SALIVARY BIOMARKERS IN PREGNANT WOMEN: CLINICAL OUTCOMES AND METHODOLOGICAL ASPECTS

Inauguraldissertation
zur
Erlangung der Würde
eines Doktors der Philosophie
vorgelegt der
Fakultät für Psychologie
der Universität Basel

von

Julian Möller

aus Münster, Deutschland

Basel, 2016

UNI
BASEL

Genehmigt von der Fakultät für Psychologie

auf Antrag von

Prof. Dr. Roselind Lieb
Prof. Dr. Gunther Meinlschmidt

Basel, den _____

_____
Prof. Dr. Roselind Lieb

UNI
BASEL

Erklärung

Hiermit erkläre ich, Julian Möller, dass ich die vorliegende kumulative Dissertation „Salivary biomarkers in pregnant women: clinical outcomes and methodological aspects“ ohne unzulässige Hilfe verfasst und jegliche Zusammenarbeit mit Dritten gekennzeichnet habe. Alle Hilfsmittel sowie alle Zitate wurden von mir gekennzeichnet.

Basel, den 29. Februar 2016

_____

Julian Möller

Danksagung

Zuallererst möchte ich mich herzlich bei Gunther Meinlschmidt und Roselind Lieb für die hervorragende Betreuung meiner Dissertation bedanken. Ihre konstruktive Unterstützung und ihre fortwährende Diskussions- und Hilfsbereitschaft waren entscheidend für das Gelingen dieser Arbeit. Roselind Lieb bin ich außerordentlich dankbar dafür, dass sie mich in ihre Arbeitsgruppe aufgenommen und in den letzten Jahren begleitet und gefördert hat. Insbesondere bin ich sehr glücklich darüber, dass sie es mir ermöglicht hat, meine psychotherapeutische Arbeit mit Forschung und Lehre zu verbinden. Gunther Meinlschmidt danke ich sehr herzlich für seine immerwährende Bereitschaft und Geduld, meine vielen Fragen zu beantworten.

Bei den Koautoren Andrea Meyer, Katharina Quack Loetscher und Bettina Krastel möchte ich mich für das Mitwirken an den dieser kumulativen Dissertation zugrundeliegenden Manuskripten bedanken. Andrea Meyer danke ich speziell für seine maßgebliche Unterstützung bei den statistischen Analysen. Ich danke auch Rolf-Dieter Stieglitz für die Übernahme des Vorsitzes meiner Promotionskommission und Anita Todd für das Korrigieren der Manuskripte.

In der Abteilung von Roselind Lieb durfte ich viele inspirierende Menschen kennenlernen, wie Andrew Gloster, Jutta Mata, Catherine Coste, Cornelia Witthauer, Lavinia Flückiger, Eva Unternährer, Yasemin Meral, Hanna Wersebe, Marcel Miché und viele mehr. Ich verdanke ihnen wichtige Anregungen und ein wunderbares Arbeitsklima.

Für das Korrekturlesen meiner Rahmenschrift danke ich sehr herzlich Kornelia Möller, Reinert und Petra Hiller, Wilhelm Möller und Melanie Krügel.

Meine Familie, meine Partnerin und meine Freunde wissen, dass ich ihnen für ihre Unterstützung in der Zeit des Verfassens meiner Dissertation äußerst dankbar bin. Kornelia, Reinert, Johannes, Gerd, Petra, Melanie, Dorle, Wilhelm, Philipp, Valerie, Victoria und Antoinette – vielen Dank, dass ihr immer für mich da wart und mich begleitet habt.

Inhaltsverzeichnis

Abkürzungen

CAR	Cortisolaufwachreaktion
HPA-Achse	Hypothalamus-Hypophysen-Nebennierenrinden-Achse
SGA	Small for Gestational Age
UN	United Nations
USZ	Universitätsspital Zürich

Zusammenfassung

Ziel: Die vorliegende Dissertation untersucht (1) methodische Aspekte einer möglichst reliablen ambulanten Erfassung von Speichel-Testosteron- und -Cortisolkonzentrationen bei Schwangeren sowie (2) die Assoziation zwischen dem Speichel-Testosteron-Tagesrhythmus Schwangerer und der Geburtsgröße des Nachwuchses.

Methode: In einer prospektiven Studie instruierten wir 75 Frauen, während ihrer Schwangerschaft Speichelproben nach einem ambulanten Standard-Protokoll zu sammeln. Wir erfassten die Adhärenz der Schwangeren bezüglich multipler vorgesehener Messzeitpunkte für die Speichelsammlung mit einem für die Schwangeren nicht ersichtlichen elektronischen Adhärenz-Monitoring. Ob eine Vorabinformation über dieses Monitoring und/oder das zur Verfügung-Stellen eines Zeitmessers mit Alarmfunktion die Adhärenz verbessert, untersuchten wir mit einem randomisierten kontrollierten Design. Wir bestimmten die Testosteron- und Cortisolkonzentrationen im Speichel und entnahmen die Geburtsgröße des Nachwuchses aus den Patientenakten.

Resultate: (1) Bei der Speichelsammlung kam eine Nicht-Adhärenz mit vorgesehenen Messzeitpunkten vor und war mit verminderten Speichel-Testosteron- und -Cortisolwerten assoziiert. Die Vorabinformation über ein Adhärenz-Monitoring, nicht aber das zur Verfügung-Stellen eines Zeitmessers, ging mit einer verbesserten Adhärenz einher. (2) Ein abgeflachter Testosteron-Tagesrhythmus war mit einer reduzierten Geburtsgröße assoziiert.

Diskussion: (1) Die bei einer ambulanten Speichelsammlung bei Schwangeren auftretende Nicht-Adhärenz kann mit einem Bias bei der Speichel-Testosteron- und -Cortisolbestimmung einhergehen. Für eine reliable Bestimmung der Konzentrationen im Speichel von Schwangeren scheint die Berücksichtigung der Adhärenz-Problematik bedeutsam zu sein. (2) Ein abgeflachter mütterlicher Testosteron-Tagesrhythmus während der Schwangerschaft könnte ein neuer, potentiell relevanter Prädiktor für ein reduziertes fetales Wachstum sein.

Abstract

***Objective*:** This dissertation examines in pregnant women (1) methodological aspects to reliably assess salivary testosterone and cortisol concentrations in an ambulatory setting and (2) the association between women's diurnal salivary testosterone change during pregnancy and offspring size at birth.

***Methods*:** In a prospective study, we instructed 75 pregnant women to collect multiple saliva samples according to a standard ambulatory saliva-sampling protocol. We assessed women's adherence to scheduled saliva sampling times with a hidden electronic adherence-monitoring system. Using a randomized controlled design, we estimated whether a disclosure intervention (informing women about the adherence monitoring) and a reminder intervention (use of acoustical reminders) improved adherence. We assessed testosterone and cortisol concentrations in the saliva samples and collected information on offspring size at birth from medical birth records.

***Results*:** (1) Overall, pregnant women indicated nonadherence with scheduled saliva sampling times that was associated with lower salivary testosterone and cortisol concentration estimates. The disclosure intervention, but not the reminder intervention, improved adherence to the sampling schedule. (2) Women's flattened diurnal testosterone change during pregnancy was associated with reduced offspring size at birth for gestational age.

***Conclusions*:** (1) Results suggest that in pregnant women, nonadherence to scheduled ambulatory saliva sampling is common and associated with biased estimates of salivary testosterone and cortisol concentrations. Therefore, it is important to address nonadherence when utilizing ambulatory assessments of salivary testosterone and cortisol concentrations in pregnant women. (2) Moreover, women's flattened diurnal testosterone change during pregnancy may be a new, potentially relevant predictor of reduced fetal growth. This finding may contribute potentially to a better risk assessment of fetal growth restrictions.

Einleitung

All children have an explicit „right to grow and develop to their full potential and live in conditions that enable them to attain the highest standard of health through the implementation of programmes that address the underlying determinants of health" (United Nations [UN], 2013)

Nach der *Erhebung zu Familien und Generationen 2013* (Bundesamt für Statistik, 2015) gehört zur Lebensplanung fast aller jungen Schweizer Bürger der Wunsch, Kinder zu haben. Höchste Priorität messen die Schweizer Eltern dabei der Kindesgesundheit und einem Aufwachsen der Kinder ohne einschneidende gesundheitliche Probleme bei (Millward Brown, 2014). Auch de jure ist die Gesundheit von Kindern ein hohes Gut: Die UN-Kinderrechtskonvention zählt gesundheitsförderliche Lebensbedingungen zu den expliziten Rechten von Kindern (UN, 2013). Darüber hinaus gilt die Kindesgesundheit auf gesellschaftlicher Ebene als wichtiges soziales und ökonomisches Kapital (Belli, Bustreo, & Preker, 2005; Bloom, Canning, & Sevilla, 2004; Mangiaterra, Mattero, & Dunkelberg, 2006; Suhrcke, McKee, & Rocco, 2005). Vor diesem Hintergrund ergibt sich für die Forschung die Aufgabe, einen Beitrag zum gesunden Aufwachsen von Kindern zu leisten.

Vorliegende Evidenzen weisen darauf hin, dass die gesundheitliche Entwicklung eines Kindes bereits substantiell während seines Heranwachsens im Mutterleib beeinflusst wird (Meinlschmidt & Tegethoff, 2015; Tegethoff, Greene, Olsen, Schaffner, & Meinlschmidt, 2011; Wadhwa, Buss, Entringer, & Swanson, 2009). Hierbei haben sich Befunde verdichtet, dass die Gesundheit des Nachwuchses mit seinem fetalen Wachstum assoziiert ist (siehe Reviews von Barker, 2007; Goldenberg & Culhane, 2007; Mayer & Joseph, 2013; Murray et al., 2015; Salam, Das, & Bhutta, 2014; Wadhwa et al., 2009). Wie von der internationalen Fachgesellschaft für *Developmental Origins of Health and Disease (DOHaD)* formuliert,

kann ein Zusammenhang zwischen einem reduzierten fetalen Wachstum und einer erhöhten Prädisposition für verschiedene substantielle kurz- und langfristige Gesundheitsprobleme des Nachwuchses angenommen werden (Barker, 2007; Barker, Eriksson, Forsen, & Osmond, 2002; Wadhwa et al., 2009).

Aufgrund negativer Gesundheitsfolgen für den Nachwuchs, relativ hoher Prävalenz und gleichzeitig nicht hinreichend geklärter Ätiologie gilt ein reduziertes fetales Wachstum als ein globales Gesundheitsproblem der heutigen Zeit (Barker, 1995; Goldenberg & Culhane, 2007; Murray et al., 2015). So wird von den National Institutes of Health (2003) in einem Positionspapier auf die Wichtigkeit hingewiesen, potentielle Mechanismen und Prädiktoren eines reduzierten fetalen Wachstums zu identifizieren mit dem Ziel, seine Prävention, Diagnostik und Therapie zu verbessern (siehe auch Hui & Challis, 2008).

Der Erfassung des Hormonhaushalts schwangerer Frauen wird das Potential beigemessen, zugrundeliegende biologische Mechanismen und Prädiktoren eines reduzierten fetalen Wachstums zu erforschen. Hormonkonzentrationen wie Testosteron- und Cortisolkonzentrationen könnten entsprechende potentielle Prädiktoren für ein reduziertes fetales Wachstum sein (Entringer, Buss, Andersen, Chicz-DeMet, & Wadhwa, 2011; Kivlighan, DiPietro, Costigan, & Laudenslager, 2008; Manikkam et al., 2004; Voegtline, Costigan, Kivlighan, Henderson, & DiPietro, 2013).

Als Methode der Wahl zur Erfassung von Testosteron- und Cortisolkonzentrationen gilt die nicht-invasive Bestimmung in ambulant gesammelten Speichelproben. Standard-Protokollen zur ambulanten Erhebung von Speichelproben kommt in entsprechenden Forschungsdesigns daher große Bedeutung zu (Al-Dujaili & Sharp, 2012; Granger et al., 2012; Hellhammer, Wust, & Kudielka, 2009; Kirschbaum & Hellhammer, 1994; Kudielka, Gierens, Hellhammer, Wust, & Schlotz, 2012; Malamud, 2011; Streckfus & Bigler, 2002).

Im Zusammenhang mit der ambulanten Bestimmung von Speichel-Testosteron- und -Cortisolkonzentrationen liegen jedoch potentielle methodische Probleme vor, die mit einem

Bias einhergehen und zu fehlerhaften Interpretationen dieser biologischen Parameter führen können (für eine Übersicht siehe Al-Dujaili & Sharp, 2012; Granger et al., 2012; Granger et al., 2007a; Granger, Shirtcliff, Booth, Kivlighan, & Schwartz, 2004; Jessop & Turner-Cobb, 2008; Kudielka et al., 2012; Stalder et al., 2016). Viele Studien haben sich aus diesem Grund – trotz des erheblichen Forschungsbedarfs in diesem Feld – gegen eine Integration der Erhebung von Speichel-Hormonen in ihr Studiendesign entschieden (Granger et al., 2007a).

Es lässt sich zusammenfassen, dass ein verbessertes Verständnis der dem fetalen Wachstums zugrundeliegenden biologischen Mechanismen und Prädiktoren ein zentrales Forschungsanliegen ist. Forschung, die den Hormonhaushalt von Frauen während ihrer Schwangerschaft untersucht, könnte zu einem verbesserten Verständnis beitragen, wobei eine möglichst reliable Bestimmung der Hormonkonzentrationen in ambulanten Speichelproben Schwangerer einen wissenschaftlichen und klinischen Fortschritt im Feld unterstützen könnte.

Daraus abgeleitet verfolgt die vorliegende Dissertation zwei übergeordnete Ziele im Bereich der Forschung zu schwangeren Frauen. Einerseits sollen methodische Fragen geklärt werden mit dem Ziel, zu einer möglichst reliablen Erfassung der Speichel-Testosteron- und -Cortisolkonzentrationen bei schwangeren Frauen beizutragen (siehe Manuskripte A und B, vgl. Appendix). Andererseits soll unter Berücksichtigung der in den Manuskripten A und B berichteten Befunde der Frage nachgegangen werden, ob mütterliche Speichel-Testosteronkonzentrationen in der Schwangerschaft ein potentieller Prädiktor für das fetale Wachstum sind (siehe Manuskript C, vgl. Appendix).

Die vorliegende Dissertation ist wie folgt gegliedert: Das Kapitel *Theoretischer Hintergrund* gibt einen Überblick über die dieser Dissertation zugrundeliegenden theoretischen Ansätze und Befunde. Die daraus abgeleiteten Forschungsfragen werden im Kapitel *Forschungsfragen* den Manuskripten A-C zugeordnet. Methodik und Ergebnisse der den drei Manuskripten zugrundeliegenden Studie sind in den Kapiteln *Methode* und *Ergebnisse* zusammengefasst. Im Kapitel *Diskussion* werden die Kongruenz der Befunde mit

der bestehenden Forschung und die Implikationen der Befunde erörtert.[1] Die vorliegende Rahmenschrift endet mit einer Zusammenfassung der Stärken und Limitationen sowie einem Ausblick und Fazit.

Theoretischer Hintergrund

Fetales Wachstum und die Gesundheit des Nachwuchses

Das fetale Wachstum gilt als reduziert, wenn der Fetus sein ihm innewohnendes Wachstumspotential nicht erreicht. Eine reduzierte Geburtsgröße[2] im Verhältnis zum Gestationsalter gilt als Indikator für ein reduziertes fetales Wachstum. Aus klinischer Perspektive wird ein Neugeborenes beispielsweise als Small for Gestational Age (SGA) diagnostiziert, wenn Gewicht bzw. Körperlänge des Neugeborenen bei der Geburt unter dem 10. Perzentil des für das jeweilige Gestationsalter zu erwartenden Wertes liegt. Nach einer anderen Definition liegt SGA vor, wenn das Gewicht und/oder die Körperlänge bei der Geburt im Verhältnis zum Gestationsalter mindestens zwei Standardabweichungen unterhalb der bekannten Medianwerte liegen (Bamfo & Odibo, 2011; Chatelain, 2000; Mayer & Joseph, 2013; Zhang, Merialdi, Platt, & Kramer, 2010).

Vor über 20 Jahren zeigte eine Arbeitsgruppe um David Barker, dass Neugeborene mit einem reduzierten Geburtsgewicht ein erhöhtes Risiko aufweisen, in ihrem Leben an einer Herzerkrankung zu sterben (Barker, 1995; Barker, Osmond, Simmonds, & Wield, 1993; Barker, Winter, Osmond, Margetts, & Simmonds, 1989). Barker und andere Autoren postulierten in den folgenden Jahren in der *fetal origins hypothesis* einen Zusammenhang zwischen dem fetalen Wachstum und der Gesundheit des Nachwuchses (Barker, 1995, 2007; Wadhwa et al., 2009). Inzwischen zeigt eine Vielzahl von Studien, dass Indikatoren eines reduzierten fetalen Wachstums mit einer Reihe von kurz- und langfristigen negativen

[1] Eine weiterführende Diskussion der einzelnen Befunde ist den jeweiligen Manuskripten A-C (vgl. Appendix) zu entnehmen.
[2] In der vorliegenden Dissertation wird der Begriff *Geburtsgröße* äquivalent zum englischen Begriff *Offspring size at birth* verwendet. Die Geburtsgröße kann sich sowohl auf das Gewicht, die Körperlänge als auch auf den Kopfumfang des Nachwuchses bei der Geburt beziehen.

Gesundheitsfolgen für den Nachwuchs assoziiert sind (siehe Reviews von Barker, 2007; Murray et al., 2015; Salam et al., 2014) wie beispielsweise mit Diabetes (Lithell et al., 1996), koronaren Herzkrankheiten (Barker, 1995), Nierenfunktionsstörung (Chan, Morris, Leslie, Kelly, & Gallery, 2010), erhöhtem Blutdruck (Bonamy, Norman, & Kaijser, 2008; Chan et al., 2010; Cheung, Wong, Lam, & Tsoi, 2004; Law et al., 1993), arterieller Gefäßsteifigkeit (Chan et al., 2010; Cheung et al., 2004), neurologischen Entwicklungsstörungen (Murray et al., 2015), Depressionen (de Mola, de Franca, Quevedo, & Horta, 2014), Kindermorbidität (McCormick, 1985) und Kindersterblichkeit (Bernstein et al., 2000; McCormick, 1985).

Während der Zusammenhang zwischen dem fetalen Wachstum und der Gesundheit des Nachwuchses als etabliert gilt, besteht bezüglich der Aufklärung der einem reduzierten fetalen Wachstum zugrundeliegenden biologischen Mechanismen und Prädiktoren noch großer Forschungsbedarf (National Institutes of Health, 2003). Forschung, die auf die Untersuchung der Hormone schwangerer Frauen, wie z.B. Testosteron und Cortisol zielt, wird hierbei als vielversprechend angesehen (Chatelain, 2000; Entringer et al., 2011; Kivlighan et al., 2008; Manikkam et al., 2004; Voegtline et al., 2013).

Hormonhaushalt schwangerer Frauen und das fetale Wachstum

Testosteron ist ein Sexualhormon, das bei Frauen überwiegend in den Eierstöcken sowie auch in der Nebennierenrinde freigesetzt wird (Al-Dujaili & Sharp, 2012; Burger, 2002). Cortisol ist dagegen ein Stresshormon, welches durch das sogenannte menschliche Stresssystem, die Hypothalamus-Hypophysen-Nebennierenrinden-Achse (HPA-Achse), ausgeschüttet wird (Hellhammer et al., 2009; Kirschbaum & Hellhammer, 1989, 1994). Die Konzentrationen beider Hormone können im Speichel gemessen werden und unterliegen einem natürlichen Tagesrhythmus (zirkadianen Rhythmus), der bei Frauen auch während einer Schwangerschaft erhalten bleibt (Kivlighan et al., 2008; Voegtline et al., 2013). Speichel-Testosteronkonzentrationen zeigen dabei die höchsten Werte am frühen Morgen und eine kontinuierliche Abnahme der Werte im weiteren Tagesverlauf (Al-Dujaili & Sharp,

2012; Dabbs, 1990). Speichel-Cortisolkonzentrationen unterliegen einem starken Anstieg in der ersten Stunde nach dem Erwachen (Cortisolaufwachreaktion, CAR) und fallen dann kontinuierlich im weiteren Tagesverlauf ab (Hellhammer et al., 2009; Pruessner et al., 1997). Eine vergleichsweise abgeflachte Veränderung der Konzentrationen im Tagesverlauf wird als möglicher Indikator für einen dysregulierten Testosteron- und Cortisol-Tagesrhythmus diskutiert (Karatsoreos, Vernov, & Romeo, 2007; Kivlighan et al., 2008).

Befunde aus Tierstudien sowie vorläufige Befunde aus Studien mit schwangeren Frauen deuten auf einen Zusammenhang zwischen mütterlichen Testosteron- bzw. Cortisolkonzentrationen während der Schwangerschaft und dem fetalen Wachstum hin (Bolten et al., 2011; Buss et al., 2009; Carlsen, Jacobsen, & Romundstad, 2006; D'Anna-Hernandez et al., 2012; Kivlighan et al., 2008; Manikkam et al., 2004; Sathishkumar et al., 2011; Veiga-Lopez et al., 2011; Voegtline et al., 2013). Während die Assoziation zwischen Cortisolkonzentrationen schwangerer Frauen und dem fetalen Wachstum immer mehr in den Forschungsmittelpunkt rückt (z.B. Bolten et al., 2011; Buss et al., 2009; D'Anna-Hernandez et al., 2012; Entringer et al., 2011; Kivlighan et al., 2008), wurde Forschung, die auf den Testosteronhaushalt von schwangeren Frauen zielt, bisher weitgehend vernachlässigt. Insbesondere scheint gegenwärtig unklar, ob der Testosteron-Tagesrhythmus schwangerer Frauen mit der Geburtsgröße des Nachwuchses assoziiert ist.

Die vorliegende Dissertation setzt an dieser Forschungslücke an. Sie untersucht den Zusammenhang zwischen dem Speichel-Testosteron-Tagesrhythmus schwangerer Frauen und der Geburtsgröße des Nachwuchses. Da ein männlicher Fetus bezüglich dysfunktionaler pränataler Einflüsse vulnerabler zu sein scheint als ein weiblicher Fetus (Di Renzo, Rosati, Sarti, Cruciani, & Cutuli, 2007; Voegtline et al., 2013; Zeitlin et al., 2002), wird des Weiteren untersucht, ob ein potentieller Zusammenhang zwischen dem Speichel-Testosteron-Tagesrhythmus schwangerer Frauen und der Geburtsgröße des Nachwuchses durch das fetale Geschlecht moderiert wird.

Methodische Aspekte der Bestimmung von Speichel-Testosteron- und -Cortisolkonzentrationen

Erhebung ambulanter Speichelproben und das Adhärenz-Problem. Die Erhebung von Speichelproben gilt gegenwärtig als Methode der Wahl zur ambulanten Bestimmung von Testosteron- und Cortisolkonzentrationen. Probanden werden hierfür üblicherweise instruiert, mittels sogenannter ambulanter Standard-Protokolle in ihrem natürlichen Umfeld an einem Tag bzw. an mehreren Tagen eine bestimmte Anzahl von Speichelproben zu zuvor festgelegten Messzeitpunkten zu sammeln. Dieses Verfahren gilt als nicht-invasiv, kosteneffizient, relativ einfach durchführbar und hiermit auch für großangelegte Studien als besonders gut geeignet (Al-Dujaili & Sharp, 2012; Kirschbaum & Hellhammer, 1994; Kudielka et al., 2012; Nunes, Mussavira, & Bindhu, 2015). Eine ambulante Erhebung von Speichelproben ist darüber hinaus im Vergleich zur Laborforschung mit einer höheren ökologischen Validität assoziiert (Kirchner & Shiffman, 2013; Robbins & Kubiak, 2014; Shiffman, Stone, & Hufford, 2008).

Bei der ambulanten Erhebung von Speichelproben sind verschiedene methodische Aspekte zu berücksichtigen (siehe Reviews von Adam & Kumari, 2009; Al-Dujaili & Sharp, 2012; Granger et al., 2012; Granger et al., 2007b; Granger et al., 2004; Jessop & Turner-Cobb, 2008; Kirschbaum & Hellhammer, 1994; Kudielka et al., 2012; Nunes et al., 2015; Stalder et al., 2016), u.a. die sogenannte Adhärenz[3]-Problematik. So kommt es vor, dass Probanden vorgesehene Messzeitpunkte in Standard-Protokollen zur ambulanten Erhebung von Speichelproben nicht einhalten und diese Nicht-Adhärenz nicht im Selbstbericht angeben, wie Studien mit einem für die Probanden nicht ersichtlichen elektronischen Adhärenz-Monitoring zeigen konnten (Broderick, Arnold, Kudielka, & Kirschbaum, 2004; Jacobs et al., 2005; Kudielka, Broderick, & Kirschbaum, 2003). Weisen Hormonkonzentrationen einen

[3] Wird der Begriff *Adhärenz* in der vorliegenden Dissertation verwendet, bezieht er sich stets auf die Adhärenz bezüglich vorgesehener Messzeitpunkte bei der ambulanten Erhebung von Speichelproben.

Tagesrhythmus auf, kann eine solche Nicht-Adhärenz mit einem Bias bei der Bestimmung von Hormonkonzentrationen assoziiert sein (Broderick et al., 2004; Kudielka et al., 2003).

Die Förderung der Adhärenz ist daher ein zentrales methodisches Anliegen. Als Adhärenz-fördernd hat sich in experimentellen Studien die Vorabinformation über die Verwendung eines für die Probanden nicht ersichtlichen elektronischen Adhärenz-Monitorings gezeigt (Broderick et al., 2004; Kudielka et al., 2003). Andere Studien verwenden elektronische Zeitmesser mit Alarmfunktion, um die Probanden akustisch an die vorgesehenen Messzeitpunkte zu erinnern (z.B. Kraemer et al., 2006; Robles et al., 2011); der Effekt solcher elektronischer Zeitmesser auf die Adhärenz wurde allerdings nach unserem besten Wissen noch nicht untersucht.

Die bisherige Forschung zur oben genannten Adhärenz-Problematik basiert auf Stichproben von gesunden Freiwilligen (Jacobs et al., 2005; Kudielka et al., 2003) sowie von Patienten und gesunden Freiwilligen (Broderick et al., 2004); dagegen wurde die Adhärenz-Problematik in Stichproben schwangerer Frauen nach unserem besten Wissen noch nicht untersucht. Da Schwangere im Vergleich zu Nicht-Schwangeren spezifische Verhaltensmuster zeigen (z.B. reduzierte und verlangsamte physische Aktivität, siehe Lof, 2011; Lof & Forsum, 2006; Poudevigne & O'Connor, 2006; Rousham, Clarke, & Gross, 2006), scheint gegenwärtig unklar, wie adhärent Schwangere vorgesehene Messzeitpunkte innerhalb eines Standard-Protokolls zur ambulanten Erhebung von Speichelproben einhalten und ob die Adhärenz der Schwangeren durch die Vorabinformation über ein nicht ersichtliches elektronisches Adhärenz-Monitoring und/oder das zur Verfügung-Stellen eines elektronischen Zeitmessers mit Alarmfunktion verbessert werden kann. Die Beantwortung dieser Fragen ist Gegenstand der vorliegenden Dissertation.

Bestimmung von Hormonkonzentrationen im Speichel und die Adhärenz-Problematik. Speichel-Testosteron- und -Cortisolkonzentrationen weisen einen natürlichen Tagesrhythmus auf. Eine Nicht-Adhärenz mit den vorgesehenen Messzeitpunkten bei der

Speichelsammlung könnte daher möglicherweise bei einer zu früh bzw. zu spät gesammelten Speichelprobe – aufgrund der üblichen tagesrhythmischen Abnahme dieser Konzentrationen – mit einem vergleichsweise höheren bzw. geringeren Konzentrationslevel im Speichel und daher mit einem Bias bei der Konzentrationsbestimmung einhergehen.

In Studien mit Nicht-Schwangeren wurde ein Zusammenhang zwischen der Adhärenz und den im Speichel gemessenen Cortisolkonzentrationen nachgewiesen (Broderick et al., 2004; Kudielka et al., 2003; Kudielka, Hawkley, Adam, & Cacioppo, 2007). Da Schwangere im Vergleich zu Nicht-Schwangeren bzw. im Vergleich zum Status vor der Schwangerschaft einen abweichenden Cortisolhaushalt aufweisen (de Weerth & Buitelaar, 2005; Kammerer, Adams, Castelberg, & Glover, 2002; Meinlschmidt, Martin, Neumann, & Heinrichs, 2010), ist gegenwärtig aber unklar, ob auch bei Schwangeren die Adhärenz mit den gemessenen Speichel-Cortisolkonzentrationen zusammenhängt. Bezüglich Testosteron liegen nach unserem besten Wissen bisher keine Untersuchungen zum Zusammenhang zwischen der Adhärenz und Speichel-Testosteronkonzentrationen vor.

Hieraus abgeleitet, untersucht die vorliegende Dissertation, ob in einer Stichprobe von Schwangeren die Adhärenz mit den gemessenen Speichel-Testosteron- und -Cortisolkonzentrationen assoziiert ist.

Forschungsfragen

Zusammengefasst könnten Studien, die einem potentiellen Zusammenhang zwischen Speichel-Testosteron- bzw. -Cortisolkonzentrationen schwangerer Frauen und der Geburtsgröße des Nachwuchses nachgehen, einen Beitrag zur Untersuchung der biologischen Mechanismen und Prädiktoren eines reduzierten fetalen Wachstums liefern. Eine möglichst reliable Bestimmung der Speichel-Testosteron- und -Cortisolkonzentrationen bei schwangeren Frauen scheint in diesem Kontext von großer Bedeutung. Die vorliegende Dissertation zielt daher auf die Untersuchung der beiden folgenden übergeordneten Fragestellungen:

1. Wie lässt sich der Adhärenz-Problematik bei der Bestimmung von ambulanten Speichel-Testosteron- und -Cortisolkonzentrationen schwangerer Frauen methodisch begegnen?
2. Zeigt sich unter Berücksichtigung der Adhärenz-Problematik ein Zusammenhang zwischen dem Speichel-Testosteron-Tagesrhythmus schwangerer Frauen und der Geburtsgröße des Nachwuchses?

In den der Dissertation zugrundeliegenden Manuskripten A-C (vgl. Appendix) sind diese übergeordneten Fragestellungen wie folgt differenziert:

Manuskript A: *Improving Ambulatory Saliva-Sampling Compliance in Pregnant Women: A Randomized Controlled Study* (veröffentlicht in *PLOS ONE*)

1. Wie adhärent halten schwangere Frauen die vorgesehenen Messzeitpunkte in einem Standard-Protokoll zur ambulanten Erhebung von Speichelproben ein?
2. Kann die Adhärenz durch Vorabinformation über ein für die Schwangeren nicht ersichtliches elektronisches Adhärenz-Monitoring und/oder durch das zur Verfügung-Stellen eines elektronischen Zeitmessers mit Alarmfunktion zur akustischen Erinnerung an die vorgesehenen Messzeitpunkte verbessert werden?
3. Zeigt sich ein Zusammenhang zwischen der Adhärenz und den im Speichel der Schwangeren gemessenen Cortisolkonzentrationen?

Manuskript B: *Nonadherence with ambulatory saliva sampling is associated with biased salivary testosterone estimates* (veröffentlicht in *Psychoneuroendocrinology*)

1. Zeigt sich ein Zusammenhang zwischen der Adhärenz und den im Speichel der Schwangeren gemessenen Testosteronkonzentrationen?

Manuskript C: *Women's diurnal salivary testosterone change during pregnancy is associated with offspring size at birth* (eingereicht in *Psychoneuroendocrinology*)

1. Zeigt sich ein Zusammenhang zwischen dem Speichel-Testosteron-Tagesrhythmus schwangerer Frauen und Gewicht, Körperlänge bzw. Kopfumfang des Nachwuchses bei der Geburt?
2. Wird dieser Zusammenhang durch das Geschlecht des Fetus moderiert?

Methode

Den drei Manuskripten der vorliegenden Dissertation liegen die Daten einer prospektiven Studie mit zwei Erhebungsphasen zugrunde. Die Manuskripte A und B basieren auf den Daten der ersten Erhebungsphase, in der schwangere Frauen ein erweitertes Standard-Protokoll zur ambulanten Erhebung von Speichelproben absolvierten. Das Manuskript C basiert auf Daten der ersten und zweiten Erhebungsphase, wobei in der zweiten Erhebungsphase die Geburtsparameter des Nachwuchses erhoben wurden.

Im Folgenden wird ein Überblick über die verwendeten Methoden gegeben; die detaillierte Methodendarstellung ist den jeweiligen Manuskripten A-C (vgl. Appendix) zu entnehmen.

Stichprobe

Die Stichprobe besteht aus 75 schwangeren Frauen, die während einer gynäkologischen Routineuntersuchung in der Klinik für Geburtshilfe des Universitätsspitals Zürich (USZ, Schweiz) rekrutiert wurden. Aus dieser Stichprobe verwendeten wir die Daten von 64 Frauen für die statistischen Analysen von Manuskript A, die Daten von 60 Frauen für die Analysen von Manuskript B sowie die Daten von 52 Frauen für die Analysen von Manuskript C.

Studiendesign und Ablauf der Untersuchung

In der ersten Erhebungsphase instruierten wir die Schwangeren, vor ihrer anstehenden Routineuntersuchung in dem USZ einem Standard-Protokoll zur ambulanten Erhebung von Speichelproben zu folgen. Dieses Standardprotokoll umfasste die Erhebung von 16 Speichelproben an zwei aufeinanderfolgenden Tagen mithilfe sogenannter Salivetten

(Eppendorf, Hamburg, Deutschland). Die Speichelproben sollten zu den folgenden Messzeitpunkten abgegeben werden: 0, 30, 45 und 60 Minuten nach dem Aufwachen und um 11:00, 15:00, 20:00 und 22:00 Uhr. Schriftlich wie auch mündlich erläuterten wir den Schwangeren die Notwendigkeit, die Speichelproben möglichst exakt zu den vorgesehenen Messzeitpunkten abzugeben. Wir instruierten die Schwangeren zudem, nach jeder Speichelsammlung die exakte Zeit der Speichelentnahme schriftlich festzuhalten. Zusätzlich erfassten wir die objektiven Zeiten der Speichelsammlung mit einem für die Schwangeren nicht ersichtlichen elektronischen Adhärenz-Monitoring.

Das Standardprotokoll zur ambulanten Speichelsammlung variierten wir im Rahmen eines randomisierten kontrollierten Designs, um den Effekt zweier auf die Adhärenz-Förderung abzielender Maßnahmen zu untersuchen. Die Schwangeren wiesen wir hierfür nach ihrer Rekrutierung randomisiert vier experimentellen Bedingungen in einem 2X2-faktoriellen Design zu; dabei unterschieden wir die Faktoren *Vorabinformation über ein nicht ersichtliches elektronisches Adhärenz-Monitoring* (ja vs. nein) und *Zur Verfügung-Stellen eines elektronischen Zeitmessers mit Alarmfunktion zur akustischen Erinnerung an die vorgesehenen Messzeitpunkte* (ja vs. nein).

In der zweiten Erhebungsphase erhoben wir die relevanten Geburtsparameter der Neugeborenen.

Datenerhebung

Adhärenz. Zur Messung der Adhärenz verwendeten wir den von den Schwangeren abgegebenen Selbstbericht (Adhärenz-Protokoll) sowie die mit einem elektronischen Adhärenz-Monitoring-System (Medication Event Monitoring Systems, Aardex Ltd. Schweiz) gemessenen objektiven Zeiten der Speichelsammlung.

Für die statistische Auswertung der Adhärenz-Daten (siehe Manuskripte A und B) klassifizierten wir eine Speichelprobe als *subjektiv adhärent*, wenn der im Adhärenz-Protokoll selbst berichtete Messzeitpunkt dieser Probe darauf hinwies, dass sie innerhalb

eines spezifischen Zeitfensters im Verhältnis zu ihrem vorhergesehenen Messzeitpunkt erhoben wurde. Eine Speichelprobe klassifizierten wir als *objektiv adhärent*, wenn der elektronisch bestimmte Messzeitpunkt dieser Probe darauf hinwies, dass sie innerhalb eines spezifischen Zeitfensters im Verhältnis zu ihrem vorhergesehenen Messzeitpunkt erhoben wurde. Je nachdem, wie viele Speichelproben einer Schwangeren eine objektiv adhärente Erhebung anzeigten, klassifizierten wir diese Probandin als gut, moderat oder niedrig objektiv adhärent (siehe Manuskript A) beziehungsweise als objektiv adhärent oder objektiv nicht-adhärent (siehe Manuskript B). Die Klassifizierung basierte, wie in den Manuskripten beschrieben, auf den im Forschungsfeld üblicherweise angewandten Kriterien.

Testosteron- und Cortisolkonzentrationen im Speichel. Die Testosteron- und Cortisolkonzentrationen wurden in den gesammelten Speichelproben der Schwangeren mittels eines Enzymimmunoassays für menschlichen Speichel (Testosteron ELISA und Cortisol ELISA, IBL, Hamburg, Deutschland) quantitativ bestimmt.

Deskriptive Informationen. Relevante deskriptive Informationen erfassten wir mittels Fragebogen sowie durch das Auslesen der elektronischen Patientenakten.

Geburtsparameter der Neugeborenen. Gewicht (g), Körperlänge (cm), Kopfumfang (cm) und Gestationsalter (Tage) des Nachwuchses bei der Geburt erhoben wir aus den elektronischen Patientenakten.

Statistische Analysen

Für die Auswertung der Daten verwendeten wir Mixed-Model-Analysen sowie Varianz- und Kovarianzanalysen. Mixed-Model-Analysen gelten als Methode der Wahl für die Analyse ambulanter Daten mit wiederholten und individuell variierenden Messzeitpunkten (siehe Kudielka et al., 2012; Lane, 2008; Singer & Willett, 2003).

Ergebnisse

Adhärenz mit einem ambulanten Standard-Protokoll zur Speichelerhebung

Die schwangeren Frauen absolvierten das Standard-Protokoll zur ambulanten Erhebung von Speichelproben in der 24. Gestationswoche (median). Die aus den Adhärenz-Protokollen der Schwangeren ermittelte subjektive Adhärenz betrug 91%, während das elektronische Adhärenz-Monitoring eine objektive Adhärenz von 70% anzeigte. Selbst die Schwangeren mit einer niedrigen objektiven Adhärenz (max. 31% objektive Adhärenz) berichteten im Adhärenz-Protokoll eine hohe subjektive Adhärenz von 80%. Die elektronisch gemessenen Zeiten der Speichelsammlung zeigten, dass die als nicht adhärent klassifizierten Speichelproben im Verhältnis zum vorgesehenen Messzeitpunkt vornehmlich verspätet abgegeben wurden.

Es zeigte sich ein Zusammenhang zwischen der Maßnahme *Vorabinformation über ein nicht ersichtliches elektronisches Adhärenz-Monitoring* und einer verbesserten objektiven Adhärenz. Schwangere, die wir vorab über das Adhärenz-Monitoring informierten, zeigten eine höhere objektive Adhärenz (81%) als die nicht informierten Schwangeren (60%). Dagegen zeigte sich zwischen der Maßnahme *Zur Verfügung-Stellen eines elektronischen Zeitmessers mit Alarmfunktion* und der objektiven Adhärenz der Schwangeren kein Zusammenhang (siehe Manuskript A).

Adhärenz und die Bestimmung der Hormonkonzentrationen im Speichel

Cortisolkonzentrationen. Es zeigte sich ein Zusammenhang zwischen der von den schwangeren Frauen gezeigten objektiven Adhärenz und den in ihrem Speichel gemessenen Cortisolkonzentrationen. Schwangere mit einer guten bis moderaten objektiven Adhärenz zeigten höhere Speichel-Cortisolkonzentrationen als Schwangere, die eine niedrige objektive Adhärenz aufwiesen.

Es zeigte sich des Weiteren ein Zusammenhang zwischen der zeitlichen Abweichung des elektronisch gemessenen Messzeitpunktes vom vorgesehenen Messzeitpunkt und der in

der entsprechenden Speichelprobe gemessenen Cortisolkonzentration. Je später eine Speichelprobe im Vergleich zu ihrem vorgesehenen Messzeitpunkt gesammelt wurde, desto niedriger war die in der Speichelprobe gemessene Cortisolkonzentration. Die ermittelten Zusammenhänge blieben auch nach der Adjustierung für a priori festgelegte potentielle Confounder bestehen (siehe Manuskript A).

Testosteronkonzentrationen. Es zeigte sich ein Zusammenhang zwischen der von den Schwangeren gezeigten objektiven Adhärenz und den in ihrem Speichel bestimmten Testosteronkonzentrationen. Adhärente Schwangere zeigten über den Tag hinweg höhere Speichel-Testosteronkonzentrationen als die von uns als nicht-adhärent klassifizierten Schwangeren.

Es zeigte sich des Weiteren ein Zusammenhang zwischen der zeitlichen Abweichung des elektronisch gemessenen Messzeitpunktes vom vorgesehenen Messzeitpunkt und der in der entsprechenden Speichelprobe gemessenen Testosteronkonzentration. Je später eine Speichelprobe im Vergleich zu ihrem vorgesehenen Messzeitpunkt gesammelt wurde, desto niedriger war die in der Speichelprobe gemessene Testosteronkonzentration. Die ermittelten Zusammenhänge blieben auch nach der Adjustierung für a priori festgelegte potentielle Confounder bestehen (siehe Manuskript B).

Speichel-Testosteron-Tagesrhythmus Schwangerer und die Geburtsgröße des Nachwuchses

Nach einer Adjustierung für das Gestationsalter und eine Reihe potentieller Confounder zeigte sich ein Zusammenhang zwischen dem mütterlichen Speichel-Testosteron-Slope[4] während der Schwangerschaft und dem Gewicht bzw. der Körperlänge des Nachwuchses bei der Geburt. Je flacher sich der mütterliche Testosteron-Slope in der

[4] Als Indikator für den Speichel-Testosteron-Tagesrhythmus verwendeten wir in den statistischen Analysen den Testosteron-Slope, d.h. die individuelle lineare Variation in den Speichel-Testosteronkonzentrationen im Tagesverlauf. Aufgrund der in den Manuskripten A und B berichteten Befunde verwendeten wir zur Slope-Berechnung die elektronisch gemessenen Uhrzeiten der Speichelsammlung anstatt der von den Schwangeren selbstberichteten Uhrzeiten.

Schwangerschaft darstellte, desto niedriger waren das Gewicht und die Körperlänge des Nachwuchses. Unter Berücksichtigung potentieller Confounder zeigte sich dagegen kein stabiler Zusammenhang zwischen mütterlichem Testosteron-Slope und dem Kopfumfang des Nachwuchses bei der Geburt. Es zeigten sich zudem jeweils keine Moderationseffekte durch das fetale Geschlecht (siehe Manuskript C).

Diskussion

Unsere Befunde zeigen zunächst, dass bei schwangeren Frauen im Rahmen eines Standard-Protokolls zur ambulanten Erhebung von Speichelproben eine substantielle objektive Nicht-Adhärenz mit den vorgesehenen Messzeitpunkten zur Speichelsammlung auftrat. Diese objektive Nicht-Adhärenz blieb im Selbstbericht weitgehend unerwähnt. Sie war mit verminderten und somit gebiasten Speichel-Testosteron- und -Cortisolwerten assoziiert. Die Vorabinformation über ein nicht ersichtliches, elektronisches Adhärenz-Monitoring, nicht aber das zur Verfügung-Stellen eines Zeitmessers mit Alarmfunktion, ging zudem mit einer verbesserten objektiven Adhärenz einher. Die Befunde zeigen des Weiteren, dass ein abgeflachter Speichel-Testosteron-Tagesrhythmus schwangerer Frauen mit einem reduzierten Geburtsgewicht und einer reduzierten Geburtskörperlänge des Nachwuchses im Verhältnis zum Gestationsalter assoziiert war, unabhängig vom Geschlecht des Nachwuchses.

Methodische Aspekte der ambulanten Bestimmung von Speichel-Testosteron- und -Cortisolkonzentrationen

Adhärenz mit einem ambulanten Standard-Protokoll zur Speichelerhebung.

Unser Befund, dass schwangere Frauen in einem Standard-Protokoll zur ambulanten Erhebung von Speichelproben eine substantielle objektive Nicht-Adhärenz zeigten und diese weitgehend nicht im Selbstbericht erwähnten, stimmt mit den Resultaten früherer Studien in Populationen nicht schwangerer Frauen überein (Broderick et al., 2004; Kudielka et al., 2003; Kudielka et al., 2007). Das substantielle Vorkommen einer objektiven Nicht-Adhärenz konnte damit auch in einer Stichprobe schwangerer Frauen belegt werden.

Maßnahmen zur Adhärenz-Förderung. Unser Befund, dass die Vorabinformation über ein nicht ersichtliches elektronisches Adhärenz-Monitoring mit einer erhöhten objektiven Adhärenz einherging, steht im Einklang mit Befunden früherer Studien (Broderick et al., 2004; Kudielka et al., 2003). Damit konnten die Befunde früherer Studien auf eine Stichprobe schwangerer Frauen erweitert werden.

Ob das zur Verfügung-Stellen eines Zeitmessers mit Alarmfunktion die objektive Adhärenz mit den vorgesehenen Messzeitpunkten bei einer ambulanten Erhebung von Speichelproben erhöht, wurde nach bestem Wissen von uns als erstes untersucht. Es zeigte sich keine Assoziation zwischen dem zur Verfügung-Stellen eines Zeitmessers und der objektiven Adhärenz.

Adhärenz und Speichel-Cortisol- bzw. -Testosteronkonzentrationen. Der von uns in einer Stichprobe schwangerer Frauen aufgezeigte Zusammenhang zwischen einer objektiven Nicht-Adhärenz und einem Bias bei der Bestimmung von Speichel-Cortisolkonzentrationen stimmt mit den Ergebnissen bisheriger Studien in Populationen nicht schwangerer Frauen überein (Broderick et al., 2004; Golden et al., 2014; Kudielka et al., 2003; Kudielka et al., 2007; Smith & Dougherty, 2014). Damit konnten die Ergebnisse bisheriger Studien auf eine Stichprobe schwangerer Frauen erweitert werden.

Nach bestem Wissen haben wir dagegen als erstes untersucht, ob eine objektive Nicht-Adhärenz auch bei der Bestimmung von Speichel-Testosteronkonzentrationen mit einem Bias assoziiert ist. Hierbei fanden wir eine Assoziation zwischen einer objektiven Nicht-Adhärenz und einem Bias bei der Bestimmung der Speichel-Testosteronkonzentrationen. Eine objektive Nicht-Adhärenz scheint demnach nicht nur bei der Bestimmung von Speichel-Cortisolkonzentrationen, sondern auch bei der Bestimmung von Speichel-Testosteronkonzentrationen mit einem Bias einherzugehen.

Implikationen. Zusammenfassend weisen diese Befunde darauf hin, dass die von schwangeren Frauen bei einer ambulanten Erhebung von Speichelproben gezeigte objektive

Nicht-Adhärenz mit einem Bias bei der Bestimmung ihrer Speichel-Testosteron- und -Cortisolkonzentrationen einhergeht und hiermit zu falschen Interpretationen der Daten führen könnte. Für eine möglichst reliable Bestimmung ambulanter Speichel-Testosteron- und -Cortisolkonzentrationen bei schwangeren Frauen könnte daher die Berücksichtigung der Adhärenz-Problematik essentiell sein. Basierend auf unseren Befunden lassen sich folgende methodische Empfehlungen zur Berücksichtigung der Adhärenz-Problematik formulieren:

- *Elektronische Adhärenz-Monitoring-Systeme:* Um bei Schwangeren in einem Standard-Protokoll zur ambulanten Erhebung von Speichelproben die Messzeitpunkte der Speichelsammlung möglichst reliabel zu erfassen, ist aufgrund der Adhärenz-Problematik die Verwendung eines elektronischen Adhärenz-Monitorings zu empfehlen (vgl. Broderick et al., 2004; Kudielka et al., 2003; Kudielka et al., 2007).
- *Maßnahmen zur Adhärenz-Förderung:* Um die Adhärenz zu fördern, sollten schwangere Frauen bereits vor der Speichelsammlung über ein elektronisches Adhärenz-Monitoring informiert werden. Der Einsatz elektronischer Zeitmesser zur akustischen Erinnerung an vorgesehene Messzeitpunkte kann dagegen aufgrund unseres Befundes nicht empfohlen werden. Sollten Studien auf ein elektronisches Adhärenz-Monitoring verzichten, könnte versucht werden, eine Adhärenz-Förderung durch eine Scheininformation über die Verwendung eines Monitorings zu erreichen. Eine andere, ethisch besser legitimierbare Möglichkeit wäre, das Monitoring nur bei einer Subgruppe durchzuführen und alle Schwangeren über die Möglichkeit eines Monitorings zu informieren. Ob hiermit eine vergleichbare Wirkung erreicht werden könnte, sollte allerdings zuvor überprüft werden (vgl., Adam & Kumari, 2009; Kudielka et al., 2012; Stalder et al., 2016).
- *Statistischer Umgang mit der Adhärenz-Problematik:* Um den individuellen Tagesrhythmus der Speichel-Testosteron- und -Cortisolkonzentrationen bei Schwangeren zu schätzen, bietet sich aufgrund der Adhärenz-Problematik als Methode

der Wahl ein statistisches Modell an, welches variierende, durch ein elektronisches Adhärenz-Monitoring gemessene Zeiten der Speichelsammlung berücksichtigen kann (siehe Manuskript C, siehe auch Singer & Willett, 2003; Stalder et al., 2016).

Speichel-Testosteron-Tagesrhythmus Schwangerer und die Geburtsgröße des Nachwuchses

Testosteron-Tagesrhythmus Schwangerer. Dass die Speichel-Testosteronkonzentrationen von Frauen einem Testosteron-Tagesrhythmus unterliegen, gilt als etablierter Befund (Al-Dujaili & Sharp, 2012). Unsere Ergebnisse weisen darauf hin, dass die Speichel-Testosteronkonzentrationen von Frauen auch während der Schwangerschaft einem Tagesrhythmus unterliegen. Sie replizieren damit den Befund von Voegtline et al. (2013).

Zusammenhang zwischen dem Testosteron-Tagesrhythmus Schwangerer und der Geburtsgröße des Nachwuchses. Unser Befund, dass – nach der Adjustierung für das Gestationsalter und verschiedene potentielle Confounder – ein abgeflachter mütterlicher Speichel-Testosteron-Tagesrhythmus während der Schwangerschaft mit einem reduzierten Gewicht bzw. einer reduzierten Körperlänge des Nachwuchses bei der Geburt assoziiert war, steht im Gegensatz zu dem Befund von Voegtline et al. (2013), die keinen Zusammenhang zwischen dem Speichel-Testosteron-Tagesrhythmus Schwangerer und dem Gewicht des Nachwuchses bei der Geburt fanden. Eine mögliche Erklärung für die inkonsistenten Befunde ist, dass in der Studie von Voegtline et al. kein elektronisches Adhärenz-Monitoring eingesetzt wurde und somit für die Schätzung des Speichel-Testosteron-Tagesrhythmus der Schwangeren keine objektiven Zeiten der Speichelsammlung verwendet werden konnten.

Ob der Speichel-Testosteron-Tagesrhythmus schwangerer Frauen mit dem Kopfumfang des Nachwuchses bei der Geburt assoziiert ist, wurde nach bestem Wissen von uns als erstes untersucht. Es zeigte sich hierbei kein stabiler Zusammenhang zwischen dem Testosteron-Tagesrhythmus und dem Kopfumfang.

Implikation. Zusammenfassend weisen diese Befunde darauf hin, dass ein abgeflachter mütterlicher Speichel-Testosteron-Tagesrhythmus während der Schwangerschaft mit einem reduzierten Gewicht bzw. einer reduzierten Körperlänge des Nachwuchses bei der Geburt im Verhältnis zum Gestationsalter assoziiert ist. Ein abgeflachter Speichel-Testosteron-Tagesrhythmus schwangerer Frauen könnte damit einen neuen, potentiell relevanten Prädiktor für ein reduziertes fetales Wachstum darstellen. Für künftige Forschung ließe sich hieraus die Empfehlung ableiten, Testosteronkonzentrationen in Populationen schwangerer Frauen zu untersuchen und hierbei speziell auch den Testosteron-Tagesrhythmus zu berücksichtigen. Eine entsprechende Forschung könnte zu einem verbesserten Verständnis der zugrundeliegenden biologischen Mechanismen und zu einer verbesserten Diagnostik eines reduzierten fetalen Wachstums beitragen.

Stärken und Limitationen

Stärken.

- *Umfangreiches Standard-Protokoll zur ambulanten Erhebung von Speichelproben*: Unser Standard-Protokoll umfasste 16 Messzeitpunkte und wies hiermit bezüglich einer Speichel-Testosteron- und -Cortisolbestimmung methodische Vorteile auf im Vergleich zu Erhebungen, die auf nur wenigen Messzeitpunkten basieren (Al-Dujaili & Sharp, 2012; Kudielka et al., 2012).
- *Elektronisches Adhärenz-Monitoring:* Durch die Verwendung eines elektronischen Adhärenz-Monitorings konnten wir die objektiven Zeiten der Speichelsammlung erheben. Dies scheint relevant, da unsere Befunde (vgl. Manuskript A) sowie frühere Studien (Broderick et al., 2004; Kudielka et al., 2003) zeigen, dass die im Selbstbericht mitgeteilten Zeiten der Speichelsammlung häufig unkorrekt sind.
- *Mixed-Model-Analysen:* Die Verwendung spezifischer Mixed-Model-Analysen ermöglichte es uns, Speichel-Testosteron- und -Cortisolkonzentrationen basierend auf den elektronisch gemessenen Zeiten der Speichelsammlung zu schätzen (vgl.

Manuskripte A-C). Mixed-Model-Analysen gelten des Weiteren als Methode der Wahl für die Analyse ambulanter Daten mit multiplen Messzeitpunkten, in denen gewöhnlich fehlende Werte vorkommen (Kudielka et al., 2012; Lane, 2008; Singer & Willett, 2003).

- *Randomisiertes kontrolliertes Design*: Ob spezifische Maßnahmen die objektive Adhärenz der schwangeren Frauen erhöhen können, untersuchten wir mit einem randomisierten kontrollierten Design. Ein solches Design gilt als Methode der Wahl, um den Effekt von Interventionen zu überprüfen (Sibbald & Roland, 1998).

Limitationen.

- *Stichprobengröße:* Die Größe unserer Stichprobe war eher klein. Es ist demzufolge angezeigt, unsere Befunde in einer größeren Stichprobe schwangerer Frauen zu replizieren.
- *Kausalität:* Von der in einem randomisierten kontrollierten Design untersuchten Fragestellung abgesehen, können zu den gefundenen Assoziationen aufgrund der Anlage der Studie keine kausalen Überlegungen angestellt werden.
- *Adhärenz*: Wir haben spezifisch die Adhärenz mit vorgesehenen Messzeitpunkten in einem Standard-Protokoll zur ambulanten Erhebung von Speichelproben und die Auswirkung einer Nicht-Adhärenz auf die gemessenen Hormonkonzentrationen im Speichel zu untersucht. Es ist aber nicht auszuschließen, dass noch andere methodische Aspekte mit einem Bias bei der ambulanten Bestimmung von Speichel-Testosteron- und -Cortisolkonzentrationen einhergehen könnten, was durch zukünftige Studien in der Zielgruppe schwangerer Frauen untersucht werden sollte (vgl., Granger et al., 2012; Granger et al., 2004; Kudielka et al., 2012).
- *Generalisierbarkeit:* Unsere Befunde basieren auf einer in der Schweiz rekrutierten Stichprobe schwangerer Frauen. Rückschlüsse auf andere Populationen sind daher nicht möglich. In unserer Stichprobe entsprach des Weiteren kein Neugeborenes den

Kriterien einer klinisch relevanten fetalen Wachstumsretardierung. Es würde die Generalisierbarkeit unserer Befunde erweitern, wenn zukünftige Studien den Zusammenhang zwischen dem Testosteron-Tagesrhythmus schwangerer Frauen und der Geburtsgröße des Nachwuchses auch in einer Stichprobe mit wachstumsretardierten Feten aufzeigen könnten.

Ausblick

Sollten sich Forschungsbefunde verdichten, dass ein abgeflachter Testosteron-Tagesrhythmus ein Prädiktor für ein reduziertes fetales Wachstum ist, würde sich als nächstes die Frage stellen, ob dieser Prädiktor den Kriterien eines Biomarkers[5] entsprechend einerseits reliabel messbar ist[6] und andererseits eine klinische Relevanz aufweist (Atkinson et al., 2001; Strimbu & Tavel, 2010). Sollte sich der Testosteron-Tagesrhythmus als ein Biomarker erweisen, könnte er für die Frühdiagnostik eines reduzierten fetalen Wachstums relevant sein.

Für einen klinischen Fortschritt könnte es zudem erfolgversprechend sein, in multimodalen Modellen verschiedene Biomarker, Doppler-sonographische Untersuchungen und bekannte Risikofaktoren (z.B. mütterliches Alter) für eine möglichst gute Vorhersage eines reduzierten fetalen Wachstums zu kombinieren (Kane, Costa, & Brennecke, 2014; Poon, Syngelaki, Akolekar, Lai, & Nicolaides, 2013).

In einem weiteren Forschungsstrang könnte aus klinischer Perspektive untersucht werden, ob ein abgeflachter Testosteron-Tagesrhythmus schwangerer Frauen durch spezifische Interventionen rhythmisiert werden kann und ob eine solche Rhythmisierung des Testosteron-Tagesrhythmus mit einer Normalisierung des fetalen Wachstums einhergeht.

[5] Biomarker sind objektiv messbare biologische Merkmale, die auf dysfunktionale oder normale biologische Prozesse eines Organismus hinweisen und beispielsweise eine diagnostische Aussage bezüglich eines Krankheitszustandes oder eines abnormalen Outcomes geben können (Atkinson et al., 2001; Strimbu & Tavel, 2010).

[6] Vergleiche hierzu die Implikationen zu den Befunden der übergeordneten ersten Fragestellung.

Fazit

Nach bestem Wissen sind die vorliegenden Befunde die ersten, die auf einen Zusammenhang zwischen einem abgeflachten mütterlichen Speichel-Testosteron-Tagesrhythmus während der Schwangerschaft und einer reduzierten Geburtsgröße des Nachwuchses im Verhältnis zum Gestationsalter hinweisen. Ein abgeflachter Speichel-Testosteron-Tagesrhythmus schwangerer Frauen könnte hiermit ein neuer, potentiell relevanter Prädiktor und Biomarker für ein reduziertes fetales Wachstum sein. Dieser Befund erscheint hoch relevant und könnte zu einem verbesserten Verständnis der zugrundeliegenden biologischen Mechanismen und zu einer verbesserten Vorhersage eines reduzierten fetalen Wachstums beitragen.

Unsere Befunde weisen des Weiteren darauf hin, dass für eine möglichst reliable ambulante Bestimmung von Speichel-Testosteron- und -Cortisolkonzentrationen bei schwangeren Frauen die Adhärenz-Problematik berücksichtigt werden sollte, da eine substanzielle objektive Nicht-Adhärenz mit vorgesehenen Messzeitpunkten bei einer ambulanten Speichelsammlung als Möglichkeit in Betracht gezogen werden muss und mit einem Bias bei der Bestimmung dieser Konzentrationen einhergehen kann. Basierend auf unseren Resultaten kann in diesem Zusammenhang empfohlen werden, ein elektronisches Adhärenz-Monitoring einzusetzen, die Schwangeren über dieses Monitoring zu informieren und ein statistisches Modell zu verwenden, welches den Tagesrhythmus der Hormonkonzentrationen basierend auf den elektronisch gemessenen Zeiten der Speichelsammlung schätzt.

Literatur

Adam, E. K., & Kumari, M. (2009). Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology, 34*(10), 1423-1436. doi: 10.1016/j.psyneuen.2009.06.011

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

Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., . . . Grp, B. D. W. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics, 69*(3), 89-95. doi: 10.1067/mcp.2000.113989

Bamfo, J. E., & Odibo, A. O. (2011). Diagnosis and management of fetal growth restriction. *J Pregnancy, 2011*, 640715. doi: 10.1155/2011/640715

Barker, D. J. P. (1995). Fetal origins of coronary heart-disease. *British Medical Journal, 311*(6998), 171-174.

Barker, D. J. P. (2007). The origins of the developmental origins theory. *Journal of Internal Medicine, 261*(5), 412-417. doi: 10.1111/j.1365-2796.2007.01809.x

Barker, D. J. P., Eriksson, J. G., Forsen, T., & Osmond, C. (2002). Fetal origins of adult disease: strength of effects and biological basis. *International Journal of Epidemiology, 31*(6), 1235-1239. doi: Doi 10.1093/Ije/31.6.1235

Barker, D. J. P., Osmond, C., Simmonds, S. J., & Wield, G. A. (1993). The relation of small head circumference and thinness at birth to death from cardiovascular-disease in adult life. *British Medical Journal, 306*(6875), 422-426.

Barker, D. J. P., Winter, P. D., Osmond, C., Margetts, B., & Simmonds, S. J. (1989). Weight in infancy and death from ischemic heart-disease. *Lancet, 2*(8663), 577-580.

Belli, P. C., Bustreo, F., & Preker, A. (2005). Investing in children's health: what are the economic benefits? *Bulletin of the World Health Organization, 83*(10), 777-784.

Bernstein, I. M., Horbar, J. D., Badger, G. J., Ohlsson, A., Golan, L., & Network, V. O. (2000). Morbidity and mortality among very-low-birth-weight neonates with intrauterine growth restriction. *American Journal of Obstetrics and Gynecology, 182*(1), 198-206. doi: Doi 10.1016/S0002-9378(00)70513-8

Bloom, D. E., Canning, D., & Sevilla, J. (2004). The effect of health on economic growth: A production function approach. *World Development, 32*(1), 1-13. doi: 10.1016/j.worlddev.2003.07.002

Bolten, M. I., Wurmser, H., Buske-Kirschbaum, A., Papousek, M., Pirke, K. M., & Hellhammer, D. (2011). Cortisol levels in pregnancy as a psychobiological predictor for birth weight. *Archives of Womens Mental Health, 14*(1), 33-41. doi: 10.1007/s00737-010-0183-1

Bonamy, A. K. E., Norman, M., & Kaijser, M. (2008). Being born too small, too early, or both: does it matter for risk of hypertension in the elderly? *American Journal of Hypertension, 21*(10), 1107-1110. doi: 10.1038/ajh.2008.241

Broderick, J. E., Arnold, D., Kudielka, B. M., & Kirschbaum, C. (2004). Salivary cortisol sampling compliance: comparison of patients and healthy volunteers. *Psychoneuroendocrinology, 29*(5), 636-650. doi: 10.1016/S0306-4530(03)00093-3S0306453003000933 [pii]

Bundesamt für Statistik. (2015). *Erhebung zu Familien und Generationen 2013: Erste Ergebnisse*. Neuchâtel: Bundesamt für Statistik (BFS).

Burger, H. G. (2002). Androgen production in women. *Fertility and Sterility, 77*(4), S3-S5.

Buss, C., Entringer, S., Reyes, J. F., Chicz-DeMet, A., Sandman, C. A., Waffarn, F., & Wadhwa, P. D. (2009). The maternal cortisol awakening response in human pregnancy is associated with the length of gestation. *American Journal of Obstetrics*

and Gynecology, 201(4), 398.e391-398. doi: S0002-9378(09)00734-0 [pii]10.1016/j.ajog.2009.06.063

Carlsen, S. M., Jacobsen, G., & Romundstad, P. (2006). Maternal testosterone levels during pregnancy are associated with offspring size at birth. *European Journal of Endocrinology, 155*(2), 365-370. doi: Doi 10.1530/Eje.1.02200

Chan, P. Y., Morris, J. M., Leslie, G. I., Kelly, P. J., & Gallery, E. D. (2010). The long-term effects of prematurity and intrauterine growth restriction on cardiovascular, renal, and metabolic function. *Int J Pediatr, 2010*, 280402. doi: 10.1155/2010/280402

Chatelain, P. (2000). Children born with intra-uterine growth retardation (IUGR) or small for gestational age (SGA): long term growth and metabolic consequences. *Endocrine Regulations, 34*(1), 33-36.

Cheung, Y. F., Wong, K. Y., Lam, B. C. C., & Tsoi, N. S. (2004). Relation of arterial stiffness with gestational age and birth weight. *Archives of Disease in Childhood, 89*(3), 217-221. doi: 10.1136/adc.2003.025999

D'Anna-Hernandez, K. L., Hoffman, M. C., Zerbe, G. O., Coussons-Read, M., Ross, R. G., & Laudenslager, M. L. (2012). Acculturation, maternal cortisol, and birth outcomes in women of Mexican descent. *Psychosomatic Medicine, 74*(3), 296-304. doi: Doi 10.1097/Psy.0b013e318244fbde

Dabbs, J. M. (1990). Salivary testosterone measurements—reliability across hours, days, and weeks. *Physiology and Behavior, 48*(1), 83-86.

de Mola, C. L., de Franca, G. V. A., Quevedo, L. D. A., & Horta, B. L. (2014). Low birth weight, preterm birth and small for gestational age association with adult depression: systematic review, and meta-analysis. *British Journal of Psychiatry, 205*(5), 340-347. doi: 10.1192/bjp.bp.113.139014

de Weerth, C., & Buitelaar, J. K. (2005). Physiological stress reactivity in human pregnancy—a review. *Neuroscience and Biobehavioral Reviews, 29*(2), 295-312. doi: S0149-7634(04)00127-7 [pii]10.1016/j.neubiorev.2004.10.005

Di Renzo, G. C., Rosati, A., Sarti, R. D., Cruciani, L., & Cutuli, A. M. (2007). Does fetal sex affect pregnancy outcome? *Gender Medicine, 4*(1), 19-30. doi: Doi 10.1016/S1550-8579(07)80004-0

Entringer, S., Buss, C., Andersen, J., Chicz-DeMet, A., & Wadhwa, P. D. (2011). Ecological momentary assessment of maternal cortisol profiles over a multiple-day period predicts the length of human gestation. *Psychosomatic Medicine, 73*(6), 469-474. doi: Doi 10.1097/Psy.0b013e31821fbf9a

Golden, S. H., Sanchez, B. N., DeSantis, A. S., Wu, M. H., Castro, C., Seeman, T. E., . . . Roux, A. V. D. (2014). Salivary cortisol protocol adherence and reliability by socio-demographic features: the Multi-Ethnic Study of Atherosclerosis. *Psychoneuroendocrinology, 43*, 30-40. doi: 10.1016/j.psyneuen.2014.01.025

Goldenberg, R. L., & Culhane, J. F. (2007). Low birth weight in the United States. *American Journal of Clinical Nutrition, 85*(2), 584S-590S.

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. *Journal of Adolescence, 35*(4), 1081-1095. doi: Doi 10.1016/J.Adolescence.2012.01.005

Granger, D. A., Kivlighan, K. T., Fortunato, C., Harmon, A. G., Hibel, L. C., Schwartz, E. B., & Whembolua, G. L. (2007a). Integration of salivary biomarkers into developmental and behaviorally-oriented research: problems and solutions for collecting specimens. *Physiology and Behavior, 92*(4), 583-590. doi: S0031-9384(07)00173-4 [pii]10.1016/j.physbeh.2007.05.004

Granger, D. A., Kivlighan, K. T., Fortunato, C., Harmon, A. G., Hibel, L. C., Schwartz, E. B., & Whembolua, G. L. (2007b). Integration of salivary biomarkers into developmental and behaviorally-oriented research: problems and solutions for collecting specimens. *Physiology & Behavior, 92*(4), 583-590. doi: Doi 10.1016/J.Physbeh.2007.05.004

Granger, D. A., Shirtcliff, E. A., Booth, A., Kivlighan, K. T., & Schwartz, E. B. (2004). The "trouble" with salivary testosterone. *Psychoneuroendocrinology, 29*(10), 1229-1240. doi: Doi 10.1016/J.Psyneuen.2004.02.005

Hellhammer, D. H., Wust, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology, 34*(2), 163-171. doi: S0306-4530(08)00299-0 [pii]10.1016/j.psyneuen.2008.10.026

Hui, L., & Challis, D. (2008). Diagnosis and management of fetal growth restriction: the role of fetal therapy. *Best Practice & Research Clinical Obstetrics & Gynaecology, 22*(1), 139-158. doi: 10.1016/i.bpobgyn.2007.06.004

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 Sciences, 76*(21), 2431-2443. doi: S0024-3205(05)00048-2 [pii]10.1016/j.lfs.2004.10.045

Jessop, D. S., & Turner-Cobb, J. M. (2008). Measurement and meaning of salivary cortisol: a focus on health and disease in children. *Stress-the International Journal on the Biology of Stress, 11*(1), 1-14. doi: 10.1080/10253890701365527

Kammerer, M., Adams, D., Castelberg, B. v., & Glover, V. (2002). Pregnant women become insensitive to cold stress. *BMC Pregnancy Childbirth, 2*(1), 8.

Kane, S. C., Costa, F. D., & Brennecke, S. (2014). First trimester biomarkers in the prediction of later pregnancy complications. *Biomed Research International, 2014*. doi: Artn80719610.1155/2014/807196

Karatsoreos, I. N., Vernov, M., & Romeo, R. D. (2007). Testosterone and the brain: implications for cognition, biological rhythms and aging. In L. I. Ardis (Ed.), *Testosterone Research Trends* (pp. 91-103). New York: Nova Science Publishers.

Kirchner, T. R., & Shiffman, S. (2013). Ecological Momentary Assessment. *The Wiley-Blackwell Handbook of Addiction Psychopharmacology*, 541-565.

Kirschbaum, C., & Hellhammer, D. H. (1989). Salivary cortisol in psychobiological research - an overview. *Neuropsychobiology, 22*(3), 150-169.

Kirschbaum, C., & Hellhammer, D. H. (1994). Salivary cortisol in psychoneuroendocrine research—recent developments and applications. *Psychoneuroendocrinology, 19*(4), 313-333.

Kivlighan, K. T., DiPietro, J. A., Costigan, K. A., & Laudenslager, M. L. (2008). Diurnal rhythm of cortisol during late pregnancy: associations with maternal psychological well-being and fetal growth. *Psychoneuroendocrinology, 33*(9), 1225-1235. doi: Doi 10.1016/J.Psyneuen.2008.06.008

Kudielka, B. M., Broderick, J. E., & Kirschbaum, C. (2003). Compliance with saliva sampling protocols: electronic monitoring reveals invalid cortisol daytime profiles in noncompliant subjects. *Psychosomatic Medicine, 65*(2), 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. *Psychosomatic Medicine, 74*(4), 418-431. doi: Doi 10.1097/Psy.0b013e31825434c7

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. *Annals of Behavioral Medicine, 34*(2), 209-216.

Lane, P. (2008). Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharmaceutical Statistics, 7*(2), 93-106. doi: Doi 10.1002/Pst.267

Law, C. M., de Swiet, M., Osmond, C., Fayers, P. M., Barker, D. J., Cruddas, A. M., & Fall, C. H. (1993). Initiation of hypertension in utero and its amplification throughout life. *BMJ (Clinical Research Ed.), 306*(6869), 24-27.

Lithell, H. O., McKeigue, P. M., Berglund, L., Mohsen, R., Lithell, U. B., & Leon, D. A. (1996). Relation of size at birth to non-insulin dependent diabetes and insulin concentrations in men aged 50-60 years. *British Medical Journal, 312*(7028), 406-410.

Lof, M. (2011). Physical activity pattern and activity energy expenditure in healthy pregnant and non-pregnant Swedish women. *European Journal of Clinical Nutrition, 65*(12), 1295-1301. doi: Doi 10.1038/Ejcn.2011.129

Lof, M., & Forsum, E. (2006). Activity pattern and energy expenditure due to physical activity before and during pregnancy in healthy Swedish women. *British Journal of Nutrition, 95*(2), 296-302. doi: Doi 10.1079/Bjn20051497

Malamud, D. (2011). Saliva as a diagnostic fluid. *Dental Clinics of North America, 55*(1), 159-178.

Mangiaterra, V., Mattero, M., & Dunkelberg, E. (2006). Why and how to invest in neonatal health. *Seminars in Fetal & Neonatal Medicine, 11*(1), 37-47. doi: 10.1016/j.siny.2005.11.010

Manikkam, M., Crespi, E. J., Doop, D. D., Herkimer, C., Lee, J. S., Yu, S., . . . Padmanabhan, V. (2004). Fetal programming: prenatal testosterone excess leads to fetal growth retardation and postnatal catch-up growth in sheep. *Endocrinology, 145*(2), 790-798. doi: 10.1210/en.2003-0478

Mayer, C., & Joseph, K. S. (2013). Fetal growth: a review of terms, concepts and issues relevant to obstetrics. *Ultrasound in Obstetrics and Gynecology, 41*(2), 136-145. doi: 10.1002/uog.11204

McCormick, M. C. (1985). The contribution of low birth-weight to infant-mortality and childhood morbidity. *New England Journal of Medicine, 312*(2), 82-90. doi: Doi 10.1056/Nejm198501103120204

Meinlschmidt, G., Martin, C., Neumann, I. D., & Heinrichs, M. (2010). Maternal cortisol in late pregnancy and hypothalamic-pituitary-adrenal reactivity to psychosocial stress postpartum in women. *Stress, 13*(2), 163-171. doi: Doi 10.3109/10253890903128632

Meinlschmidt, G., & Tegethoff, M. (2015). How life before birth affects human health and what we can do about it. *European Psychologist, 20*(2), 85-89. doi: 10.1027/1016-9040/a000233

Millward Brown. (2014). Nestlé Studie: Was bewegt Eltern in der Schweiz. Retrieved from https://http://www.nestle.ch/de/documents/nestle%C2%A6%C3%BC_studie_was_bewegt_eltern_in_der_schweiz_reporting.pdf website: https://http://www.nestle.ch

Murray, E., Fernandes, M., Fazel, M., Kennedy, S. H., Villar, J., & Stein, A. (2015). Differential effect of intrauterine growth restriction on childhood neurodevelopment: a systematic review. *Bjog-an International Journal of Obstetrics and Gynaecology, 122*(8), 1062-1072. doi: 10.1111/1471-0528.13435

National Institutes of Health. (2003). *Pregnancy and perinatology branch strategic plan 2005-2010* Washington, DC: National Institutes of Health.

Nunes, L. A. S., Mussavira, S., & Bindhu, O. S. (2015). Clinical and diagnostic utility of saliva as a non-invasive diagnostic fluid: a systematic review. *Biochemia Medica, 25*(2), 177-192.

Poon, L. C. Y., Syngelaki, A., Akolekar, R., Lai, J., & Nicolaides, K. H. (2013). Combined Screening for Preeclampsia and Small for Gestational Age at 11-13 Weeks. *Fetal Diagnosis and Therapy, 33*(1), 16-27. doi: 10.1159/000341712

Poudevigne, M. S., & O'Connor, P. J. (2006). A review of physical activity patterns in pregnant women and their relationship to psychological health. *Sports Medicine, 36*(1), 19-38.

Pruessner, J. C., Wolf, O. T., Hellhammer, D. H., Buske-Kirschbaum, A., Auer, K. v., Jobst, S., . . . Kirschbaum, C. (1997). Free cortisol levels after awakening: a reliable biological marker for the assessment of adrenocortical activity. *Life Sciences, 61*(26), 2539-2549.

Robbins, M. L., & Kubiak, T. (2014). Ecological Momentary Assessment in Behavioral Medicine Research and Practice. *Handbook of Behavioral Medicine*, 429-446. doi: Book_Doi 10.1002/9781118453940

Rousham, E. K., Clarke, P. E., & Gross, H. (2006). Significant changes in physical activity among pregnant women in the UK as assessed by accelerometry and self-reported activity. *European Journal of Clinical Nutrition, 60*(3), 393-400. doi: Doi 10.1038/Sj.Ejcn.1602329

Salam, R. A., Das, J. K., & Bhutta, Z. A. (2014). Impact of intrauterine growth restriction on long-term health. *Current Opinion in Clinical Nutrition and Metabolic Care, 17*(3), 249-254. doi: 10.1097/MCO.0000000000000051

Sathishkumar, K., Elkins, R., Chinnathambi, V., Gao, H. J., Hankins, G. D. V., & Yallampalli, C. (2011). Prenatal testosterone-induced fetal growth restriction is associated with down-regulation of rat placental amino acid transport. *Reproductive Biology and Endocrinology, 9*. doi: Artn 11010.1186/1477-7827-9-110

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1-32. doi: Doi 10.1146/Annurev.Clinpsy.3.022806.091415

Sibbald, B., & Roland, M. (1998). Understanding controlled trials: why are randomised controlled trials important? *BMJ (Clinical Research Ed.), 316*(7126), 201.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurence*. Oxford: Oxford University Press.

Smith, V. C., & Dougherty, L. R. (2014). Noisy Spit: Parental Noncompliance with Child Salivary Cortisol Sampling. *Developmental Psychobiology, 56*(4), 647-656. doi: 10.1002/dev.21133

Stalder, T., Kirschbaum, C., Kudielka, B. M., Adam, E. K., Pruessner, J. C., Wust, S., . . . Clow, A. (2016). Assessment of the cortisol awakening response: expert consensus guidelines. *Psychoneuroendocrinology, 63*, 414-432. doi: 10.1016/j.psyneuen.2015.10.010

Streckfus, C. F., & Bigler, L. R. (2002). Saliva as a diagnostic fluid. *Oral Diseases, 8*(2), 69-76. doi: Doi 10.1034/J.1601-0825.2002.1o834.X

Strimbu, K., & Tavel, J. A. (2010). What are biomarkers? *Current Opinion in Hiv and Aids, 5*(6), 463-466. doi: 10.1097/COH.0b013e32833ed177

Suhrcke, M., McKee, M., & Rocco, L. (2005). Health: a vital investment for economic development and poverty reduction in Eastern Europe and Central Asia. *European Journal of Public Health, 15*, 110-110.

Tegethoff, M., Greene, N., Olsen, J., Schaffner, E., & Meinlschmidt, G. (2011). Stress during pregnancy and offspring pediatric disease: a national cohort study. *Environmental Health Perspectives, 119*(11), 1647-1652. doi: Doi 10.1289/Ehp.1003253

UN. (2013). Convention on the Rights of the Child: general comment No. 15 (2013) on the right of the child to the enjoyment of the highest attainable standard of health (art. 24). Retrieved from http://www2.ohchr.org/english/bodies/crc/docs/GC/CRC-C-GC-15_en.doc website: http://www.ohchr.org

Veiga-Lopez, A., Steckler, T. L., Abbott, D. H., Welch, K. B., MohanKumar, P. S., Phillips, D. J., . . . Padmanabhan, V. (2011). Developmental programming: impact of excess

prenatal testosterone on intrauterine fetal endocrine milieu and growth in sheep. *Biology of Reproduction, 84*(1), 87-96. doi: 10.1095/biolreprod.110.086686

Voegtline, K. M., Costigan, K. A., Kivlighan, K. T., Henderson, J. L., & DiPietro, J. A. (2013). Sex-specific associations of maternal prenatal testosterone levels with birth weight and weight gain in infancy. *Journal of Developmental Origins of Health and Disease, 4*(4), 280-284. doi: Doi 10.1017/S2040174413000135

Wadhwa, P. D., Buss, C., Entringer, S., & Swanson, J. M. (2009). Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms. *Seminars in Reproductive Medicine, 27*(5), 358-368. doi: 10.1055/s-0029-1237424

Zeitlin, J., Saurel-Cubizolles, M. J., de Mouzon, J., Rivera, L., Ancel, P. Y., Blondel, A., & Kaminski, M. (2002). Fetal sex and preterm birth: are males at greater risk? *Human Reproduction, 17*(10), 2762-2768. doi: Doi 10.1093/Humrep/17.10.2762

Zhang, J., Merialdi, M., Platt, L. D., & Kramer, M. S. (2010). Defining normal and abnormal fetal growth: promises and challenges. *American Journal of Obstetrics and Gynecology, 202*(6), 522-528. doi: 10.1016/j.ajog.2009.10.889

Appendix A-D

Appendix A (Manuskript A)

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(1), e86204. doi: 10.1371/journal.pone.0086204

Appendix B (Manuskript B)

Moeller, J., Lieb, R., Meyer, A. H., Quack Loetscher, K., Krastel, B., & Meinlschmidt, G. (2014). Nonadherence with ambulatory saliva sampling is associated with biased salivary testosterone estimates. Psychoneuroendocrinology, 44(0), 13-19. doi: http://dx.doi.org/10.1016/j.psyneuen.2014.02.012

Appendix C (Manuskript C)

Moeller, J., Lieb, R., Meyer, A. H., Quack Loetscher, K., Krastel, B., & Meinlschmidt, G. (submitted to *Psychoneuroendocrinology*). Women's diurnal salivary testosterone change during pregnancy is associated with offspring size at birth.

Appendix D

Curriculum Vitae von Julian Möller

Appendix A

Improving ambulatory saliva-sampling compliance in pregnant women: a randomized controlled study

(Published in PLoS ONE)

OPEN ACCESS Freely available online

PLOS ONE

Improving Ambulatory Saliva-Sampling Compliance in Pregnant Women: A Randomized Controlled Study

Julian Moeller[1,2], Roselind Lieb[1], Andrea H. Meyer[1], Katharina Quack Loetscher[3], Bettina Krastel[4], Gunther Meinlschmidt[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, University of Basel, Basel, Switzerland, **3** Department of Obstetrics, University Hospital Zurich, Zurich, Switzerland, **4** National Centre of Competence in Research (NCCR), Swiss Etiological Study of Adjustment and Mental Health (sesam), University of Basel, Basel, Switzerland, **5** Faculty of Medicine, Ruhr-University Bochum, Bochum, Germany

Abstract

Objective: Noncompliance with scheduled ambulatory saliva sampling is common and has been associated with biased cortisol estimates in nonpregnant subjects. This study is the first to investigate in pregnant women strategies to improve ambulatory saliva-sampling compliance, and the association between sampling noncompliance and saliva cortisol estimates.

Methods: We instructed 64 pregnant women to collect eight scheduled saliva samples on two consecutive days each. Objective compliance with scheduled sampling times was assessed with a Medication Event Monitoring System and self-reported compliance with a paper-and-pencil diary. In a randomized controlled study, we estimated whether a disclosure intervention (informing women about objective compliance monitoring) and a reminder intervention (use of acoustical reminders) improved compliance. A mixed model analysis was used to estimate associations between women's objective compliance and their diurnal cortisol profiles, and between deviation from scheduled sampling and the cortisol concentration measured in the related sample.

Results: Self-reported compliance with a saliva-sampling protocol was 91%, and objective compliance was 70%. The disclosure intervention was associated with improved objective compliance (informed: 81%, noninformed: 60%), $F(1,60) = 17.64$, $p<0.001$, but not the reminder intervention (reminders: 68%, without reminders: 72%), $F(1,60) = 0.78$, $p = 0.379$. Furthermore, a woman's increased objective compliance was associated with a higher diurnal cortisol profile, $F(2,64) = 8.22$, $p<0.001$. Altered cortisol levels were observed in less objective compliant samples, $F(1,705) = 7.38$, $p = 0.007$, with delayed sampling associated with lower cortisol levels.

Conclusions: The results suggest that in pregnant women, objective noncompliance with scheduled ambulatory saliva sampling is common and is associated with biased cortisol estimates. To improve sampling compliance, results suggest informing women about objective compliance monitoring but discourage use of acoustical reminders.

Citation: Moeller J, Lieb R, Meyer AH, Loetscher KQ, Krastel B, et al. (2014) Improving Ambulatory Saliva-Sampling Compliance in Pregnant Women: A Randomized Controlled Study. PLoS ONE 9(1): e86204. doi:10.1371/journal.pone.0086204

Editor: Harpal Singh Randeva, University of Warwick – Medical School, United Kingdom

Received May 4, 2013; **Accepted** December 8, 2013; **Published** January 22, 2014

Funding: This work is part of the National Centre of Competence in Research (NCCR) Swiss Etiological Study of Adjustment and Mental Health (sesam). The Swiss National Science Foundation (SNSF) (project no. 51A240-104890), the University of Basel, the F. Hoffmann-La Roche Corp., and the Freie Akademische Gesellschaft provided core support for the NCCR sesam. Additionally, G.M. receives SNSF funding under project no. 100014_135328. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: This work is part of the National Centre of Competence in Research (NCCR) Swiss Etiological Study of Adjustment and Mental Health (sesam). The Swiss National Science Foundation (SNSF) (project no. 51A240-104890), the University of Basel, the F. Hoffmann-La Roche Corp., and the Freie Akademische Gesellschaft provided core support for the NCCR sesam. Additionally, G.M. receives SNSF funding under project no. 100014_135328. The funding sources had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. Moreover, the funding sources do not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: gunther.meinlschmidt@unibas.de

Introduction

Maternal stress during pregnancy can adversely affect birth outcomes and offspring development (e.g. [1,2]). Cortisol as a stress marker, released by the maternal hypothalamic–pituitary–adrenal (HPA) axis, may partly explain these effects (for reviews see [3–6]). Therefore, psychoneuroendocrine research examining cortisol in pregnant women has been given high priority and may contribute to a better understanding of the biochemical mechanism underlying the adverse effects described above [4]. Research examining cortisol in pregnant women may also improve risk assessment for adverse birth outcomes [7,8].

In ambulatory settings, cortisol concentrations are commonly assessed noninvasively in saliva. Subjects are instructed to collect a certain number of saliva samples at scheduled sampling times on one or more consecutive study days [9–11]. Compared to laboratory research, ambulatory research results in higher ecological validity [12]. However, ambulatory saliva sampling

may be biased by noncompliance: Subjects may not follow scheduled sampling times and may fail to self-report this noncompliance in paper-and-pencil diaries – even if study collaborators stress the importance of both requirements. Indeed, such patterns have been observed in studies using hidden electronic compliance-monitoring systems, comparing subjects' objective compliance to their self-reported compliance [13–15]. Noncompliance with scheduled saliva-sampling times has been associated with biased cortisol estimates due to the cortisol circadian rhythm [13,14,16]. Biased estimates of cortisol concentrations may cause invalid interpretations of the data. In contrast, Jacobs et al. [15] reported no biased cortisol estimates when noncompliant saliva samples were included in analyses. Saliva-sampling noncompliance has also been associated with additional study costs and, in the case of missing samples, with reduced statistical power [17].

To deal with the compliance problem, experimental studies tested whether informing subjects about objective compliance monitoring improves compliance with scheduled saliva sampling. In these studies, subjects who were informed about monitoring displayed higher compliance with the sampling protocols compared to noninformed subjects [13,14]. Acoustical reminders such as preprogrammed wristwatches have been used to improve saliva sampling compliance [18,19], although, to our knowledge, their effect on saliva-sampling compliance has not yet been experimentally tested, and experimental evidence that they improve compliance comes from other research fields only: For example, in a review, reminders improved medication compliance in antiretroviral therapy in four of eight studies [20]. Moreover, electronic reminders improved participants' compliance with paper pain diaries, but still, according to Broderick and colleagues [21], the compliance rates were unsatisfactory.

In sum, ambulatory saliva sampling has gained great importance in psychoneuroendocrine research, being used to examine cortisol concentrations in pregnant women. However, findings that noncompliance with saliva sampling is common and can bias cortisol estimates, and that informing subjects about objective compliance monitoring improves saliva-sampling compliance, are based on ambulatory research in healthy volunteers, patients, and an older population [13,14,16]. It is unclear whether these findings can be generalized to pregnant women, especially as pregnant women display different behavioral patterns (e.g. reduced physical activity, more sitting, lying, sleeping, and slower walking pace) [22–24], altered cortisol levels, and altered cortisol responses to stress [5,25,26] compared to nonpregnant controls or to nonpregnant state. To our knowledge, it has not yet been investigated in a sample of pregnant women whether informing them about objective compliance monitoring improves compliance with scheduled saliva sampling, or whether noncompliance biases cortisol estimates. Moreover, to our knowledge, whether acoustical reminders improve saliva-sampling compliance has not yet been experimentally tested at all.

The goals of the present study were a) to estimate compliance rates with a standard ambulatory saliva-sampling protocol; b) to estimate whether the strategies of informing subjects about objective compliance monitoring and using acoustical reminders improve compliance with scheduled saliva sampling; and c) to estimate the association between saliva-sampling noncompliance and saliva cortisol concentrations in pregnant women.

Methods

Ethics Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committees of Basel (Ethikkommission beider Basel, Basel, Switzerland) and Zurich (Kantonale Ethikkommission Zuerich, Zuerich, Switzerland). All participants gave written informed consent.

Subjects

We recruited pregnant women during antenatal visits at the outpatient service of the Department of Obstetrics, University Hospital Zurich, Switzerland. Eligible women were in their 12^{th} to 32^{nd} week of gestation, had sufficient German language skills, and underwent regular antenatal visits at the outpatient service. Exclusion criteria were diseases potentially affecting the neuroendocrine system, high-risk pregnancy, human immunodeficiency virus (HIV) infection, and the use of hormone-containing medication. These criteria were chosen to minimize distortions in the women's cortisol concentrations.

Experimental interventions and design

We told all the women to collect eight saliva samples on two consecutive days. Scheduled sampling times were: 0, 30, 45, and 60 min after awakening, and at 1100, 1500, 2000, and 2200 h. Objective compliance with the time of sampling was monitored, in all women, with a hidden Medication Event Monitoring System (MEMS 6 TrackCap Monitor, Aardex Ltd., Switzerland). Two interventions were tested by creating four experimental groups: Group 1 received information about the objective compliance monitoring at the beginning of the study and received timers (Kuechentimer, Zyliss, Switzerland) and alarm clocks (basic alarm clock, Intertronic, Switzerland) to provide acoustical reminders at the scheduled sampling times. The second group received only the information about the objective compliance monitoring. The third group received the acoustical reminders alone, and the fourth group received neither the information about the objective compliance monitoring nor acoustical reminders. Women in the groups receiving the acoustical reminders were advised to use the timer to time the samples at +30, +45, and +60 min after awakening and the clock to time the 1100 h, 1500 h, 2000 h, and 2200 h samples, even if they possessed their own watch.

In sum, a randomized 2 (disclosure intervention: informed vs. noninformed) ×2 (reminder intervention: acoustical reminders vs. no acoustical reminders) ×2 (days) ×8 (sampling times) design was applied. A blocked randomization sequence was created with a computerized random number generator and applied with a 1:1:1:1 assignment to the four experimental groups. Study collaborators using sealed envelopes conducted the assignment. All women were blinded to the true nature of the present study; laboratory staff analyzing the saliva samples were blinded to the women's experimental group assignment.

Procedure

During an antenatal visit, obstetricians told women who met the inclusion criteria about the study. A study collaborator gave standardized information to interested women. Women who agreed to participate were assigned to the experimental groups as described above and received packages with the respective study materials. A study collaborator, stressing the importance of high compliance with the sampling protocol, explained the use of the material in detail. The women were instructed to collect saliva samples on two consecutive days right before their next scheduled antenatal visit. Three days before sampling, a study collaborator

contacted them by telephone as a reminder and to answer any questions. The women provided demographic information, including age, height, employment status, number of hours worked per week, prepregnancy body weight, current body weight, gestational age of their fetus, gravidity, and parity, via questionnaire. They were instructed not to brush their teeth, eat, consume caffeine, or smoke during the first hour after awakening and 1 h before each scheduled saliva sampling. They were also advised to avoid physical exercise but to otherwise follow their daily routine. These restrictions were meant to minimize distortions in the women's cortisol concentrations. The women handed over the study material to a study collaborator the day after sampling, at the antenatal visit. On this occasion, we asked the women, by questionnaire, whether there was anything that had attracted particular attention or was particularly noticeable during the study. In doing so, we sought to test whether noninformed women obtained knowledge about objective compliance monitoring. Finally, the women were debriefed about the true nature of the study.

Saliva sampling

Straws and 2.0-mL safe-lock tubes (Eppendorf, Hamburg, Germany), labeled with scheduled sampling time, were clearly arranged until usage in a transparent MediDispenser (Wiegand, Buelach, Switzerland). Women were instructed to place the tubes immediately after saliva sampling into small nontransparent medicine containers (Wiegand, Buelach, Switzerland) fitted with MEMS 6 caps. As a cover story, all the women were told that this procedure was important to maintain sample quality by minimizing light exposure. For the same reason, they were advised to open the medicine container only to insert the saliva samples. This container was to be stored overnight and, when possible, in a refrigerator.

Biochemical analyses

We froze returned saliva samples at $-20^{\circ}C$ until biochemical analysis. Thawed samples were centrifuged at 3000 g for 10 min. Salivary free cortisol was analyzed using a commercial enzyme immunoassay for human saliva (cortisol ELISA, IBL, Hamburg, Germany). Analytical assay sensitivity was 2.0 pg/mL. The intra- and interassay coefficients of variation were $\leq$7.3% and $\leq$9.3%, respectively.

Compliance with saliva sampling

We assessed self-reported and objective compliance with scheduled saliva-sampling times. Self-reported compliance was assessed with a paper-and-pencil diary, in which the women were asked to record the exact time and date of each saliva sampling. Objective compliance was assessed with the MEMS 6 caps that recorded the moment of each opening and closing of the medical container. The opening times of the MEMS 6 caps were processed with PoverView (Aardex Ltd., Switzerland). Compliance criteria were adapted from Kudielka et al. [14] and applied for both self-report and objective compliance. Accordingly, we classified the +0-min sample as compliant if collected within $\pm$10 min of the self-reported wake-up time, the +30-, +45-, and +60-min samples as compliant if collected within $\pm$7 min, and the 1100 h, 1500 h, 2000 h, and 2200 h samples as compliant if collected within $\pm$1 h of the scheduled sampling time. In the case of multiple MEMS 6 cap openings around the scheduled sampling times, we selected the most compliant. If a women delivered more saliva samples than recorded MEMS 6 cap opening times, we selected the most compliant opening times for the delivered samples and classified the remaining samples as noncompliant.

Statistical analyses

In the first set of analyses, we estimated the association of the two interventions with objective compliance using general linear models (GLMs). Disclosure intervention (informed vs. noninformed) and reminder intervention (acoustical reminders vs. no acoustical reminders) were the two fixed independent factors, and objective compliance (percentage of compliant samples) was the dependent variable. The +0-min samples were excluded, as we could not objectively determine whether the women reported their wake-up times accurately. We repeated the analysis described above with objective morning compliance (+30-, +45-, and +60-min samples only; percentage of compliant samples) as the dependent variable. High compliance with the morning samples is considered especially important as the cortisol awakening response (CAR) is often used for research purposes as indicator of HPA reactivity [11]. The CAR represents the rapid steep increase of cortisol concentrations within the first 30 min of awakening [11,27].

In the second set of analyses, we estimated the association of objective compliance with saliva cortisol concentrations using a random coefficient model, a type of linear mixed model [28]. This type of model has been shown to provide more efficient and less biased results in data where missing values occur, compared with complete case analyses or analyses in which missing values are imputed using the last observation carried forward method [29]. Further, linear mixed models do not require omitting subjects with missing data from the analyses, thereby minimizing data loss and risk of bias while increasing power. Our model included a random intercept as well as a random slope parameter when this improved model fit (based on Akaike's Information Criterion, AIC) [28].

The random coefficient model allowed us to differentiate between objective state and trait compliance. Trait compliance is a time-invariant predictor and measuring it allowed us to estimate whether the women's objective compliance with the sampling schedule was associated with their diurnal profiles of cortisol concentrations. We estimated the effect of the women's objective compliance by dividing the women into a low (0–5 compliant samples; 0–31% of all scheduled samples), a moderate (6–12 compliant samples; >31–80%), and a high (13–16 compliant samples; >80–100%) compliance group. We used the categorical compliance predictor instead of the continuous compliance predictor "number of compliant samples" because preliminary analyses revealed better model fits for the former. For the high compliance group, we chose the 80% cut-off because prior research used this cut-off to classify compliance in a cortisol-sampling protocol [30] and because medical research usually classifies patients with compliance of more than 80% as compliant [30,31]. To enlarge the small sample size in the low compliance group, we chose a 31% cut-off (0–5 compliant samples) instead of the 20% cut-off applied by Hall et al. [30]. For the predictor time we assumed linear trajectories for each of two intervals covering the time points +0- to +30-min and +30-min to the last time point (2200 h), respectively.

State compliance relates to individual saliva samples and allowed us to estimate whether a deviation from a scheduled sampling time was associated with the cortisol concentration measured in the related sample. We estimated this association by entering the time-varying predictor "deviation from scheduled sampling time in minutes" into our model. Again, we excluded the +0-min samples for the same reasons as stated above. The predictor time was again assumed to be linear, but we only considered the interval +30-min to 2200 h. We repeated our mixed model analyses, adjusting for several a priori defined potential time-invariant confounders, including the continuous

covariates age, gestational age of fetus, parity [32], and current body weight [33].

The percentage of compliant samples was arcsine transformed, cortisol data were square root transformed, and deviations from scheduled sampling times in minutes were log transformed to approximate normal distributions. An alpha level of 0.05 determined statistical significance. Data analysis was carried out using IBM SPSS Statistics 20 for Mac OS X.

Results

We included 75 eligible women in the present study. Six women declined further participation before saliva sampling. Two women were excluded because of a MEMS 6 cap defect and two because they collected saliva samples without using the MEMS 6 caps. Another woman was excluded because she took part only on the first day of the study because she delivered prematurely on the second day. Thus, the final sample consisted of 64 women. Demographic information is presented in Table 1. None of the noninformed women reported any knowledge of the objective compliance monitoring on the questionnaire before the debriefing.

Compliance with the saliva-sampling protocol and interventions

Self-reported compliance and objective compliance refer to all samples (+30-, +45-, +60-min, 1100 h, 1500 h, 2000 h, and 2200 h samples), and objective morning compliance refers to the +30-, +45-, and +60-min samples only. Across all the women, self-reported compliance with the saliva-sampling protocol was 91% and objective compliance was 70%. Self-reported compliance was high in all experimental groups (range 88–94%). Objective compliance was highest in the informed group without acoustical reminders (86%) and lowest in the noninformed group without acoustical reminders (58%). The women's objective morning compliance was lower (59%) compared to their objective compliance reported above (70%). Moreover, self-reported compliance in women with low objective compliance (0–31% compliant samples) was 80%. Descriptive compliance data are presented in Table 2.

Objective compliance in informed and noninformed women was 81% and 60%, and in women with and without acoustical reminders 68% and 72%, respectively. The GLM showed significant main effects of disclosure intervention (informed vs. noninformed) on both objective compliance, $F(1,60) = 17.64$, $p<0.001$, and objective morning compliance, $F(1,60) = 9.27$, $p = 0.003$. However, there was no significant main effect of reminder intervention (acoustical reminders vs. no acoustical reminders) on either compliance type [objective compliance, $F(1,60) = 0.78$, $p = 0.379$; objective morning compliance, $F(1,60) = 0.80$, $p = 0.374$]. Interaction effects between disclosure intervention and reminder intervention were also nonsignificant for both

Table 1. Demographic variables[a] across experimental groups.

Variable	Total	Compliance monitoring			
	(*n* = 64)	Informed		Noninformed	
		Acoustical reminders		Acoustical reminders	
		With	Without	With	Without
		(*n* = 16)	(*n* = 17)	(*n* = 15)	(*n* = 16)
Age (years)	33 (28;36)	33 (27;39)	33 (28;35)	34 (32;40)	31 (26;35)
Height (cm)	167 (163;172)	163 (162;172)	164 (161;170)	170 (166;173)	168 (165;172)
Employed[b]					
Yes	45 (70.3)	10 (62.5)	14 (82.4)	9 (60)	12 (75)
No	16 (25)	5 (31.3)	2 (11.8)	5 (33.3)	4 (0)
Unknown	3 (4.7)	1 (6.3)	1 (5.9)	1 (6.7)	0 (0)
Hours worked per week	28 (17;42)	35 (11;42)	25 (15;41)	40 (20;51)	28 (24;42)
Prepregnancy body weight (kg)	60 (56;68)	59 (56;71)	59 (54;64)	61 (58;69)	59 (55;67)
Current body weight (kg)	70 (62;75)	67 (63;76)	68 (58;74)	71 (64;85)	72 (60;75)
Gestational age of fetus	26 (17;31)	26 (16;30)	22 (17;31)	29 (21;34)	24 (17;29)
Gravidity[b]					
0	28 (43.8)	7 (43.8)	6 (35.3)	6 (40)	9 (56.3)
1–2	21 (32.8)	4 (25)	6 (35.3)	5 (33.4)	6 (37.5)
≥3	9 (14.2)	2 (12.6)	3 (17.7)	3 (20)	1 (6.3)
Unknown	6 (9.4)	3 (18.8)	2 (11.8)	1 (6.7)	0 (0)
Parity[b]					
0	34 (53.1)	9 (56.3)	8 (47.1)	7 (46.7)	10 (62.5)
1–2	21 (32.8)	3 (18.8)	6 (35.3)	6 (40)	6 (37.5)
≥3	3 (4.7)	1 (6.3)	1 (5.9)	1 (6.7)	0 (0)
Unknown	6 (9.4)	3 (18.8)	2 (11.8)	1 (6.7)	0 (0)

[a]If not otherwise specified, median (25 percentile; 75 percentile) is reported.
[b]Number of pregnant women (percent) is reported.
doi:10.1371/journal.pone.0086204.t001

Table 2. Self-reported and objective compliance with scheduled saliva sampling across experimental groups.

Descriptive data	Self-reported compliance					Objective compliance				
	Informed		Non-informed		Total	Informed		Non-informed		Total
	Acoustical reminders		Acoustical reminders			Acoustical reminders		Acoustical reminders		
	With	Without	With	Without		With	Without	With	Without	
	(*n* = 15)	(*n* = 16)	(*n* = 14)	(*n* = 15)	(*n* = 60)[a]	(*n* = 16)	(*n* = 17)	(*n* = 15)	(*n* = 16)	(*n* = 64)
Morning samples[b]										
Number of scheduled samples	90	96	84	90	360	96	102	90	96	384
Number of compliant samples	73	88	76	74	311	60	76	43	46	225
Compliant samples in % (SD)	81% (25.9)	92% (25.1)	90% (18.2)	82% (30.5)	86% (25.2)	62% (24.0)	75% (25.1)	48% (28.8)	48% (30.3)	59% (28.8)
All samples[c]										
Number of scheduled samples	210	224	196	210	840	224	238	210	224	896
Number of compliant samples	185	210	177	192	764	168	204	129	130	631
Compliant samples in % (SD)	88% (13.4)	94% (12.2)	90% (21.3)	91% (13.5)	91% (15.1)	75% (20.2)	86% (16.4)	61% (25.6)	58% (25.5)	70% (24.3)
Summed absolute deviations from scheduled sampling times in minutes[c]										
Median	34	51	53	25	47	145	92	248	185	147
25th percentile; 75th percentile	12; 120	28; 92	40; 107	0; 80	15; 99	60, 281	56, 190	147; 553	87; 369	74; 274

[a]Number in sample for self-reported compliance is four less than number in sample for objective compliance because of missing data in self-report questionnaires.
[b]Including the +30-min, +45-min, and +60-min samples on two consecutive days.
[c]Including the +30-min, +45-min, +60-min samples and the 1100 h, 1500 h, 2000 h, and 2200 h samples on two consecutive days.
SD, standard deviation.
doi:10.1371/journal.pone.0086204.t002

compliance types [objective compliance, $F(1,60) = 2.46$, $p = 0.122$; objective morning compliance, $F(1,60) = 0.77$, $p = 0.385$].

Objective compliance and cortisol concentrations

Objective compliance information was used to estimate the associations between compliance with scheduled sampling and cortisol concentrations.

Cortisol concentrations in high-, moderate-, and low-compliance women. Twenty-eight women (44%) showed high objective compliance with the saliva-sampling protocol, 29 (45%) moderate objective compliance, and seven (11%) low objective compliance. Using random coefficient models, we compared the cortisol concentrations of women with high, moderate, and low objective compliance. We found a main effect for objective compliance on cortisol concentrations, $F(2,64) = 8.22$, $p<0.001$, which was due almost entirely to the difference in cortisol concentrations between objective low-compliance women on the one hand and the combined objective moderate- and high-compliance women on the other, $F(1,74) = 16.14$, $p<0.001$, for contrast. However, there was no difference in cortisol concentrations between objective moderate- and high-compliance women, $F(1,63) = 0.41$, $p = 0.525$. We also found an interaction effect between objective compliance and time of saliva measurement on cortisol concentrations, $F(2,64) = 5.26$, $p = 0.008$. As shown in Figure 1, cortisol levels of women in the low-compliance group were lower than cortisol levels of women in the moderate- and high-compliance groups, but this effect slowly disappeared throughout the day, such that at the 2200 h scheduled sampling time, women at the three levels of compliance displayed comparable cortisol levels. Accordingly, if we considered morning cortisol concentrations only (+0-, +30-, +45-, and +60-min samples), we still found a main effect of objective compliance on morning cortisol concentrations, $F(2,64) = 10.24$, $p<0.001$. Intraindividual variation in these morning concentrations is indicative of the CAR. The associations stated above did not change significantly after adjustment for age, current body weight, gestational age of fetus, and parity (data available on request).

Deviation from scheduled sampling and the cortisol concentration in the related sample. A total of 753 saliva samples were included in this analysis. An objective deviation from a scheduled sampling time was associated with the cortisol level measured in the related sample, $F(1,705) = 7.38$, $p = 0.007$; that is, the longer the time delay from a scheduled sampling, the lower the cortisol level. The cortisol levels (on the square-root scale) decreased per minute deviation (on the natural logarithm scale) by a value of 0.82 (SE 0.30). The nonsignificant interaction effect between deviation from scheduled sampling time and time of

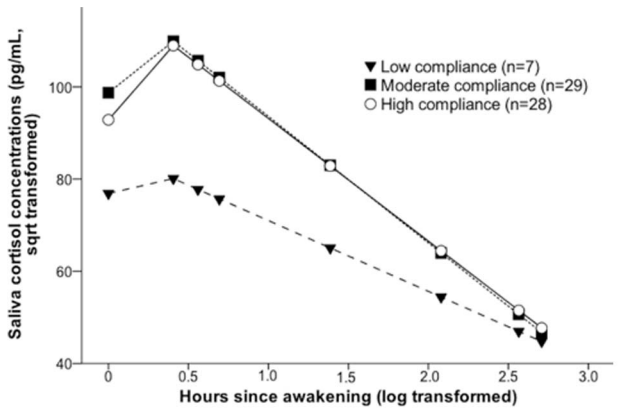

Figure 1. Cortisol concentrations in women with low, moderate, and high compliance. Saliva cortisol concentrations were averaged across two sampling days representing estimated values from a linear mixed model.
doi:10.1371/journal.pone.0086204.g001

saliva measurement on cortisol concentrations, $F(1,713) = 2.50$, $p = 0.115$, indicated a stable association of deviation from scheduled sampling time with cortisol concentrations throughout the day. The associations stated above did not change significantly after adjustment for age, current body weight, gestational age of fetus, and parity (data available on request).

Discussion

Objective noncompliance with scheduled saliva sampling was associated with biased cortisol estimates in pregnant women. Informing women about the compliance monitoring improved objective compliance with scheduled saliva sampling. In contrast, the use of acoustical reminders had no effect on objective compliance.

Compliance with the saliva-sampling protocol

The women's self-reported compliance was higher than their objective compliance with the saliva-sampling protocol, especially when they were not informed about the compliance monitoring. Even the women with low objective compliance (0–31% compliant samples) self-reported, on average, high compliance. These findings are in line with prior studies in nonpregnant subjects showing a possible bias in self-reported saliva-sampling compliance [13–15].

In the present study, women informed about compliance monitoring displayed objective compliance rates of 75% (with acoustical reminders) and 86% (without acoustical reminders). By comparison, prior studies found objective compliance rates in informed subjects of 89–97% [13,14]. Objective compliance rates in our noninformed women were 61% (with acoustical reminders) and 58% (without acoustical reminders). In contrast, noninformed subjects in prior studies had objective compliance rates of 62–84% [13,14]. Several reasons may explain our somewhat lower objective compliance rates compared to the prior studies: First, there may be specific behavioral patterns in pregnant women [22–24]. Second, there is the issue of saliva-sampling burden. Kudielka et al. [16] hypothesized that a higher sampling burden related to the number of scheduled saliva samples per day may lead to lower saliva-sampling compliance. In line with this, the daily saliva-sampling burden in the present study (eight samples per day) was higher than that in the studies of Kudielka et al. (six samples per day; [14]) and Broderick et al. (five samples per day; [13]). Third, other differences in study design may account for lower compliance rates in the present study (e.g. we applied more conservative compliance criteria regarding the morning samples compared to Broderick et al.) [13]. Moreover, we cannot exclude further factors in pregnant women associated with lower objective saliva-sampling compliance.

Associations of the interventions with objective compliance

In this randomized controlled trial, the disclosure intervention (informing about compliance monitoring) was associated with higher objective compliance with the saliva-sampling protocol. This finding is in line with prior studies in nonpregnant subjects [13,14] and suggests that informing about compliance monitoring improves saliva-sampling compliance in pregnant women. Thus, the present study extends earlier results to a sample of pregnant women. In the present study, informing the women was associated with higher objective compliance with respect to both all scheduled samples and the scheduled morning samples. High compliance with the morning samples is particularly relevant because the CAR has been used extensively as an indicator of HPA activity, and because the CAR is increasingly relevant in endocrine research in pregnant women (e.g., [7,26,34,35]).

In contrast, we did not find any positive effect of the reminder intervention on objective compliance. Using acoustical reminders was not associated with improved objective saliva-sampling compliance. One possible explanation for this observation is that carrying timers and alarm clocks was rather inconvenient. Indeed, several women reported this during the debriefing at the end of the study. To our knowledge, the present study is the first to investigate the association of acoustical reminders with saliva-sampling compliance. However, in the research field of antiretroviral therapy in HIV treatment, a review described that the use of electronic reminders improved medication compliance in four of eight studies [20]. This review, however, included studies relying on self-report measures. In a recent randomized controlled trial, the use of pocket digital alarms had no effect on objective medication compliance, as measured by the percentage of dispensed drug doses [36]. Thus, in line with the latter, our data discourages the use of acoustical reminders to improve saliva-sampling compliance in pregnant women: While having no positive effect on compliance, the use of acoustical reminders increases the study burden on women and generates additional study costs.

Associations of objective compliance with cortisol concentrations

The women's objective compliance with the saliva-sampling protocol was associated with their cortisol concentrations. Women with low compliance displayed lower cortisol levels compared to those with moderate or high compliance. In detail, women with low compliance showed lower CARs and downward-shifted day slopes of cortisol compared to women with moderate and high compliance. One explanation could be that women with low compliance deliver samples more often with a delay, which – due to the diurnal decline of cortisol concentrations – is likely to be associated with lower cortisol levels, resulting in lower levels on average. Alternatively, being low compliant may be related to certain trait characteristics, which in turn may be associated with lower cortisol levels. Cortisol levels did not differ between women with moderate and high compliance. Hence, low compliance may bias cortisol estimates, but moderate compliance may have less impact. The finding that noncompliance may bias cortisol results is in line with prior studies in nonpregnant subjects [13,14,16].

Without objective compliance monitoring, we would not have been able to identify the biased cortisol estimates of women with low compliance, as they incorrectly self-reported high compliance. Thus, without objective compliance information, cortisol slopes or CARs, biased by low compliance, could lead to invalid interpretations. For example, prior research has associated lower morning cortisol levels with cumulative stress in pregnant women [37]. Without objective compliance information, it might be difficult to conclude whether lower morning cortisol levels are directly associated with cumulative stress or with a bias introduced by saliva-sampling noncompliance related to stress (compare [14]). In the present study, the association between women's objective compliance and their cortisol levels decreased through the day. This finding could be important as prior research used late afternoon or evening cortisol data instead of the CAR to examine specific research questions in pregnant women [38,39]. Based on our findings, when objective compliance information is not available, using evening samples may reduce the potential bias in cortisol estimates introduced by noncompliance.

In the present study, women's compliance was associated with their diurnal profiles of cortisol concentrations. Moreover, we

observed altered cortisol levels in less-compliant samples. The larger a time delay from a scheduled saliva sampling, the stronger was the bias by reduced cortisol levels. This finding confirms the importance of saliva-sampling compliance in pregnant women.

Limitations and strengths

The present study has several limitations. First, the sample size of our low-compliance group was small. Hence, our findings should be replicated in a larger sample. Second, the following hampered calculating a priori power analyses: To our knowledge, the present study is the first to estimate whether acoustical reminders improve saliva-sampling compliance; hence, estimation of effect sizes was difficult. Moreover, power analysis for mixed models requires information regarding estimates of intraclass correlations, which were not available in our case. Third, we could not control whether the women actually used the acoustical reminders. Thus, the nonsignificant association between reminder intervention and saliva-sampling compliance could be due to women not having applied the intervention (compare [36]). However, we were less interested in whether the use of acoustical reminders improved compliance, but mainly interested in whether the distribution of acoustical reminders and the recommendation of their use improved compliance. Fourth, without actigraph monitoring, we could not objectively define whether women reported their wake-up times accurately. However, evidence suggests that self-reported wake-up times are reasonably accurate, compared with objectively measured wake-up times [40,41]. Fifth, women may have collected saliva samples at scheduled times without storing them in the MEMS container. Putting several compliant saliva samples into the MEMS container at the same time would have led to missing MEMS 6 cap opening times and, thus, to objective compliant samples being classified as noncompliant. In the present study, this would have led to an underestimation of objective compliance rates. Last, findings regarding acoustical reminders may not be generalizable to other reminder systems. Further studies might examine whether other reminder systems (e.g. handheld computers, mobile apps) improve saliva-sampling compliance.

Despite the study's limitations, the present study has important strengths. First, we applied a standard two-day ambulatory saliva-sampling protocol, and second, we used a randomized controlled trial to test strategies to improve saliva-sampling compliance. Third, we used mixed model analyses to estimate the associations between compliance and cortisol concentrations. Mixed model analysis is considered the method of choice for analyzing ambulatory saliva cortisol data [11].

Conclusions

Our study findings indicate that informing about the use of objective compliance monitoring substantially improved saliva-sampling compliance in pregnant women. In contrast, using acoustical reminders had no positive effect. They should inform future studies examining cortisol in pregnant women within ambulatory saliva-sampling designs and are highly important for several reasons. First, noncompliance with a standard ambulatory saliva-sampling protocol was common in pregnant women and occurred more frequently than in prior studies with nonpregnant subjects. Second, noncompliant women could not be identified by self-report data. Third, objective noncompliance biased estimates of women's cortisol concentrations and, hence, may have led to invalid interpretations. Thus, the present study encourages using objective compliance monitoring to identify noncompliance with a saliva-sampling protocol in pregnant women. Moreover, it suggests informing women about objective compliance monitoring to improve compliance.

Acknowledgments

We thank Rene Angst, Tanja Angst, Monica Bachmann, Simone Briner Meier, Susan C.A. Burkhardt, Sigrid Falk, Gabriela Hunziker, Melanie Knabe, Laura Landi Degen, Cyrill Martin, Fabian Peter, Michael Pluess, Vera Schumacher and Anna Wiener for their valuable help with the project.

Author Contributions

Conceived and designed the experiments: GM RL. Performed the experiments: GM KQL BK. Analyzed the data: JM AHM. Wrote the paper: JM. Critically revised the manuscript for important intellectual content: GM RL AHM KQL BK.

References

1. Tegethoff M, Greene N, Olsen J, Schaffner E, Meinlschmidt G (2011) Stress during pregnancy and offspring pediatric disease: a national cohort study. Environ Health Perspect 119: 1647–1652.
2. Tegethoff M, Greene N, Olsen J, Meyer AH, Meinlschmidt G (2010) Maternal psychosocial adversity during pregnancy is associated with length of gestation and offspring size at birth: evidence from a population-based cohort study. Psychosom Med 72: 419–426.
3. Paarlberg KM, Vingerhoets AJ, Passchier J, Dekker GA, Van Geijn HP (1995) Psychosocial factors and pregnancy outcome: a review with emphasis on methodological issues. J Psychosom Res 39: 563–595.
4. Wadhwa PD (2005) Psychoneuroendocrine processes in human pregnancy influence fetal development and health. Psychoneuroendocrinology 30: 724–743.
5. de Weerth C, Buitelaar JK (2005) Physiological stress reactivity in human pregnancy – a review. Neurosci Biobehav Rev 29: 295–312.
6. Schetter CD (2011) Psychological science on pregnancy: stress processes, biopsychosocial models, and emerging research issues. Annu Rev Psychol 62: 531–558.
7. Buss C, Entringer S, Reyes JF, Chicz-DeMet A, Sandman CA, et al. (2009) The maternal cortisol awakening response in human pregnancy is associated with the length of gestation. Am J Obstet Gynecol 201: 398.e391–398.
8. Entringer S, Buss C, Andersen J, Chicz-DeMet A, Wadhwa PD (2011) Ecological momentary assessment of maternal cortisol profiles over a multiple-day period predicts the length of human gestation. Psychosom Med 73: 469–474.
9. Hellhammer DH, Wust S, Kudielka BM (2009) Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinology 34: 163–171.
10. Kirschbaum C, Hellhammer DH (1994) Salivary cortisol in psychoneuroendocrine research – recent developments and applications. Psychoneuroendocrinology 19: 313–333.
11. Kudielka BM, Gierens A, Hellhammer DH, Wust S, Schlotz W (2012) Salivary cortisol in ambulatory assessment – some dos, some don'ts, and some open questions. Psychosom Med 74: 418–431.
12. Shiffman S, Stone AA, Hufford MR (2008) Ecological momentary assessment. Annu Rev Clin Psychol 4: 1–32.
13. Broderick JE, Arnold D, Kudielka BM, Kirschbaum C (2004) Salivary cortisol sampling compliance: comparison of patients and healthy volunteers. Psychoneuroendocrinology 29: 636–650.
14. Kudielka BM, Broderick JE, Kirschbaum C (2003) Compliance with saliva sampling protocols: electronic monitoring reveals invalid cortisol daytime profiles in noncompliant subjects. Psychosom Med 65: 313–319.
15. Jacobs N, Nicolson NA, Derom C, Delespaul P, van Os J, et al. (2005) Electronic monitoring of salivary cortisol sampling compliance in daily life. Life Sci 76: 2431–2443.
16. Kudielka BM, Hawkley LC, Adam EK, Cacioppo JT (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.
17. Granger DA, Kivlighan KT, el-Sheikh M, Gordis EB, Stroud LR (2007) Salivary alpha-amylase in biobehavioral research: recent developments and applications. Ann N Y Acad Sci 1098: 122–144.
18. Kraemer HC, Giese-Davis J, Yutsis M, O'Hara R, Neri E, et al. (2006) Design decisions to optimize reliability of daytime cortisol slopes in an older population. Am J Geriatr Psychiatry 14: 325–333.
19. Robles TF, Shetty V, Zigler CM, Glover DA, Elashoff D, et al. (2011) The feasibility of ambulatory biosensor measurement of salivary alpha amylase:

relationships with self-reported and naturalistic psychological stress. Biol Psychol 86: 50–56.
20. Wise J, Operario D (2008) Use of electronic reminder devices to improve adherence to antiretroviral therapy: a systematic review. AIDS Patient Care STDS 22: 495–504.
21. Broderick JE, Schwartz JE, Shiffman S, Hufford MR, Stone AA (2003) Signaling does not adequately improve diary compliance. Ann Behav Med 26: 139–148.
22. Poudevigne MS, O'Connor PJ (2006) A review of physical activity patterns in pregnant women and their relationship to psychological health. Sports Med 36: 19–38.
23. Lof M (2011) Physical activity pattern and activity energy expenditure in healthy pregnant and non-pregnant Swedish women. Eur J Clin Nutr 65: 1295–1301.
24. Rousham EK, Clarke PE, Gross H (2006) Significant changes in physical activity among pregnant women in the UK as assessed by accelerometry and self-reported activity. Eur J Clin Nutr 60: 393–400.
25. Kammerer M, Adams D, Castelberg Bv, Glover V (2002) Pregnant women become insensitive to cold stress. BMC Pregnancy Childbirth 2: 8.
26. Meinlschmidt G, Martin C, Neumann ID, Heinrichs M (2010) Maternal cortisol in late pregnancy and hypothalamic-pituitary-adrenal reactivity to psychosocial stress postpartum in women. Stress 13: 163–171.
27. Pruessner JC, Wolf OT, Hellhammer DH, Buske-Kirschbaum A, Auer Kv, et al. (1997) Free cortisol levels after awakening: a reliable biological marker for the assessment of adrenocortical activity. Life Sci 61: 2539–2549.
28. Singer JD, Willett JB (2003) Applied Longitudinal Data Analysis. Oxford: Oxford University Press.
29. Lane P (2008) Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. Pharm Stat 7: 93–106.
30. Hall DL, Blyler D, Allen D, Mishel MH, Crandell J, et al. (2011) Predictors and patterns of participant adherence to a cortisol collection protocol. Psychoneuroendocrinology 36: 540–546.
31. Ho PM, Bryson CL, Rumsfeld JS (2009) Medication adherence: its importance in cardiovascular outcomes. Circulation 119: 3028–3035.
32. Goedhart G, Vrijkotte TG, Roseboom TJ, van der Wal MF, Cuijpers P, et al. (2010) Maternal cortisol and offspring birthweight: results from a large prospective cohort study. Psychoneuroendocrinology 35: 644–652.
33. Bjorntorp P, Rosmond P (2000) Obesity and cortisol. Nutrition 16: 924–936.
34. Entringer S, Buss C, Shirtcliff EA, Cammack AL, Yim IS, et al. (2010) Attenuation of maternal psychophysiological stress responses and the maternal cortisol awakening response over the course of human pregnancy. Stress 13: 258–268.
35. Shea AK, Streiner DL, Fleming A, Kamath MV, Broad K, et al. (2007) The effect of depression, anxiety and early life trauma on the cortisol awakening response during pregnancy: preliminary results. Psychoneuroendocrinology 32: 1013–1020.
36. Chung MH, Richardson BA, Tapia K, Benki-Nugent S, Kiarie JN, et al. (2011) A randomized controlled trial comparing the effects of counseling and alarm device on HAART adherence and virologic outcomes. PLoS Med 8: e1000422.
37. Suglia SF, Staudenmayer J, Cohen S, Enlow MB, Rich-Edwards JW, et al. (2010) Cumulative stress and cortisol disruption among Black and Hispanic pregnant women in an urban cohort. Psychol Trauma 2: 326–334.
38. Kivlighan KT, DiPietro JA, Costigan KA, Laudenslager ML (2008) Diurnal rhythm of cortisol during late pregnancy: associations with maternal psychological well-being and fetal growth. Psychoneuroendocrinology 33: 1225–1235.
39. Obel C, Hedegaard M, Henriksen TB, Secher NJ, Olsen J, et al. (2005) Stress and salivary cortisol during pregnancy. Psychoneuroendocrinology 30: 647–656.
40. DeSantis AS, Adam EK, Mendelsohn KA, Doane LD (2010) Concordance between self-reported and objective wakeup times in ambulatory salivary cortisol research. Int J Behav Med 17: 74–78.
41. Dockray S, Bhattacharyya MR, Molloy GJ, Steptoe A (2008) The cortisol awakening response in relation to objective and subjective measures of waking in the morning. Psychoneuroendocrinology 33: 77–82.

Appendix B

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

(Published in Psychoneuroendocrinology)

Psychoneuroendocrinology (2014) 44, 13—19

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/psyneuen

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

Julian Moeller [a,b], Roselind Lieb [a], Andrea H. Meyer [a], Katharina Quack Loetscher [c], Bettina Krastel [d], Gunther Meinlschmidt [a,d,e,*]

[a] *University of Basel, Department of Psychology, Division of Clinical Psychology and Epidemiology, Basel, Switzerland*
[b] *Diagnostic and Crisis Intervention Centre, Department of Psychiatry (UPK), University of Basel, Basel, Switzerland*
[c] *Department of Obstetrics, University Hospital Zurich, Zurich, Switzerland*
[d] *National Centre of Competence in Research, Swiss Etiological Study of Adjustment and Mental Health (sesam), Basel, Switzerland*
[e] *Faculty of Medicine, Ruhr-University Bochum, Bochum, Germany*

Received 16 September 2013; received in revised form 17 February 2014; accepted 24 February 2014

KEYWORDS
Compliance;
Adherence;
Saliva;
Salivary;
Saliva sampling;
Testosterone;
Pregnancy;
Women

Summary
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, 1100 h, 1500 h, 2000 h, and 2200 h. 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,

* Corresponding author at: University of Basel, Missionsstrasse 60/62, CH-4055 Basel, Switzerland. Tel.: +41 61 26 70271; fax: +41 61 26 70659.
E-mail address: gunther.meinlschmidt@unibas.de (G. Meinlschmidt).

http://dx.doi.org/10.1016/j.psyneuen.2014.02.012
0306-4530/

$F(1,58) = 5.41$, $p = 0.023$, 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.

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 1100 h, 1500 h, 2000 h, and 2200 h), 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 1100 h, 1500 h, 2000 h, and 2200 h 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 inter-assay 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

Table 1 Demographic and other descriptive information in the total sample and in adherent and nonadherent participants[a].

Characteristic	Total	Adherence	
	(*n* = 60)	Adherent (*n* = 25)	Nonadherent (*n* = 35)
Age (years)	33 (28; 36)	34 (27; 37)	33 (28; 36)
Current body weight (kg)	70 (61; 75)	70 (62; 74)	66 (60; 75)
Gestational age of fetus (weeks)	26 (17; 31)	27 (19; 31)	25 (17; 31)
Parity[b]			
0	32 (53.3)	17 (68)	15 (42.9)
1–2	20 (33.3)	7 (28)	13 (37.1)
≥3	2 (3.4)	1 (4)	1 (2.9)
Unknown	6 (10)	0 (0)	6 (17.1)
Employed[b]			
Yes	43 (71.7)	19 (76)	24 (68.6)
No	14 (23.3)	6 (24)	8 (22.9)
Unknown	3 (5)	0 (0)	3 (8.6)
Number of cigarettes smoked on study days[b]			
0	55 (91.7)	23 (92)	32 (91.4)
≥1	5 (8.3)	2 (8)	3 (8.6)

[a] If not otherwise specified, median (25th percentile; 75th percentile) is reported.
[b] Number of participants (percentage) is reported.

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.

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 Fig. 1, we depicted for each sample the deviation from scheduled sampling time in minutes against the scheduled sampling time.

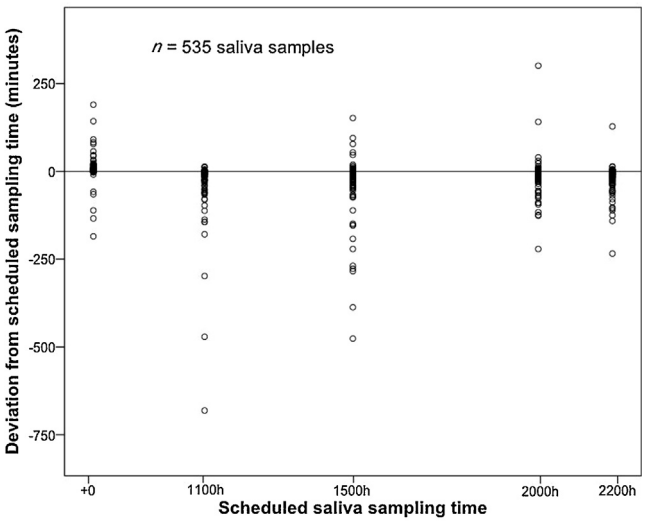

Figure 1 Deviation from scheduled saliva sampling in minutes at the five scheduled sampling times. Note: negative and positive values on the *y* axis indicate delayed and premature saliva sampling, respectively, relative to the scheduled sampling time; +0, at self-reported awakening time.

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 Fig. 2, adherent participants had higher diurnal sT concentrations compared with nonadherent participants.

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

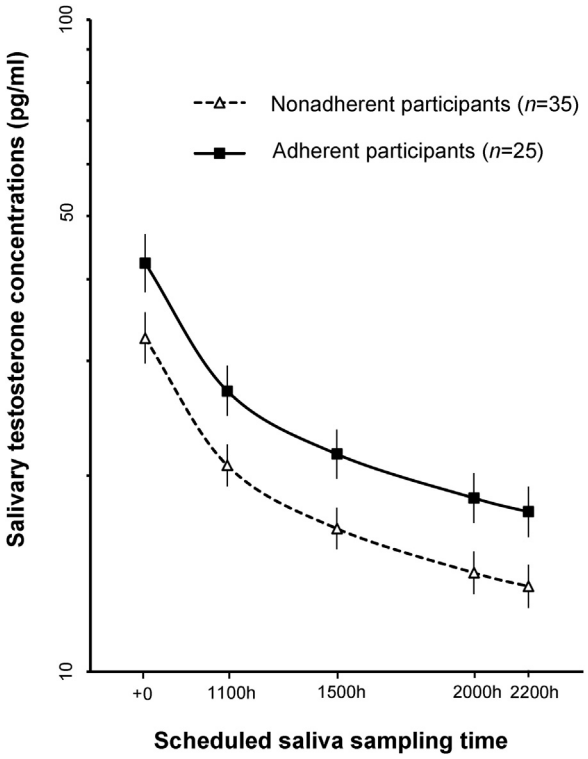

Figure 2 Salivary testosterone concentrations on a logarithmized scale stratified by trait adherent (>8 adherent samples of the total of 10 scheduled samples) and trait nonadherent (≤8 adherent samples) participants. Note: salivary testosterone concentrations were averaged across two collection days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

prior studies that indicated that ambulatory nonadherence with scheduled sampling is associated with biased salivary cortisol estimates (Kudielka et al., 2003, 2007; Broderick et al., 2004; 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, 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, 2012; Al-Dujaili and Sharp, 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 et al. (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 Section 2, 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.

Role of funding source

This work is part of the National Centre of Competence in Research (NCCR) Swiss Etiological Study of Adjustment and Mental Health (sesam). The Swiss National Science Foundation (SNSF) (project no. 51A240-104890), the University of Basel, the F. Hoffmann-La Roche Corp., and the Freie Akademische Gesellschaft provided core support for the NCCR sesam. Additionally, G.M. receives SNSF funding under project no. 100014_135328. The funding sources had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We thank Rene Angst, Tanja Angst, Monica Bachmann, Simone Briner Meier, Susan C.A. Burkhardt, Sigrid Falk, Gabriela Hunziker, Melanie Knabe, Melanie Kruegel, Laura Landi Degen, Cyrill Martin, Fabian Peter, Michael Pluess, Vera Schumacher and Anna Wiener for their valuable help with the project.

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 (Suppl. 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., von Auer, K., 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.

Appendix C

Women's diurnal salivary testosterone change during pregnancy is associated with offspring size at birth

(Submitted to Psychoneuroendocrinology)

Women's diurnal salivary testosterone change during pregnancy is associated with offspring size at birth

Julian Moeller, MSc[1,2]; Roselind Lieb, PhD[1]; Andrea H. Meyer, PhD[1]; Katharina Quack Loetscher, MD[3]; Bettina Krastel, PhD[4]; Gunther Meinlschmidt, PhD[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 70961, fax: +41 61 26 70659, email: gunther.meinlschmidt@unibas.de

Running title: Maternal testosterone during pregnancy and offspring size at birth

Word count: 3074

Number of tables: 3

For submission to *Psychoneuroendocrinology*

Abstract

Objective: The objective of this study was to estimate the association of women's diurnal salivary testosterone change during pregnancy with offspring size at birth.

Methods: In this prospective study, pregnant women followed a standard ambulatory saliva-sampling protocol, consisting of 10 saliva sample collections scheduled on two consecutive days between awakening and 2200 h. We measured testosterone concentrations in the samples and collected information on offspring weight, length, and head circumference at birth from medical birth records. To assess women's diurnal salivary testosterone change during pregnancy, we used a linear mixed-effect model, estimating for each woman the temporal linear testosterone slope. We used analysis of covariance models to estimate the association of testosterone slope with offspring size at birth, and to estimate whether offspring sex is moderating the observed associations.

Results: We included 52 women with singleton pregnancies in this study. Median gestational age at saliva sampling was 27 weeks. A flatter testosterone slope was associated with lower offspring weight (b = –3.9, standard error = 1.3, $P = 0.004$, explained variance $R^2 = 12.0\%$) and length at birth (b = –1.3, standard error = 0.4, $P = 0.003$, explained variance $R^2 = 14.3\%$), after adjusting for gestational age at birth and several potential confounders. There was no indication of moderation by offspring sex.

Conclusions: This study is the first to suggest that women's diurnal testosterone change during pregnancy is a predictor of offspring size at birth. An altered diurnal testosterone change—indicated by a flattened testosterone slope—may represent a biomarker for reduced fetal growth.

Keywords: birth outcome, circadian rhythm, intrauterine programming, prenatal exposure delayed effects, prenatal programming, testosterone

1. Introduction

Fetal growth restrictions, as indicated by low offspring size at birth for gestational age, are a major public health concern: they have a substantial adverse effect on the subsequent health of the offspring, their prevalence across countries is relatively high, and the understanding of their etiology is limited (for reviews see de Mola et al., 2014; Goldenberg and Culhane, 2007; McCormick, 1985; Murray et al., 2015; Salam et al., 2014; compare Wadhwa et al., 2004). Accordingly, it has become a high priority to improve risk assessment and to gain insight into the biological mechanism underlying fetal growth restrictions (Grivell et al., 2009; National Institutes of Health, 2003).

A growing number of animal studies suggest that maternal testosterone during pregnancy may be a potential predictor and biological mechanism underlying this phenomenon, indicating that experimentally induced changes in the maternal testosterone milieu during pregnancy result in reduced fetal growth (Manikkam et al., 2004; Recabarren et al., 2005; Sathishkumar et al., 2011; Veiga-Lopez et al., 2011). However, in spite of these animal-based observations, only few studies examined testosterone levels in normal pregnant women.

In healthy women in childbearing age, a normal testosterone concentration pattern is indicated by a diurnal change, with a testosterone peak early in the morning and a following decline across the day (Al-Dujaili and Sharp, 2012). This diurnal change appears to be driven by a circadian rhythm and is maintained during pregnancy (Al-Dujaili and Sharp, 2012; Bammann et al., 1980; Voegtline et al., 2013). A dysregulated diurnal testosterone pattern may be indicated by a flatter-than-average diurnal testosterone slope (Karatsoreos et al., 2007).

To our knowledge, to date, only one study has estimated whether maternal diurnal testosterone change during pregnancy predicts offspring size at birth in normal pregnant women. In this study, Voegtline et al. (2013) instructed pregnant women to follow an

ambulatory salivary-sampling protocol for testosterone assessment and did not report an association between diurnal salivary testosterone slope and offspring weight at birth. However, they found women's elevated morning salivary testosterone concentrations—measured at a single time point—associated with reduced birth weight in male, but not female, offspring. This sex-specific finding may be explained by a greater vulnerability of the male fetus to potential prenatal exposures (Di Renzo et al., 2007; Eriksson et al., 2010; Voegtline et al., 2013; Zeitlin et al., 2002). Furthermore, a previous study reported reduced weight and length at birth in the offspring of women with elevated serum testosterone concentrations during pregnancy, with testosterone concentrations measured again at a single time point (Carlsen et al., 2006).

The aim of the present study was to extend prior research on women's testosterone concentrations during pregnancy and offspring size at birth by estimating women's diurnal testosterone change based on multiple saliva sample collections and saliva sampling times assessed with an electronic adherence-monitoring system. Multiple measures throughout the day are needed to capture the diurnal testosterone change (Al-Dujaili and Sharp, 2012). Electronic adherence monitoring reduced bias in ambulatory salivary testosterone assessment (Moeller et al., 2014b).

Moreover, to our knowledge, the present study is the first to examine in normal pregnant women whether women's saliva testosterone concentrations during pregnancy predict offspring head circumference at birth. The trajectory of fetal head growth, as indicated by head circumference at birth, may represent a distinct fetal growth trajectory, which is associated with specific health outcomes in the offspring (Bloomfield et al., 2006; Cooke et al., 1977; Gale et al., 2006).

Taken together, although animal studies suggest that maternal testosterone during pregnancy is a biological mechanism of fetal growth, little attention has been paid to women's testosterone concentrations during pregnancy and human fetal growth. To date, it is unclear

whether women's diurnal testosterone change during pregnancy predicts fetal growth in normal pregnant women. In the present longitudinal study, we sought to address this gap: we aimed to estimate (1) whether women's diurnal salivary testosterone change during pregnancy predicts offspring weight, length, and head circumference at birth, and (2) whether offspring sex is moderating the observed associations.

2. Methods

2.1. Participants

We recruited pregnant women at the outpatient service of the Department of Obstetrics, University Hospital Zurich, Switzerland, during their regular antenatal visits. Recruitment took place in the context of a study published elsewhere (Moeller et al., 2014a). Eligible women underwent regular antenatal visits at the outpatient service, were in their 12th to 32nd week of gestation, and had sufficient German language skills. Their medical records contained no mention of high-risk pregnancy, diseases potentially affecting the neuroendocrine system, human immunodeficiency virus (HIV) infection, or use of hormone-containing medication. We conducted this study in accordance with the Declaration of Helsinki. Ethics committees of Zurich (Kantonale Ethikkommission Zuerich, Zurich, Switzerland) and Basel (Ethikkommission Nordwest- und Zentralschweiz (EKNZ), Basel, Switzerland, formerly Ethikkommission Beider Basel, Basel, Switzerland) approved the study, and all participants provided written informed consent before participating.

2.2. Women's testosterone concentrations during pregnancy

We instructed the women to follow a standard ambulatory saliva-sampling protocol during their pregnancy, consisting of 10 saliva sample collections scheduled on two consecutive days at awakening and at 1100 h, 1500 h, 2000 h, and 2200 h. We gave them standardized instructions for accurate saliva sampling, including information about the

importance of high adherence with the sampling schedule and about restrictions (e.g., not to brush teeth, consume caffeine, eat, or smoke 1 hour before each scheduled sampling) to minimize bias in salivary testosterone concentration measurement. The women collected saliva with straws and 2.0 mL safe-lock tubes (Eppendorf, Hamburg, Germany) that were prelabeled with scheduled sampling time. We measured the women's exact sampling times with electronic Medication Event Monitoring Systems (MEMS 6 TrackCap Monitor, Aardex Ltd., Switzerland). We informed about half of the women about the electronic compliance monitoring and gave about half of the women timers and alarm clocks to remind them at scheduled sampling times. Women were instructed to store the saliva samples on the two consecutive sampling days in a refrigerator (for further details, see Moeller et al., 2014a, b).

After receiving the samples, we stored them at –20°C until biochemical analysis. We centrifuged thawed samples at 3000 g for 10 min and assayed salivary testosterone with 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. Demographic and other descriptive information

On days of saliva sample collection, women completed a questionnaire, providing information on maternal age, height, pre-pregnancy body weight, and current body weight.

2.4. Obstetric conditions and information on offspring size at birth

We retrieved information on gestational age at saliva sampling, parity, offspring sex, and gestational age at birth, as well as offspring weight, length, and head circumference at birth from medical records.

Gestational age was determined based on the last menstrual period and first ultrasound examination. Offspring length and head circumference at birth were measured to the nearest

0.5 cm using a standard length board and a plastic nonelastic measuring tape, respectively. Birth weight was measured in g, using an electronic weighing scale (Seca Weighing and Measuring Systems, Hamburg, Germany).

2.5. Statistical analyses

We eliminated women with multiple pregnancies from the statistical analyses, because multiple pregnancies are linked to different fetal growth patterns and maternal testosterone concentration patterns as compared to singleton pregnancies (Stirrup et al., 2015; Toriola et al., 2011). To estimate the associations between women's diurnal salivary testosterone change during pregnancy and offspring size at birth, we used analysis of covariance models. Birth size outcomes were offspring weight (g), length (cm), and head circumference (cm) at birth. Each outcome was analyzed in a separate model. Predictor was the maternal testosterone slope estimate, that is, the linear change in salivary testosterone concentrations across the saliva samples collected between awakening and 2200 h, pooled across the two consecutive sampling days. Linear change was thereby defined as the individual slope coefficient as obtained from a random coefficient model with salivary testosterone concentrations as outcome, time (in minutes) as predictor, allowing for random intercepts and slopes. The so-obtained slope coefficient is often called an "empirical Bayes" estimator and can be seen as a weighted average of the fixed effect of the slope (i.e., the slope estimate for the entire sample) and the individual slope estimates as obtained when performing linear regression models of testosterone concentration against time for each woman separately (Singer and Willett, 2003).

In the first set of analyses, models were adjusted a priori for gestational age at birth (days, continuous; Bammann et al., 1980; Ott, 1993) and for several potential confounders, including gestational age at saliva sampling (weeks, continuous; Bammann et al., 1980), offspring sex (male, female; categorical; Di Renzo et al., 2007; Meulenberg and Hofman, 1991), and parity (categorical, as depicted in Table 2; Ong et al., 2002; Toriola et al., 2011).

In the second set of analyses, we adjusted the models additionally for pre-pregnancy body mass index (BMI) (kg/m2, continuous; Sowers et al., 2001) or maternal age (years, continuous; Granger et al., 2004; Toriola et al., 2011). In the third set, we repeated the analyses described above but included offspring sex as a potential moderator variable (Voegtline et al., 2013).

Offspring weight, height, and head circumference at birth variables were log-transformed to approximate normal distributions. An alpha level of 0.05 indicated statistical significance. As an indicator of effects size, we used the explained variance of the predictor maternal testosterone slope estimate after controlling for all confounders in the model. Analyses were performed using R software (version 3.1.0; R Core Team, 2014), including the package lme (version 1.1-7; Bates et al., 2014).

3. Results

Sixty-nine pregnant women participated in this study. We excluded five women because of major protocol deviations (e.g., MEMS defect, see Moeller et al., 2014) and four because of multiple pregnancies. Moreover, we excluded seven women because they did not deliver in the University Hospital Zurich and one because she collected an insufficient amount of saliva, which hampered birth outcome information collection and salivary testosterone concentration measurement, respectively. Hence, the sample on which the subsequent analyses were based comprised 52 women with singleton pregnancies. Table 1 shows demographic and other descriptive information.

-Insert Table 1 approximately here-

3.1. Women's salivary testosterone concentrations during pregnancy

Median (25th percentile, 75th percentile) salivary testosterone concentrations in the samples scheduled at awakening and at 1100 h, 1500 h, 2000 h, and 2200 h were 40.68

(27.30; 60.70), 22.99 (15.42; 33.98), 19.90 (12.86; 29.62), 16.21 (9.18; 23.38), and 15.44 (10.16; 22.43) pg/ml, respectively. The maternal testosterone concentrations indicated the expected diurnal decline throughout the day (b = –91.3, standard error = 7.2, $P < 0.001$).

3.2. Associations between women's diurnal testosterone change during pregnancy and offspring size at birth

We used analysis of covariance models to estimate the associations between women's salivary testosterone slope estimates during pregnancy and offspring size at birth. As depicted in Table 2, women's testosterone slope estimates were associated with offspring size at birth. The flatter the women's testosterone slope estimates, the lower the offspring weight, length, and head circumference at birth. These associations were adjusted for a priori defined potential confounders, including gestational age at saliva sampling, parity, offspring sex, and gestational age at birth.

-Insert Table 2 approximately here-

When we additionally adjusted the covariance models stated above for maternal age or maternal pre-pregnancy BMI, women's testosterone slope estimates remained associated with offspring weight and length at birth. However, the association between women's testosterone slope estimates and head circumference at birth disappeared after additional adjustment for maternal age or maternal pre-pregnancy BMI (depicted in Table S1 in the Supplemental Digital Content).

3.3. Moderator role of infant sex

We used interaction analyses to estimate the potential moderator role of offspring sex in the associations between women's testosterone slope estimates during pregnancy and offspring size at birth. They indicated that the associations stated above were not moderated by offspring sex (see Table 2).

4. Discussion

To the best of our knowledge, this is the first study linking women's diurnal salivary testosterone change during pregnancy with offspring size at birth in normal pregnant women; women's flattened diurnal testosterone change was associated with reduced offspring weight and length but not head circumference at birth, after adjusting for gestational age at birth and a range of potential confounders and irrespective of offspring sex.

In this study, we replicated the prior observation that women's diurnal testosterone change is maintained during pregnancy (Voegtline et al., 2013). Prior research suggested that offspring weight (Carlsen et al., 2006; Voegtline et al., 2013) and length at birth (Carlsen et al., 2006) are predicted by women's testosterone concentrations during pregnancy. However, these findings were partly restricted to women carrying a male fetus (Voegtline et al., 2013) and relied on testosterone levels measured at a single time point throughout the day (Carlsen et al., 2006; Voegtline et al., 2013). The present study is an important extension of these observations, suggesting that women's flattened diurnal testosterone change during pregnancy, as indicated by a flatter-than-average diurnal testosterone slope, may be a predictor for reduced offspring size at birth. Diurnal testosterone change is a distinct measure of women's testosterone concentrations, specifically able to represent the testosterone milieu with its underlying circadian rhythmicity (Al-Dujaili and Sharp, 2012). Moreover, as offspring sex did not moderate our results, our findings suggest that diurnal testosterone change during pregnancy predicts offspring size at birth, regardless of whether a woman is carrying a male or female fetus.

In contrast, Voegtline et al. (2013) reported in the study mentioned above no association between women's diurnal testosterone change during pregnancy and offspring weight at birth. We assume that methodological differences in the estimation of women's diurnal saliva testosterone change, including the use of electronically measured saliva

sampling times and more saliva sample collections in our study, may underlie these inconsistent findings (see section methodological strengths below; see also Al-Dujaili and Sharp, 2012; Broderick et al., 2004; Granger et al., 2004; Kudielka et al., 2003; Moeller et al., 2014a; Moeller et al., 2014b).

In our study, women's diurnal testosterone change during pregnancy was associated with offspring weight and length at birth after adjusting for a range of potential confounders. In contrast, the association of diurnal testosterone change with offspring head circumference at birth disappeared, once we adjusted additionally for maternal age (see Granger et al., 2004; Toriola et al., 2011) or pre-pregnancy BMI (see Sowers et al., 2001). This observation underlines the relevance of adjusting for these variables, when examining the association described above, to reduce the risk of confounding.

If the associations found in our study are causal, a woman's flattened diurnal testosterone change during pregnancy will result in reduced fetal growth. Although we cannot answer the question of causality due to our study design, findings of experimental animal studies suggest a causal link between maternal testosterone levels during pregnancy and fetal growth (Manikkam et al., 2004; Recabarren et al., 2005; Sathishkumar et al., 2011; Veiga-Lopez et al., 2011). As a putative underlying mechanism, it has been discussed that maternal testosterone concentrations during pregnancy may modify placental function and amino acid transporter activity, potentially resulting in a decreased transport of nutrients to the fetus, and hence reduced fetal growth (for further discussion, see Manikkam et al., 2004; Sathishkumar et al., 2011; Voegtline et al., 2013). Alternatively, residual confounding may explain the link described above. For example, a flattened diurnal testosterone change during pregnancy may indicate a dysfunctional alteration of a biological characteristic or other specific characteristics in pregnant women, which in turn may be associated with reduced offspring size at birth. Taken together, our findings suggest women's diurnal testosterone change during pregnancy as a new potentially relevant predictor of reduced fetal growth. Our findings may

contribute to a better understanding of the mechanisms underlying human fetal growth and potentially to a better prediction of reduced fetal growth. It is important to note that apparently salivary testosterone concentrations can be reliably assessed in women when specific methodological aspects are considered (Al-Dujaili and Sharp, 2012; Granger et al., 2004; Liening et al., 2010; Moeller et al., 2014b; Schultheiss, 2013; Schultheiss and Stanton, 2009).

This study has some limitations. First, we estimated women's diurnal testosterone change during pregnancy at a median gestational age of 27 weeks. It would be important to know whether women's diurnal testosterone change during pregnancy predicts offspring size at birth at different stages of gestation, especially at the early stage of gestation (see Kane et al., 2014). This should be scrutinized in future studies. Second, our sample size was relatively small. Hence, our findings should be replicated with a larger sample.

This study also has several important strengths: we rigorously estimated women's diurnal testosterone change based on (1) a two-day ambulatory saliva-sampling protocol consisting of 10 scheduled saliva measures, (2) women's electronically measured saliva sampling times, and (3) a linear mixed-effect model. The use of multiple measurements throughout the day was needed to capture the diurnal testosterone change (Al-Dujaili and Sharp, 2012). Moreover, the use of repeated testosterone measurements is associated with various methodological strengths compared to the utilization of single testosterone measures (Al-Dujaili and Sharp, 2012; Granger et al., 2012; Granger et al., 2004). The use of electronic sampling time measurements was important, as nonadherence with an ambulatory saliva-sampling schedule is common, is often not captured by self-report (Broderick et al., 2004; Kudielka et al., 2003; Kudielka et al., 2012; Moeller et al., 2014a; Stalder et al., 2016), and may introduce a substantial bias in salivary testosterone estimates (Moeller et al., 2014b). Moreover, the application of a linear mixed-effect model allowed us to estimate the temporal

linear testosterone slope for each woman as an indicator of diurnal testosterone change (Singer and Willett, 2003). Another strength is our prospective study design.

As none of the offspring in our sample showed indications of clinically relevant growth restrictions, further studies may scrutinize whether the association between women's diurnal testosterone change during pregnancy with offspring size at birth will also be found in samples of growth-restricted fetuses, which would expand the generalizability of our findings.

Taken together, to the best of our knowledge, this is the first study linking a women's flattened diurnal testosterone change during pregnancy with an offspring size at birth lower than the average for gestational age. Our findings are highly important, as they suggest that women's diurnal testosterone change during pregnancy might be a new predictor for human fetal growth. They encourage further studies to scrutinize maternal testosterone concentrations during pregnancy when addressing the underlying mechanisms and risk assessment of fetal growth restrictions.

References

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.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2014. lme4: linear mixed-effects models using Eigen and S4. R package version 1.

Bloomfield, F.H., Oliver, M.H., Harding, J.E., 2006. The late effects of fetal growth patterns. Arch Dis Child-Fetal 91, F299-F304.

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.

Cooke, R.W., Lucas, A., Yudkin, P.L., Pryse-Davies, J., 1977. Head circumference as an index of brain weight in the fetus and newborn. Early Hum. Dev. 1, 145-149.

de Mola, C.L., de Franca, G.V.A., Quevedo, L.D.A., Horta, B.L., 2014. Low birth weight, preterm birth and small for gestational age association with adult depression: systematic review, and meta-analysis. Br. J. Psychiatry 205, 340-347.

Di Renzo, G.C., Rosati, A., Sarti, R.D., Cruciani, L., Cutuli, A.M., 2007. Does fetal sex affect pregnancy outcome? Gend. Med. 4, 19-30.

Eriksson, J.G., Kajantie, E., Osmond, C., Thornburg, K., Barker, D.J.P., 2010. Boys live dangerously in the womb. Am J Hum Biol 22, 330-335.

Gale, C.R., O'Callaghan, F.J., Bredow, M., Martyn, C.N., 2006. The influence of head growth in fetal life, infancy, and childhood on intelligence at the ages of 4 and 8 years. Pediatrics 118, 1486-1492.

Goldenberg, R.L., Culhane, J.F., 2007. Low birth weight in the United States. Am. J. Clin. Nutr. 85, 584S-590S.

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.

Grivell, R., Dodd, J., Robinson, J., 2009. The prevention and treatment of intrauterine growth restriction. Best Pract Res Cl Ob 23, 795-807.

Kane, S.C., Costa, F.D., Brennecke, S., 2014. First trimester biomarkers in the prediction of later pregnancy complications. Biomed Res Int 2014.

Karatsoreos, I.N., Vernov, M., Romeo, R.D., 2007. Testosterone and the brain: implications for cognition, biological rhythms and aging, in: Ardis, L.I. (Ed.), Testosterone Research Trends. Nova Science Publishers, New York, pp. 91-103.

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.

Liening, S.H., Stanton, S.J., Saini, E.K., Schultheiss, O.C., 2010. Salivary testosterone, cortisol, and progesterone: two-week stability, interhormone correlations, and effects of time

of day, menstrual cycle, and oral contraceptive use on steroid hormone levels. Physiology & Behavior 99, 8-16.

Manikkam, M., Crespi, E.J., Doop, D.D., Herkimer, C., Lee, J.S., Yu, S., Brown, M.B., Foster, D.L., Padmanabhan, V., 2004. Fetal programming: prenatal testosterone excess leads to fetal growth retardation and postnatal catch-up growth in sheep. Endocrinology 145, 790-798.

McCormick, M.C., 1985. The contribution of low birth-weight to infant-mortality and childhood morbidity. N. Engl. J. Med. 312, 82-90.

Meulenberg, P.M.M., Hofman, J.A., 1991. Maternal testosterone and fetal sex. J. Steroid Biochem. Mol. Biol. 39, 51-54.

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

Moeller, J., Lieb, R., Meyer, A.H., Quack Loetscher, K., Krastel, B., Meinlschmidt, G., 2014b. Nonadherence with ambulatory saliva sampling is associated with biased salivary testosterone estimates. Psychoneuroendocrinology 44, 13-19.

Murray, E., Fernandes, M., Fazel, M., Kennedy, S.H., Villar, J., Stein, A., 2015. Differential effect of intrauterine growth restriction on childhood neurodevelopment: a systematic review. Bjog-Int J Obstet Gy 122, 1062-1072.

National Institutes of Health, 2003. Pregnancy and Perinatology Branch Strategic Plan 2005–2010. National Institutes of Health, Washington DC.

Ong, K.K.L., Preece, M.A., Emmett, P.M., Ahmed, M.L., Dunger, D.B., Team, A.S., 2002. Size at birth and early childhood growth in relation to maternal smoking, parity and infant breast-feeding: longitudinal birth cohort study and analysis. Pediatr. Res. 52, 863-867.

Ott, W.J., 1993. Intrauterine growth-retardation and preterm delivery. Am. J. Obstet. Gynecol. 168, 1710-1717.

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Recabarren, S.E., Padmanabhan, V., Codner, E., Lobos, A., Duran, C., Vidal, M., Foster, D.L., Sir-Petermann, T., 2005. Postnatal developmental consequences of altered insulin sensitivity in female sheep treated prenatally with testosterone. Am J Physiol-Endoc M 289, E801-E806.

Salam, R.A., Das, J.K., Bhutta, Z.A., 2014. Impact of intrauterine growth restriction on long-term health. Curr. Opin. Clin. Nutr. Metab. Care 17, 249-254.

Sathishkumar, K., Elkins, R., Chinnathambi, V., Gao, H.J., Hankins, G.D.V., Yallampalli, C., 2011. Prenatal testosterone-induced fetal growth restriction is associated with down-regulation of rat placental amino acid transport. Reprod. Biol. Endocrinol. 9.

Schultheiss, O.C., 2013. Effects of sugarless chewing gum as a stimulant on progesterone, cortisol, and testosterone concentrations assessed in saliva. Int. J. Psychophysiol. 87, 111-114.

Schultheiss, O.C., Stanton, S.J., 2009. Assessment of salivary hormones, in: Harmon-Jones, E., Beer, J.S. (Eds.), Methods in social neuroscience. The Guiilford Press, New York, pp. 17-44.

Singer, J.D., Willett, J.B., 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurence. 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.

Stalder, T., Kirschbaum, C., Kudielka, B.M., Adam, E.K., Pruessner, J.C., Wust, S., Dockray, S., Smyth, N., Evans, P., Hellhammer, D.H., Miller, R., Wetherell, M.A., Lupien, S.J., Clow, A., 2016. Assessment of the cortisol awakening response: expert consensus guidelines. Psychoneuroendocrinology 63, 414-432.

Stirrup, O.T., Khalil, A., D'Antonio, F., Thilaganathan, B., Res, S.T.O., 2015. Fetal growth reference ranges in twin pregnancy: analysis of the Southwest Thames Obstetric Research Collaborative (STORK) multiple pregnancy cohort. Ultrasound Obst Gyn 45, 301-307.

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.

Veiga-Lopez, A., Steckler, T.L., Abbott, D.H., Welch, K.B., MohanKumar, P.S., Phillips, D.J., Refsal, K., Padmanabhan, V., 2011. Developmental programming: impact of excess prenatal testosterone on intrauterine fetal endocrine milieu and growth in sheep. Biol. Reprod. 84, 87-96.

Voegtline, K.M., Costigan, K.A., Kivlighan, K.T., Henderson, J.L., DiPietro, J.A., 2013. Sex-specific associations of maternal prenatal testosterone levels with birth weight and weight gain in infancy. J Dev Orig Hlth Dis 4, 280-284.

Wadhwa, P.D., Garite, T.J., Porto, M., Glynn, L., Chicz-DeMet, A., Dunkel-Schetter, C., Sandman, C.A., 2004. Placental corticotropin-releasing hormone (CRH), spontaneous preterm birth, and fetal growth restriction: a prospective investigation. Am. J. Obstet. Gynecol. 191, 1063-1069.

Zeitlin, J., Saurel-Cubizolles, M.J., de Mouzon, J., Rivera, L., Ancel, P.Y., Blondel, A., Kaminski, M., 2002. Fetal sex and preterm birth: are males at greater risk? Hum. Reprod. 17, 2762-2768.

Tables

Table 1. Descriptive information of women and their offspring.[a]

Variable	Total (n = 52)
Maternal and pregnancy information	
Age (years)	33 (28; 37)
Pre-pregnancy BMI (kg/m^2)	22 (20; 23)
Parity[b]	
0	25 (48)
1–2	18 (35)
≥3	3 (6)
Unknown	6 (12)
Pregnancy BMI at saliva sampling (kg/m^2)	24 (23; 26)
Gestational age at saliva sampling (weeks)	27 (17; 31)
Birth outcomes in the offspring	
Offspring sex[b]	
Female	24 (46)
Male	28 (54)
Gestational age at birth (days)	278 (268; 285)
Birth weight (g)	3360 (3058; 3602)
Birth length (cm)	50 (48; 51)
Head circumference at birth (cm)	35 (34; 36)

[a] If not otherwise specified, the median (25th percentile; 75th percentile) is reported.
[b] Number of participants (percentage)
Note: BMI = body mass index. Percentages may not add up to 100 due to rounding.

Table 2. Covariance models of maternal salivary testosterone slope during pregnancy predicting offspring weight, length, and head circumference at birth.

Outcome	b [x10-3]	SE [x10-3]	t (df)	*P*-value	explained variance (%)
Birth weight	-3.93	1.29	-3.05 (39)	0.004	12.0
Birth length	-1.27	0.40	-3.18 (39)	0.003	14.3
Head circumference at birth	-0.88	0.42	-2.11 (38)	0.041	6.7
Birth weight x infant sex	1.21	2.62	0.46 (38)	0.646	0.3
Birth length x infant sex	-1.00	0.80	-1.26 (38)	0.214	2.2
Head circumference at birth x infant sex	1.14	0.83	1.37 (37)	0.179	2.8

Note: Analyses were adjusted for gestational age at saliva sampling, parity, offspring sex, and gestational age at birth.
b = slope coefficient from linear regression analysis, SE = standard error, df = degrees of freedom, $P < 0.05$.

Table S1. Covariance models of maternal salivary testosterone slope during pregnancy predicting offspring weight, length, and head circumference at birth, with additional adjustments.

Outcome	b [x10-3]	SE [x10-3]	t (df)	*P*-value	explained variance (%)
Models additionally adjusted for pre-pregnancy BMI					
Birth weight	-3.75	1.31	-2.86 (37)	0.007	10.8
Birth length	-1.28	0.41	-3.10 (37)	0.004	14.3
Head circumference at birth	-0.77	0.41	-1.89 (36)	0.067	5.0
Birth weight x infant sex	0.40	2.75	0.15 (36)	0.884	0.0
Birth length x infant sex	-1.04	0.85	-1.23 (36)	0.228	2.2
Head circumference at birth x infant sex	0.77	0.84	0.92 (35)	0.365	1.2
Models additionally adjusted for maternal age					
Birth weight	-3.54	1.37	-2.59 (38)	0.014	8.7
Birth length	-1.08	0.42	-2.58 (38)	0.014	9.3
Head circumference at birth	-0.66	0.43	-1.53 (37)	0.135	3.4
Birth weight x infant sex	1.09	2.63	0.41 (37)	0.681	0.2
Birth length x infant sex	-1.07	0.79	-1.36 (37)	0.182	2.5
Head circumference at birth x infant sex	1.08	0.82	1.31 (36)	0.198	2.5

Note: All analyses were adjusted for gestational age at saliva sampling, parity, offspring sex, and gestational age at birth.
Each outcome was analyzed in a separate model. b = slope coefficient from linear regression analysis, BMI = body mass index, SE = standard error, df = degrees of freedom, $P < 0.05$.

Acknowledgments:

We thank Rene Angst, Tanja Angst, Monica Bachmann, Simone Briner Meier, Susan C.A. Burkhardt, Sigrid Falk, Gabriela Hunziker, Melanie Knabe, Laura Landi Degen, Cyrill Martin, Fabian Peter, Michael Pluess and Vera Schumacher for their valuable help with the project.

Conflict of interest:

G.M. is a consultant for Janssen Research & Development, LLC., receiving a moderate personal fee. J.M., R.L., A.H.M., K.Q.L. and B.K. declare that they have no conflicts of interest.

Contributors:

G.M. and R.L. designed the study and wrote the protocol; K.Q.L., B.K., G.M. and R.L. organized the data collection; G.M. and J.M. managed the literature searches; A.H.M. and J.M. undertook the statistical analyses; J.M. wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Role of funding source:

This work is part of the National Centre of Competence in Research (NCCR) Swiss Etiological Study of Adjustment and Mental Health (sesam). The Swiss National Science Foundation (SNSF) (project no. 51A240-104890), the University of Basel, the F. Hoffmann-La Roche Corp., and the Freie Akademische Gesellschaft provided core support for the NCCR sesam. Additionally, G.M. receives SNSF funding under project no. 100014_135328. The funding sources had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.