# Within-person changes of cortisol, dehydroepiandrosterone, testosterone, estradiol, and progesterone in hair across pregnancy, with comparison to a non-pregnant reference group

Kristine Marceau [a,*], Emily Rolan [b], Olivia C. Robertson [a], Wen Wang [c], Elizabeth A. Shirtcliff [c]

[a] Purdue University, USA
[b] Michigan State University, USA
[c] Iowa State University, USA

A B S T R A C T

The literature on hormone changes in pregnancy has focused largely on cortisol, and changes in sample average concentrations. Within-person changes and variability in hormone concentrations are less commonly reported, particularly for sex hormones, and especially measured in hair. Using a prospective sample of pregnant women and a non-pregnant comparison group, we examined changes in five steroid hormones in hair. Non-pregnant women were recruited from the same area with parallel procedures and assessment timeline. Participants include 68 women (34 pregnant, average age = 29.14, and 34 non-pregnant; average age = 27.18) who were predominately non-Hispanic White (83%), and above the 2020 poverty line (75%). Pregnant women provided 3 cm hair samples and completed questionnaires three times during pregnancy: 1) at 12 weeks, 2) at 26 weeks, and 3) at 38 weeks. Non-pregnant women provided 3 cm hair samples and completed questionnaires three times, at baseline, 14 weeks later, and 12 weeks after that to mirror the assessment schedule of the pregnant group. There was clear evidence that progesterone was higher initially and increased dramatically across pregnancy whereas non-pregnant patterns showed no systematic change. There was suggestive evidence that cortisol and estradiol increased over pregnancy and in non-pregnant women similarly across the same time course. There was suggestive evidence that DHEA decreased across pregnancy, particularly early in pregnancy, differently from patterns in non-pregnant women over the same time course. Most importantly, there was substantial variability of hormone concentrations and many different within-person patterns of changes in these hormones over time, with little evidence of systematic change or stability within-individuals. Moving beyond discussing sample averages to including within-person and non-linear changes in studies of hormones-behavior associations during pregnancy is an important future direction for further investigation.

## 1. Introduction

Pregnancy is an important period of time for understanding the dynamics of changing hormone concentrations. In order to understand the nature of hormone changes across pregnancy, it is critical to ascertain hormone concentrations in a way that a) reliably maps onto the time-scale of interest, and b) describes the extent to which there are or are not individual differences in individuals' trajectories of change. Hair hormone concentrations are better suited to the time-scale of pregnancy than other common biospecimens like saliva or serum because they are not subject to short-term fluctuations (e.g., minutes, hours, days). Although there is a growing body of literature examining average

concentrations of cortisol in hair across trimesters of pregnancy (see Ref. [1], how sex hormone concentrations in hair change across pregnancy has not been adequately documented. Therefore, the present study examines hormone changes of five steroid hormones in hair: cortisol, dehydroepiandrosterone (DHEA), testosterone, estradiol, and progesterone, using a prospectively measured sample of pregnant women and a comparison sample of non-pregnant women recruited from the same area with matching procedures and assessed on the same timeline. To our knowledge, this is one of the first studies in humans that tracks these hormones prospectively across pregnancy (see Ref. [2,3] for work with rhesus macaques [4]; and [5] for humans), and is amongst a very small number of studies that examine these hormones longitudinally in human

---

hair (see, for example, [6,7].

Drawing from other biospecimens [8], synthesized reference ranges for many laboratory values including steroid hormones across normal pregnancies into a succinct and accessible reference table. The reported average concentrations of cortisol, testosterone, estradiol and progesterone slowly increased across pregnancy from trimester 1 to trimester 3. Further, the concentrations and ranges (i.e. spread) decreased somewhat for dehydroepiandrosterone-sulphate (DHEA-S). In all cases, the ranges were highly overlapping across trimesters, and were typically overlapping with the non-pregnant reference groups, with the exception of progesterone and estradiol. After trimester 2, progesterone and estradiol were higher in pregnant women than in non-pregnant women. Although this type of data aggregation effort is commendable, there are also limitations, including that the reported values were taken from multiple studies analyzed in different laboratories, and laboratory values can vary widely, presumably due to assay procedure and technology differences [9]. Further, many key factors such as biospecimen and timing of assessment are unavailable in these data, and the non-pregnant comparisons are from potentially quite different populations.

Several studies have examined hair cortisol concentrations at multiple times over the course of pregnancy, often showing increases in hair cortisol on average across pregnancy (e.g., Refs. [10,11]). However, the reported means and standard deviations suggest that there is a greater degree of variation within any given trimester than there are increases across trimesters, and some studies did not find this average cortisol increase (e.g., Ref. [11]; see also [1]). Overall, sample average patterns of change have been reported for sex hormones across pregnancy as well. In serum, testosterone and progesterone concentrations increased over pregnancy, estradiol concentrations increased particularly in the first trimester, and DHEA-S decreased over the three trimesters [12]. Cross-sectionally, in serum, DHEA also decreased across pregnancy [13]. Testosterone in saliva was also found to increase on average across pregnancy, but the reported ranges were highly variable [14]. Estradiol and progesterone has also been shown to increase in saliva across pregnancy, again with the pattern of widening variability across pregnancy [15]. This widening variability underscores the importance of considering individual trajectories to better understand hormone changes across pregnancy. Critically, sample average trajectories do not necessarily reflect the patterns of change that would be observed within individuals in the sample – that is, an increase across trimesters in terms of the sample average concentration does not mean that every individual experiences, or even most individuals experience an increase over pregnancy. Higher variability in concentrations within trimesters than across trimesters in these studies indicates that many individual patterns of change are likely, and focusing solely on the sample average can miss critical individual differences in hormone trajectories that may prove important for behavior or development. A novel contribution of this study is the characterization of multiple patterns of within-person changes in multiple hormones across pregnancy and in a non-pregnant comparison group.

Moreover, by examining a biospecimen that is relatively robust against momentary fluctuations, we are in a position to estimate the true magnitude of within-individual change across three-month durations of time. Specifically, a substantive limitation of the work assessing sex hormones across pregnancy is that hormone concentrations were assayed in serum or saliva, and these biospecimens show substantial variation due to specific environments and cyclic variation on several timescales. Thus, any single assessment is unlikely to provide a reliable snapshot of hormone concentrations at that time, and is not ideal for tracking longer-term changes [16]. Hair hormone concentrations are better measures of basal levels than single-assessments of saliva or serum, as they are not subject to short-term fluctuations. While hair has emerged as a common biospecimen for cortisol, less work has investigated hair androgens such as testosterone and DHEA [17,18]. Further, estradiol and progesterone remain critically understudied [6,19,20], particularly over time. Describing concentrations and changes in a range of steroid hormones in

hair is an important contribution of this study which few prior studies have accomplished in humans [4,5].

The purpose of the present study was 1) to present within-person patterns of change in five hormones in hair over the course of pregnancy, and 2) to compare the concentration, pattern and magnitude of within-person changes between a sample of pregnant and non-pregnant women. Strengths of the design include repeated prospective assessments, one per trimester, in a sample of pregnant women and in a sample of non-pregnant women assessed by the same team on the same assessment schedule. Thus, we are able not only to compare concentrations across pregnant and non-pregnant participants, but also to compare within-person changes across groups.

## 2. Materials and methods

### 2.1. Participants and procedures

Participants include 68 women (34 pregnant, average age = 29.14, SD = 5.06 years), and 34 non-pregnant; average age = 27.18, SD = 3.87 years). The sample was predominately (83%) non-Hispanic White, and (75%) above the poverty line as defined by the 2020 poverty guidelines [21]; see Table 1 for full demographics. To be eligible for enrollment in the study, pregnant women had to be less than 12 weeks pregnant (self-reported) and live within about a 1-h driving radius from Purdue University. Pregnant women were recruited from July 2017 through October 2018. Non-pregnant women were recruited from June 2018 through April 2019, from the same catchment areas and age ranges using the same strategies. Exclusion criteria for both samples included (1) failing to understand the elements of informed consent, (2) failing to understand English at an 8th grade level, (3) being a minor (under the age of 18 years).

Pregnant women were followed longitudinally, and assessed three times during pregnancy (see Supplemental Fig. S1): 1) at 12 weeks (1st trimester, n = 34, M = 12.47, SD = 1.21), 2) at 26 weeks (during 2nd trimester, n = 33, M = 26.16, SD = 1.41), and 3) at 38 weeks (during the 3rd trimester, n = 31, M = 37.62, SD = 1.17). Non-pregnant women were assessed three times, at baseline (n = 34), 14 weeks later (n = 31, $M_{T1-T2}$ = 14.17, $SD_{T1-T2}$ = 0.52), and 12 weeks later (n = 29, $M_{T2-T3}$ = 12.23, $SD_{T2-T3}$ = 0.54) to mirror the assessment schedule of the pregnant group (aside from the six-month postpartum follow-up, not presented in this manuscript; see Supplementary Table S1 for details of assessment timelines). Each assessment consisted of hair collection and all assessments included an online questionnaire (approximately 45min for pregnant women and 30min for non-pregnant women). A trained research assistant collected hair samples at each visit, either in the participants' home or in the lab. Questionnaires included demographic information described below. In addition, although not analyzed here due to the scope of this analysis, we also assessed mental health history (assessment 1 only), pregnancy experiences (pregnant sample only), psychological stress and distress, and life events; questionnaires were completed independently by participants online at the time and location of their choosing, within a week of the hair collection. Participants were compensated with cash ($25 per visit for pregnant women, $20 for non-pregnant women). The study was approved by the Purdue University IRB (#1704019124), and all participants provided informed consent. Please see Appendix 1 Part A "Recruitment, Attrition, and Missing Data" for further details.

### 2.2. Hair hormones

#### 2.2.1. Hair collection

Hair samples were collected by a trained research assistant at the end of each trimester or corresponding time in the non-pregnant sample. The majority of the non-pregnant sample chose for hair collection to occur in the lab (79–84% across assessments), whereas the majority of the pregnant sample had hair collection occur in the home (58–64% across

**Table 1**
Sample demographic statistics.

| | Pregnant Sample | | Non-Pregnant Sample | |
|---|---|---|---|---|
| | Mean(SD) | Min-Max | Mean(SD) | Min-Max |
| **Household Income** | $65,000 ($48,000) | 0-$230,000 | $55,000 ($49,000) | 0-$150,000 |
| **Age at first visit** | 29.14 (5.06) | 19.55–39.74 | 27.18 (3.87) | 18.13–41.78 |
| **Number of Children** | 1.14 (1.22) | 0–4 | 0.72 (1.67) | 0–3 |
| **Parity** | | | | |
| Number of Births | 1.12(1.23) | 0–4 | 0.71 (1.19) | 0–5 |
| Previous Miscarriages | 0.82 (1.64)[c] | 0–9 | 0.31 (0.59) | 0–2 |
| Never Pregnant | – | – | 0.59 (0.50) | |
| **Financial Need** | 2.24 (2.03 | 0–7 | 2.38 (2.67) | 0–11 |
| **Financial Deprivation** | 2.00 (0.92) | 1–4 | 2.21 (0.80) | 1–4.67 |
| **Race**[a] | **N (%)** | | **N (%)** | |
| White | 25 (83.3) | | 24 (77.4) | |
| Black or African American | 3 (10.0) | | 1 (3.2) | |
| Asian | 1 (3.3) | | 3 (10.0) | |
| Latinx or Hispanic | 1 (3.3) | | 1 (3.2) | |
| More than one/ Other | 0 (0.0) | | 2 (6.4) | |
| **Education** | | | | |
| Less than high school degree | 1 (2.9) | | 0 (0.0) | |
| High School degree/GED | 11 (32.4) | | 13 (38.2) | |
| 2-year college degree | 3 (8.8) | | 1 (2.9) | |
| 4-year college or university degree | 11 (32.4) | | 15 (44.1) | |
| Graduate Degree | 8 (23.5) | | 5 (14.7) | |
| **Employment Status** | | | | |
| Unemployed/ Student | 12 (35.29) | | 20 (60.61) | |
| Part Time | 5 (14.71) | | 5 (15.15) | |
| Full Time | 17 (50.00) | | 8 (24.24) | |
| **Marital Status** | | | | |
| Single, never married | 6 (17.6) | | 17 (50.0) | |
| Married/ Committed Living Together | 28 (81.4) | | 14 (41.2) | |
| Separated/ Divorced | 0 (0.0) | | 3 (8.8) | |
| **Contraception**[b] | | | | |
| Hormonal Contraception | – | – | 19 (57.6) | |
| None | – | – | 13 (39.4) | |

*all contraception was hormonally based, and most was continually taken, however, one non-pregnant participant took a plan B pill between the second and third visits. One non-pregnant participant was breastfeeding throughout the study.
[a] These are out of 31 because of missing data for race/ethnicity.
[b] Out of 33 because one participant missing on contraception.
[c] For the pregnant sample, 10 women experienced 1, 4 experienced 2–3, and one woman experienced 9 miscarriages (44% experienced at least one). For the non-pregnant sample, only 25% experienced at least one miscarriage (6 women experienced 1, two women experienced 2).

assessments), although this is not expected to affect hormone concentration results based on prior studies [22]. Participants' hair was parted horizontally along the posterior vertex (center, back of the head). Hair bands and clips were used to hold hair away from the part and then the end of a comb or scissor tips were used to isolate the hair sample. Hair was cut as close to the scalp as possible without grazing the scalp and

then wrapped in a foil sheet, sealed in an envelope and stored in a cool and dry filing cabinet until ready to be mailed to the Iowa State Stress Physiology Investigative Team (SPIT) laboratory for assay. Availability of data on specific hormones varied based on the amount of hair sampled. Assays were prioritized in the following order: cortisol, DHEA/testosterone, estradiol, and progesterone; insufficient hair quantity led to missingness. See Appendix 1 Part A "Other Missing Data" for details.

*2.2.2. Assay*

Each of the three hair samples was segmented by 3 cm from the scalp end, each segment to reflect long-term hormone concentrations within approximately one trimester based on the common rule of thumb of 1 cm of growth per month [23]. Thus, with three prospective assessments, we capture hair hormone concentrations across the entire pregnancy. During hair collection, we also measured the hair growth for a small subset of individuals. Research assistants used a piece of paper touching the scalp to mark the length of previously cut hair and later measured the length from the edge of the paper to the mark using a ruler. Hair growth was measured in this way for six pregnant women from trimester 1 to 2 (M = 3.93 cm, SD = 0.43, range = 3.5 cm to 4.7 cm), and seven pregnant women from trimester 2 to 3 (M = 3.49 cm, SD = 0.65 cm, range = 2.7 cm to 4.3 cm). Further, we measured the hair growth for 11 non-pregnant women between assessments (M = 4.32 cm, SD = 0.32, range = 3.8 cm to 4.8 cm). Considering the length of time between assessments (calculated by subtracting the assessment dates, measured in weeks), we estimated the rate of growth as cm/month as (cm growth)/(assessment 2 – assessment 1)/(4.34524 weeks per month). Our estimated rates per growth in cm/month were 1.30 (SD = 0.23, range = 1.08–1.62) for pregnant women between trimester 1 and 2, 1.23 (SD = 0.13, range = 1.08–1.46) for pregnant women between trimester 2 and 3, and 1.31 (SD = 0.10, range = 1.11–1.46) for non-pregnant women between assessments 1 and 2. These rates of growth on a very small subsample are consistent with the 1 cm/month rule of thumb albeit with potentially slightly faster growth indicating slightly larger catchment windows.

See Appendix 1 Part B "Additional Assay Procedural Details" for specifics on washing, drying, and extraction steps. Segmenting and assays were performed by the SPIT lab at Iowa State University, in Ames, IA. In brief, segmented samples were washed in isopropanol with rotation and then dried. Three 15 mg samples (used to assay cortisol, DHEA and testosterone together, estradiol) and one 10 mg sample (used to assay progesterone) were weighed and ground in a ball mill. Powdered hair was extracted in methanol (except for estradiol which was in ethyl acetate, see Appendix 1 Part B), centrifuged, dried, and then reconstituted with kit-provided assay diluent and assayed immediately using commercially available enzyme immune-assay kits (Salimetrics, PA). For estradiol a double-extraction method was applied, and for progesterone, a distinct agitation method was conducted for extraction (see Appendix 1 Part B for details).

Our experimental design is aimed at tracking within-person variability and changes. Thus, all three assessments for any given individual were assayed on the same tray in order to reduce intra-individual variability due to assay batch-related error. There were only two instances where individuals had two assessments assayed on one tray and the third on another. For all the assays, a tray was considered reliable if the standard curve had $R^2 > 0.996$. All samples were tested in duplicate and samples were re-assayed if duplicate test values that varied by more than 7–10% error (CV: 7% for cortisol, DHEA, and progesterone, 10% for testosterone and estradiol). There were 4 samples reanalyzed for testosterone, 14 samples reanalyzed for estradiol, and 27 samples reanalyzed for progesterone (most due to being above sensitivity rather than having an intra-assay CV > 10%).

*2.3. Covariates*

*2.3.1. Socio-economic status*

Socio-economic status (SES) was conceptualized as a formative latent

variable, operationalized using principal component analysis in SAS (PROC FACTOR, extracting exactly 1 factor). Indicators included the following: *Household income* was continuously measured in pre-tax Dollars. *Employment status* at the first visit was scored as 0 = not employed, 1 = employed part time, 2 = employed full time. The highest level of *education* was coded on a scale of 1 (less than high school) to 7 (graduate program). These three indicators have been identified as the "big three" important for measuring SES, however, subjective measures of financial hardship have also been recommended [24] and were therefore included here. Participants reported yes (1) or no (0) on a series of 12 questions with the prompt "During the past 12 months, have you made any of the following adjustments because of financial need?" (e.g., taken an extra job to help meet basic living expenses; sold possessions; cashed out a retirement account), which were summed into a *financial need* measure (Chronbach's alpha = .79; e.g., Ref. [25]). Participants reported on a scale from 1 (Strongly agree) to 5 (strongly disagree) for 6 items such as "I have enough money to afford … the place where I want to live", which was averaged into a *financial deprivation* score (Chronbach's alpha = .91; e.g., Ref. [25]). There were no differences between the pregnant and non-pregnant sample on any indicator (data available upon author request); thus, to improve power and precision, pregnant and non-pregnant participants were analyzed together to create the SES variable. The SES factor (Eigenvalue = 2.29) explained 45% of the variance in these measures and indexed higher SES (factor loadings: income = .79; employment = .22; education = .68; financial need = -.67; financial deprivation = -.87).

### 2.3.2. Additional covariates

Additional covariates included pre-pregnancy BMI (or current, for the non-pregnant sample) assessed at the first visit, calculated from self-reported height and weight. Self-reported race/ethnicity was a nominal variable including non-Hispanic White, Latina, Black or African American, Asian, and Other. Season of assessment was a time-varying covariate such that an assessment date occuring in December-February was coded as "Winter", March–May was "Spring", June–August was "Summer" and September–November was "Fall". Age was also a time-varying covariate, calculated from visit dates and reported (mother/participant) birth dates.

For the non-pregnant sample, hormonal birth control use was also considered. For the pregnant sample, gestational age at the visit was also considered as an alternate indicator of time than assessment. Assay batch was controlled during the data preparation steps (see below).

### 2.4. Analytic strategy

### 2.4.1. Data cleaning

Data were examined for outliers, and windsorized at +3 standard deviations of the grand mean concentrations (see Appendix 1 Part C "Additional Data Cleaning Details" for specifics, including decisions for handling outliers and cut-off values). Next, we removed batch effects by regressing a nominal variable indicating the tray on which each hormone was assayed on the hormone concentration and saving the residuals. Because in the vast majority of cases all assessments per individual were assayed on the same tray, these batch effects would contribute to between-person differences rather than within-person variability.[1] Batch effects were present ($R^2$ ranged from .04 to .34 across hormones and pregnant and non-pregnant samples, see Supplemental Table S2 for results, and Appendix 1 Part C "Controlling for Batch Effects" for additional discussion). For all analyses, batch corrected data were used except when noted explicitly.

---

[1] Except progesterone, because redo trays and one initial tray included both pregnant and non-pregnant samples.

### 2.4.2. Aim 1: description of within-person change over time for pregnant and non-pregnant samples

We first examined the distributions and plots of individuals' trajectories over time. Next, we created 'pattern' variables to quantify the number of individuals showing each trajectory of change over time by creating differences scores from trimester/assessment 1 to trimester/assessment 2 and trimester/ assessment 2 to trimester/assessment 3 (e.g., increasing, increasing; or stable, decreasing; 9 total possible categories). This allowed us to visually inspect how many different patterns of change from trimester/assessment 1 to trimester/assessment 2 and trimester/assessment 2 to trimester/assessment 3 were present in the data, as well as visually inspect which patterns of within-person change are most prevalent (e.g., describe inter-individual differences in intra-individual change). Stable patterns were identified by samples that were within the CV range for that hormone to incorporate expected assay error, rather than only allowing stability to reflect exactly the same values across two assessments (see Appendix 1 Part C "Quantifying Stability" for further explanation, logic, and decision-making).

Next, we conducted unconditional multi-level models separately by group (pregnant vs. not) to quantify the amount of between-person versus within-person variability in hormone concentrations over time (intra-class correlations, ICC's). Then, we conducted multilevel models of change separately by group with a random intercept and slope in order to test for linear change over time. Specifically, we entered assessment coded 0 = trimester/assessment 1, 1 = trimester/assessment 2, 2 = trimester/assessment 3 as a level 1 predictor of the repeated measures of cortisol, DHEA, testosterone, estradiol, and progesterone – each hormone in a separate model. We included a random intercept and slope. Percent reduction in error is presented to quantify the variance explained by the linear slope as a metric of effect size.

### 2.4.3. Aim 2: differences in concentrations and within-person changes in pregnant vs. non-pregnant samples

Because batch effects are confounded with pregnant vs. non-pregnant status, a four-step approach was taken. The first three steps were a series of multilevel models of change on the entire sample, with assessment coded as described above, adding the indicator variable identifying participants as pregnant or non-pregnant (a level 2, person-level predictor) as well as the cross-level interaction of assessment and group (pregnant vs. not) in order to formally test for group differences in slopes. First, multi-level models of change were run on the raw, windsorized data, completely ignoring batch effects. Second, multi-level models of change were run on the raw, windsorized data including both the pregnant vs. non-pregnant status indicator and the nominal batch variable (included as a level 1 predictor). This controlled for batch within pregnant and non-pregnant groups, as PROC MIXED solves the equation in such a way that the variance accounted for by pregnancy was estimated first and batch effect second (with functionally two reference groups-one per sub-sample). Third, multi-level models of change were run on the batch corrected data. Finally, as a more descriptive exercise, change scores (difference scores) were subjected to t-tests separately for pregnant and non-pregnant groups (to judge difference from zero) and in independent samples t-tests (to judge differences across groups). To mitigate the limitations in our experimental design, differences in pregnant vs. non-pregnant samples were interpreted based on converging and diverging evidence from these approaches.

### 2.4.4. Adjusted models

We conducted bivariate correlations of covariates (SES, pre-pregnancy/visit 1 BMI, race/ethnicity, season, and age; hormonal contraceptive use for non-pregnant women only, gestational age at visit for pregnant women only) with levels at each assessment and changes (difference scores) of each hormone. Any that showed evidence of nominal correlation (i.e., $p < .05$) were included in adjusted models in sensitivity analyses for the multilevel models described for aims 1 and 2 (deviation from main study results are noted; full results available upon author request).

## 3. Results

### 3.1. Description of change over time

Raw, windsorized hormone concentrations are presented in Table 2. Plots of the batch corrected data revealed substantial heterogeneity in trajectories of change in all hormones (i.e., diverse patterns of change; see Fig. 1). Pattern variables confirmed heterogeneity in the patterns of change in all hormones, with very little stability within individuals over time (see Figs. 2 and 3; Supplemental Table S3). The exception is for progesterone, which showed relatively more stability in the non-pregnant sample and homogenous increases in the pregnant sample. Fisher's exact tests show no differences (e.g., all $p > .05$) in proportions in each pattern category for pregnant vs non-pregnant groups for cortisol, DHEA, testosterone, or estradiol (although we are underpowered to detect true but small differences; results available upon author request). However, there were group (pregnant vs. not) differences in progesterone both overall and in the piecewise pattern categories, Table Probability (P)'s $< .0007$, $p$'s $< .05$.

### 3.1.1. Patterns of change

Patterns of change from trimester (or assessment) 1 to 3 (i.e., skipping trimester/assessment 2) are presented in Fig. 2, and are referred to in text as "overall" change. Patterns of change from trimester/assessment 1 to 2 and trimester/assessment 2 to 3 are presented in Fig. 3.

For cortisol, most (pregnant: 71%, non-pregnant: 64%) participants showed an overall cortisol increase from trimester/assessment 1 to 3, and the rest showed an overall decrease (pregnant: 29%, non-pregnant: 36%); there were no stable patterns. Considering all three trimesters/assessments, seven of the nine possible patterns of change were observed for pregnant and non-pregnant women. The most frequent change patterns among pregnant women and non-pregnant women was a decrease from trimester/assessment 1 to 2 and then subsequent increase from trimester/assessment 2 to 3 (pregnant: 29%, non-pregnant: 32%), an increase from trimester/assessment 1 to 2 and 2 to 3 (pregnant: 25%, non-pregnant: 21%), or an increase from trimester/assessment 1 to 2 and subsequent decrease from 2 to 3 (pregnant: 25%, non-pregnant: 21%).

For DHEA, there was a relatively even split between participants that showed overall increases (pregnant: 45%, non-pregnant: 52%) vs. overall decreases (pregnant: 50%, non-pregnant: 44%) from trimester 1 to 3, with one pregnant woman and one non-pregnant woman showing a pattern of stability. Considering all three trimesters/assessments, six of the nine possible patterns of change were observed for pregnant and non-pregnant women. The most frequent change patterns among pregnant women and non-pregnant women was a decrease from trimester/assessment 1 to 2 and then subsequent increase from trimester/assessment 2 to 3 (pregnant: 32%, non-pregnant: 30%), and a constant increase from trimester/assessment 1 to 2 and 2 to 3 (pregnant: 21%, non-pregnant: 22%).

For testosterone, some participants showed overall increases (pregnant: 67%, non-pregnant: 52%) and others showed decreases (pregnant: 24%, non-pregnant: 41%) from trimester/assessment 1 to 3, with two pregnant and two non-pregnant women showing a pattern of stability. Considering all three trimesters/assessments, five of the nine possible patterns of change were observed for pregnant women and six for non-pregnant women. The most frequent change patterns were a decrease from trimester 1 to 2 and then increase from 2 to 3 for pregnant women (39%), followed by a constant increase (e.g., from trimester 1 to 2 and 2 to 3; 33%). A constant increase was the most frequent change pattern for non-pregnant women (33%) followed by a decrease from assessment 1 to 2 and then subsequent increase from assessment 2 to 3 or increase from assessment 1 to 3 and then subsequent increase from assessment 2 to 3 (22% each).

For estradiol, most participants showed overall increases (pregnant: 88%, non-pregnant: 61%) from trimester/assessment 1 to 3. One pregnant woman showed an overall decrease and one a pattern of stability, whereas 32% of non-pregnant women showed an overall decrease and two (7%) showed a pattern of stability. Considering all three trimesters/assessments, only three of the nine possible patterns of change were observed for pregnant women, whereas six were observed for non-pregnant women. For non-pregnant women, the most common pattern of change was a decrease from assessment 1 to 2 and then subsequent increase from trimester 2 to 3 (32%), followed by a constant increase or increase from assessment 1 to 2 and subsequent decrease from 2 to 3 (25% each). For pregnant women, the most common change pattern for estradiol was a constant increase across trimesters 1 to 2 and 2 to 3 (73%).

For progesterone, there were large group differences. More of the non-pregnant sample decreased (54%) than increased (36%) overall, with some remaining stable (10%). All of the pregnant sample (100%) showed overall increases in progesterone. Considering all three trimesters/assessments, only three of the nine possible patterns of change were observed for pregnant women, whereas eight were observed for non-pregnant women. For non-pregnant women, there was substantial heterogeneity in patterns of change, with the most prevalent pattern (29%) being a decrease from assessment 1 to 2 then increase from

**Table 2**
Hormone concentrations.

| Cortisol (pg/mg) | Pregnant Sample | | | | | Non-Pregnant Sample | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | (Std. Dev.) | Min | Max | N | Mean | (Std. Dev.) | Min | Max |
| Trimester 1 | 33 | 9.31 | (9.35) | 0.60 | 39.30 | 34 | 8.78 | (5.17) | 3.90 | 34.05 |
| Trimester 2 | 32 | 8.74 | (11.70) | 1.05 | 68.11 | 31 | 10.66 | (9.81) | 4.05 | 47.55 |
| Trimester 3 | 29 | 12.80 | (15.07) | 2.55 | 68.11 | 28 | 10.24 | (6.30) | 4.80 | 37.05 |
| **DHEA (pg/mg)** | | | | | | | | | | |
| Trimester 1 | 30 | 14.10 | (9.56) | 2.65 | 32.71 | 33 | 22.75 | (8.62) | 7.13 | 39.16 |
| Trimester 2 | 26 | 10.86 | (6.45) | 1.47 | 23.87 | 31 | 24.85 | (13.90) | 7.51 | 66.85 |
| Trimester 3 | 24 | 11.53 | (4.74) | 3.51 | 21.71 | 28 | 26.70 | (14.61) | 11.36 | 66.85 |
| **Testosterone (pg/mg)** | | | | | | | | | | |
| Trimester 1 | 29 | 1.88 | (1.06) | 0.67 | 4.80 | 34 | 1.79 | (0.66) | 0.82 | 3.28 |
| Trimester 2 | 25 | 1.96 | (1.00) | 0.21 | 4.80 | 31 | 1.78 | (0.58) | 1.04 | 3.25 |
| Trimester 3 | 24 | 2.08 | (0.89) | 0.81 | 4.80 | 27 | 1.82 | (0.58) | 1.10 | 3.53 |
| **Estradiol (pg/g)** | | | | | | | | | | |
| Trimester 1 | 26 | 35.12 | (18.05) | 9.23 | 80.23 | 34 | 44.60 | (17.75) | 25.52 | 100.91 |
| Trimester 2 | 19 | 43.29 | (16.20) | 8.93 | 81.72 | 31 | 46.23 | (14.42) | 25.88 | 86.77 |
| Trimester 3 | 17 | 52.67 | (21.39) | 16.05 | 100.91 | 28 | 51.12 | (14.52) | 31.90 | 80.02 |
| **Progesterone (pg/mg)** | | | | | | | | | | |
| Trimester 1 | 12 | 44.79 | (18.70) | 19.35 | 74.87 | 31 | 6.81 | (5.65) | 0.23 | 18.12 |
| Trimester 2 | 13 | 102.82 | (49.82) | 29.96 | 217.80 | 31 | 5.95 | (4.49) | 0.21 | 17.54 |
| Trimester 3 | 10 | 189.48 | (39.83) | 121.18 | 228.98 | 28 | 5.25 | (4.29) | 0.44 | 17.40 |

Note. Raw, windsorized values are presented (prior to removing batch effects).
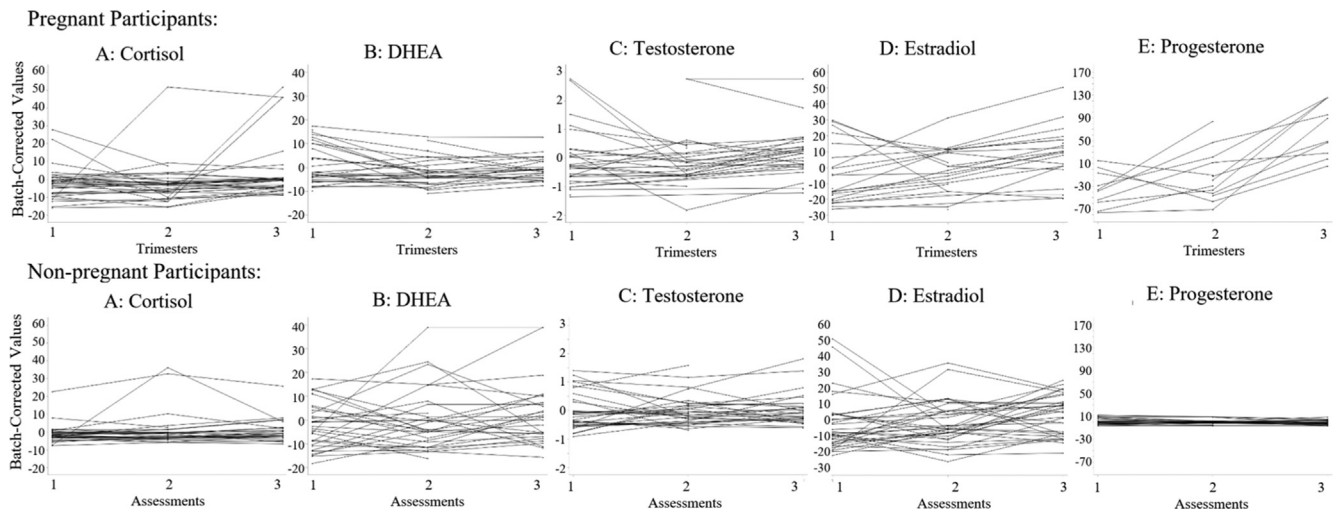
Pregnant Participants:



Non-pregnant Participants:



**Fig. 1.** Plots of the batch corrected dataNote

The batch-corrected concentrations unadjusted for other covariates are plotted over time on the same scale for pregnant and non-pregnant participants for each hormone in order to compare levels across groups. Values are in pg/mg for cortisol, DHEA, testosterone, and progesterone (panels A–C and E), and pg/g for estradiol (panel D), *relative to the assay batch average* (i.e., they can be interpreted as within-batch centered values in the original unit). See Supplemental Fig. S2 for a "zoomed in" plot of progesterone in the non-pregnant sample that offers a clearer illustration of stability and change within that group.
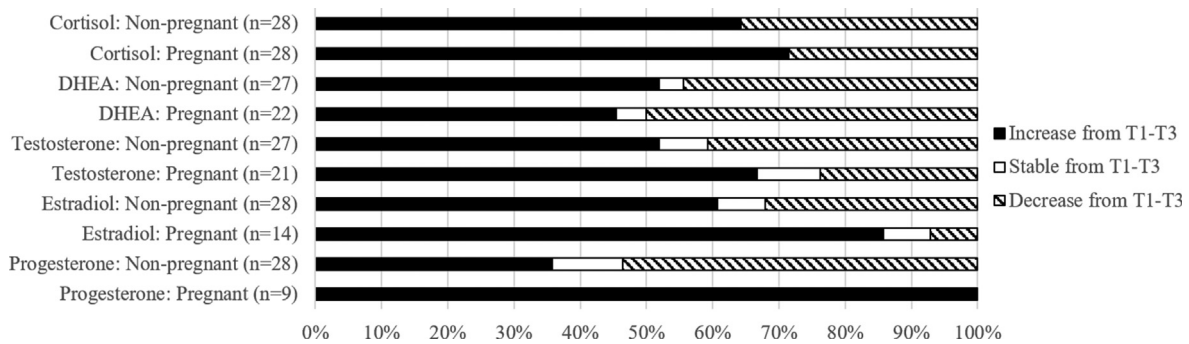


**Fig. 2.** Within-person Patterns of change from T1 to T3.
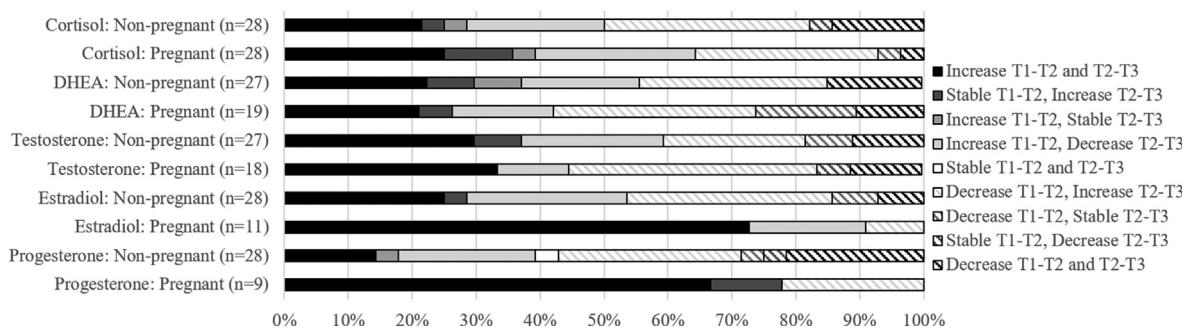


**Fig. 3.** Within-person Patterns of change from T1 to T2 and T2 to T3.

assessment 2 to 3. However, for the pregnant sample, most women showed a consistent increasing pattern from trimester 1 to 2 and trimester 2 to 3 (67%).

*3.1.2. Multi-level models of change*

Intra-class correlations are presented in Supplemental Table S4. ICC's indicated that each hormone showed substantial within-person variability over time in both samples, as the within-person variability accounted for more than half (>50%) of the variance in each hormone within each group (pregnant and non-pregnant), except for cortisol in the non-pregnant sample cortisol (44% of the variance was within-person)

and testosterone in the pregnant sample (28% was within-person). Table 3 presents multi-level models of change for each hormone in each group. For cortisol, there was no systematic change in the pregnant sample, whereas for the non-pregnant sample, there was a positive linear slope. However, there was significant variance in the linear slope in the pregnant sample. Adding a linear slope explained 23% of the residual variance for the pregnant sample (but not for the non-pregnant sample, PRE = .01). For DHEA and testosterone, there was no systematic change in the pregnant or non-pregnant sample. However, there was significant variance in the linear slope in the non-pregnant sample; the linear slope term explained 26% of the variance in DHEA and 30% of the variance in

testosterone (but not in the pregnant sample, percent reduction in error [PRE] = 2% and 0%, respectively). For estradiol and progesterone, there was an increase in the pregnant but not non-pregnant sample. Additionally, there was significant variance in the linear slope in the pregnant sample, which explained 41% of the variance in estradiol and 59% of the variance in progesterone (but not in the non-pregnant sample, PRE = 1% and 6%, respectively).

### 3.1.3. Covariates

Associations with covariates were sparse. Out of 30 correlations (i.e., (three change scores + three samples) x five hormones), SES was negatively related to only trimester/assessment 1 cortisol, $r = -.31$, $p = .014$, and so was included in adjusted models. Out of 30 correlations, trimester/assessment 1 BMI was positively related to only trimester/assessment 1 cortisol, $r = .32$, $p = .008$, and so was included in adjusted models. Out of 90 correlations, age at trimester/assessment 3 was related to a larger increase in estrogen from trimester/assessment 2 to trimester/assessment 3, $r = .34$, $p = .037$, and age at trimester/assessment 1 was negatively related to trimester/assessment 1 DHEA, $r = -.29$, $p = .023$. Because within-person centered age was correlated with assessment at $r = .93$, each person's average age (indexing between-person differences in age only) was included in adjusted models, as a level 2 covariate. We began recruitment and assessment of the pregnant sample first, and later completed the non-pregnant sample; thus the pregnant sample was more likely to have trimester/assessment 1 assessments in summer/Fall (29/34), and the non-pregnant sample was more likely to have trimester/assessment 1 assessments in Winter (25/34). Controlling for assessment (coded as 0, 1, 2), in a repeated measures ANOVA there were no seasonal differences hormones except that DHEA was lower in non-pregnant women in the Fall and Winter than Spring and Summer. Nonetheless, season was retained as a covariate in sensitivity analyses.

We did not include race/ethnicity in adjusted models because there were no race/ethnicity differences in change scores or hormone levels.

There were no differences in levels or change scores for non-pregnant women who used hormonal contraceptives vs. did not. Gestational age at each visit was correlated with assessment at $r = .997$, and was therefore not included as a covariate. These group-specific covariates were not included in adjusted models.

Because of the sparseness of covariate effects and because missing data on covariates reduces the sample size using SAS PROC MIXED, the unadjusted models are preferred and presented. Effects reported in Table 3 for cortisol, testosterone, estradiol and progesterone were consistent with adjustment for covariates with one exception: There was a linear increase in DHEA for non-pregnant women in the adjusted model that was not found in the unadjusted analysis. Findings for covariates (available upon author request) included: For pregnant women, lower SES was associated with higher cortisol. For non-pregnant women, older age, higher BMI, and season (assessments in spring and summer) were associated with lower DHEA. For pregnant women, cortisol was relatively lower in summer, estradiol was relatively higher in summer and fall, and progesterone was relatively lower in spring but higher in summer. For non-pregnant women, progesterone was relatively lower in fall. For pregnant women, older age was associated with higher testosterone levels. For non-pregnant women, older age was associated with lower testosterone levels.

### 3.2. Differences in pregnant vs. non-pregnant samples

Across the multilevel models of change, cortisol showed no evidence of differences in intercept levels, $\beta$'s $-0.55$ to $-3.27$, S.E.'s 1.56–2.64, or change over time across groups, $\beta$'s 1.07–1.10, S.E.'s 1.22–1.25 (Supplementary Table S5). Change scores suggested that cortisol increased from trimester/assessment 1 to 3 more in pregnant than non-pregnant women ($M_{pregnant} = 7.42$, SD = 16.49; $M_{non-pregnant} = 1.19$, SD = 4.24, $p_{difference} < .05$). This finding was driven by change from trimester/assessment 2 to 3 ($M_{pregnant} = 7.25$, SD = 17.42; $M_{non-pregnant} = -.53$, SD

**Table 3**
Multi-level models of change.

| Pregnant Sample | Cortisol | | | DHEA | | | Testosterone | | | Estradiol | | | Progesterone | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Est | | S.E. | Est | | S.E. | Est | | S.E. | Est | | S.E. | Est | | S.E. |
| *Fixed effects* | | | | | | | | | | | | | | | |
| Intercept | −1.78 | | 1.26 | 1.13 | | 1.41 | −0.02 | | 0.2 | −5.37 | + | 3.09 | −54.29 | *** | 10.26 |
| Linear slope | 1.8 | | 1.17 | −1.16 | | 0.8 | 0.05 | | 0.09 | 5.71 | * | 2.26 | 57.84 | *** | 9.04 |
| *Random effects: Individual-level* | | | | | | | | | | | | | | | |
| Variance intercept | 20.54 | | 16.74 | 19.73 | *** | 8.71 | 0.82 | | 0.26 | 147.24 | | 63.48 | 39.73 | | 402.78 |
| Variance linear slope | 26.45 | * | 13.31 | 0 | | . | 0 | | . | 33.99 | | 35.2 | 170.45 | | 284.8 |
| Residual variance | 65.76 | *** | 13.22 | 27.19 | *** | 5.78 | 0.32 | *** | 0.07 | 103.53 | ** | 40.79 | 1648.18 | ** | 560.34 |
| *Model Fit Statistics* | | | | | | | | | | | | | | | |
| −2LL | 700.9 | | | 519 | | | 198.4 | | | 505 | | | 349.4 | | |
| AIC | 706.9 | | | 523 | | | 202.4 | | | 511 | | | 355.4 | | |
| BIC | 707.1 | | | 523.2 | | | 202.5 | | | 511.4 | | | 356.2 | | |
| Psuedo $R^2$ | 0.23 | | | 0.02 | | | 0.00 | | | 0.41 | | | 0.59 | | |

| Non-pregnant sample | Cortisol | | | DHEA | | | Testosterone | | | Estradiol | | | Progesterone | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Est | | S.E. | Est | | S.E. | Est | | S.E. | Est | | S.E. | Est | | S.E. |
| *Fixed effects* | | | | | | | | | | | | | | | |
| Intercept | −0.75 | | 0.92 | −1.74 | | 1.65 | −0.04 | | 0.1 | −2.61 | | 2.62 | 0.76 | | 0.87 |
| Linear slope | 0.8 | * | 0.37 | 1.84 | | 1.36 | 0.05 | | 0.07 | 2.44 | | 1.61 | −0.81 | + | 0.43 |
| *Random effects: Individual-level* | | | | | | | | | | | | | | | |
| Variance intercept | 28.41 | *** | 8.93 | 40.27 | * | 20.51 | 0.17 | ** | 0.07 | 70.94 | * | 32.37 | 9.19 | ** | 3.18 |
| Variance linear slope | 0 | | . | 24.83 | * | 13.08 | 0.04 | | 0.04 | 0 | | . | 0 | | . |
| Residual variance | 21.69 | *** | 3.97 | 61.70 | *** | 14.03 | 0.13 | ** | 0.04 | 150.31 | *** | 27.59 | 8.44 | *** | 1.57 |
| *Model Fit Statistics* | | | | | | | | | | | | | | | |
| −2LL | 596.4 | | | 690.2 | | | 146.5 | | | 750 | | | 488.6 | | |
| AIC | 600.4 | | | 696.2 | | | 152.5 | | | 754 | | | 492.6 | | |
| BIC | 600.5 | | | 696.5 | | | 152.8 | | | 754.2 | | | 492.8 | | |
| Psuedo $R^2$ | 0.01 | | | 0.26 | | | 0.30 | | | 0.01 | | | 0.06 | | |

Note. ***$p < .001$, **$p < .01$, *$p < .05$, +$p < .10$. Est. = Estimate; S.E. = Standard Error; DHEA = Dehydroepiandrosterone; Variance estimates of 0 with. for S.E. indicate that there was insufficient variance in the linear slope for convergence of the estimation of that parameter. Psuedo $R^2$ = Percent Reduction in Error (PRE). The PRE was calculated by subtracting the residual variance of the multilevel models of change from the residual variance of the unconditional model, and dividing by the residual variance of the unconditional model [26].

= 6.15, $p_{difference} < .05$); there was no detectable change from trimester/assessment 1 to 2 (Supplementary Table S6).

For DHEA there was inconsistent evidence of differences in intercept levels: pregnant women had reduced DHEA compared to non-pregnant women in models using the raw windsorized data, $\beta's = -9.57$ to $-10.08$, S.E.'s = 2.24–3.46, but differences were not present when using batch-corrected data, $\beta = 2.85$, S.E. = 2.17. There was a suggestive but not consistently statistically significant difference in change over time across groups, with linear slopes being more negative for pregnant women than non-pregnant women ($\beta's = -2.99$ to $-3.06$, S.E.'s = 1.57–1.61, $p < .10$ in all three versions; Supplementary Table S7). Change scores suggested that DHEA decreased from trimester/assessment 1 to 2 in pregnant but not non-pregnant women ($M_{pregnant} = -5.16$, SD = 7.58; $M_{non-pregnant} = 2.09$, SD = 13.44, $p_{difference} < .05$). Neither group showed change in DHEA from trimester/assessment 2 to 3, or 1 to 3 (Supplementary Table S6).

Across the multilevel models of change, testosterone showed no evidence of differences in intercept levels or change over time, $\beta's = -.02$ to 0.17, S.E.'s = 0.11–0.25 (Supplementary Table S8). Examining the change scores, there was evidence of an increase in testosterone from trimester/assessment 2 to 3 for pregnant but not for non-pregnant women ($M_{pregnant} = 0.31$, SD = 0.53; $M_{non-pregnant} = 0.05$, SD = 0.42, $p_{difference} < .05$; Supplementary Table S6), but no evidence of change in either group from trimester/assessment 1 to 2 or 1 to 3.

For estradiol, in the first analysis, there was evidence that pregnant women had lower intercept levels than non-pregnant women, $\beta = -9.51$, S.E. = 4.09, though this effect was not recovered after controlling for batch effects or the covariates, $\beta's = -2.75$ to 5.55, S.E.'s = 4.04–5.18, suggesting this effect may be driven by batch effects (Supplementary Table S9). There were no indications of differences in linear slopes for pregnant vs. non-pregnant women, $\beta's = 0.00$ to 4.62, S.E.'s = 0.00–2.87. There was, however, evidence of increases in estradiol from trimester 1 to 3 based on change scores in pregnant women ($M_{pregnant} = 14.33$, SD = 15.89; $M_{non-pregnant} = 6.39$, SD = 18.96), driven by change from trimester 2 to 3 ($M_{pregnant} = 9.79$, SD = 8.79; $M_{non-pregnant} = 5.33$, SD = 14.06; but not 1 to 2; Supplementary Table S6). Notably these changes in pregnant women did not differ in magnitude from non-pregnant women.

For progesterone, across multilevel models, there was evidence that pregnant women had higher intercept levels of progesterone, $\beta's = 30.80$–32.37, S.E.'s = 7.39–9.38, and steeper increases over time relative to non-pregnant women, $\beta's = 58.61$–74.85, S.E.'s = 5.12–9.05 (Supplementary Table S10). Change scores also showed evidence of increases from trimester/assessment 1 to 3 ($M_{pregnant} = 145.79$, SD = 39.88; $M_{non-pregnant} = -0.96$, SD = 28.62, $p_{difference} < .05$), as well as trimester/assessment 2 to 3 ($M_{pregnant} = 91.00$, SD = 43.93; $M_{non-pregnant} = -2.46$, SD = 20.94, $p_{difference} < .05$) and 1 to 2 ($M_{pregnant} = 57.78$, SD = 40.22; $M_{non-pregnant} = 1.516$, SD = 20.97, $p_{difference} < .05$; Supplementary Table S6). Each difference score indicated larger increases in pregnant relative to non-pregnant women.

In summary, there was clear evidence that progesterone is higher initially and increases dramatically across pregnancy in a way that is very distinct from non-pregnant patterns. There was suggestive evidence that cortisol and estradiol increase over pregnancy, but the magnitude of the changes were not likely different than changes noted in non-pregnant women over the same time course. There was suggestive evidence that DHEA decreases across pregnancy, particularly early in pregnancy, differently from patterns in non-pregnant women over the same time course. There was little evidence of change in testosterone.

## 4. Discussion

We used a prospective longitudinal study with a control group of non-pregnant women recruited with the same strategies from the same catchment area and age ranges to assess changes in five hair hormones across pregnancy. To our knowledge, this is one of the first studies to examine these five hair steroids in pregnant humans [4,5] and the only

study to do so with a non-pregnant control group. It is innovative to examine a suite of steroid hormones in a relatively novel biospecimen [18,27], as hair is suited to capturing changes on the months-long time scale of pregnancy and, as we illustrate here, longitudinal hormone changes are not limited to cortisol only. Further innovations of our study include a focus on within-person changes and variability which we illustrate despite the stability of this biospecimens [18]. We showed that there is substantial heterogeneity in individual trajectories of change in cortisol, DHEA, testosterone, and estradiol in hair across pregnancy, and that for these four hormones, concentrations are similar in pregnant and non-pregnant participants (except perhaps for DHEA which appears suppressed during early-to-mid pregnancy). The increase in progesterone during pregnancy expected based on prior findings in biofluids including saliva [8,15] was very strong and consistently increasing; in contrast, findings involving estradiol were weaker than expected based on findings in biofluids [8,15]. We found no evidence of increases in hair testosterone across pregnancy that have been found previously in biofluids including saliva [8,14].

Our findings suggest that the often-referenced increase in cortisol across pregnancy is observed in the sample average, but that cortisol is unlikely to increase linearly for most women. Cortisol is by far the best-studied hormone in pregnancy, including in hair, and our findings fit well within the literature on hair cortisol concentrations during pregnancy. A recent systematic review of hair cortisol concentrations across pregnancy found that across nine samples, there were increasing sample average values across trimesters, although these tended to be retrospectively measured [1], with potential for washout effects in distal segments of hair collected further from the scalp [10]. In contrast, seven samples showed an average decline or stability in the first two trimesters but rebound in the third trimester either back to trimester 1 levels (in three studies), or to higher levels at trimester 3 (in four studies), and an additional three samples showed declines or trimester 2 peaks in sample average hair cortisol levels [1]. It is important to note that the reviewed findings pertain to sample average levels of hair cortisol and not the within-person trajectories of change assessed here. Here, beyond replicating the majority of studies that found increases in hair cortisol across pregnancy, we found specifically that cortisol increases more during the latter part of pregnancy for many pregnant women and that the most prevalent pattern of within-person change was non-linear.

We also highlighted the heterogeneity in patterns of change across pregnancy, and that the frequently cited sample average linear increase across trimesters in multiple hormones in pregnant women applied to a low proportion of the sample. Across models, there was most support for a sample-average linear increase in cortisol for both pregnant and non-pregnant samples. This was reflected in 71% of pregnant women and 64% of non-pregnant women having higher concentrations at the third visit than the first, although only 25% and 21% respectively showed the constant (e.g., trimester 1 to 2 and trimester 2 to 3) increase produced by the linear model. In the pregnant sample, 50% showed an overall decrease in DHEA, with ~46% showing either decreasing and then stable or decrease and then increasing patterns that matched the average change scores. In contrast, there was little heterogeneity in patterns of change in progesterone for the pregnant sample, with 100% showing an overall increase, and 2/3 showing a constant (e.g., trimester 1 to 2 and trimester 2 to 3) increasing pattern. In addition to the lack of linear change at the sample average in the majority of hormones across trimesters, there was also very little evidence of stability. This substantial diversity in the pattern and magnitude of within-person changes highlights a need for future studies to assess within-person hormone trajectories and to consider non-linearity when establishing associations of hormone changes with phenotypes of interest (e.g., perceived stress and cortisol changes, for example).

Our use of hair to prospectively measure changes in hormone levels was a strength, as the specific levels at each assessment index cumulative hormone levels rather than other biospecimens that are potentially highly influenced by shorter-term fluctuations (e.g., stressor-specific,

diurnal, menstrual for the non-pregnant sample). While hair cortisol has been increasingly popular to track pregnancy-related hormone changes [1], we are one of the first to measure estradiol, progesterone, DHEA, and testosterone in hair (see also [2,4]). And, to our knowledge, we are the first to repeatedly and prospectively measure these steroids in pregnant women with a non-pregnant comparison group.

Given the novelty of the sex hormone assays in hair, in particular, it is important to consider whether comparable concentration levels and ranges of hormone values are being returned across studies. For estradiol and progesterone, this entailed establishing comparable units of measurement by converting results from pmol/gr to pg/mg reported in a premenopausal sample of women similar in age and BMI to our non-pregnant sample [19]. They found, using immunoassays, average estradiol levels of 36,149.20 pg/g (whereas we found sample averages in the range of 45–51 pg/g) and average progesterone values of .029 pg/mg (whereas we found sample averages in the range of 5–7 pg/mg). However, in another sample of slightly older women (mean age 41y), hair progesterone measured by via liquid chromatography tandem mass spectrometry (LC-MS/MS) was 5.94 pg/mg [20], very similar to the non-pregnant sample here. In a sample of rural Indian women (mean age 39y, 1.6% pregnant) assessed via LC–MS/MS, average DHEA concentrations were 5.81 pg/mg (whereas we recovered values in the range of 22–27 pg/mg for non-pregnant women), progesterone was 6.03 pg/mg (similar to our non-pregnant sample), and testosterone was 0.40 pg/mg (whereas we recovered values in the range of 1.5–2 pg/mg) [28]. Finally, in a subsample of pregnant participants without post-partum depression, third trimester levels (measured by LC-MS/MS, which we converted from nmol/L using information provided in the manuscript) were reported of 327.06 pg/mg for cortisol (much higher than our average of 12 pg/mg), 7.15 pg/mg for testosterone (higher than our average of 2 pg/mg), and 1826.97 pg/mg for progesterone (much higher than our average of 190 pg/mg), and 61.21 pg/mg for DHEA (much higher than our average of 11.5 pg/mg) [4]. It is critical to note that these comparison values vary widely from each other, which has been documented in the cortisol literature [1,9]. In sum, our cortisol assay results fit within the observed range in the broader literature. Our testosterone and DHEA results were higher than one comparison sample but lower than the other. Our estradiol results were higher than the other available comparison sample. Our progesterone assay results were consistent with two of the four comparison studies, lower than one and higher than one. It is likely, as in the cortisol literature, that the laboratory and specific assay procedures cloud the comparability of these results [1,9,27].

Our novel findings regarding heterogeneity in levels and both the pattern and magnitude of within-person changes, and non-systematic fluctuations in particular are important for advancing our understanding of hair as a biospecimen for hormone research. Presuming that hair is the gold standard for basal, stable levels of any hormone is likely too simplistic, and a continued discussion of how hair hormone concentrations can be leveraged vs. limited is important. While the time-scale for hair hormone concentrations may be more stable than many biospecimens, the concentrations nonetheless do fluctuate over time. For example, hair may be a particularly useful biospecimen for understanding the importance of seasonal effects and moderately-chronic stressors (e.g., pandemic-related stress) that unfold across months. Similar to other biospecimens (e.g., saliva), a single assessment of two or three cm of hair may suffer severe limitations when attempting to assess basal levels – namely an inadequate number of assessments to arrive at a stable average for an individual. Alternatively, fluctuations over time within hair hormone concentrations may also point to the endocrine system as a system that fundamentally shows variability over time and is responsive to social contextual influences on a wide variety of time-scales, with hair excelling at longer, slower influences than other biospecimens. This potential strength is not without unique limitations, which can be mitigated through the use of key covariates that may contribute to fluctuations, but more work is needed to identify the critical covariates for specifically hair as a biospecimen. Acknowledging fluctuations and heterogeneity in

patterns of within-person change of hair hormone concentrations, which we found across an approximately nine-month timeframe that is consistent in pregnant and non-pregnant women, is important in that it may shift the focus of future research using hair hormone concentrations to individual differences in the dynamics of change rather than descriptions of sample-level averages.

This study has limitations which are important to acknowledge, such as a small sample size, worsened by insufficient quantity of hair for sex hormones in the pregnant sample. The use of $p < .05$ may be overly restrictive for this small sample size, due to limited power in some instances. We were unable to assess non-linear change in growth models due to our assessment schedule (i.e., only 3 timepoints). Further, we relied on the common benchmark of 1 cm of hair corresponding to 1 month of growth, on average, despite known wide ranges in rates of growth across individuals and warnings that blind following of this rule of thumb will increase measurement error and cloud interpretation (e.g., Refs. [23,29]). Our small sub-sample measurement suggested that the rates of growth were consistent with this estimate, although perhaps slightly faster. Further, we did see variability in the growth rate in our small sub-sample, introducing noise into the trimester distinctions, as expected. Thus although we interpret our measures as cumulative hair hormone concentrations in each trimester, there may be a wider variation in the amount of time actually captured that could explain some of the heterogeneity in patterns of change [30]. Further, it is likely that there are fluctuations within individuals in the rate of growth across this 9-month time scale, given evidence of seasonal changes in human hair growth [31] and in growth rates during pregnancy, given differences in scalp hair diameter during pregnancy [32], and in hair growth rates surrounding reproductive transitions [33]. Diminishing this limitation, our findings did not suggest large differences in average growth rates from trimester/assessment 1 to 2 vs. 2 to 3 in pregnant women, or in pregnant and non-pregnant women. However, these findings should be interpreted with caution given the tiny sample for which hair growth was measured in this study (e.g., underpowered to detect any true, small differences). In the future, directly measuring the rate of growth from the repeated samples could greatly improve the accuracy of interpretation and timing of hair hormone dynamics.

We showed substantial batch effects. It should be noted that batch is infrequently controlled for in hormone data – or at least rarely reported, because researchers in the field of behavioral endocrinology have largely assumed such effects to be nominal based on CV % indicating high-quality assay. However, there is wide recognition of this source of variation among laboratories conducting the assays and among groups assaying other specimens (e.g., epigenome wide association studies, [34]). Here, batch effects contribute to between-person differences because of our experimental design. Unfortunately, batch effects contribute to group differences in this study as well, since group was a between-person variable that was completely confounded with batch. This is a limitation of comparing across groups in this study, such that we cannot tell whether group differences that are found in the raw, windsorized data but not the cleaned data represent true group differences or error related to batch effects. A stronger experimental design would have been to include both pregnant and non-pregnant participants on the same trays. By treating assays as an experiment, and taking care to counter balance participants and samples on trays, effects of covariates can be minimized through design as well as by covariate adjustment. Careful experimental assay design then also allows for statistical control of batch effects unconfounded with other study features (i.e., in this case, seasonality and pregnant/non-pregnant status). Despite imperfect experimental design, we did design our trays to include all assessments for an individual on the same tray to minimize the role that batch effects could play on our key focus: within-individual changes over time. This adds strength to our conclusions that heterogeneity, or individual differences in concentrations and patterns and magnitudes of changes in concentrations over time within individuals is a real phenomenon, and not a statistical artifact. Thus, we show that controlling for batch effects and

considering experimental assay design even with high quality assays is critically important for researchers in this field to consider.

## 5. Conclusions

Hair is a viable biospecimen for assessing multiple steroid hormones. Our examination of hair testosterone, DHEA, estradiol, and progesterone adds to a small but growing literature, and reporting hair concentrations of these hormones at multiple times in pregnant and non-pregnant women is particularly novel. Our extraction techniques for estradiol and progesterone allowed for the ascertainment of even very low concentrations without sacrificing reliability. Using a rigorous experimental and analytic design, we recovered the strong expected pattern of progesterone being higher in the first trimester and further increasing dramatically across pregnancy. However, we found weaker evidence that cortisol and estradiol increases over pregnancy and in non-pregnant women over the same time course. There was suggestive evidence that DHEA decreases across pregnancy, particularly early in pregnancy, differently from patterns in non-pregnant women over the same time course. Most importantly, there was very great degrees of heterogeneity of concentrations and changes in these hormones over time – in both pregnant and non-pregnant participants. Although hair hormone concentrations are touted as a relatively stable biospecimen, these results highlight that there are fluctuations in all five hormones on this relatively longer time-scale. It is critical for future research to gain insights into the processes that this time-scale so that we can better understand how hair hormone concentrations can be leveraged and limited. Moving beyond discussing sample averages to including within-person and non-linear changes in studies of hormones-behavior associations during pregnancy is an important future direction for further investigation.

## Funding

## CRediT authorship contribution statement

**Kristine Marceau:** Formal analysis, Writing - original draft, Formal analysis, Writing - review & editing, designed and implemented the study, analyzed the data, and drafted the manuscript. **Emily Rolan:** Formal analysis, Writing - original draft, Writing - review & editing, collected the data, and assisted with data analysis and drafted portions of the manuscript. **Olivia C. Robertson:** Formal analysis, Writing - original draft, Writing - review & editing, collected the data, and assisted with data analysis and drafted portions of the manuscript. **Wen Wang:** Writing - review & editing, were instrumental in hormone assay and interpretation. **Elizabeth A. Shirtcliff:** Writing - original draft, Writing - review & editing, were instrumental in hormone assay and interpretation. All authors contributed to the writing and editing of the manuscript, and approved the final version for submission.

## 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.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpnec.2020.100024.

## References

[1] K. Marceau, W. Wang, O. Robertson, E.A. Shirtcliff, A Systematic Review of Hair Cortisol during Pregnancy: Reference Ranges and Methodological Considerations, Psychoneuroendocrinology (2020) 104904.

[2] A. Kapoor, G.R. Lubach, T.E. Ziegler, C.L. Coe, Hormone levels in neonatal hair reflect prior maternal stress exposure during pregnancy, Psychoneuroendocrinology 66 (2016) 111–117.

[3] A.M. Dettmer, K.L. Rosenberg, S.J. Suomi, J.S. Meyer, M.A. Novak, Associations between parity, hair hormone profiles during pregnancy and lactation, and infant development in rhesus monkeys (Macaca mulatta), PloS One 10 (7) (2015), e0131692.

[4] L. Jahangard, T. Mikoteit, S. Bahiraei, M. Zamanibonab, M. Haghighi, D. Sadeghi Bahmani, S. Brand, Prenatal and postnatal hair steroid levels predict post-partum depression 12 weeks after delivery, J. Clin. Med. 8 (9) (2019) 1290.

[5] K. Schury, A.M. Koenig, D. Isele, A.L. Hulbert, S. Krause, M. Umlauft, S. Kolassa, U. Ziegenhain, A. Karabatsiakis, F. Reister, H. Guendel, Alterations of hair cortisol and dehydroepiandrosterone in mother-infant-dyads with maternal childhood maltreatment, BMC Psychiatr. 17 (1) (2017) 213.

[6] E. Tisdale, L. Williams, J.J. Schultz, S.M. Wheeler, Detection of cortisol, estradiol, and testosterone in archaeological human hair from the Dakhleh Oasis, Egypt, J. Archaeol. Sci.: Report 27 (2019) 101968.

[7] H.Z. Yang, J. Lan, Y.J. Meng, X.J. Wan, D.W. Han, A preliminary study of steroid reproductive hormones in human hair, J. Steroid Biochem. Mol. Biol. 67 (5–6) (1998) 447–450.

[8] M. Abbassi-Ghanavati, L.G. Greer, F.G. Cunningham, Pregnancy and laboratory studies: a reference table for clinicians, Obstet. Gynecol. 114 (6) (2009) 1326–1331.

[9] W.F. Albar, E.W. Russell, G. Koren, M.J. Rieder, S.H. Van Umm, Human hair cortisol analysis: comparison of the internationally-reported ELISA methods, Clin. Investigative Med. 36 (6) (2013) E312–E316.

[10] O.R. Orta, S.S. Tworoger, K.L. Terry, B.A. Coull, B. Gelaye, C. Kirschbaum, S.E. Sanchez, M.A. Williams, Stress and hair cortisol concentrations from preconception to the third trimester, Stress-the International Journal on the Biology of Stress 22 (1) (2019) 60–69.

[11] H.M. Schreier, M.B. Enlow, T. Ritz, C. Gennings, R.J. Wright, Childhood abuse is associated with increased hair cortisol levels among urban pregnant women, J. Epidemiol. Community Health 69 (12) (2015) 1169–1174.

[12] P. O'Leary, P. Boyne, P. Flett, J. Beilby, I. James, Longitudinal assessment of changes in reproductive hormones during normal pregnancy, Clin. Chem. 37 (5) (1991) 667–672.

[13] N. Tagawa, Y. Hidaka, T. Takano, Y. Shimaoka, Y. Kobayashi, N. Amino, Serum concentrations of dehydroepiandrosterone and dehydroepiandrosterone sulfate and their relation to cytokine production during and after normal pregnancy, Clin. Chim. Acta 340 (1) (2004) 187–193.

[14] D.E. Saxbe, C.D. Schetter, C.D. Simon, E.K. Adam, M.U. Shalowitz, High paternal testosterone may protect against postpartum depressive symptoms in fathers, but confer risk to mothers and children, Horm. Behav. 95 (2017) 103–112.

[15] J. Darne, H.H. McGarrigle, G.C. Lachelin, Saliva oestriol, oestradiol, oestrone and progesterone levels in pregnancy: spontaneous labour at term is preceded by a rise in the saliva oestriol:progesterone ratio, Br. J. Obstet. Gynaecol. 94 (3) (1987) 227–235.

[16] S.J. Short, T. Stalder, K. Marceau, S. Entringer, N.K. Moog, E.A. Shirtcliff, P.D. Wadhwa, C. Buss, Correspondence between hair cortisol concentrations and 30-day integrated daily salivary and weekly urinary cortisol measures, Psychoneuroendocrinology 71 (2016) 12–18.

[17] A.D. Grotzinger, F.D. Mann, M.W. Patterson, J.L. Tackett, E.M. Tucker-Drob, K.P. Harden, Hair and salivary testosterone, hair cortisol, and externalizing behaviors in adolescents, Psychol. Sci. 29 (5) (2018) 688–699.

[18] 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.

[19] E. Kadour-Peero, S. Khoury, J. Awad, L. Shpritz, L. Chen-Konak, C. Shechner, L. Saiegh, Determination of estradiol and progesterone concentrations in human scalp hair, Clin. Exp. Obstet. Gynecol. 47 (2) (2020) 268–271.

[20] C.D. Voegel, M. Hofmann, T. Kraemer, M.R. Baumgartner, T.M. Binz, Endogenous steroid hormones in hair: investigations on different hair types, pigmentation effects and correlation to nails, Steroids 154 (2020) 108547.

[21] Annual Update of the HHS Poverty Guidelines, Department of Health and Human Services, 2020, pp. 3060–3061. https://www.federalregister.gov/documents/2020/01/17/2020-00858/annual-update-of-the-hhs-poverty-guidelines.

[22] I. Ouellet-Morin, M. Laurin, M.P. Robitaille, M. Brendgen, S.J. Lupien, M. Boivin, F. Vitaro, Validation of an adapted procedure to collect hair for cortisol determination in adolescents, Psychoneuroendocrinology 70 (2016) 58–62.

[23] R. Wennig, Potential problems with the interpretation of hair analysis results, Forensic Sci. Int. 107 (1–3) (2000) 5–12.

[24] C.D. Cowan, R.M. Hauser, R.A. Kominski, H.M. Levin, S.R. Lucas, S.L. Morgan, C. Chapman, Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation, National Center for Education Statistics, 2012.

[25] R.D. Conger, X. Ge, G.H. Elder, F.O. Lorenz, et al., Economic stress, coercive family process, and developmental problems of adolescents, Child Dev. 65 (2) (1994) 541–561.

[26] T. Snijders, R. Bosker, An Introduction to Basic and Advanced Multilevel Modeling, Sage, London, 1999.

[27] E.A. Shirtcliff, W. Wang, S.N. Moody, J.D. Smith, J.G. Simmons, Letter to the editor: a call for transparency in immunoassay techniques to enhance rigor and reproducibility, Dev. Psychobiol. 61 (6) (2019) 971–973.

[28] A. Walther, C. Tsao, R. Pande, C. Kirschbaum, E. Field, L. Berkman, Do dehydroepiandrosterone, progesterone, and testosterone influence women's depression and anxiety levels? Evidence from hair-based hormonal measures of 2105 rural Indian women, Psychoneuroendocrinology 109 (2019) 104382.

[29] M.J. Greff, J.M. Levine, A.M. Abuzgaia, A.A. Elzagallaai, M.J. Rieder, S.H. van Uum, Hair cortisol analysis: an update on methodological considerations and clinical applications, Clin. Biochem. 63 (2019) 1–9.

[30] G. Loussouarn, I. Lozano, S. Panhard, C. Collaudin, C. El Rawadi, G. Genain, Diversity in human hair growth, diameter, colour and shape. An in vivo study on young adults from 24 different ethnic groups observed in the five continents, Eur. J. Dermatol. 26 (2) (2016) 144–154.

[31] V.A. Randall, F.J.G. Ebling, Seasonal changes in human hair growth, Br. J. Dermatol. 124 (2) (1991) 146–151.

[32] J. Nissimov, U. Elchalal, Scalp hair diameter increases during pregnancy, Clin. Exp. Dermatol.: Exp. Dermatol. 28 (5) (2003) 525–530.

[33] P. Mirmirani, F. Luo, S.R. Youngquist, B.K. Fisher, J. Li, J. Oblong, T.L. Dawson, Hair growth parameters in pre-and postmenopausal women, in: Aging Hair, Springer, Berlin, Heidelberg, 2010, pp. 49–60.

[34] Y. Zhang, D.F. Jenkins, S. Manimaran, W.E. Johnson, Alternative empirical Bayes models for adjusting for batch effects in genomic studies, BMC Bioinf. 19 (1) (2018) 262.