HHS Public Access

A Systematic Review of Hair Cortisol during Pregnancy: Reference Ranges and Methodological Considerations

Kristine Marceau, **Wen Wang**, **Olivia Robertson**, **Elizabeth A. Shirtcliff**

Abstract

The present study systematically reviewed 56 articles that assessed hair cortisol concentrations during pregnancy collected from PubMed, Scopus, and Web of Science on 8/9/19 and updated on 6/29/20. Our goals were to establish reference ranges by trimester based on published studies. The majority of any given sample (e.g., 70%, the range of −1SD to +1SD) is expected to fall between 0 and 34.15 pg/mg in trimester 1 and 2, and between 8.59 and 44 pg/mg in trimester 3, with very wide ranges (e.g., values of >250 pg/mg) and substantially higher values (e.g., averages of 200's-300's reaching as high as 768 pg/mg) coming out of one specific lab. Delineating a reference range for hair cortisol concentrations across pregnancy is challenging because of known factors like differences in values returned by different laboratories and assay types. We observed inconsistency in descriptions of the data and data preparation steps post-assay. Key findings include that only half of the studies examining all three trimesters showed a constant increase in mean levels (most retrospectively assessed via segmenting), with considerable variability in patterns of change. None of the studies reported individual patterns of change. Examining within-person changes are an important next step for the field. We conclude that researchers should more clearly report decisions around outliers, units, and specifics of data transformations in the future in order to improve our ability to compare findings across studies, to understand differences in HCC values reported, and potentially to understand differences in reported associations of HCC with other phenotypes in the literature.

Keywords

Hair Cortisol Concentrations; Systematic Review; Pregnancy; Prenatal Stress

1. Introduction

The number of studies examining maternal hair cortisol concentrations (HCC) as a marker of prenatal stress has dramatically increased in recent years. Reference ranges have been reported in children (Noppe et al., 2014), and healthy adults (Albar et al., 2013) but not yet in pregnant women. This study conducted a systematic review of studies reporting hair

cortisol concentrations in pregnant women in order to determine whether there is a robust reference range and pattern of change across pregnancy, to document norms for collection and assay, and to propose standards for data preparation and reporting.

HCC is a non-invasive measure of cumulative hormone production. In brief, (see Stalder and Kirschbaum, 2012 for a more detailed explanation) it is thought that free cortisol is incorporated into the hair shaft as it grows, primarily through passive diffusion, but also through sweat and/or sebum secretions, and small amounts of production of cortisol from the follicle itself. These are presumed mechanisms - the exact mechanisms by which cortisol is embedded in hair are unknown. Because hair contains free cortisol, the levels should be comparable to that present in other commonly used biofluids, like saliva, and should represent bioavailable cortisol. These hypotheses are typically confirmed (e.g., Short et al., 2016). Since hair grows an average of 1 cm per month (Wenning, 2000), fluctuations in cortisol at the minutes- and hours- level are controlled through analyzing cortisol from hair samples. As such, HCC is thought to indicate a stable basal level of cortisol production rather than stress responsivity. Indeed, HCC levels were not linked to specific reports of psychosocial stress, but have been associated with chronic and ongoing stress (Stalder et al., 2017).

1.1 Methodological Considerations

There are several methodological considerations to interpreting HCC values that have been discussed extensively in the literature. Of particular relevance to pregnancy, the timing and meaning of hair samples has implications for resulting HCC. The benchmark of 1 cm of hair corresponding to one month of growth and therefore representing cumulative HCC exposure over one month has become a nearly ubiquitous rule of thumb. This is despite wide reported variations (e.g., .5 to 3.6 cm) in rates of growth and cautions against this metric in the literature (e.g., Loussouarn et al., 2016; Greff et al., 2019; Wennig, 2000). In pregnancy, there is evidence that a higher percentage of scalp hairs actively grow and that scalp hair diameter is thicker during pregnancy in comparison to non-pregnant women (Nissimov & Elchalal, 2003). Hair growth rates change during menopause (Mirmirani et al., 2010), another time of hormonal change in the lives of women. Further, there is evidence of seasonal changes in human hair growth (Randall & Ebling, 1991) which could contribute to fluctuations within individuals in the rate of growth across the 9-month time scale of pregnancy. Thus, the widespread use of 9 cm of hair to index all of pregnancy, or a 3 cm of hair to index particular trimesters likely introduces error variability to measures of HCC, and this error may be increased during pregnancy relative to other life periods.

In general, HCC has been found to be less affected by covariates/confounding influences than cortisol measured in other biospecimens (Stalder and Kirschbaum, 2012). However, a unique feature of hair cortisol (as opposed to other biospecimens) is the need for washing samples: Studies have shown that HCC values can be contaminated by cortisol in sweat or sebum, which could substantially elevate concentrations, and thus washing is considered a critical step in the pre-processing of hair prior to assay (Stalder and Kirschbaum, 2012). Accordingly, gently shaking or inverting samples in isopropanol in order to remove cortisol

deposited on the surface of the hair from sweat or sebum has been incorporated into most hair assay procedures (Meyer et al., 2014).

1.1.1 Covariates.—Hair color and texture have largely not shown systematic differences in HCC values, nor have smoking status, oral contraceptives, or general medicine use (Stalder and Kirschbaum, 2012; except a recent study, Fischer et al., 2017, found associations with anxiolytics/antidepressants). However, body fat and potentially age, sex, heat-based treatments, season (summer vs. winter months) and frequency of hair washing are potentially related to HCC levels (Stalder and Kirschbaum, 2012; Stalder et al., 2017). This may be particularly true for more distal segments, which are particularly common in the pregnancy literature, although no effects were found in a more recent study (Fischer et al., 2017). Overall, it is thought that for retrospective assessments, the greater exposure to damage from the sun, products, or frequent washing in hair further from the scalp could lead to greater wash-out of cortisol from those hair segments (Stalder and Kirschbaum, 2012; Orta et al., 2018). In addition to these covariates, other studies have shown BMI, single mother status, medical conditions (e.g., gestational diabetes mellitus), cardiovascular medications, sample storage, and race have been linked to HCC levels in pregnant and non-pregnant samples as well (Liu and Doan, 2019). Thus, in this review we cataloged the use of covariates in studies of hair cortisol during pregnancy.

1.1.2 Assay Procedures.—Other sources of methodological variability in hormone concentrations may come from assay procedures. There is evidence to suggest that there is little standardization across methods of extraction and analysis, and kits used (Albar et al., 2013). In healthy adults, the range of average HCC levels varied from 20pg/mg to 46.1pg/mg, perhaps not as widely as may be expected based on myriad differences in participant characteristics as well as several key aspects of extraction. Further, average concentrations differed across four major laboratories that perform hair cortisol assays (Albar et al., 2013). Although there is evidence of rank-order stability (i.e., high correlations) across assay manufacturers, HCC values vary widely (Russell et al., 2015). Differences in extraction and assay included preferred hair mass, washing method, pulverization method, extraction solvent, duration of extraction and condition (i.e., temperature), centrifugation, and evaporation processes, reconstitution, vortexing, and the assaying technology. In a meta-analysis, Stalder et al., (2017) replicated large lab-based differences in HCC levels and reported that immunoassay protocols tended to yield higher HCC levels than Liquid Chromatography (LC)-based methods, with LC-based methods likely to be less biased, potentially because immunoassays overestimate steroid content due to antibody cross-reactivity.

1.2 Present study

The present study sought to conduct a review of studies that have assessed HCC during pregnancy. Our goal was to establish reference ranges based on published studies. As hair has caught on as a biospecimen in which to measure cortisol during pregnancy, several differences in studies (e.g., collection timing, assay procedures, covariate usage, data handling) have begun to emerge. Variability due to error or systematically introduced by other factors can be particularly problematic in that it can lead to mixed findings in

associations with phenotypes of interest, as we have learned from the vast literature on salivary cortisol using similar assay procedures and often identical assay kits (e.g., Jessop et al., 2008). We were motivated to identify whether there are systematic differences in ranges recovered from prospective measures vs. retrospective segmentation across studies (e.g., Orta et al., 2019), to ascertain the extent to which trajectories of change are assessed at the sample average vs. individual levels, and to assess the quality of reporting of analytic steps that could aid in interpretation of HCC values. These features in particular are important for establishing quality reference ranges, as poor quality reporting or methods would have downstream ramifications for measurement error and errors in interpretation that could in turn affect associations of HCC with phenotypes of interest. Thus, in order to understand the differences in values and ranges across studies in a narrative sense, we assessed details of the study (e.g., sample information, timing of collection, covariates, exclusion criteria), and details of the assay (e.g., hair length and treatment, assay type, and assay process). Through our observations of the literature, we also noted inconsistency in descriptions of the data and data preparation steps post-assay, and so details of the data cleaning process post-assay (e.g., outliers, transformations, covariates) were also documented and summarized.

2. Materials and Method

2.1 Literature Search

We conducted a systematic review of the literature using PRISMA guidelines (Moher et al., 2009) in order to collect articles that assess HCC during pregnancy. Our search terms were as follows (conducted on 8/9/19 with no date restriction, and updated on 6/29/20 restricted to articles published since 8/9/19):

PubMed. All Fields: ((hair cortisol) AND (pregnan* OR prenatal)) Filters: Humans

Scopus. (TITLE-ABS-KEY ("hair cortisol") AND TITLE-ABS-KEY (pregnan* OR prenatal))

Web of Science. TOPIC: (((hair cortisol) AND (pregnan* OR prenatal))) Refined by: DOCUMENT TYPES: (ARTICLE OR REVIEW OR PROCEEDINGS PAPER) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

Citations were downloaded into a reference management software (Endnote X9). First, duplicate references from the three searches were identified and removed. Next, abstracts were reviewed, and references were excluded if they were: 1) animal studies, 2) studies that were not written in English, 3) assessed child but not maternal hair cortisol, 4) assessed cortisol from a different vehicle, not hair, 5) were reviews, or 6) otherwise did not assess prenatal maternal hair cortisol concentrations.

2.2. Full Text Review and Recording

Next, full texts were obtained for each of the relevant articles. Full texts were reviewed and coded for the sample characteristics (N, study name), hair sample information (weight, segment length, number of segments per sample, timing of collection, location of collection), analysis, HCC information, and goal of the study. In case-control or multiple-

group studies, cases and controls/groups were recorded separately. Extraction and assay procedure steps were recorded, including washing, drying, weighing, grinding, extraction, incubation, evaporation, reconstitution, and assay kit and timing (Meyer et al., 2014).

3. Results

3.1 Search Results

Figure 1 displays the PRISMA flow chart for article selection. Searches occurred in two waves. The initial endnote library contained 259 references (PubMed = 67; Scopus = 74; Web of Science = 118). Removing duplicate records yielded 152 articles. References were excluded if they were: a) animal studies (n = 41), b) studies that were not written in English (n = 6), c) assessed child but not maternal hair cortisol (n = 14), d) assessed cortisol from a different biospecimen, not hair (n = 9), e) were reviews (n = 10), or f) otherwise did not assess prenatal maternal hair cortisol concentrations (n = 30). This yielded 41 relevant articles. The updated search yielded 36 additional references after removing duplicates across platforms and from the prior search (PubMed = 33, Scopus = 33, Web of Science = 33). From these additional references we excluded a) animal studies (n = 5), b) studies that were not written in English (n = 0), c) assessed child but not maternal hair cortisol (n = 9), d) assessed cortisol from a different vehicle, not hair (n = 1), e) were reviews (n = 1), or f) otherwise did not assess prenatal maternal hair cortisol concentrations (n = 5), yielding 15 additional articles for inclusion. Together, 56 references were included (see Table 1), including participants from 15 different countries: United States (n = 16), Germany (n = 13), Spain (n = 9), Canada, Finland, Peru (n = 3 each), Australia, China, Denmark, Israel, Iran, Kenya, the Netherlands, Norway, and Sweden (n = 1 each). Supplemental Table A.1 includes additional sample information including study goals and collection details.

There were a few participant samples duplicated across the studies. To the best of our knowledge, there were 3 instances of publications coming from the same sample of women - (Braig et al., 2015 and 2016; Orta et al., 2018 and 2019; Schreier et al., 2016; Schreier et al., 2015). There were also some closely related participant samples. For example, Bosquent Enlow et al., 2017 and 2019a and 2019b are from the same overarching study as the Schreier studies. Further, the Caparros-Gonzalez, Garcia-Leon, and Romero-Gonzalez studies shared an IRB approval and protocol. Braig et al., 2017 and 2020 are drawn from the same sample as Braig et al., 2015, 2016. Finally, Mustonen et al, 2019 and Kajanoja et al., 2020 appear to be drawn from the same participant sample. For these studies, we note in Tables 1 and A.1–A.4 when the same participant sample or overarching study was used. However, each manuscript is included in the Tables and narrative review particularly when reporting standards are discussed, and because not all details were identical (e.g., hair samples or segments used, covariates included, focus of the paper, exact N used). When quantifying assay procedures or sample sizes in terms of proportions of the literature, we use a different n, noted in the text, based on the 47 unique participant samples. Based on this n of 47 unique participant samples, studies ranged in sample size from one to 768: 14 studies had fewer than 80 participants, 12 had 80–100 participants, and 21 had more than 100 participants. There was a wide range in study goals, from establishing assays and

patterns over pregnancy to establishing associations with various maternal characteristics (e.g., mental health, preferred delivery mode) and birth and infant outcomes.

Considering the manuscripts that had non-identical samples (47), 10 manuscripts (21%) examined pregnancy as a whole, either by using 9cm of hair at delivery, or by having variable gestational ages at hair sample collection that covered most of the pregnancy. Fourteen (30%) examined only trimester 3, whereas two examined trimesters 2 and 3 only (4%; one with prospectively collected hair samples, and one with retrospective, segmented hair samples), two (4%) examined trimester 2 only, and three (7%) examined trimester 1 only. Just over a third, 16 (34%), examined all three trimesters: of those, half (8, 17%) were prospectively collected hair samples and the other half (7, 15%) were retrospectively segmented (and one was a combination with T1 prospective but T2 retrospective).

3.2   Hair Collection Information

Information on hair collection and storage are also presented in Table A.1. Storage of hair was typically in aluminum foil at room temperature, but sometimes at −80°C, −20°C or in Eppendorff tubes. Few manuscripts (14, though some of these could be inferred by referenced articles by the same group/protocol) did not report storage conditions. The location (i.e., hospital, laboratory, home) of collection was rarely reported (data not shown). The vast majority of manuscripts reported segmenting of the hair sample – when segmenting was done, it was most often a hair sample that was collected shortly after delivery, which was segmented into 3×3cm segments in order to retrospectively represent pregnancy. All but one manuscript reported the segment length, and this was most typically 3cm (presumably corresponding to approximately one trimester) ranging as short as 1cm. The timing of collection was almost always reported, and in many cases (21 studies) a hair sample was collected shortly after delivery. In 28 manuscripts hair samples were taken at some point during pregnancy, with only 9 manuscripts using prospectively collected hair samples in each trimester.

3.3   Extraction and Assay Methods

Information on cortisol extraction, and assay are presented in Supplemental Table A.2. Of the non-identical samples that reported the assay type (45 studies), LC-based methods (21, 45%) and/or ELISA methods (15, 32%) are most commonly used, with some of the less precise enzyme immunoassay (7, 15%) and/or radioimmunoassay (RIA) methods (3, 6%). All but 7 studies reported the kit/manufacturer. One study compared LC-MS and ELISA assays, and one reported using RIA and ELISA. Three studies did not report the assay method used.

Hair processing procedures include washing, drying, grinding, extraction (and condition), evaporation, and reconstitution (Meyer et al., 2014). Ideally, details should be reported for each step, especially for some important steps such as extraction and reconstitution. There was a wide variety in reporting. Some groups reported detailed assay procedures in every study, whereas other reported them once and cited back to the original article; both strategies are considered high quality reporting. However, some studies failed to report sufficient detail and cited back to a more general protocol, or a study that contained discussion of multiple

protocols without specifying which was followed. The most frequently reported assay steps were washing, grinding, extraction, evaporation, and reconstitution. The least frequently reported steps were assay timing, drying after washing, weighing (including weight used and order of weighing in the protocol), incubation, and lab in which the assay was performed. These are considered lower quality reporting. Only three studies failed to provide most of the processing procedures. The most frequently used lab was the Kirschbaum laboratory at the Technical University of Dresden, Germany (15/47), though several studies were assayed at the University of Granada (9/47); in many cases, a lab was not explicitly named (11/47).

All of the studies that reported washing the hair as part of the assay procedure used isopropanol, with varying levels of specificity in terms of amount (e.g., 2.5mL of isopropanol), length of wash time (e.g., two vs. three minutes if specified) and number of times (e.g., once vs. twice). Drying was reported less frequently, and typically included drying for at least 12h or overnight, sometimes with more specificity (e.g., covered with Kim wipes and left on top of a table to dry overnight; at room temperature; under a fume hood). Weight of hair used for assay (7.5 vs 10–25mg of hair) and when it was weighed (i.e., before or after washing or grinding if applicable) also were infrequently reported and varied. Grinding methods were reported in many studies, and included finely mincing with scissors, pulverized using a ball mill, using liquid nitrogen, using a tissue lyser, or using whole non-pulverized hair. Extraction typically was done using methanol, with variations in amount (e.g., ranging from 1mL to 2mL if specified) and grade (e.g., HPLC-grade or not specified). Incubation ranged in length from 18hr to 72hr, and in specifics (e.g., in the dark, with constant inversion/rotation or shaking, at room temperature vs. 50 degrees Celsius, sonicated vs. not). Evaporation was typically under a stream of nitrogen or using a vacuum evaporator. Rarely, details such as the temperature, amount of supernatant, and amount of time for evaporation was reported. Reconstitution (i.e., liquid and amount of liquid) varied with the kit used, often including distilled or double-distilled water, phosphate-buffered saline, or assay buffer. Few studies reported on assay timing or whether the reconstituted liquid was frozen (at −20°C if reported) for later assay vs. assayed immediately.

3.4 Data Cleaning Procedures, Covariates, and Exclusions

Available sample and data cleaning details are presented in Supplemental Table A.3. Generally, the most common pattern was for studies to report log transforming HCC values but not to mention how outliers were handled or not handled. 37/56 manuscripts (66%) reported log transforming HCC values. This creates a challenge in observing a reference range: only 19 of the 37 studies provided the function (e.g., natural log [ln, base e] or log10 [base 10]); none of these studies reported whether or not a constant was added prior to log transforming. With both pieces of information, the arithmetic mean can be calculated (e.g., as e^x or 10^x).

3.4.1   Covariates.—Handling of covariates varied widely by study. Studies often cited Stalder and Kirschbaum (2012) or other literature as a rationale for the inclusion of HCC covariates. Due to the wide variability in whether or not and how covariates were considered, we provide a broad and narrative overview (Supplemental Table A.3). Nine manuscripts failed to note covariates. Many manuscripts examined bivariate correlations

between HCC and *potential* covariates or ran ANOVAs/t-tests and then only included those that were significantly associated with HCC in finals models. There was a large variety of covariates investigated across studies depending on the focus of the study (e.g., with HCC as a predictor or an outcome, see Supplemental Table A.3).

Here, we focus on covariates specifically relevant to HCC, as recommended by Stalder and Kirschbaum (2012) and those that were otherwise prevalent during our review of the literature. In line with recommendations by Stalder and Kirschbaum (2012), the most commonly investigated covariates were body composition variables (e.g. pre-pregnancy BMI, BMI, waistline size), followed by chemical treatment of hair (e.g., perms, dyes, bleaching, relaxers, chemical straightening), maternal age, and sex of the fetus. Of note, two of the manuscripts that investigated fetal sex found higher HCC in mothers gestating males or an interactive effect at high levels of HCC in mothers gestating males, although most investigations yielded null effects of sex on maternal HCC. The next most commonly investigated covariates were maternal education, hair washing frequency, glucocorticoid use, hair type (e.g. color, "structure"), and previous miscarriage. Less commonly investigated covariates were season of delivery, batch assay, hair age, ultraviolet ray exposure, and exercise frequency. The majority of covariates were ultimately excluded from models, and no systematic patterns emerged in terms of key covariates that were typically ultimately included.

3.4.2   Exclusion Criteria.—In terms of exclusion criteria, almost half of manuscripts excluded participants based on glucocorticoid use (25/56). Most made these exclusions from the recruitment phase of the study although some dropped participants later on who reported glucocorticoid use. Just under half excluded based on participants having a major chronic diseases or conditions (24/56) and almost a quarter excluded based on having a psychiatric condition or who reported heavy drug use (12/56). Many excluded participants who had a multiple gestation pregnancy (21/56), and some excluded on having severe pregnancy or fetal complication (16/56). Fewer manuscripts excluded participants based on shift work (4/56), having bleached hair (3/56), having been induced or delivery by cesarean section (4/56) or previous miscarriage (2/56).

3.5   Hair Cortisol Concentrations – Is there a reference range?

3.5.1   Quality of the data.—Generally, the reviewed articles reported HCC consistently in pg/mg units. Typically, authors reported either pg/mg or "log transformed" pg/mg. In Supplemental Table A.4, we calculated the arithmetic mean assuming no constant was added, in the cases in which the function was reported. Many manuscripts that included log transformation also provided the raw mean/standard deviation (SD) or median/inter-quartile-range (IQR) for comparison across manuscripts - in those cases, the raw values were recorded in Supplemental Table A.4.

There were five manuscripts reporting units other than pg/mg. One manuscript reported in pmol/g and another in nmol/g. The molar mass of cortisol is 362.46 g/mol, so unit conversions were used such that 1 nmol/g = 362.46 pg/mg, and .001 pmol/g = 362.46 pg/mg. One manuscript reported ng/g which is equivalent to pg/mg. Another reported mg, which we

presume may be a typo that should have been pg/mg, and three others were ng/ml, pg/mL, and nmol/L. We exclude these four from further discussion. Six of the manuscripts failed to provide the units in which they measured cortisol, but based on the other work from those labs, the units are likely pg/mg or log-transformed pg/mg. In terms of metrics for HCC values, most reported mean and SD, some only reported the range, and some reported the median and IQR. Few manuscripts provided multiple metrics to better define sample distributions.

3.5.2   HCC reference ranges.—Due to the wide variation in the quality of reporting, establishing a reference range is challenging. In Supplemental Table A.4, we present the reported mean or median HCC value, unit, and either a calculated value for −1 SD and +1 SD or the reported IQR. If range data were presented this was also recorded (i.e., in the notes column E of Supplemental Table A.4). Finally, another source of error in this qualitative review is that several manuscripts did not provide descriptive data in numerical format, and thus the first author judged the approximate levels from the included Figures by eye (in 7 studies, noted in the notes column E of Supplemental Table A.4).

Ten manuscripts, including 17 subsamples – i.e., HCC presented for mothers who had pre-term births vs. term births separately, reported HCC for the entire pregnancy. The (sub)sample averages were between 6.6 and 190.6 pg/mg. In the manuscript reporting an average of 6.6 pg/mg, a high cortisol outlier of 67.7 pg/mg was removed prior to calculating the mean. The range where approximately 70% of the sample is expected to fall (+/−1 SD) varied between 2.1 pg/mg and 289.6 pg/mg across the studies.

Among trimester 1 (sub)samples, the sample averages ranged from 1.88 (a calculated arithmetic mean, the lowest reported value in pg/mg was 3) to 601.86 pg/mg. Samples from the University of Granada research group tended to be high, with sample averages ranging from 289.69 – 601.86 pg/mg with majority sample (i.e., +/−1SD) ranges from 0–948.89 pg/mg. Two of the four very low estimates of 3 and 4 pg/mg were estimated by eye from the case studies. The remaining samples had averages ranging from 5.1 to 27 pg/mg, with reported ranges between 0 and 123.7 pg/mg, and +/−1 SD calculations indicating that the majority (i.e., 70%) of individuals are expected to fall in the 0 to 34.15 pg/mg range. This pattern was entirely consistent when examining trimester 2, except that the highest +1SD value increased to 1135.42 pg/mg in one study from the University of Granada.

Among trimester 3 samples, as noted in the previous trimesters, the assays from the University of Granada had higher reported concentrations, with sample averages ranging from 267.06 – 482.69 pg/mg, and +/−1SD ranges of 0 to 768.66 pg/mg. One other sample reported a high range – a sample of women with opioid use disorder (average 131.8 pg/mg, +/−1SD range = 7.1–256.5 pg/mg). The case studies reported an (eyeballed) 5–6 pg/mg. Three samples reported median values of 2.8 to 86.6 pg/mg (interquartile ranges encompassing 2.6 to 169.2 pg/mg); the other studies reported mean estimates between 8.59 and 44 pg/mg and +/− 1SD ranges from 0 to 93.4 pg/mg.

3.5.3   Change across pregnancy.—Prospectively collected data (8 manuscripts) showed little average change (e.g., T1: 5.78, T2: 5.76, T3: 5.76 for logarithmized

(pg/mg) means in Romero-Gonzalez et al., 2018). Retrospectively segmented samples (7 manuscripts) included more (6) instances of increases across trimesters (e.g., T1: .72, T2: .74, T3: .92 for logarithmized (pg/mg) means in Bosquet Enlow et al., 2019; T1: 15, T2: 20, T3: 44 pg/mg for Kirschbaum et al., 2009), which is likely evidence of wash-out effects in longer (e.g., 9cm) hair samples particularly using ELISA assays (Orta et al., 2018). Examining the studies that had values reported for all 3 trimesters regardless of units, overall HCC estimates, or collection/assay procedures, there were multiple average patterns of change. Nine studies reported the often cited increase over time, although it should be noted that these were very small increases relative to the +/− 1SD ranges reported at each trimester. For example, the group of women who delivered at term in Hoffman et al., 2016 had average HCC in log(pg/mg) of 1.8, 2, and 2.4 in respect to trimester, with the +/− 1SD range spanning 1.1–2.5 log(pg/mg) in trimester 1 increasing to 1.7 to 3.1 log(pg/mg) in trimester 3. This indicates that the sample-level increases are smaller than the sample-level variation at any given time-point. Three samples showed an average decline from trimester 1 to 2 with some rebound in trimester 3, to at or near trimester 1 levels. Four samples showed an average decline or stability from trimester 1 to 2, with a relatively steeper increase to a peak in trimester 3. One sample showed on average decline across pregnancy. Two samples showed average peak levels in trimester 2. It is important to note that sample average trajectories are not equivalent to within-person change, as they confound between and within-person variance over time. None of the manuscripts reported individual trajectories of change across trimesters. Thus, it is currently as yet unknown what the predominant trajectory of within-individual change in HCC over pregnancy is, and how much variability in trajectories exist. It is likely that there is variability in trajectories based on the available data, and evidence of variability in trajectories from Galbally et al., 2018, which fit growth curves of HCC from three T3 segments and one 12-months post-partum sample.

4. Discussion

In an age when open science is increasingly embraced, and scientists and funding agencies are increasingly concerned about rigor and reproducibility, it is critical to explore the strengths and limitations of bodies of literature. HCC has generated excitement in the field for many reasons, notably because hormone concentrations in hair are more stable and less influenced by time-specific influences at the minutes/hours/days level than many other common biospecimens. Thus, findings from HCC should be more replicable than findings in saliva or blood because more of the context-specific fluctuations are controlled by the nature of the biospecimen itself. We sought to review the literature in hope of understanding whether there is a general reference range across pregnancy and within trimesters for HCC, which would allow for comparisons across studies. We found that delineating a reference range for HCC across pregnancy is challenging, due not only to large differences in values returned by different laboratories and assay types (Albar et al., 2013), but also incomplete reporting of data preparation steps (e.g., outliers, units, transformations). Disentangling the causes of the differences in values is difficult because of the number of variables that differ in collection, extraction and assay, and data cleaning procedures across studies. Moving forward, clearer reporting of these factors will improve our ability to compare findings

across studies, and to understand differences in HCC values recovered, and to potentially clarify mixed findings across the literature.

4.1 HCC levels across pregnancy

The reviewed studies included several retrospectively collected HCC levels, and several prospectively collected studies. The patterns of findings across these studies calls into question the ubiquity of increasing cortisol across pregnancy, as relatively few studies discovered monotonic increases at the sample level. Further, the presence of wider variation within trimester than increases across trimesters suggests substantial within-person variation in HCC trajectory. Unfortunately, none of the studies reported on the variations in patterns of change within individuals. Presumably, within-individual changes are of interest, and disentangling within and between-person variation and changes are an important avenue for future research. Therefore, for repeated measures, description of the trajectories is an important area in which to gain knowledge, along with some metric of individual differences in trajectories.

4.2 Recommendations

4.2.1 Assay Procedures.—Generally, reporting of assay procedures were detailed and appropriately referenced the protocols adopted. However, there were some steps of the process that were not as frequently detailed (e.g., assay timing, lab). Continuing consistent and detailed reporting of assay procedures are recommended in order to increase our ability to test whether specific components of assay procedures (i.e., times washed vs. evaporation time/temperature vs. grinding method) are associated with results (e.g., in future meta-analyses). Further, many but not all studies reported the ranges of detection for assays; range of detection is critically important because it indicates whether concentrations are plausible values and, depending on whether values fall at the highest or lowest ends of the standard curve, reliability of those concentrations. Researchers should look to other studies using the same method, and using the same lab for guidance in terms of the plausibility of specific values.

4.2.2 Data cleaning and reporting.—It is imperative for studies to report units for HCC levels. Most, but not all did. Although it is common to transform data prior to analysis in order to improve the distribution, the untransformed data should also be presented for interpretability, at minimum in an online supplement. Ideally, when reporting hormone levels, measures of sample dispersion should also be reported (e.g., standard deviation and/or interquartile range, and total range). Our review of the literature found that although many aspects of collection and assay are frequently reported, decisions around cleaning the data are far less frequently reported despite the fact that such data treatments are highly impactful for rigor and reproducibility. As one example, managing outliers and influential observations is reasonable even within small samples, but the concentrations of outlying values could be within a normal range in a broader population and thus should be reported. Further, some high outliers, particularly in small samples, may represent a group of individuals with extremely high cortisol after multiple assays – a phenomenon that has been noted across several labs although infrequently reported in manuscripts. With clearer communication of outliers, phenomena like these that could represent mechanistic issues in

hair cortisol assays or real phenotypic variation could be investigated and better understood. Omission of information about data cleaning procedures could indicate that word count limitations preclude the details, or that researchers are not systematic in cleaning the data. Regardless, decisions about how outliers are defined and treated should be reported so that the field can establish standards for data cleaning procedures.

We propose that the best practices for outliers include a) examining the distributions of the sample for natural cut-offs, b) examining the 3SD cut point of the data, c) leveraging results reported from other, ideally larger samples in the literature, and in discussion with the lab completing the assay, and d) understanding the detection limits of the assay. Using these pieces of information together, a reasonable value for outliers can be arrived at and justified in the publication. Sensitivity analyses that probe whether decisions to remove, windsorize, or include outliers substantially change the pattern of findings are also recommended to ensure that the most robust results are reported. Further, specificity in the transformation is required for comparisons across samples, including what transformation was applied, as well as whether a constant was applied.

4.2.3 Covariates.—Several publications provide guidelines for reporting covariates (e.g., Stadler et al., 2017). We recommend that studies continue to engage in the process most common in this literature: to examine bivariate associations of covariates with HCC levels, and include those that matter for the sample. Attention to effect sizes, particularly in small samples, may be more important than using a $p<.05$ threshold for inclusion. Clear reporting of the covariates examined in this way, as well as those that mattered and were included is critical information for streamlining the recommended covariates in the future.

Two covariates that stood out as receiving relatively little attention despite clear and often large effects and ease of measurement are season and assay batch. Seasonality is important not only for differences in hormone concentrations (e.g., Hadlow et al., 2014), but also due to sources of error variation specific to hair samples (e.g., sunlight exposure). Batch effects are controlled for in the quality control pipelines for other biological data (e.g., genetic and epigenetic analyses), and are typically found in the few hormone studies accounting for batch. Controlling for assay batch could improve the quality of hormone data in the future.

5. Conclusion

The present study sought to review the literature on HCC during pregnancy. Considering pregnancy as a whole, across samples, 70% of the sample is expected to fall between 2.1 and 289.6 pg/mg; trimester 1 and 2 HCC is mainly in the 0–34.15 pg/mg range, and trimester 3 between 8.59 and 44 pg/mg, with substantially higher values (e.g., 200–768 pg/mg) coming out of the University of Granada samples assayed using ELISA. Further, only half of the studies examining all three trimesters showed a constant increase in mean levels. Even based on sample averages, there was considerable variability in patterns of change, and none of the studies reported individual patterns of change. Examining within-person changes are an important next step for the field.

Although our goal at the outset was to delineate a reference range for HCC across pregnancy, known factors like differences in values returned by different laboratories and assay types and due to incomplete reporting of data preparation steps made such an endeavor problematic (Albar et al., 2013). Researchers should more clearly report decisions around outliers, units, and specifics of data transformations in the future. Doing so will improve our ability to compare findings across studies in the future, to understand the sources whether real or methodological, of differences in HCC values reported, and potentially to understand differences in reported associations of HCC with important biobehavioral phenotypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Aatsinki AK, Keskitalo A, Laitinen V, Munukka E, Uusitupa HM, Lahti L, Kortesluoma S, Mustonen P, Rodrigues AJ, Coimbra B, Huovinen P, Karlsson H, Karlsson L, 2020. Maternal prenatal psychological distress and hair cortisol levels associate with infant fecal microbiota composition at 2.5 months of age. Psychoneuroendocrinology 119, 104754. [PubMed: 32531627]

Albar WF, Russell EW, Koren G, Rieder MJ, Van Umm SH, 2013. Human hair cortisol analysis: Comparison of the internationally-reported ELISA methods. Clinical and Investigative Medicine 36, E312–E316. [PubMed: 24309228]

Bosquet Enlow M, Devick KL, Brunst KJ, Lipton LR, Coull BA, Wright RJ, 2017. Maternal Lifetime Trauma Exposure, Prenatal Cortisol, and Infant Negative Affectivity. Infancy 22, 492–513. [PubMed: 28983193]

Bosquet Enlow M, Sideridis G, Bollati V, Hoxha M, Hacker MR, Wright RJ, 2019a. Maternal cortisol output in pregnancy and newborn telomere length: Evidence for sex-specific effects. Psychoneuroendocrinology 102, 225–235. [PubMed: 30590340]

Bosquet Enlow M, Sideridis G, Chiu YM, Nentin F, Howell EA, Le Grand BA, Wright RJ, 2019b. Associations among maternal socioeconomic status in childhood and pregnancy and hair cortisol in pregnancy. Psychoneuroendocrinology 99, 216–224. [PubMed: 30265918]

Bowers K, Ding L, Gregory S, Yolton K, Ji H, Meyer J, Ammerman RT, Van Ginkel J, Folger A, 2018. Maternal distress and hair cortisol in pregnancy among women with elevated adverse childhood experiences. Psychoneuroendocrinology 95, 145–148. [PubMed: 29859342]

Braig S, Grabher F, Ntomchukwu C, Reister F, Stalder T, Kirschbaum C, Genuneit J, Rothenbacher D, 2015. Determinants of maternal hair cortisol concentrations at delivery reflecting the last trimester of pregnancy. Psychoneuroendocrinology 52, 289–296. [PubMed: 25553388]

Braig S, Grabher F, Ntomchukwu C, Reister F, Stalder T, Kirschbaum C, Rothenbacher D, Genuneit J, 2016. The Association of Hair Cortisol with Self-Reported Chronic Psychosocial Stress and Symptoms of Anxiety and Depression in Women Shortly after Delivery. Paediatric and perinatal epidemiology 30, 97–104. [PubMed: 26525484]

Braig S, Weiss JM, Stalder T, Kirschbaum C, Rothenbacher D, Genuneit J, 2017. Maternal prenatal stress and child atopic dermatitis up to age 2 years: The Ulm SPATZ health study. Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology 28, 144–151. [PubMed: 27801949]

Braig S, Logan CA, Reister F, Rothenbacher D, Genuneit J, 2020. Psychosocial stress and longitudinally measured gestational weight gain throughout pregnancy: The Ulm SPATZ Health Study. Sci Rep 10(1), 1996. [PubMed: 32029794]

Caparros-Gonzalez RA, Romero-Gonzalez B, Gonzalez-Perez R, Lara-Cinisomo S, Martin-Tortosa PL, Oliver-Roig A, Peralta-Ramirez MI, 2019a. Maternal and Neonatal Hair Cortisol Levels and Psychological Stress Are Associated With Onset of Secretory Activation of Human Milk Production. Adv Neonatal Care 19(6), E11–e20. [PubMed: 31764138]

Caparros-Gonzalez RA, Romero-Gonzalez B, Gonzalez-Perez R, Lucena-Prieto L, Perez-Garcia M, Cruz-Quintana F, Peralta-Ramirez MI, 2019b. Maternal and Neonatal Hair Cortisol Levels Are Associated with Infant Neurodevelopment at Six Months of Age. J Clin Med 8(11).

Caparros-Gonzalez RA, Romero-Gonzalez B, Strivens-Vilchez H, Gonzalez-Perez R, Martinez-Augustin O, Peralta-Ramirez MI, 2017. Hair cortisol levels, psychological stress and psychopathological symptoms as predictors of postpartum depression. PloS one 12, e0182817. [PubMed: 28846691]

Caparros-Gonzalez RA, Romero-Gonzalez B, Quesada-Soto JM, Gonzalez-Perez R, Marinas-Lirola JC, Peralta-Ramírez MI, 2019c. Maternal hair cortisol levels affect neonatal development among women conceiving with assisted reproductive technology. Journal of Reproductive and Infant Psychology.

Conradt E, Shakiba N, Ostlund B, Terrell S, Kaliush P, Shakib JH, Crowell SE, 2020. Prenatal maternal hair cortisol concentrations are related to maternal prenatal emotion dysregulation but not neurodevelopmental or birth outcomes. Developmental psychobiology.

D'Anna-Hernandez KL, Ross RG, Natvig CL, Laudenslager ML, 2011. Hair cortisol levels as a retrospective marker of hypothalamic-pituitary axis activity throughout pregnancy: comparison to salivary cortisol. Physiology & behavior 104, 348–353. [PubMed: 21397617]

de Kruijff I, Noppe G, Kieviet N, Choenni V, Lambregtse-van den Berg MP, Begijn DGA, Tromp E, Dorst K, van Rossum EFC, de Rijke YB, van den Akker ELT, 2020. LC-MS/MS-based reference intervals for hair cortisol in healthy children. Psychoneuroendocrinology 112, 104539. [PubMed: 31841987]

Duffy AR, Schminkey DL, Groer MW, Shelton M, Dutra S, 2018. Comparison of Hair Cortisol Levels and Perceived Stress in Mothers Who Deliver at Preterm and Term. Biological research for nursing 20, 292–299. [PubMed: 29490472]

Fischer S, Duncko R, Hatch SL, Papadopoulos A, Goodwin L, Frissa S, Hotopf M, Cleare AJ, 2017. Sociodemographic, lifestyle, and psychosocial determinants of hair cortisol in a South London community sample. Psychoneuroendocrinology 76, 144–153. [PubMed: 27923182]

Flom JD, Chiu YM, Hsu HL, Devick KL, Brunst KJ, Campbell R, Enlow MB, Coull BA, Wright RJ, 2018. Maternal Lifetime Trauma and Birthweight: Effect Modification by In Utero Cortisol and Child Sex. J Pediatr 203, 301–308. [PubMed: 30197200]

Galbally M, van Rossum EFC, Watson SJ, de Kloet ER, Lewis AJ, 2019. Trans-generational stress regulation: Mother-infant cortisol and maternal mental health across the perinatal period. Psychoneuroendocrinology 109, 104374. [PubMed: 31394490]

Garcia-Leon MA, Caparros-Gonzalez RA, Romero-Gonzalez B, Gonzalez-Perez R, Peralta-Ramirez I, 2019. Resilience as a protective factor in pregnancy and puerperium: Its relationship with the psychological state, and with Hair Cortisol Concentrations. Midwifery 75, 138–145. [PubMed: 31102974]

Garcia-Leon MA, Peralta-Ramirez MI, Arco-Garcia L, Romero-Gonzalez B, Caparros-Gonzalez RA, Saez-Sanz N, Santos-Ruiz AM, Montero-Lopez E, Gonzalez A, Gonzalez-Perez R, 2018. Hair cortisol concentrations in a Spanish sample of healthy adults. PloS one 13, e0204807. [PubMed: 30265722]

Gelaye B, Kirschbaum C, Zhong QY, Sanchez SE, Rondon MB, Koenen KC, Williams MA, 2019. Chronic HPA activity in mothers with preterm delivery: A pilot nested case-control study. Journal of neonatal-perinatal medicine.

Greff MJ, Levine JM, Abuzgaia AM, Elzagallaai AA, Rieder MJ, & van Uum SH 2019. Hair cortisol analysis: An update on methodological considerations and clinical applications. Clinical biochemistry, 63, 1–9. [PubMed: 30261181]

Hadlow NC, Brown S, Wardrop R and Henley D, 2014. The effects of season, daylight saving and time of sunrise on serum cortisol in a large population. Chronobiology international, 31(2), 243–251. [PubMed: 24156521]

Hoffman MC, Mazzoni SE, Wagner BD, Laudenslager ML, Ross RG, 2016. Measures of maternal stress and mood in relation to preterm birth. Obstetrics & Gynecology 127, 545–552.

Jahangard L, Mikoteit T, Bahiraei S, Zamanibonab M, Haghighi M, Sadeghi Bahmani D, Brand S, 2019. Prenatal and Postnatal Hair Steroid Levels Predict Post-Partum Depression 12 Weeks after Delivery. J Clin Med 8(9).

Jessop DS, & Turner-Cobb JM (2008). Measurement and meaning of salivary cortisol: a focus on health and disease in children. Stress, 11(1), 1–14. [PubMed: 17853059]

Kalra S, Einarson A, Karaskov T, Van Uum S, Koren G, 2007. The relationship between stress and hair cortisol in healthy pregnant women. Clinical & Investigative Medicine 30, 103–107.

Kajanoja J, Karukivi M, Mustonen P, Scheinin NM, Kortesluoma S, Rodrigues AJ, Karlsson H, Karlsson L, 2020. Alexithymic Traits and Hair Cortisol Concentrations in Pregnant Women. Front Psychiatry 11, 421. [PubMed: 32477193]

Karakash SD, Tschankoshvili N, Weedon J, Schwartz RM, Kirschbaum C, Minkoff H, 2016. Hypocortisolism and preterm birth. Journal of neonatal-perinatal medicine 9, 333–339. [PubMed: 28009336]

Karlen J, Frostell A, Theodorsson E, Faresjo T, Ludvigsson J, 2013. Maternal influence on child HPA axis: a prospective study of cortisol levels in hair. Pediatrics 132, e1333–1340. [PubMed: 24101769]

Khoury JE, Bosquet Enlow M, Patwa MC, Lyons-Ruth K, 2020. Hair cortisol in pregnancy interacts with maternal depressive symptoms to predict maternal disrupted interaction with her infant at 4 months. Developmental psychobiology.

Kirschbaum C, Tietze A, Skoluda N, Dettenborn L, 2009. Hair as a retrospective calendar of cortisol production-Increased cortisol incorporation into hair in the third trimester of pregnancy. Psychoneuroendocrinology 34, 32–37. [PubMed: 18947933]

Koenig AM, Ramo-Fernandez L, Boeck C, Umlauft M, Pauly M, Binder EB, Kirschbaum C, Gundel H, Karabatsiakis A, Kolassa IT, 2018. Intergenerational gene x environment interaction of FKBP5 and childhood maltreatment on hair steroids. Psychoneuroendocrinology 92, 103–112. [PubMed: 29674169]

Kramer MS, Lydon J, Séguin L, Goulet L, Kahn SR, McNamara H, Genest J, Dassa C, Chen MF, Sharma S, 2009. Stress pathways to spontaneous preterm birth: the role of stressors, psychological distress, and stress hormones. American journal of epidemiology 169, 1319–1326. [PubMed: 19363098]

Krumbholz A, Anielski P, Reisch N, Schelling G, Thieme D, 2013. Diagnostic value of concentration profiles of glucocorticosteroids and endocannabinoids in hair. Therapeutic drug monitoring 35, 600–607. [PubMed: 23942543]

Krumbholz A, Schonfelder M, Hofmann H, Thieme D, 2018. The plasma protein binding of the endogenous glucocorticosteroids is of vital importance for the concentrations in hair and saliva. Forensic science international 286, 23–30. [PubMed: 29547783]

Liu CH, Doan SN, 2019. Innovations in biological assessments of chronic stress through hair and nail cortisol: Conceptual, developmental, and methodological issues. Developmental psychobiology 61, 465–476. [PubMed: 30740655]

Lobmaier SM, Müller A, Zelgert C, Shen C, Su PC, Schmidt G, Haller B, Berg G, Fabre B, Weyrich J, Wu HT, Frasch MG, Antonelli MC, 2020. Fetal heart rate variability responsiveness to maternal stress, non-invasively detected from maternal transabdominal ECG. Archives of gynecology and obstetrics 301(2), 405–414. [PubMed: 31781889]

Loussouarn G, Lozano I, Panhard S, Collaudin C, El Rawadi C, & Genain G 2016. 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. European Journal of Dermatology, 26(2), 144–154. [PubMed: 27019510]

Mesner O, Davis A, Casman E, Simhan H, Shalizi C, Keenan-Devlin L, Borders A, Krishnamurti T, 2019. Using graph learning to understand adverse pregnancy outcomes and stress pathways. PloS one 14(9), e0223319. [PubMed: 31568495]

Meyer J, Novak M, Hamel A, Rosenberg K, 2014. Extraction and analysis of cortisol from human and monkey hair. J Vis Exp, e50882. [PubMed: 24513702]

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group, 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097 [PubMed: 19621072]

Musana JW, Cohen CR, Kuppermann M, Gerona R, Wanyoro A, Aguilar D, Santos N, Temmerman M, Weiss SJ, 2019. Association of differential symptoms of stress to hair cortisol and cortisone concentrations among pregnant women in Kenya. Stress (Amsterdam, Netherlands), 1–11.

Mustonen P, Karlsson L, Kataja EL, Scheinin NM, Kortesluoma S, Coimbra B, Rodrigues AJ, Sousa N, Karlsson H, 2019. Maternal prenatal hair cortisol is associated with prenatal depressive symptom trajectories. Psychoneuroendocrinology 109, 104383. [PubMed: 31400561]

Noppe G, Van Rossum EF, Koper JW, Manenschijn L, Bruining GJ, De Rijke YB, & Van Den Akker EL (2014). Validation and reference ranges of hair cortisol measurement in healthy children. Hormone research in paediatrics, 82(2), 97–102. [PubMed: 25115629]

Nystrom-Hansen M, Andersen MS, Khoury JE, Davidsen K, Gumley A, Lyons-Ruth K, MacBeth A, Harder S, 2019. Hair cortisol in the perinatal period mediates associations between maternal adversity and disrupted maternal interaction in early infancy. Developmental psychobiology 61, 543–556. [PubMed: 30747450]

Orta OR, Tworoger SS, Terry KL, Coull BA, Gelaye B, Kirschbaum C, Sanchez SE, Williams MA, 2018. An evaluation of distal hair cortisol concentrations collected at delivery. Stress (Amsterdam, Netherlands) 21, 355–365.

Orta OR, Tworoger SS, Terry KL, Coull BA, Gelaye B, Kirschbaum C, Sanchez SE, Williams MA, 2019. Stress and hair cortisol concentrations from preconception to the third trimester. Stress-the International Journal on the Biology of Stress 22, 60–69.

Romero-Gonzalez B, Caparros-Gonzalez RA, Gonzalez-Perez R, Coca-Arco S, Peralta-Ramirez MI, 2019. Hair cortisol levels, psychological stress and psychopathological symptoms prior to instrumental deliveries. Midwifery 77, 45–52. [PubMed: 31254963]

Romero-Gonzalez B, Caparros-Gonzalez RA, Gonzalez-Perez R, Delgado-Puertas P, Peralta-Ramirez MI, 2018. Newborn infants' hair cortisol levels reflect chronic maternal stress during pregnancy. PloS one 13, e0200279. [PubMed: 29979751]

Romero-Gonzalez B, Puertas-Gonzalez JA, Strivens-Vilchez H, Gonzalez-Perez R, Peralta-Ramirez MI, 2020. Effects of cognitive-behavioural therapy for stress management on stress and hair cortisol levels in pregnant women: A randomised controlled trial. J Psychosom Res 135, 110162. [PubMed: 32485622]

Scharlau F, Pietzner D, Vogel M, Gaudl A, Ceglarek U, Thiery J, Kratzsch J, Hiemisch A, Kiess W, 2018. Evaluation of hair cortisol and cortisone change during pregnancy and the association with self-reported depression, somatization, and stress symptoms. Stress (Amsterdam, Netherlands) 21, 43–50.

Schonblum A, Arnon L, Ravid E, Salzer L, Hadar E, Meizner I, Wiznitzer A, Weller A, Koren L, 2018. Can hair steroids predict pregnancy longevity? Reproductive biology 18, 410–415. [PubMed: 30219333]

Schreier HM, Enlow MB, Ritz T, Coull BA, Gennings C, Wright RO, Wright RJ, 2016. Lifetime exposure to traumatic and other stressful life events and hair cortisol in a multi-racial/ethnic sample of pregnant women. Stress (Amsterdam, Netherlands) 19, 45–52.

Schreier HM, Enlow MB, Ritz T, Gennings C, Wright RJ, 2015. Childhood abuse is associated with increased hair cortisol levels among urban pregnant women. J Epidemiol Community Health 69, 1169–1174. [PubMed: 26219886]

Schury K, Koenig AM, Isele D, Hulbert AL, Krause S, Umlauft M, Kolassa S, Ziegenhain U, Karabatsiakis A, Reister F, Guendel H, Fegert JM, Kolassa IT, 2017. Alterations of hair cortisol and dehydroepiandrosterone in mother-infant-dyads with maternal childhood maltreatment. BMC Psychiatry 17, 213. [PubMed: 28587668]

Short SJ, Stalder T, Marceau K, Entringer S, Moog NK, Shirtcliff EA, Wadhwa PD, Buss C, 2016. Correspondence between hair cortisol concentrations and 30-day integrated daily salivary and weekly urinary cortisol measures. Psychoneuroendocrinology 71, 12–18. [PubMed: 27235635]

Smy L, Shaw K, Amstutz U, Smith A, Berger H, Carleton B, Koren G, 2016. Hair cortisol as a hypothalamic-pituitary-adrenal axis biomarker in pregnant women with asthma: a retrospective observational study. BMC pregnancy and childbirth 16, 176. [PubMed: 27440139]

Stalder T, Kirschbaum C, 2012. Analysis of cortisol in hair – State of the art and future directions. Brain, behavior, and immunity 26, 1019–1029.

Stalder T, Steudte-Schmiedgen S, Alexander N, Klucken T, Vater A, Wichmann S, Kirschbaum C, Miller R, 2017. Stress-related and basic determinants of hair cortisol in humans: A meta-analysis. Psychoneuroendocrinology 77, 261–274. [PubMed: 28135674]

Stickel S, Eickhoff S, Goecke TW, Schneider F, Quinete NS, Lang J, Habel U, Chechko N, 2019. Cumulative cortisol exposure in the third trimester correlates with postpartum mothers' neural response to emotional interference. Biological psychology 143, 53–61. [PubMed: 30797948]

Swales DA, Stout-Oswald SA, Glynn LM, Sandman C, Wing DA, Davis EP, 2018. Exposure to traumatic events in childhood predicts cortisol production among high risk pregnant women. Biological psychology 139, 186–192. [PubMed: 30359722]

Tuschy B, Berlit S, Stutzer P, Lis S, Schmahl C, Baumgartner U, Sutterlin M, 2018. Evaluation of psychosocial and biological parameters in women seeking for a caesarean section and women who are aiming for vaginal delivery: a cross-sectional study. Archives of gynecology and obstetrics 297, 897–905. [PubMed: 29344846]

van der Voorn B, Hollanders JJ, Kieviet N, Dolman KM, de Rijke YB, van Rossum EFC, Rotteveel J, Honig A, Finken MJJ, 2018. Maternal Stress During Pregnancy Is Associated with Decreased Cortisol and Cortisone Levels in Neonatal Hair. Hormone research in paediatrics 90, 299–307. [PubMed: 30541006]

Wachman EM, Hunter RG, Shrestha H, Lapp HE, Meyer J, Alvarez CD, Tronick E, 2019. Maternal hair cortisol levels as a novel predictor of neonatal abstinence syndrome severity: A pilot feasibility study. Developmental psychobiology.

Wennig R, 2000. Potential problems with the interpretation of hair analysis results. Forensic science international, 107(1–3), 5–12. [PubMed: 10689559]

Wikenius E, Moe V, Kjellevold M, Smith L, Lyle R, Waagbo R, Page CM, Myhre AM, 2016. The Association between Hair Cortisol and Self-Reported Symptoms of Depression in Pregnant Women. PloS one 11, e0161804. [PubMed: 27584584]

Zhang H, Liu S, Si Y, Zhang S, Tian Y, Liu Y, Li H, Zhu Z, 2019. Natural sunlight plus vitamin D supplementation ameliorate delayed early motor development in newborn infants from maternal perinatal depression. Journal of Affective Disorders 257, 241–249. [PubMed: 31301627]
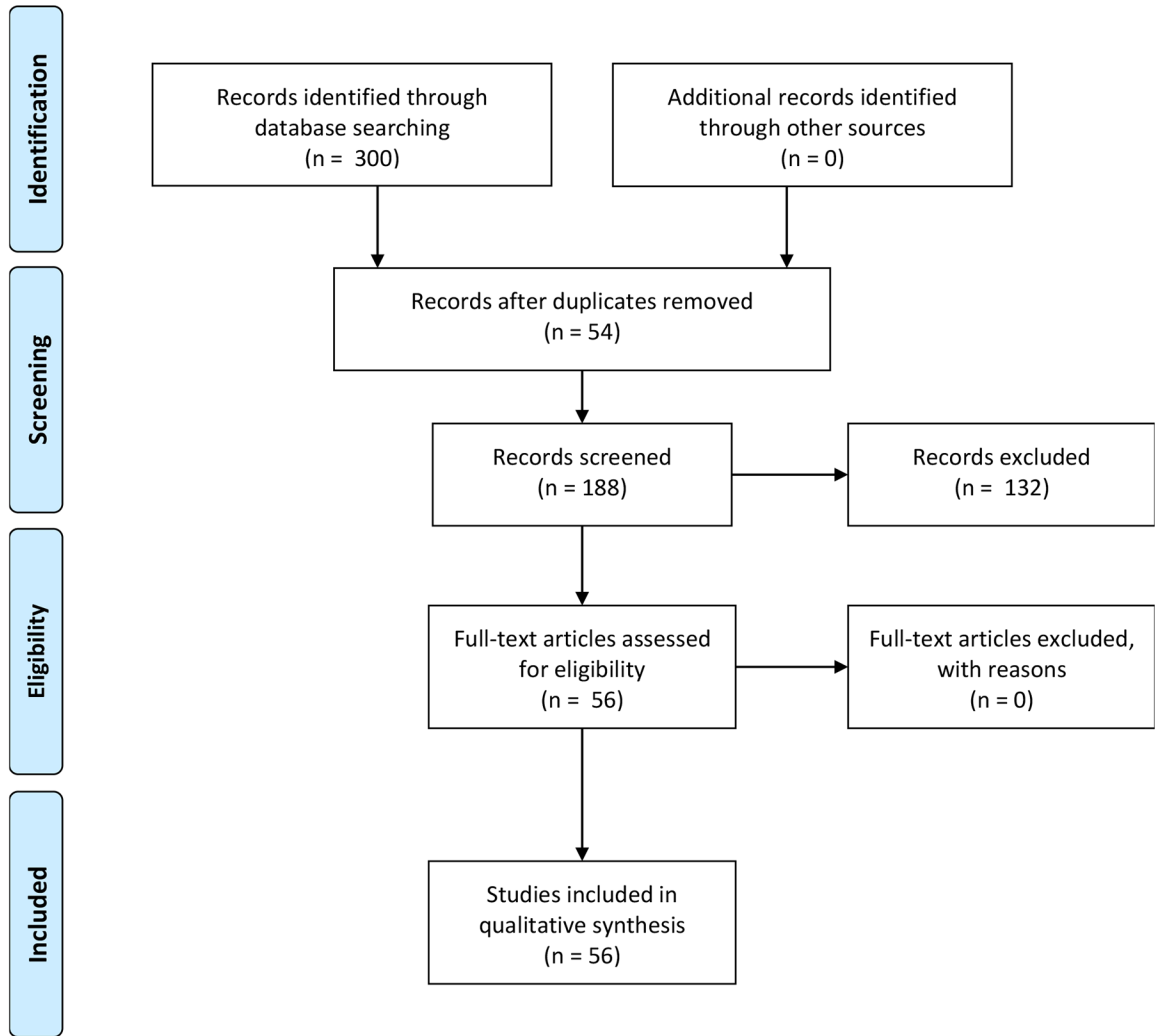
Figure 1.
PRISMA Flow Diagram
Note. Searches included PubMed: All Fields: ((hair cortisol) AND (pregnan* OR
prenatal)) Filters: Human; Scopus: (TITLE-ABS-KEY ("hair cortisol") AND TITLE-ABS-
KEY (pregnan* OR prenatal)); and Web of Science: TOPIC: (((hair cortisol) AND
(pregnan* OR prenatal))) Refined by: DOCUMENT TYPES: (ARTICLE OR REVIEW OR
PROCEEDINGS PAPER) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI,
CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

Table 1.

Included studies, with participant sample size and basic hair sample collection details.

Reference	N		Timing of Collection	# Samples	# Segments	Segment Length	Period Captured	Collection Strategy
Aatsinki et al., 2020	N = 115 with data		T2	1	1	5cm	T1+T2	Prospective
Bowers et al., 2018	N = 30 with data		In mid-pregnancy	1	1	3cm	T3	
Braig et al., 2015	N = 768	Same Sample	Within 3 days after delivery	1	1	3cm	T3	Prospective
Braig et al., 2016	N = 768							
Braig et al., 2017	N = 626							
Braig et al., 2020	N = 748							
Caparros-Gonzalez et al., 2017 [1]	N = 44		T1, T2, T3	3	1 each	3cm	each trimester	Prospective
Caparros-Gonzalez et al., 2019a [1]	N = 60		T3	1	1	3cm	T3	
Caparros-Gonzalez et al., 2019b [1]	N = 41		T1, T2, T3, Postnatal	4	1 each	3cm	each trimester	Prospective
Caparros-Gonzalez et al., 2019c [1]	N = 91		T1, T2, T3, Postnatal	4	1 each	3cm	each trimester	Prospective
Conradt et al., 2020	N = 137		Early T3	1	1 each	6 cm	T1+T2	Prospective
D'Anna-Hernandez et al., 2011	N = 21		T1, T2, T3, postpartum, an additional full length sample was cut at T3	5	1 each	3cm	each trimester	Prospective
Duffy et al., 2018	N = 52		After delivery	1	3	3cm	each trimester	Retrospective via segments
Flom et al., 2018	N = 314		Within 1 week of delivery	1	1	3cm	T3	Retrospective via segments
Galbally et al., 2019	N = 241		Delivery	1	3	1cm	T3	Retrospective via segments
Garcia-Leon et al., 2018 [1]	N = 62		T1, T2, T3	3	1 each	3cm	each trimester	Prospective
Garcia-Leon et al., 2019 [1]	N = 151		T3, postnatal	2	1 each	3cm	T3	Prospective
Gelaye et al., 2019	N = 137		T1 (< 16 weeks)	1	3	3cm	T1	Prospective
Hoffman et al., 2016	N = 90		T1, T2, T3	3	1 each	3cm	each trimester	Prospective
Jahangard et al., 2019	N = 48 with and 50 without PPD		12 weeks postpartum	1	2	3cm	T3	Retrospective via segments
Kalra et al., (2007)	N = 25		Before 2nd trimester	1	1	1–1.5cm	T1	Prospective
Karakash et al., 2016	N = 29 who had preterm birth + N = 29 who delivered at term		Within 72 hours of preterm birth for term infants; same	1	3–9	3cm	T2 & T3	Retrospective via segments

Reference	N	Timing of Collection	# Samples	# Segments	Segment Length	Period Captured	Collection Strategy
		gestational age for term controls					
Karlen et al., 2013	N = 100	At the time of child delivery	1 (9cm)	3	3cm	each trimester	Retrospective via segments
Khoury et al., 2020	N = 51	T3	1	1	4cm	T2 - T3	
Kirschbaum et al., 2009	N = 103 mothers of newborn + N = 19 mothers of 3–9mo infants, + N = 20 control women		1	2–4	3cm	each trimester	Retrospective via segments
Koenig et al., 2018	N = 474	Within one week after delivery	1	1	3cm	T3	Retrospective
Kramer et al., 2009	N = 117 with sufficient maternal hair for cortisol analyses		1	1	9cm	overall	Retrospective
Krumbholz et al., 2013	N = 1	From 6th week of pregnancy, and collected monthly after that till the 9th week after delivery	11	4–5 each	2cm	each trimester	Prospective
Krumbholz et al., 2018	N = 1	1st samples before pregnancy, 8 samples during pregnancy and 5 samples postpartum	14	5 each	2cm	each trimester	Prospective
Lobmaier et al., 2020	N = 110	On the day of parturition	1	1	3cm	T3	
Mesner et al., 2019	N = 744	T1 or T2 (12–21 weeks)	1	1	3cm		Prospective
Musana et al., 2019	N = 133	T2 (22–28 weeks)	1	1	3cm	T2	
Mustonen et al., 2019	N = 689	T2 or 1–3 days after birth	1 or 2	1 each	5cm	T2 and/or delivery	Both
Kajanoja et al., 2020	N = 130	1–3 days after birth	1	1	5cm	Delivery	
Nystrom-Hansen et al., 2019	N = 40 with mental disorders; N = 14 controls	3rd trimester, 4 months postpartum	2	1 each	4cm	overall	Prospective
Orta et al., 2018	N = 97	At enrollment (M=13.1 week) and at full-term delivery (9cm)	2	1 each	3cm	T1	Comparison: retrospective & prospective
Orta et al., 2019	N = 97		2	1 for sample 1, 2 for sample 2	3cm	each trimester	T1 and T3 prospective, T2 retrospective
Romero-Gonzalez et al., 2019 [1]	N = 88	After delivery	1	1	3cm	T3	
Romero-Gonzalez et al., 2018 [1]	N = 80	T1, T2, T3, Postpartum	4	1 each	3cm	Each trimester	Prospective

Annotations: Mustonen et al., 2019 / Kajanoja et al., 2020 — "Same Study, different N"; Orta et al., 2018 / Orta et al., 2019 — "Same Sample"

Reference	N	Timing of Collection	# Samples	# Segments	Segment Length	Period Captured	Collection Strategy
Romero-Gonzalez et al., 2020 [1]	N = 78	T1 or T2 (12–28wks)	1	1	3cm	T1 or T2	Prospective
Scharlau, 2018	N = 62	2nd and 3rd trimester	2	1 each	1cm	T2 & T3	Prospective
Schonblum et al., 2018	N = 125	"completed T1", presumed T2	1	2	3cm	T1	Prospective
Schreier et al., 2015	N = 180 (Same Sample)	Near the time of delivery	1	3	3cm	each trimester	retrospective via segments
Schreier et al., 2016	N = 180 (Same Sample)						
Bosquet Enlow et al., 2017	N = 194 with data (Same Study, Different N)	Within 1 week of delivery	1	used T3 segment only	3cm	T3	
Bosquet Enlow et al., 2019a	N = 93 subset from main study	Within 1 week of delivery	1	3	3cm	each trimester	retrospective via segments
Bosquet Enlow et al., 2019b	N = 343	Within 1 week of delivery	1	3	3cm	each trimester	retrospective via segments
Schury et al, 2017	N = 94	within 6 days after delivery	1	1	3cm	T3	retrospective via segments
Smy et al, 2016	N= 31 controls; N=31 no chronic inhaled corticosteroid; N=56 treated	between T1 and 6mo postpartum	1	Up to 4	2–3.6 cm (majority = 3 cm long)	each trimester	retrospective via segments
Stickel et al., 2019	N = 76	Within 1–6 days of childbirth	1	1	3cm	T3	
Swales et al., 2018	N = 90 predominantly Hispanic (68.9%)	assessed once during pregnancy (M=28.6, SD=3.2)	1	1	3cm	T2	Prospective
Tuschy et al., 2018	N = 93 study; N = 109 control	Between 34 and 37 completed weeks of pregnancy at delivery	1	1	3cm	T3	
van der Voorn et al., 2018	N = 172	On the first day postpartum	1	1	1cm (6–1.4cm)	T3	
Wachman et al., 2019	N = 70 postpartum women with opioid use disorder	Within 72 hr of delivery	1	1	3cm	T3	
Wikenius et al., 2016	N = 181	Varying gestational ages at sample (17–32)	1	1	1–3cm	overall	
Zhang et al., 2019	N = 120 healthy women N = 229 depressed women	At 2 days, one-month, two-month postpartum	3	1 each	3cm	T3	

Note. T1 = Trimester 1; T2 = Trimester 2; T3 = Trimester 3; Prospective/retrospective assessment only coded for T1 and T2 assessments. N refers to the included sample size. Timing of collection is the reported time relative to pregnancy that the hair sample(s) were collected. # Samples refers to the number of distinct hair samples collected (and included in the analysis) across the duration of the study. # Segments refers to how many segments from any given hair sample were assayed and included in the analysis. Segment length refers to the length of hair used for analysis. Period captured is the reported timeframe during pregnancy that the HCC values are expected to apply to, based on the study's rationale. Collection strategy refers to whether repeated measures/multiple periods captured was accomplished via repeated sampling (i.e., a unique hair sample in each trimester) vs. multiple segmenting of the same hair sample (i.e., one sample with multiple segments retrospectively used to capture unique trimesters).