their last, their penultimate and the estimated date of their next menstrual onset were assessed. Finally, demographic data was collected.

Sessions two to five (the testing sessions) took place during two (mostly consecutive) ovulatory cycles per participant. All participants attended two sessions while being in their fertile phase (validated with

LH tests) and two sessions while being in their luteal phase. Importantly, both sessions of the same cycle phase were scheduled across two separate ovulatory cycles (i.e. fertile and luteal sessions alternated). To control for possible effects of diurnal changes in hormone levels [30], sessions two to five were scheduled in the second half of the day (mainly between 11.30 a.m. and 6 p.m.). When arriving at the lab, participants first completed a screening questionnaire, assessing their eligibility and some control variables for saliva sampling [17]. Then we collected saliva samples in each of the four sessions. Hair samples were collected in session three and five. Each participant was enrolled in the study for about two months. Out of all participants who finished every testing session, 66 participants started with the first testing session in their luteal phase, 91 started testing in the fertile phase.

## 2.3. Measures

### 2.3.1. Ovulatory cycle phase

Women's cycle phase was determined by the reverse cycle day method, based on the estimated day of the next menstrual onset [39] and confirmed by highly sensitive (10 mIU/ml) urine ovulation test strips from Purbay®, which measure the luteinizing hormone (LH). More precisely, the reverse cycle day method relies on backward counting, i.e. counting from the day of the (estimated) next menstrual onset backward to the day the session takes place. It is assumed that ovulation occurs on average 15-days prior to the next menstrual onset [39]. Thus, we scheduled the fertile phase sessions on the estimated days preceding ovulation (between day 16 and 18 before the next estimated menstrual onset). The luteal phase sessions were scheduled on the days after ovulation (between day 4 and 11 before the next estimated menstrual onset). Scheduling was mainly done via email and was individual for each session and each participant, depending on the participant's ovulatory cycles and schedules. Participants self-reported the actual day of their menstrual onset (via email to the experimenter) and used LH tests during their estimated fertile window to validate our cycle phase estimates. This gave us the possibility to reschedule sessions spontaneously (e.g. if ovulation or menstrual onset occurred earlier than expected). For this purpose, each participant received a minimum of 10 LH tests in the introductory session that were used at home at the estimated day of ovulation and the four days prior (resulting in five LH tests per cycle, although some participants did more tests when the five obligatory tests did not cover an LH surge). Participants were instructed to take the tests at approximately the same time each day and to self-report the results to the experimenter via email. LH test results suggested that fertile phase sessions were scheduled on average 1.12 days before ovulation ( $SD = 1.39$ , median = one day before ovulation, range four days before to three days after ovulation). Further, the actual days of the menstrual onset following the testing session suggested that fertile phase sessions were scheduled on average 17.66 days before the next menstrual onset ( $SD = 4.25$ , median = 17 days before menstrual onset, range 41 days to six days before), whereas luteal phase sessions were scheduled on average 7.69 days before the next menstrual onset ( $SD = 5.11$ , median = 7 days before menstrual onset, range 37 days before to five days after menstrual onset).<sup>1</sup>

The reverse cycle day method is superior to the forward counting method (counting from the last menstrual onset forward to the day of the session), as the luteal phase is less variable in length than the follicular phase. Aside of ovulation assessment via ultrasound, a follow-up to the actual day of menstrual onset in combination with using LH

<sup>1</sup> Note that the scheduled days suggest that, while the vast majority of participants were sampled appropriately, some participants have been missampled (e.g. due to very irregular cycles), as reported in detail in Ref. [10]. As the current manuscript focusses on whether hormones in saliva and hair are associated, not on effects of fertility, we decided to include all hormonal data in the analyses.

tests is currently regarded as the gold standard in the literature on ovulatory cycle research (e.g. Ref. [40]).

### 2.3.2. Saliva hormones

We collected four saliva samples from each participant, one per testing session. Participants rinsed their mouths with water and provided at least 2 ml of saliva via unstimulated passive drool. Contamination of saliva samples was minimized by asking participants to abstain from eating, drinking (except plain water), smoking, chewing gum, or brushing teeth for at least 1 h before each session. Moreover, participants were asked to refrain from ingesting caffeine for at least 3 h before each session and from drinking alcohol, exercising, taking recreational or non-prescribed clinical drugs on the day of each session. To check participants' adherence to these instructions and to assess further potential influences on the saliva samples and hormonal levels, a screening questionnaire was administered at the beginning of the session [17]. Samples were visually inspected for blood contamination and stored at  $-80\text{ }^{\circ}\text{C}$  directly after collection until shipment on dry ice to the Kirschbaum Lab at Technical University of Dresden, Germany (one freeze-thaw cycle). There, progesterone, testosterone and cortisol were assessed via LC-MS/MS (for details see Ref. [41]). The lab reported the following coefficients of variation (CV): progesterone (nominal 0.01 ng/ml) intra-assay CV = 10.8%, inter-assay CV = 9.7%; testosterone (nominal 0.01 ng/ml) intra-assay CV = 7.2%, inter-assay CV = 8.6%; cortisol (nominal 0.01 ng/ml) intra-assay CV = 9.2%, inter-assay CV = 7.7%. The lab reported the following limits of quantification (LOQs): progesterone 5 pg/ml, testosterone 1 pg/ml, cortisol 5 pg/ml.

### 2.3.3. Hair hormones

Two hair samples were taken per participant, one during the third and one during the fifth session (approx. one month apart). For this purpose, participants' hair was pinned up and two wisps of hair from the back of participant's head (each with a diameter of min. 3 mm) were separated and tightened with a thread. Then, both strands were cut as close as possible to the scalp, packed up in aluminium foil and the scalp-near ends were marked on the foil. Samples were stored under dry conditions until shipment to the Kirschbaum Lab, where progesterone, testosterone and cortisol were assessed via LC-MS/MS (for details see Ref. [42]). Importantly, for each hair sample, the last grown 1 cm of hair (closest to the scalp) was analyzed, corresponding to the average hormone levels of the last month. Our and the labs aim was to obtain 7.5 mg of hair for this 1 cm segment, which worked out for the majority of samples. However, as some participants had very short or very thin hair (or only agreed to provide a smaller amount of hair), we did not manage to reach the optimal weight level for all samples (median 7.5 mg, range 0.6 mg–7.5 mg<sup>2</sup>). The lab reported the following LOQs: progesterone 0.09 pg/mg, testosterone 0.08 pg/mg, cortisol 0.09 pg/mg. The following CVs were reported: progesterone (nominal 0.5 pg/mg) intra-assay CV = 7.1%, inter-assay CV = 8.3%; testosterone (nominal 0.5 pg/mg) intra-assay CV = 8.1%, inter-assay CV = 8.8%; cortisol (nominal 0.5 pg/mg) intra-assay CV = 8.4%, inter-assay CV = 8.8%.

## 2.4. Statistical analyses

All analyses in the current manuscript were computed with the statistical software R 4.0.5 [43]. The following packages were used: psych 2.1.6 [44], dplyr 1.0.7 [45], reshape2 1.4.4 [46], ggplot2 3.3.5 [47], ggpubr 0.4.0.999 [48], cocor 1.1–3 [49], DescTools 0.99.43 [50], nopaco 1.0.6 [51], RVAideMemoire 0.9–80 [52]. Open data and analysis script are available at <https://osf.io/nxvrb/>.

<sup>2</sup> Note that 7.5 mg is the maximum value here, as 7.5 mg were analyzed even if the sample was heavier due to the standard procedure of the lab.

### 3. Results

#### 3.1. Preliminary analyses and analytical decisions

We first checked hormone concentration ranges for all assayed hormones. A number of assays dropped out from further analyses because of “not detectable” hormone concentrations (likely because the levels were below the limit of quantification), resulting in 304 observations for saliva progesterone (−4), 307 for saliva testosterone (−1), 308 for saliva cortisol (−0), 305 for hair progesterone (−3), 224 for hair testosterone (−84) and 305 for hair cortisol (−3). Ranges for the remaining saliva hormone levels were: 0.26–1480 pg/ml for progesterone, 1.39–302 pg/ml for testosterone, 0.06–33.98 nmol/l for cortisol. Ranges for the remaining hair hormone levels were: 0.08–223.65 pg/mg for progesterone, 0.09–838 pg/mg for testosterone, 0.15–73.58 pg/mg for cortisol. The reported values suggest that there were a few substantial outliers and that the data was not normally distributed. Thus, we decided to a) compute non-parametric tests to analyze our data and b) log-transform our data and repeat all analyses with parametric tests for robustness checks.<sup>3</sup> Log transforming the data substantially increased the normality of the data (see qq plots in the open script for comparisons). Although some of our variables were not normally distributed after log transformation, we still decided to keep these analyses to show whether they lead to different results compared to the non-parametric tests, as log transforming hormone variables and performing parametric tests afterwards is standard practice in psychoneuroendocrinology.

#### 3.2. Stability of hormone measures

To analyze the stability of hormone concentrations across cycles, we split our data into hormone samples from the two investigated months per person. Then, we computed Spearman rank correlations between assays of both months. For this purpose, the two saliva samples from each month were averaged (following [27]). Saliva samples of both months were moderately to highly correlated (for progesterone,  $\rho = 0.42$ ,  $p < .001$ , 95% CI = [0.26; 0.56], for testosterone,  $\rho = 0.76$ ,  $p < .001$ , 95% CI = [0.68; 0.82], for cortisol,  $\rho = 0.41$ ,  $p < .001$ , 95% CI = [0.27; 0.54]), suggesting stability, but also variability of averaged saliva hormones between cycles. As described above, we also computed Pearson correlation coefficients for log-transformed data. Results were comparable to the Spearman rank correlations (for progesterone,  $r = 0.34$ ,  $p < .001$ , 95% CI = [0.19; 0.47], for testosterone,  $r = 0.78$ ,  $p < .001$ , 95% CI = [0.70; 0.83], for cortisol,  $r = 0.44$ ,  $p < .001$ , 95% CI = [0.31; 0.56]; Fig. 1).

Hair hormones of both months were substantially correlated as well (for progesterone,  $\rho = 0.67$ ,  $p < .001$ , 95% CI = [0.56; 0.77], for testosterone,  $\rho = 0.57$ ,  $p < .001$ , 95% CI = [0.38; 0.73], for cortisol,  $\rho = 0.68$ ,  $p < .001$ , 95% CI = [0.56; 0.78]). Parametric Pearson correlations of the log-transformed data that were conducted for robustness checks revealed comparable results (for progesterone,  $r = 0.71$ ,  $p < .001$ , 95% CI = [0.62; 0.78], for testosterone,  $r = 0.76$ ,  $p < .001$ , 95% CI = [0.66; 0.83], for cortisol,  $r = 0.62$ ,  $p < .001$ , 95% CI = [0.51; 0.71]; Fig. 2).

As correlation coefficients for stability of hair and saliva samples differed, we tested whether stability of hormone levels in hair and saliva were significantly different from each other. Fisher's z tests (based on [53] revealed that stability of hair progesterone ( $z = 12.93$ ,  $p < .001$ ; log-transformed data:  $z = 4.53$ ,  $p < .001$ ) was significantly higher compared to the respective saliva levels. Whether testosterone and cortisol levels were more stable in hair than in saliva depended on data transformation: with testosterone being more stable in hair in the

untransformed dataset ( $z = 17.00$ ,  $p < .001$ ; log-transformed data:  $z = -0.37$ ,  $p = .711$ ), but cortisol in hair being more stable in the log-transformed dataset ( $z = -0.39$ ,  $p = .695$ ; log-transformed data:  $z = 2.10$ ,  $p = .036$ ).

Next, we also computed non-parametric concordance coefficients (CCCs [58]) and parametric intraclass correlation coefficients (ICCs, for the log-transformed dataset) for all four saliva samples per participant, as averaged values might inflate the correlation coefficients due to reduced session-to-session fluctuations. Results suggest moderately to highly stable levels of testosterone (CCC = 0.80, ICC = 0.85) and cortisol (CCC = 0.73, ICC = 0.65): Levels of progesterone expectedly showed stronger fluctuations across the sessions, especially for the parametric ICCS (CCC = 0.65, ICC = 0.22), as two samples come from fertile and two from luteal cycle phase sessions.

#### 3.3. Correlations between hormone concentrations in hair and saliva

Saliva hormone levels from all sessions were averaged and correlated with the average of both hair hormone levels. Results show that saliva and hair values for progesterone ( $\rho = 0.48$ ,  $p < .001$ , 95% CI = [0.34; 0.60]) and testosterone ( $\rho = 0.38$ ,  $p < .001$ , 95% CI = [0.21; 0.51]) are significantly correlated, whereas the correlation between saliva and hair cortisol was small, with the 95% confidence interval including zero ( $\rho = 0.12$ ,  $p = .131$ , 95% CI = [−0.06; 0.28]). Results were virtually identical when computing Pearson correlation coefficients for the log-transformed dataset (for progesterone,  $r = 0.46$ ,  $p < .001$ , 95% CI = [0.33; 0.58], for testosterone,  $r = 0.48$ ,  $p < .001$ , 95% CI = [0.34; 0.60], for cortisol,  $r = 0.08$ ,  $p = .316$ , 95% CI = [−0.08; 0.24]; Fig. 3).

Again, as averaging the values might inflate the correlation coefficients, we decided to additionally analyze correlations of single samples (e.g. first saliva sample with first hair sample), which is also in line with the analyses reported by Wang and colleagues (2019). Results (displayed in Table 1) suggest medium sized associations between progesterone or testosterone samples (for progesterone from the luteal phase sessions only) and the corresponding hair sample. Effects for associations of cortisol from saliva with the corresponding hair samples were small (and non-significant).

#### 3.4. Robustness checks

Besides the already reported robustness checks, we decided to repeat all of our analyses, but with not detectable hormone levels replaced by the level of quantification value (the smallest value that could still be detected by the analysis method), to account for the large number of missing values for some hormones in some sample materials (especially missing testosterone levels from hair samples). Results were virtually identical to the results reported above.

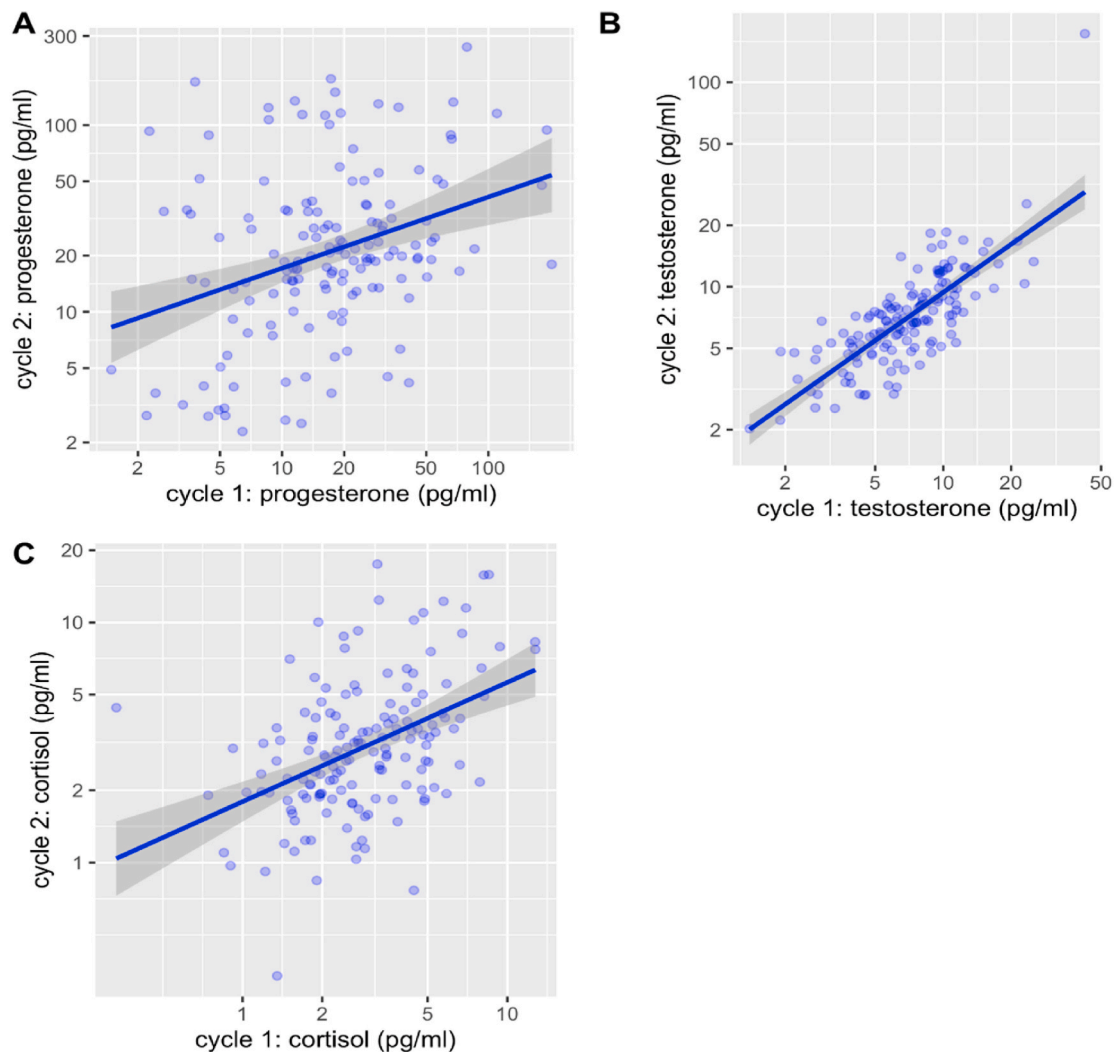
## 4. Discussion

The present study investigated whether steroid hormone levels in hair and saliva correlate with each other and whether levels are stable across two ovulatory cycles, by employing a large dataset and analyzing samples via the current gold standard method LC-MS/MS. Results suggest that average hormone levels of progesterone, testosterone and cortisol in saliva and hair are moderately stable across two cycles. Progesterone levels seem to be more stable in hair than in saliva. Hormone concentrations of averaged as well as single session testosterone and luteal phase progesterone in saliva and hair were moderately correlated as well, whereas the correlation between levels of salivary and hair cortisol was rather weak.

#### 4.1. Stability of hormone measures

Regarding the stability of saliva samples, progesterone, testosterone and cortisol were somewhat stable across two ovulatory cycles. As saliva

<sup>3</sup> We also checked whether different outlier exclusion criteria lead to differences in our results. However, results remained virtually identical when different outlier exclusion criteria were applied.



*Note:* A = association between averaged salivary progesterone levels from both cycles, B = association between averaged salivary testosterone levels from both cycles, C = association between averaged salivary cortisol levels from both cycles.

**Fig. 1.** Associations between two averaged hormone levels from saliva samples across two ovulatory cycles

*Note:* A = association between averaged salivary progesterone levels from both cycles, B = association between averaged salivary testosterone levels from both cycles, C = association between averaged salivary cortisol levels from both cycles.

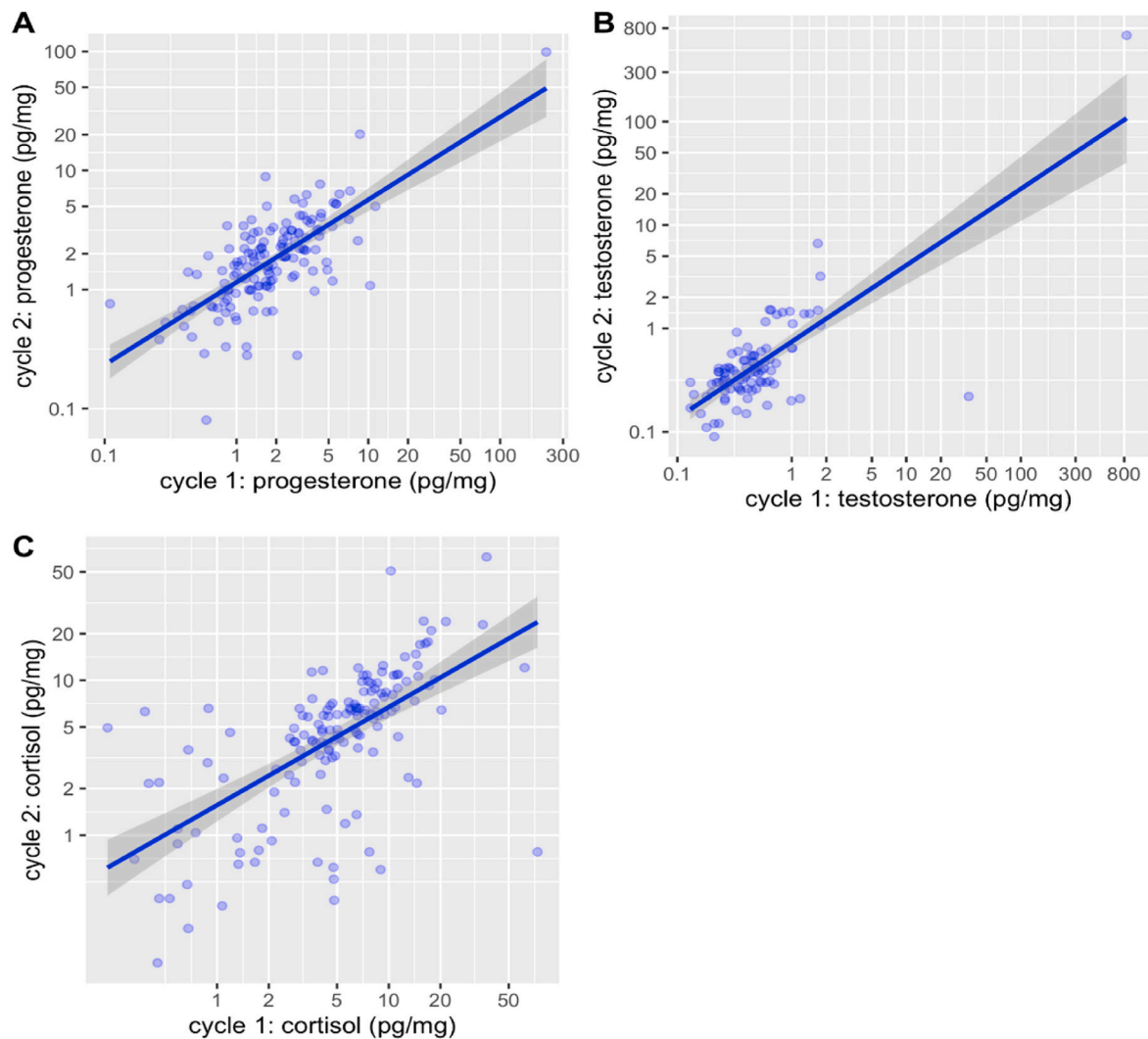
samples also capture moment-to-moment fluctuations, are easily influenced by emotional states and stress, and progesterone in particular fluctuates strongly across the ovulatory cycle, the reported correlation coefficients are in a range that can be expected for measurement stability of these hormones. Testosterone levels in saliva seem to be highly stable, as suggested by the averaged associations, as well as the strong concordance coefficients and intraclass correlations from single sessions. The finding that progesterone levels in women's hair samples showed stronger test-retest associations across two ovulatory cycles than in saliva suggest that levels of progesterone in women's hair indeed reflect stable long-term hormonal concentrations and could be used to study between-women differences in hormones and to more accurately adjust for between-women differences in studies of within-woman effects. Still, two averaged saliva samples across the cycle, especially for salivary testosterone, could be a good indicator for between-women baseline hormone levels.

Our findings fit to the previous literature reporting that single hair

cortisol assessments comprise a strong trait component and are stable across multiple assays [31]. We provide first evidence for the stability of hair progesterone. Further, our findings regarding the stability of testosterone levels in hair are in line with those reported by Wang and colleagues (2019) and suggest that hair samples are appropriate for studying between-women hormone associations of testosterone as well.

#### 4.2. Validity of hormone measures in hair and saliva

While our results suggest small-to-medium sized associations between averaged hair and saliva progesterone and testosterone, averaged cortisol levels from saliva samples seem to be only weakly correlated with levels from hair samples from the respective time span. Associations between hair and saliva samples from single sessions suggest that the cycle phase in which the sample was assayed matters, at least for progesterone levels, as luteal phase saliva samples showed significant higher associations between hair and salivary hormone levels compared



*Note:* A = association between hair progesterone levels from both cycles, B = association between hair testosterone levels from both cycles, C = association between hair cortisol levels from both cycles.

**Fig. 2.** Associations between hormone levels from hair samples across two ovulatory cycles

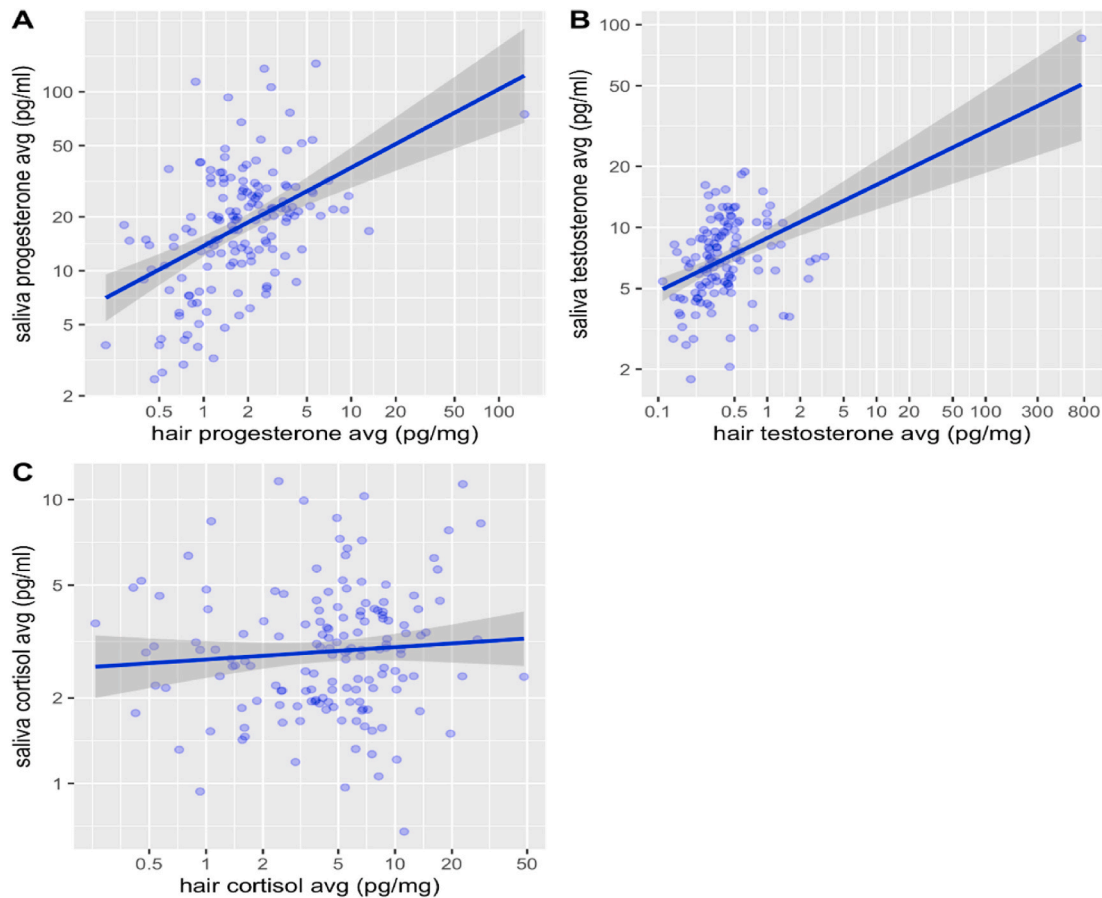
*Note:* A = association between hair progesterone levels from both cycles, B = association between hair testosterone levels from both cycles, C = association between hair cortisol levels from both cycles.

to fertile phase samples. A potential explanation for this result is that progesterone levels fluctuate profoundly across the ovulatory cycle [7]. More precisely, whereas progesterone levels are usually low and not very variable in the follicular phase, they rise substantially across the luteal phase, often being more than ten times higher as compared to the follicular phase. Indeed, in our sample, progesterone levels across the luteal phase samples had a standard deviation that was more than four times higher as compared to fertile phase progesterone levels. The variability of observations substantially affects the size of correlation coefficients, with more variability leading to higher correlation coefficients (e.g. Ref. [54]).

Whereas this is the first study that investigated the links between salivary and hair progesterone (to our knowledge), our results on cortisol fit into the range of previously reported correlation coefficients between hair and salivary cortisol [29]. Cortisol secretion in saliva shows substantial situational variability, e.g. through exercising,

smoking, food intake or circadian rhythmicity [29]. Although we tried to control for these influences, and cortisol across both hair and saliva samples was considerably stable, our results suggest that two salivary cortisol samples per cycle are not a sufficient indicator for long-term cortisol levels (which is in line with Stalder and colleagues, 2017, interpretations). Rather, it seems that more saliva samples are needed to validly match cortisol levels from hair. While Short and colleagues (2016) report that cortisol levels from saliva samples collected each day across a period of 30 days show considerable correlations with hair cortisol reflecting the same 30-day period ( $r = 0.61$ ), another study shows that averaged cortisol levels from six saliva samples already show correlations up to  $r = .57$  with hair cortisol [55].

Our reported association of averaged testosterone levels from women's hair and saliva was smaller, whereas the single sample associations were larger than the association reported by Ref. [27]. Differing methods might explain differences in results. Notably, in Ref. [27] study,



*Note:* A = association between hair and salivary progesterone, B = association between hair and salivary testosterone, C = association between hair and salivary cortisol. Avg = average, prog = progesterone, test = testosterone.

**Fig. 3.** Associations between averaged hair and saliva hormone levels

*Note:* A = association between hair and salivary progesterone, B = association between hair and salivary testosterone, C = association between hair and salivary cortisol. Avg = average, prog = progesterone, test = testosterone.

**Table 1**

Pearson correlation coefficients of the single session saliva and hair samples from the corresponding month.

	Progesterone	Testosterone	Cortisol
Fertile session 1	$\rho = .01,$	$\rho = .25,$	$\rho = .01,$
Spearman-rank	$p = .930$	$p = .010$	$p = .929$
Pearson	$r = .01,$ $p = .901$	$r = .43,$ $p < .001$	$r = .001,$ $p = .991$
Luteal session 1	$\rho = .40,$	$\rho = .44,$	$\rho = .09,$
Spearman rank	$p < .001$	$p < .001$	$p = .264$
Pearson	$r = .42,$ $p < .001$	$r = .21,$ $p = .029$	$r = -.01,$ $p = .903$
Fertile session 2	$\rho = .01,$	$\rho = .19,$	$\rho = .10,$
Spearman rank	$p = .910$	$p = .052$	$p = .232$
Pearson	$r = .07,$ $p = .421$	$r = .49,$ $p < .001$	$r = .10,$ $p = .204$
Luteal session 2	$\rho = .33,$	$\rho = .27,$	$\rho = .10,$
Spearman-rank	$p < .001$	$p = .005$	$p = .222$
Pearson	$r = .44,$ $p < .001$	$r = .41,$ $p < .001$	$r = .14,$ $p = .087$

*Note:* The Pearson correlation coefficients were computed on the log-transformed dataset.

correlations increased with the amount of salivary measures that were averaged. Please note that the same was evident for cortisol levels in a different study [56]. Given that Wang and colleagues (2019) collected four saliva samples per ovulatory cycle, whereas we collected two samples within the same time span, it is possible that correlations might be higher when more saliva samples within-person are collected. This claim is also supported by our results, as multiple averaged saliva samples showed higher associations with levels from hair samples than single saliva samples. Another potential explanation for differences between our and Wang and colleagues (2019) results is that the weight of the used hair samples differed between both studies. Our samples weighed substantially less (Wang and colleagues (2019) used 15 mg hair samples), potentially also resulting in a larger number of missing values for testosterone levels in hair. Moreover, immunoassays (as used in Ref. [27]) and LC-MS/MS require different levels of the targeted chemical to reach a certain level of precision, which may also have influenced the large number of missing values for hair testosterone in our study. However, as our sample size was much larger than those of previous studies and LC-MS/MS analyses are regarded to be more valid than immunoassays [34], especially when analyzing testosterone levels in women [57], it is also possible that the strength of the relationship between salivary and hair testosterone has been overestimated

previously. Overall, averaged values from multiple progesterone or testosterone saliva samples seem to be a somewhat valid indicator of hormone levels in hair and vice versa, whereas individual saliva samples seem to be more prone to daily fluctuations or systematic fluctuations across the cycle, as expected. Thus, hair samples might be superior to saliva samples when obtaining baseline hormone levels and investigate between-women differences in hormones and their effects.

#### 4.3. Limitations

As described above, we only collected two saliva samples per person per cycle, which is the time span covered by one hair sample in our study. We collected the samples across two distinct cycle phases to capture high within-cycle fluctuation of hormones. However, to better understand whether hair hormones provide a cumulative index of the daily variation in saliva hormones across the ovulatory cycle, more frequent (ideally daily) saliva samples per person would be desirable. Further, we did not manage to obtain enough hair from each participant to reach a sample weight of 7.5 mg of hair for all samples, which might have led to a higher rate of missing values for some hormone levels. The high rate of samples with testosterone levels below detection threshold in hair is a clear limitation and could potentially also be due to the sensitivity of the analysis method used. Ideally, we would have a third well-validated measure, such as serum levels of testosterone, to better understand the validity of testosterone levels in saliva and hair towards the lower end of the range.

#### 4.4. Implications for future studies

The results of our study provide a number of implications for future studies, especially for studies investigating potential between-women hormone effects (across the ovulatory cycle). First, although we found that hormone levels in saliva and hair were somewhat stable across two cycles, there were still (sometimes large) fluctuations. More precisely, the cycle phase in which saliva samples are collected matters for the validity of the between-women hormone measure, especially for salivary progesterone. Future studies should rather aim for collecting hair samples when interested in between-women effects of progesterone or, in case hair samples are not possible, they should aim to collect multiple saliva samples across the luteal phase to maximise the validity of their hormone measure.

Second, our results suggest that whether hair or saliva samples are collected to investigate hormonal effects may also influence the results for cortisol. Single or a few averaged saliva samples do not seem to reflect long-term cortisol levels, as measured in hair. When interested in effects of long-term cortisol or to get a precise and valid indicator of between-women differences in cortisol, future studies should aim to collect hair samples rather than saliva samples. Nevertheless, previous studies suggest that increasing the amount of collected saliva samples to at least six samples per participant could substantially increase the validity of between-women estimates for cortisol [55,56].

Third, regarding testosterone levels, saliva and hair samples were almost equally stable across two cycles. However, as the correlation between hair and saliva samples already increased when averaging multiple saliva samples (as compared to one single sample), and a previous study with four averaged saliva samples per cycle reported a stronger association with hair testosterone [27], we also recommend collecting a larger amount of saliva samples per participant when interested in between-women testosterone effects. Further, we would like to highlight the importance of a highly sensitive (LC-MS/MS) method in combination with getting an appropriate amount of weight for hair samples to reduce the number of values below the limit of quantification.

Fourth, it might be interesting for future studies to see that analyzing hormone data using non-parametric tests or log-transforming data and computing parametric tests did not meaningfully affect the results for

the vast majority of analyses. Thus, both procedures might be reasonable to consider. Fifth, we encourage other researchers to investigate the reliability and validity of hormone samples, as hormones and their effects are examined in an increasing amount of studies but standards for accurate hormone measurement continue to evolve.

## 5. Conclusion

In summary, our results provide evidence that hair and salivary hormone levels are moderately correlated for testosterone and progesterone, but rather weakly for cortisol. Hormones in hair and saliva samples seem to be moderately stable across two ovulatory cycles. At least progesterone levels show higher stability in hair samples as compared to saliva samples. Future studies should collect daily saliva samples to compare their average value to hormone levels from hair, and use the most valid methods possible to assess hormone levels.

### Author contributions

JS: conceptualization, validation, formal analysis, investigation, data curation, writing - original draft, writing - review & editing, visualization, project administration.

RCA: validation, writing - review & editing.

LP: conceptualization, resources, writing - review & editing, supervision, funding acquisition.

### Declaration of conflicting interests

The authors declare that they have no conflicts of interests.

### Acknowledgments

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number 254142454/GRK 2070.

### References

- [1] B. Ditzen, S. Palm-Fischbacher, L. Gossweiler, L. Stucky, U. Ehlert, Effects of stress on women's preference for male facial masculinity and their endocrine correlates, *Psychoneuroendocrinology* 82 (2017) 67–74.
- [2] B.C. Jones, A.C. Hahn, C.I. Fisher, H. Wang, M. Kandrik, C. Han, L.M. DeBruine, No compelling evidence that preferences for facial masculinity track changes in women's hormonal status, *Psychol. Sci.* 29 (2018) 996–1005, <https://doi.org/10.1177/0956797618760197>.
- [3] B.C. Jones, A.C. Hahn, C.I. Fisher, H. Wang, M. Kandrik, L.M. DeBruine, General sexual desire, but not desire for uncommitted sexual relationships, tracks changes in women's hormonal status, *Psychoneuroendocrinology* 88 (2018) 153–157.
- [4] J. Jünger, T. Kordsmeyer, T.M. Gerlach, L. Penke, Fertile women evaluate male bodies as more attractive, regardless of masculinity, *Evol. Hum. Behav.* 39 (2018) 412–423.
- [5] J. Jünger, N.V. Motta-Mena, R. Cardenas, D. Bailey, K.A. Rosenfield, C. Schild, L. Penke, D.A. Puts, Do women's preferences for masculine voices shift across the ovulatory cycle? *Horm. Behav.* 106 (2018) 122–134.
- [6] U.M. Marcinkowska, A. Galbarczyk, G. Jasienska, La donna è mobile? Lack of cyclical shifts in facial symmetry, and facial and body masculinity preferences - a hormone based study, *Psychoneuroendocrinology* 88 (2018) 47–53.
- [7] J.R. Roney, Z.L. Simmons, Hormonal predictors of sexual motivation in natural menstrual cycles, *Horm. Behav.* 63 (2013) 636–645.
- [8] J.R. Roney, Z.L. Simmons, Elevated psychological stress predicts reduced Estradiol concentrations in young women, *Adapt. Human Behav. Physiol.* 1 (2015) 30–40.
- [9] J.R. Roney, Z.L. Simmons, Within-cycle fluctuations in progesterone negatively predict changes in both in-pair and extra-pair desire among partnered women, *Horm. Behav.* 81 (2016) 45–52.
- [10] J. Stern, T.M. Gerlach, L. Penke, Probing ovulatory cycle shifts in women's mate preferences for men's behaviors, *Psychol. Sci.* 31 (2020) 424–436, <https://doi.org/10.1177/0956797619882022>.
- [11] J. Stern, T.L. Kordsmeyer, L. Penke, A longitudinal evaluation of ovulatory cycle shifts in women's mate attraction and preferences, *Horm. Behav.* 128 (2021), 104916, <https://doi.org/10.1016/j.yhbeh.2020.104916>.
- [12] J.M. Dabbs Jr., Salivary testosterone measurements: reliability across hours, days, and weeks, *Physiol. Behav.* 48 (1990) 83–86.
- [13] L. Dettenborn, C. Muhtz, N. Skoluda, T. Stalder, S. Steudte, K. Hinkelmann, C. Otte, Introducing a novel method to assess cumulative steroid concentrations: increased



- hair cortisol concentrations over 6 months in medicated patients with depression, *Stress* 15 (2012) 348–353.
- [14] E.A. Shirtcliff, R.L. Buck, M.J. Laughlin, T. Hart, C.R. Cole, P.D. Sloney, Salivary cortisol results obtainable within minutes of sample collection correspond with traditional immunoassays, *Clin. Therapeut.* 37 (2015) 505–514.
- [15] L.M. DeBruine, A.C. Hahn, B.C. Jones, Does the interaction between partnership status and average progesterone level predict women's preferences for facial masculinity? *Horm. Behav.* 107 (2019) 80–82, <https://doi.org/10.1016/j.yhbeh.2018.12.004>.
- [16] U.M. Marcinkowska, G. Kaminski, A.C. Little, G. Jasienska, Average ovarian hormone levels, rather than daily values and their fluctuations, are related to facial preferences among women, *Horm. Behav.* 102 (2018) 114–119, <https://doi.org/10.1016/j.yhbeh.2018.05.013>.
- [17] O.C. Schultheiss, S.J. Stanton, Assessment of salivary hormones, in: E. Harmon-Jones, J.S. Beer (Eds.), *Methods in Social Neuroscience*, Guilford Press, New York, NY, 2009, pp. 17–44.
- [18] S.F. Lipson, P.T. Ellison, Endocrinology Comparison of salivary steroid profiles in naturally occurring conception and non-conception cycles, *Hum. Reprod.* 11 (1996) 2090–2096.
- [19] K.R. Blake, B. Bastian, S.M. O'Dean, T.F. Denson, High estradiol and low progesterone are associated with high assertiveness in women, *Psychoneuroendocrinology* 75 (2017) 91–99, <https://doi.org/10.1016/j.psyneuen.2016.10.008>.
- [20] B.J. Dixon, K.R. Blake, T.F. Denson, A. Gooda-Vossos, S.M. O'Dean, D. Sulikowski, R.C. Brooks, The role of mating context and fecundability in women's preferences for men's facial masculinity and beardness, *Psychoneuroendocrinology* 93 (2018) 90–102, <https://doi.org/10.1016/j.psyneuen.2018.04.007>.
- [21] N.M. Grebe, M.E. Thompson, S.W. Gangestad, Hormonal predictors of women's extra-pair vs. in-pair sexual attraction in natural cycles: implications for extended sexuality, *Horm. Behav.* 78 (2016) 211–219, <https://doi.org/10.1016/j.yhbeh.2015.11.008>.
- [22] J.S. Lobmaier, F. Probst, D.I. Perrett, M. Heinrichs, Menstrual cycle phase affects discrimination of infant cuteness, *Horm. Behav.* 70 (2015) 1–6, <https://doi.org/10.1016/j.yhbeh.2015.02.001>.
- [23] T.N. Shirazi, J.A. Bossio, D.A. Puts, M.L. Chivers, Menstrual cycle phase predicts women's hormonal responses to sexual stimuli, *Horm. Behav.* 103 (2018) 45–53, <https://doi.org/10.1016/j.yhbeh.2018.05.023>.
- [24] T.N. Shirazi, H. Self, K. Dawood, K.A. Rosenfield, L. Penke, J.M. Carré, D.A. Puts, Hormonal predictors of women's sexual motivation, *Evol. Hum. Behav.* 40 (2019) 336–344.
- [25] D.A. Puts, D.H. Bailey, R.A. Cárdenas, R.P. Burriss, L.L. Welling, J.R. Wheatley, K. Dawood, Women's attractiveness changes with estradiol and progesterone across the ovulatory cycle, *Horm. Behav.* 63 (2013) 13–19, <https://doi.org/10.1016/j.yhbeh.2012.11.007>.
- [26] P.T. Ellison, *On Fertile Ground*, Harvard University Press, Cambridge, MA, 2001.
- [27] W. Wang, S.N. Moody, J. Kiesner, A.T. Appiani, O.C. Robertson, E.A. Shirtcliff, Assay validation of hair androgens across the menstrual cycle, *Psychoneuroendocrinology* 101 (2019) 175–181.
- [28] E. Russell, G. Koren, M. Rieder, S. Van Uum, Hair cortisol as a biological marker of chronic stress: current status, future directions and unanswered questions, *Psychoneuroendocrinology* 37 (2012) 589–601.
- [29] T. Stalder, S. Steudte-Schmiedgen, N. Alexander, T. Klucken, A. Vater, S. Wichmann, R. Miller, Stress-related and basic determinants of hair cortisol in humans: a meta-analysis, *Psychoneuroendocrinology* 77 (2017) 261–274.
- [30] S.H. Liening, S.J. Stanton, E.K. Saini, O.C. Schultheiss, Salivary testosterone, cortisol, and progesterone: two-week stability, interhormone correlations, and effects of time of day, menstrual cycle, and oral contraceptive use on steroid hormone levels, *Physiol. Behav.* 99 (2010) 8–16.
- [31] T. Stalder, S. Steudte, R. Miller, N. Skoluda, L. Dettendorfer, C. Kirschbaum, Intraindividual stability of hair cortisol concentrations, *Psychoneuroendocrinology* 37 (2012) 602–610.
- [32] D.C. Cumming, M.E. Quigley, S.C. Yen, Acute suppression of circulating testosterone levels by cortisol in men, *J. Clin. Endocrinol. Metabol.* 57 (1983) 671–673.
- [33] T.J. Dekkers, J.A.A. van Rentergem, B. Meijer, A. Popma, E. Wagemaker, H. M. Huizenga, A meta-analytical evaluation of the dual-hormone hypothesis: does cortisol moderate the relationship between testosterone and status, dominance, risk taking, aggression, and psychopathy? *Neurosci. Biobehav. Rev.* 96 (2019) 250–271.
- [34] O.C. Schultheiss, G. Dlugash, P.H. Mehta, Hormone measurement in social neuroendocrinology: a comparison of immunoassay and mass spectrometry methods, in: O.C. Schultheiss, P.H. Mehta (Eds.), *Routledge International Handbook of Social Neuroendocrinology*, Routledge, New York, NY, 2019, pp. 239–255.
- [35] Y.J. Bae, A. Gaudl, S. Jaeger, S. Stadelmann, A. Hiemisch, W. Kiess, U. Ceglarek, Immunoassay or LC-MS/MS for the measurement of salivary cortisol in children? *Clin. Chem. Lab. Med.* 54 (2016) 811–822.
- [36] D.A. Granger, E.A. Shirtcliff, A. Booth, K.T. Kivlighan, E.B. Schwartz, The "trouble" with salivary testosterone, *Psychoneuroendocrinology* 29 (2004) 1229–1240.
- [37] 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.
- [38] J. Stern, R.C. Arslan, T.M. Gerlach, L. Penke, No robust evidence for cycle shifts in preferences for men's bodies in a multivariate analysis: commentary on Gangestad, Dinh, Grebe, Del Giudice and Emery Thompson (2019), *Evol. Hum. Behav.* 40 (2019) 517–525.
- [39] K.A. Gildersleeve, M.G. Haselton, C.M. Larson, E.G. Pillsworth, Body odor attractiveness as a cue of impending ovulation in women: evidence from a study using hormone-confirmed ovulation, *Horm. Behav.* 61 (2012) 157–166, <https://doi.org/10.1016/j.yhbeh.2011.11.005>.
- [40] S.W. Gangestad, M.G. Haselton, L.L. Welling, K. Gildersleeve, E.G. Pillsworth, R. P. Burriss, D.A. Puts, How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications, *Evol. Hum. Behav.* 37 (2016) 85–96, <https://doi.org/10.1016/j.evolhumbeh.2015.09.001>.
- [41] 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.
- [42] W. Gao, T. Stalder, P. Foley, M. Rauh, H. Deng, C. Kirschbaum, Quantitative analysis of steroid hormones in human hair using a column-switching LC–APCI–MS/MS assay, *J. Chromatogr. B* 928 (2013) 1–8.
- [43] R Core Team, R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, Retrieved from, <http://www.R-project.org/>, 2016.
- [44] W. Revelle, Psych: procedures for personality and psychological research (Version 1.6.12) [Computer software]. Retrieved from, <https://CRAN.R-project.org/package=psych>, 2016.
- [45] H. Wickham, The split-apply-combine strategy for data analysis, *J. Stat. Software* 40 (2011) 1–29, <https://doi.org/10.18637/jss.v040.i01>.
- [46] H. Wickham, reshape2: Flexibly Reshape Data: a Reboot of the Reshape Package, R package version, 2012, p. 1.
- [47] H. Wickham, Ggplot2: Elegant Graphics for Data Analysis, 2009, <https://doi.org/10.1007/978-0-387-98141-3>.
- [48] A. Kassambara, M.A. Kassambara, Package 'ggpubr', 2020.
- [49] B. Diedenhofen, J. Musch, cocor: a comprehensive solution for the statistical comparison of correlations, *PLoS One* 10 (2015), e0121945.
- [50] A. Signorell, Tools for Descriptive Statistics [R Package DescTools Version 0.99.43], Comprehensive R Archive Network (CRAN), 2021.
- [51] R. Kuiper, R. Hoogenboezem, Nopaco: a non-parametric concordance coefficient, version 1.0.6. Retrieved from, <https://CRAN.R-project.org/package=nopaco>, 2019.
- [52] M. Hervé, M.M. Hervé, Package 'RVAideMemoire', See, <https://CRAN.R-project.org/package=RVAideMemoire>, 2020.
- [53] N.C. Silver, J.B. Hittner, K. May, Testing dependent correlations with nonoverlapping variables: a Monte Carlo simulation, *J. Exp. Educ.* 73 (2004) 53–69, <https://doi.org/10.3200/JEXE.71.1.53-70>.
- [54] L.D. Goodwin, N.L. Leech, Understanding correlation: factors that affect the size of r, *J. Exp. Educ.* 74 (2006) 249–266, <https://doi.org/10.3200/JEXE.74.3.249-266>.
- [55] K.L. D'Anna-Hernandez, R.G. Ross, C.L. Natvig, M.L. Laudenslager, Hair cortisol levels as a retrospective marker of hypothalamic–pituitary axis activity throughout pregnancy: comparison to salivary cortisol, *Physiol. Behav.* 104 (2011) 348–353.
- [56] S.J. Short, T. Stalder, K. Marceau, S. Entringer, N.K. Moog, E.A. Shirtcliff, C. Buss, Correspondence between hair cortisol concentrations and 30-day integrated daily salivary and weekly urinary cortisol measures, *Psychoneuroendocrinology* 71 (2016) 12–18.
- [57] V. Moal, E. Mathieu, P. Reynier, Y. Malthiery, Y. Gallois, Low serum testosterone assayed by liquid chromatography–tandem mass spectrometry. Comparison with five immunoassay techniques, *Clin. Chim. Acta* 386 (2007) 12–19.
- [58] P. Rothery, A nonparametric measure of intraclass correlation, *Biometrika* 66 (1979) 629–639.