Nonadherence with ambulatory saliva sampling is associated with biased salivary testosterone estimates

Moeller, Julian; Lieb, Roselind; Meyer, Andrea H; Quack Lötscher, Katharina; Krastel, Bettina; Meinlschmidt, Gunther

Abstract: OBJECTIVE Nonadherence with scheduled saliva sampling, as encountered in ambulatory settings, can bias the estimation of salivary cortisol concentrations. This study is the first to estimate if such nonadherence is also associated with biased salivary testosterone concentration estimates. METHODS Using a standard ambulatory saliva-sampling protocol, we instructed pregnant women to collect saliva samples on two consecutive days at awakening, 1100h, 1500h, 2000h, and 2200h. We estimated testosterone concentrations in the saliva samples and participants' actual sampling times with an electronic medication event-monitoring system. We classified a saliva sample as adherent if it was sampled within a specific time window relative to its scheduled sampling time. We used a mixed-model analysis to distinguish between trait (number of adherent saliva samples per participant) and state (adherence status of a specific sample) adherence. RESULTS We included 60 pregnant women in this study. Seventy-five percent (448 of 600) of the scheduled samples indicated adherence with the sampling schedule. Participants' trait adherence was associated with their diurnal profiles of salivary testosterone estimates; that is, adherent participants had higher salivary testosterone estimates compared with nonadherent participants, $F(1,58)=5.41$, $p=0.023$, Cohen's $d=0.67$. The state adherence of a sample was associated with the salivary testosterone estimate of the related sample, $F(1,469)=4.48$, $p=0.035$, Cohen's $d=0.20$, with delayed sampling associated with lower salivary testosterone estimates. CONCLUSIONS The results suggest that common ambulatory nonadherence with scheduled saliva sampling is associated with biased salivary testosterone estimates. They will inform further studies estimating salivary testosterone with ambulatory saliva-sampling designs and highlight the relevance of strategies to improve or confirm adherence, beyond routinely used instructions.

DOI: https://doi.org/10.1016/j.psyneuen.2014.02.012

Nonadherence with ambulatory saliva sampling is associated with biased salivary testosterone estimates

Julian Moeller, M.Sc.[1,2]; Roselind Lieb, Ph.D.[1]; Andrea H. Meyer, Ph.D.[1]; Katharina Quack Loetscher, M.D.[3]; Bettina Krastel, Ph.D.[4]; Gunther Meinlschmidt, Ph.D.[1,4,5,*]

[1] University of Basel, Department of Psychology, Division of Clinical Psychology and Epidemiology, Basel, Switzerland

[2] Diagnostic and Crisis Intervention Centre, Department of Psychiatry (UPK), University of Basel, Basel, Switzerland

[3] Department of Obstetrics, University Hospital Zurich, Zurich, Switzerland

[4] National Centre of Competence in Research, Swiss Etiological Study of Adjustment and Mental Health (sesam), Basel, Switzerland

[5] Faculty of Medicine, Ruhr-University Bochum, Bochum, Germany

*Correspondence should be addressed to

Gunther Meinlschmidt, University of Basel, Missionsstrasse 60/62, CH-4055 Basel, Switzerland; phone: ++41 61 26 70271, fax: ++41 61 26 70659, e-mail: gunther.meinlschmidt@unibas.de

Running title: Salivary testosterone varies with saliva-sampling adherence

Word count: 3171

Number of tables: 1

Number of figures: 2

For submission to *Psychoneuroendocrinology*

Objective: Nonadherence with scheduled saliva sampling, as encountered in ambulatory settings, can bias the estimation of salivary cortisol concentrations. This study is the first to estimate if such nonadherence is also associated with biased salivary testosterone concentration estimates.

Methods: Using a standard ambulatory saliva-sampling protocol, we instructed pregnant women to collect saliva samples on two consecutive days at awakening, 1100h, 1500h, 2000h, and 2200h. We estimated testosterone concentrations in the saliva samples and participants' actual sampling times with an electronic medication event-monitoring system. We classified a saliva sample as adherent if it was sampled within a specific time window relative to its scheduled sampling time. We used a mixed-model analysis to distinguish between trait (number of adherent saliva samples per participant) and state (adherence status of a specific sample) adherence.

Results: We included 60 pregnant women in this study. Seventy-five percent **(**448 of 600) of the scheduled samples indicated adherence with the sampling schedule. Participants' trait adherence was associated with their diurnal profiles of salivary testosterone estimates; that is, adherent participants had higher salivary testosterone estimates compared with nonadherent participants, $F(1,58) = 5.41$, $p = 0.023$, Cohen's $d = 0.67$. The state adherence of a sample was associated with the salivary testosterone estimate of the related sample, $F(1,469) = 4.48$, $p = 0.035$, Cohen's $d = 0.20$, with delayed sampling associated with lower salivary testosterone estimates.

Conclusions: The results suggest that common ambulatory nonadherence with scheduled saliva sampling is associated with biased salivary testosterone estimates. They will inform further studies estimating salivary testosterone with ambulatory saliva-sampling designs and highlight the relevance of strategies to improve or confirm adherence, beyond routinely used instructions.

Salivary testosterone varies with saliva-sampling adherence

Keywords: Compliance, adherence, saliva, salivary, saliva sampling, testosterone, pregnancy, women

1. Introduction

Testosterone is a commonly used and well-established biomarker in psychoneuroendocrine research (Dabbs, 1993; Granger et al., 2004). Maternal testosterone concentrations during pregnancy have been associated with a range of pregnancy outcomes, such as offspring size at birth (Carlsen et al., 2006), offspring body weight (Carlsen et al., 2006; Gutnikova et al., 2010), placental weight (Lagiou et al., 2013), and sex-specific behavior of the offspring (Hines, 2006). Given the large number of studies examining testosterone in pregnant women, and to foster scientific and clinical progress in the field, applicable and accurate testosterone measurement seems to be fundamental.

Testosterone concentrations can be successfully analyzed in saliva (salivary testosterone, sT) based on ambulatory saliva sampling (Dabbs, 1993; Granger et al., 2004), for which participants are instructed to collect a series of saliva samples at scheduled sampling times. Particularly relevant for large-scale studies, this procedure is noninvasive, ecologically valid, and relatively inexpensive (Granger et al., 2004; Shiffman et al., 2008; Giltay et al., 2012; Kudielka et al., 2012). However, a disadvantage of ambulatory saliva sampling is that nonadherence with scheduled sampling times is common and often not self-reported by participants—as observed in prior studies using covert electronic adherence-monitoring systems (Kudielka et al., 2003; Broderick et al., 2004; Jacobs et al., 2005; Moeller et al., 2014). Due to the circadian decline of sT concentrations over the course of the day (Dabbs, 1990), nonadherence with scheduled saliva sampling may bias the estimation of sT concentrations and hence cause unreliable and invalid sT data.

Such a pattern has been observed in prior studies estimating the association between ambulatory saliva-sampling nonadherence and salivary cortisol, another well-known and frequently used biomarker with a circadian rhythm (Kudielka et al., 2003; Broderick et al., 2004; Kudielka et al., 2007; Moeller et al., 2014). However, to our knowledge, studies on the association between saliva-sampling nonadherence and sT estimates are lacking. Hence, when

estimating sT with ambulatory saliva-sampling designs, the need for strategies to improve or confirm adherence is open to question. Such strategies have been successfully used in ambulatory salivary cortisol research (e.g. electronic adherence-monitoring systems; see Adam and Kumari, 2009; Granger et al., 2012; Kudielka et al., 2012; Moeller et al., 2014) but are associated with additional study costs and study burden for participants.

In sum, sT and salivary cortisol are both important and frequently assessed biomarkers in psychoneuroendocrine research. While several studies examined the association between ambulatory nonadherence with the sampling schedule and salivary cortisol estimates, there are, to our knowledge, no studies examining the association between such nonadherence and sT estimates. With a standard ambulatory saliva-sampling design and a sample of pregnant women, we sought to address this gap: First, we estimated whether the "trait adherence" (number of adherent saliva samples) of participants was associated with their diurnal profiles of sT estimates. Second, we estimated whether the "state adherence" of a specific saliva sample was associated with the sT concentration in the related sample. For this, we used electronic adherence-monitoring systems to assess participants' objective adherence with scheduled sampling.

2. Methods

2.1. Participants

Pregnant women were recruited at the outpatient service of the Department of Obstetrics, University Hospital Zurich, Switzerland, during their antenatal visits. Recruitment took place in the context of a previously published study (Moeller et al., 2014). We applied the following exclusion criteria: week of gestation <12 or >32, presence of diseases potentially affecting the neuroendocrine system, high-risk pregnancy, human immunodeficiency virus (HIV) infection, the use of hormone-containing medication, insufficient German language skills, and the absence of regular antenatal visits at the outpatient service. The present study was approved

by the ethics committees of Zurich and Basel and conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent.

2.2. Procedures

Within an elaborated standard ambulatory saliva-sampling design, including 10 scheduled saliva samples divided over two consecutive days (awakening, and at 1100h, 1500h, 2000h, and 2200h), participants received standardized information for accurate saliva sampling. This information included restrictions (e.g. not to eat or consume caffeine 1 h before each scheduled saliva sampling) to minimize distortions in sT estimates. Moreover, it emphasized the importance of high adherence with scheduled sampling times. Participants were instructed to use straws and 2.0-mL safe-lock tubes (Eppendorf, Hamburg, Germany) to collect scheduled saliva samples. We advised them to place the tubes (Wiegand, Buelach, Switzerland), each prelabeled with scheduled sampling time, into small medicine containers directly after saliva sampling and to self-report exact sampling times with a paper-and-pencil diary. Participants' objective adherence with scheduled sampling was estimated with a covert Medication Event Monitoring System (MEMS 6 TrackCap Monitor, Aardex Ltd., Switzerland), which was fitted to the medicine containers. Participants were advised to store the medicine container containing saliva samples in a refrigerator. After the samples were returned to us, we froze them at –20°C until biochemical analysis. It should be noted that participants were also instructed to collect saliva samples 30, 45, and 60 min after awakening (see Moeller et al., 2014). These samples were intended to estimate circadian awakening responses in biomarkers if required (e.g. cortisol awakening response, Pruessner et al., 1997). There is no such response for sT. Hence, these samples were not considered in this study.

2.3. Measures

2.3.1. Adherence with saliva sampling

We used objective adherence information to estimate participants' adherence with scheduled sampling, measured with the MEMS 6 caps that time stamped each opening of the medical container. The MEMS 6 data were processed with PoverView (Aardex Ltd., Switzerland). The awakening samples were classified as adherent if sampled within ±10 min of the self-reported wake-up time, and the 1100h, 1500h, 2000h, and 2200h samples as adherent if sampled within ±1 h of the scheduled time. These adherence criteria were adapted from prior research on salivary cortisol and adherence (Kudielka et al., 2003; Moeller et al., 2014). Delivered saliva samples with missing MEMS 6 time stamps were classified as nonadherent.

2.3.2. Testosterone concentrations

We centrifuged thawed samples at 3000 g for 10 min and analyzed sT using a commercial enzyme immunoassay for human saliva (Testosterone ELISA, IBL, Hamburg, Germany). Analytical assay sensitivity was 2 pg/ml. The intra- and interassay coefficients of variation were ≤ 15.1% and ≤ 6.0%, respectively.

2.3.3. Demographic and other descriptive information

Participants provided demographic and other descriptive information, including age, employment status, body weight, gestational age of their fetus at saliva sampling, parity, and number of cigarettes smoked on sampling days, via questionnaire.

2.4. Statistical analysis

We checked the data for level-one (within-subject) and level-two (between-subjects) outliers (Nieuwenhuis et al., 2012) and distribution properties. We used a random coefficient model, a type of linear mixed model (Singer and Willett, 2003), to estimate the association between adherence with scheduled sampling and sT concentrations. This model contained a random intercept and a random slope parameter when this improved model fit (based on Akaike's

Information Criterion, AIC; Singer and Willett, 2003). Crucial for this study, this model allowed us to separately address trait and state adherence.

Trait adherence relates to the overall number of adherent saliva samples of a participant and measuring it allowed us to estimate the association between participants' trait adherence and their diurnal profiles of sT concentrations. We estimated this association by using the time-invariant predictor "trait adherence," categorizing the participants into an adherent (>8 adherent samples of the total of 10 scheduled samples) and a nonadherent (≤8 adherent samples) group. This 80% cutoff was chosen because studies in medical settings usually classify patients with adherence rates of 80% as adherent (Ho et al., 2009). We used a categorical predictor rather than the continuous trait adherence predictor "number of adherent samples" because this improved model fit in preliminary analyses. To account for linear trends in sT concentrations over the course of the day, we also included time of saliva sampling as an additional time-varying predictor. We also accounted for the sampling day, but this effect was negligible in all cases and is therefore not mentioned further.

State adherence relates to the adherence status of saliva samples and measuring it allowed us to estimate whether a deviation from a scheduled sampling time was associated with the sT concentration in the related sample. We estimated this association by entering the time-varying predictor "state adherence" (deviations from scheduled sampling times in minutes) into the model.

We used Cohen's *d* (Cohen, 1977) to estimate model-based effect sizes, based on *t* values and degrees of freedom. We also ran an adjusted model, that is, the same model described above but including a priori selected potential time-invariant predictors of testosterone as covariates: age (Granger et al., 2004), body weight (Sowers et al., 2001), number of cigarettes smoked on sampling days (Sowers et al., 2001; Toriola et al., 2011), gestational age at saliva sampling (Bammann et al., 1980), and parity (Toriola et al., 2011). Notably, parameter estimates based on the adjusted model were comparable to those based on

the unadjusted model. Therefore, and because some of the covariates contained missing values, thus increasing risk of bias and reducing statistical power, we decided to report the results of the unadjusted model only. Moreover, after including employment status (Purifoy and Koopmans, 1979) in secondary analyses as an additional covariate in the adjusted model, parameter estimates were still comparable (data of the adjusted models are available on request).

"Deviations from scheduled sampling times in minutes" and sT data were log transformed to approximate normal distributions. An alpha level of 0.05 indicated statistical significance. We carried out the data analysis using IBM SPSS Statistics 20 for Mac OS X.

3. Results

Sixty-nine pregnant women participated in this ambulatory saliva-sampling study. We excluded two participants because of a MEMS 6 cap defect, two because of saliva sampling without using the MEMS 6 caps, and one because of prematurely delivering during the study. Moreover, four participants were eliminated from the statistical model because of outliers in their sT estimates: one because of several level-one outliers and the remaining three because of one level-one outlier and outliers in either the intercept or slope estimates (level-two coefficients). Thus, the final sample consisted of 60 pregnant women. At the beginning of the study, we informed 32 (53%) of these 60 participants about the electronic adherence-monitoring system and gave timers and alarm clocks to 28 (47%) of them so they could remind themselves at scheduled sampling times (see also Moeller et al., 2014). Demographic and other descriptive information is presented in Table 1.

-Insert Table 1 approximately here-

3.1. Adherence with the saliva-sampling protocol

We assessed adherence and sT estimates in 10 scheduled saliva samples of 60 participants. Four hundred and forty-eight (75%) of the total of 600 scheduled saliva samples indicated adherence with the sampling schedule (see adherence criteria described in section 2.3.1.).

Across all samples, mean deviations from scheduled sampling times were +18 min (standard deviation = 65.23 min), which indicates that the participants collected their saliva samples after rather than before the scheduled time. In Figure 1, we depicted for each sample the deviation from scheduled sampling time in minutes against the scheduled sampling time.

-Insert Figure 1 approximately here-

3.2. Association between trait adherence and sT estimates

Twenty-five participants (42%) indicated trait adherence with the sampling schedule (>8 adherent samples), and 35 (58%) indicated trait nonadherence (≤8 adherent samples). To estimate the association between trait adherence and sT concentrations, we compared sT concentrations between adherent and nonadherent participants. We found a main effect for trait adherence on sT concentrations, $F(1,58) = 5.41$, $p = 0.023$, Cohen's $d = 0.67$. As shown in Figure 2, adherent participants had higher diurnal sT concentrations compared with nonadherent participants.

-Insert Figure 2 approximately here-

3.3. Association between state adherence and sT estimates

This analysis included 535 saliva samples. The state adherence of a saliva sample was significantly associated with the sT concentration in the related sample, $F(1,469) = 4.48$, $p = 0.035$, Cohen's $d = 0.20$: the greater the time delay of a sample relative to its scheduled sampling time, the lower the sT concentration. Concentrations of sT thereby decreased per minute delay by a value of 0.91 on the natural logarithm scale [standard error 0.43, $t(469) = 2.12$, $p = 0.035$].

4. Discussion

In this study, pregnant women's sT concentrations varied with their adherence with an ambulatory sampling schedule: both trait and state nonadherence were associated with a biased estimation of sT concentrations, leading to biased results in the case of nonadherence. The finding extends results from prior studies that indicated that ambulatory nonadherence with scheduled sampling is associated with biased salivary cortisol estimates (Kudielka et al., 2003; Broderick et al., 2004; Kudielka et al., 2007; Moeller et al., 2014). With regard to cortisol, the steep variation in hormone concentration after awakening (i.e. the cortisol awakening response; Pruessner et al., 1997) has been considered mainly responsible for the bias introduced by nonadherence with the sampling schedule (Kudielka et al., 2003; Kudielka et al., 2012). Notably, unlike cortisol, sT shows no such steep hormone concentration variation (Dabbs, 1990), but still, nonadherence with scheduled sampling is associated with substantial bias in concentration estimates. In detail, the associations between trait and state adherence and sT concentration estimates indicated moderate and small effect sizes, respectively.

In this study, trait adherent participants displayed higher diurnal sT profiles compared with trait nonadherent participants. This might be because nonadherent participants collected their saliva samples on average with longer delays from the sampling schedule, which is—due to the diurnal decline of sT concentrations—likely associated with lower diurnal sT profiles. Accordingly, with regard to the state adherence of a specific sample, we found that the greater the time delay from a scheduled sampling time, the lower the sT concentration in the related sample. If the associations found between ambulatory nonadherence with the sampling schedule and sT estimates are causal, ambulatory nonadherence with the sampling schedule will result in a biased estimation of sT concentrations. Alternatively, nonadherence with the sampling schedule may be associated with specific characteristics of participants, which in

turn may be associated with lower sT estimates. However, we could not find evidence of such associated characteristics when adjusting our analyses for potential confounding variables.

Taken together, our findings indicate that nonadherence with scheduled sampling, as encountered in ambulatory settings, may lead on average to decreased and hence biased sT estimates. This may lead in turn to risk of misinterpretations, as illustrated by the following example: Patients with anxiety disorders may display decreased sT concentrations compared to healthy controls (e.g. Giltay et al., 2012). Without the option to confirm adherence, it might be difficult to conclude whether decreased sT estimates are directly associated with an anxiety disorder or rather with a bias introduced by nonadherence with the sampling schedule related to an anxiety disorder (cf. Kudielka et al., 2003; Moeller et al., 2014). Our findings underline that to reduce risk of bias when estimating sT with ambulatory saliva-sampling designs, it is important to specifically address the risk of nonadherence with scheduled sampling when designing a study. Addressing the adherence issue may be relevant not only in ambulatory assessment of salivary cortisol (Kudielka et al., 2003; Broderick et al., 2004; Kudielka et al., 2007; Moeller et al., 2014) but, based on our findings, also when estimating sT in ambulatory settings. Notably, in our study, rates of nonadherence seemed to be severe enough to bias sT estimates, even though we informed participants about the importance of high adherence with the sampling schedule. For discussions of how adherence with scheduled ambulatory saliva sampling can be improved or confirmed, see Adam and Kumari (2009), Granger et al. (2012), and Kudielka et al. (2012). Moreover, it is important to note that factors other than nonadherence with scheduled sampling could also introduce biases in the estimation of sT concentrations: for example, not following storage temperature recommendations for saliva samples (Granger et al., 2004; Durdiakova et al., 2013), blood from micro-injuries in the oral mucosa that contaminates saliva samples (Kivlighan et al., 2004), and eating or drinking right before saliva sampling (Granger et al., 2012). When estimating sT within ambulatory saliva-sampling designs, recommendations for dealing with such factors should be consistently

followed as closely as possible (for reviews see Granger et al., 2004; Al-DujailI and Sharp, 2012; Granger et al., 2012).

This study has some limitations: first, ambulatory sT research often applies saliva-sampling designs with fewer scheduled samples per day than in our study (e.g. Hamilton and Meston, 2010), which may result in higher average adherence rates (Kudielka et al., 2003) than those found in our study and hence less bias in the sT estimates. However, multiple saliva samplings, as applied in the present study, are required to capture diurnal sT profiles (Al-DujailI and Sharp, 2012). Second, we specifically addressed adherence with the saliva-sampling schedule. Obviously, trait nonadherent participants may have also adhered less strictly—compared to trait adherent participants—to the predefined storage protocol and stored their saliva samples on both study days at room temperature and not as instructed in a refrigerator. However, even in such a case, we would not expect substantial bias in our data, as Durdiakova and colleagues (2013) suggested that storing saliva samples unrefrigerated for few days does not introduce bias in sT estimates. Yet, we cannot absolutely rule out other factors related to trait nonadherence that may have partly contributed to lower diurnal sT profiles in trait nonadherent participants. Third, as described in the Methods section, participants were instructed to collect three saliva samples in the morning (30, 45, and 60 min after awakening; see Moeller et al., 2014) that were not considered in this study. We cannot rule out that adherence with the subsequent sampling schedule was impacted by the sampling burden of these morning samples. This should be scrutinized in future studies. Last, our sample consisted of pregnant women and the results may not extend to other populations.

Despite the limitations, this study has important strengths, including an ambulatory saliva-sampling design covering multiple scheduled samples on two consecutive days (see Al-DujailI and Sharp, 2012). Another strength is that we used electronic adherence monitoring instead of self-report questionnaires to estimate participants' adherence with the sampling schedule. Prior studies showed that participants' self-reported adherence with scheduled

sampling might be inaccurate (Kudielka et al., 2003; Broderick et al., 2004; Jacobs et al., 2005). Furthermore, we applied a mixed-model analysis that is the method of choice for analyzing repeated ambulatory saliva data in which missing values are usually present (cf. Singer and Willett, 2003; Lane, 2008; Kudielka et al., 2012).

In this study, sT concentrations varied with ambulatory trait and state adherence with the sampling schedule. Adherent participants had higher sT estimates compared with nonadherent participants and delayed saliva sampling was associated with lower sT concentration estimates. To our knowledge, this study is the first to suggest that ambulatory nonadherence with scheduled sampling can introduce a bias in the estimation of sT concentrations. Average delayed saliva sampling appears to be associated with an underestimation of sT concentration. Our findings will inform further studies estimating sT with ambulatory saliva-sampling designs. They highlight the importance of efforts to improve or confirm adherence with scheduled ambulatory saliva sampling.

References

Adam, E.K., Kumari, M., 2009. Assessing salivary cortisol in large-scale, epidemiological research. Psychoneuroendocrinology 34, 1423-1436.

Al-DujailI, E.A.S., Sharp, M.A., 2012. Female salivary testosterone: measurement, challenges and applications, In: Ostojic, S.M. (Ed.), Steroids—From Physiology to Clinical Medicine. InTech, Rijeka, Croatia, pp. 129-167.

Bammann, B.L., Coulam, C.B., Jiang, N.S., 1980. Total and free testosterone during pregnancy. Am. J. Obstet. Gynecol. 137, 293-298.

Broderick, J.E., Arnold, D., Kudielka, B.M., Kirschbaum, C., 2004. Salivary cortisol sampling compliance: comparison of patients and healthy volunteers. Psychoneuroendocrinology 29, 636-650.

Carlsen, S.M., Jacobsen, G., Romundstad, P., 2006. Maternal testosterone levels during pregnancy are associated with offspring size at birth. Eur. J. Endocrinol. 155, 365-370.

Cohen, J., 1977. Statistical power analysis for the behavioral sciences, Rev. ed. Academic Press, New York.

Dabbs, J.M., 1990. Salivary testosterone measurements—reliability across hours, days, and weeks. Physiol. Behav. 48, 83-86.

Dabbs, J.M., 1993. Salivary testosterone measurements in behavioral studies. Ann. N. Y. Acad. Sci. 694, 177-183.

Durdiakova, J., Fabryova, H., Koborova, I., Ostatnikova, D., Celec, P., 2013. The effects of saliva collection, handling and storage on salivary testosterone measurement. Steroids 78, 1325-1331.

Giltay, E.J., Enter, D., Zitman, F.G., Penninx, B.W.J.H., van Pelt, J., Spinhoven, P., Roelofs, K., 2012. Salivary testosterone: associations with depression, anxiety disorders, and antidepressant use in a large cohort study. J. Psychosom. Res. 72, 205-213.

Granger, D.A., Fortunato, C.K., Beltzer, E.K., Virag, M., Bright, M.A., Out, D., 2012. Focus on methodology: salivary bioscience and research on adolescence: an integrated perspective. J. Adolesc. 35, 1081-1095.

Granger, D.A., Shirtcliff, E.A., Booth, A., Kivlighan, K.T., Schwartz, E.B., 2004. The "trouble" with salivary testosterone. Psychoneuroendocrinology 29, 1229-1240.

Gutnikova, L.V., Aleksandrova, A.A., Shkurat, T.P., 2010. Hormone status of pregnant women and newborn body weight. Reprod. Biomed. Online 20, Supplement 3, 94.

Hamilton, L.D., Meston, C.M., 2010. The effects of partner togetherness on salivary testosterone in women in long distance relationships. Horm. Behav. 57, 198-202.

Hines, M., 2006. Prenatal testosterone and gender-related behaviour. Eur. J. Endocrinol. 155, 115-121.

Ho, P.M., Bryson, C.L., Rumsfeld, J.S., 2009. Medication adherence: its importance in cardiovascular outcomes. Circulation 119, 3028-3035.

Jacobs, N., Nicolson, N.A., Derom, C., Delespaul, P., van Os, J., Myin-Germeys, I., 2005. Electronic monitoring of salivary cortisol sampling compliance in daily life. Life Sci. 76, 2431-2443.

Kivlighan, K.T., Granger, D.A., Schwartz, E.B., Nelson, V., Curran, M., Shirtcliff, E.A., 2004. Quantifying blood leakage into the oral mucosa and its effects on the measurement of cortisol, dehydroepiandrosterone, and testosterone in saliva. Horm. Behav. 46, 39-46.

Kudielka, B.M., Broderick, J.E., Kirschbaum, C., 2003. Compliance with saliva sampling protocols: electronic monitoring reveals invalid cortisol daytime profiles in noncompliant subjects. Psychosom. Med. 65, 313-319.

Kudielka, B.M., Gierens, A., Hellhammer, D.H., Wust, S., Schlotz, W., 2012. Salivary cortisol in ambulatory assessment—some dos, some don'ts, and some open questions. Psychosom. Med. 74, 418-431.

Kudielka, B.M., Hawkley, L.C., Adam, E.K., Cacioppo, J.T., 2007. Compliance with ambulatory saliva sampling in the Chicago Health, Aging, and Social Relations Study and associations with social support. Ann. Behav. Med. 34, 209-216.

Lagiou, P., Hsieh, C.C., Samoli, E., Lagiou, A., Xu, B., Yu, G.P., Onoyama, S., Chie, L., Vatten, L.J., Adami, H.O., Trichopoulos, D., Williams, M.A., 2013. Associations of placental weight with maternal and cord blood hormones. Ann. Epidemiol. 23, 669-673.

Lane, P., 2008. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. Pharm. Stat. 7, 93-106.

Moeller, J., Lieb, R., Meyer, A.H., Quack Loetscher, K., Krastel, B., Meinlschmidt, G., 2014. Improving ambulatory saliva-sampling compliance in pregnant women: a randomized controlled study. PLoS ONE 9, e86204.

Nieuwenhuis, R., te Grotenhuis, M., Pelzer, B., 2012. influence.ME: tools for detecting influential data in mixed effects models. R. J. 4, 38-47.

Pruessner, J.C., Wolf, O.T., Hellhammer, D.H., Buske-Kirschbaum, A., Auer, K.v., Jobst, S., Kaspers, F., Kirschbaum, C., 1997. Free cortisol levels after awakening: a reliable biological marker for the assessment of adrenocortical activity. Life Sci. 61, 2539-2549.

Purifoy, F.E., Koopmans, L.H., 1979. Androstenedione, testosterone, and free testosterone concentration in women of various occupations. Soc. Biol. 26, 179-188.

Shiffman, S., Stone, A.A., Hufford, M.R., 2008. Ecological momentary assessment. Annu. Rev. Clin. Psychol. 4, 1-32.

Singer, J.D., Willett, J.B., 2003. Applied Longitudinal Data Analysis. Oxford University Press, Oxford.

Sowers, M.F., Beebe, J.L., McConnell, D., Randolph, J., Jannausch, M., 2001. Testosterone concentrations in women aged 25-50 years: associations with lifestyle, body composition, and ovarian status. Am. J. Epidemiol. 153, 256-264.

Toriola, A.T., Vaarasmaki, M., Lehtinen, M., Zeleniuch-Jacquotte, A., Lundin, E., Rodgers, K.G., Lakso, H.A., Chen, T.H., Schock, H., Hallmans, G., Pukkala, E., Toniolo, P., Grankvist, K., Surcel, H.M., Lukanova, A., 2011. Determinants of maternal sex steroids during the first half of pregnancy. Obstet. Gynecol. 118, 1029-1036.

Figure captions

Figure 1. Deviation from scheduled saliva sampling in minutes at the five scheduled sampling times. Note: Negative and positive values on the *y* axis indicate delayed and premature saliva sampling, respectively, relative to the scheduled sampling time; +0, at awakening.

Figure 2. Salivary testosterone concentrations on a logarithmized scale stratified by trait adherent (>8 adherent samples of the total of 10 scheduled samples) and trait nonadherent (≤8 adherent samples) participants. Note: Salivary testosterone concentrations were averaged across two collection days representing estimated values from a linear mixed model. Error bars denote estimates of the standard error of the group mean; +0, at awakening.