Psychoneuroendocrinology (2014) 39, 184—193

Available online at www.sciencedirect.com

ScienceDirect

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

How stable are diurnal cortisol activity indices in healthy individuals? Evidence from three multi-wave studies

Kharah M. Ross *, Michael L.M. Murphy, Emma K. Adam, Edith Chen, Gregory E. Miller

Department of Psychology Northwestern University, Evanston, IL 60208-2710, USA

Received 26 June 2013; received in revised form 12 September 2013; accepted 14 September 2013

KEYWORDS
Cortisol;
Hypothalamic pituitary adrenal (HPA) axis;
Multilevel modeling;
Stability;
Within-person

Summary
Background: Indices of cortisol activity, including the cortisol awakening response (CAR), diurnal slope, and cortisol output across the day (total daily output), are often studied as mechanistic indicators that could link stress with health. Yet there is a paucity of data speaking to their temporal features, particularly whether they behave in a more state- or trait-like manner across time.
Methods: To address this issue, data from 3 studies were used to assess CAR, diurnal slope and total daily output stability over different age groups and time spans: 130 healthy children and adolescents collected salivary cortisol samples 5 times/day (1, 4, 9 and 11 h after wake) over 2 days at 5 visits spaced 6 months apart (Study 1); 147 adolescent girls collected saliva 6 times/day (wake, 1, 4, 9 and 14 h after wake) for 2 days at 3 visits, each a year apart (Study 2); and 47 healthy, primarily middle age adults collected saliva 6 times/day (wake, 1, 4, 9 and 14 h after wake) for 3 days at 4 visits spaced 2—3 months apart (Study 3). Stability was estimated by multilevel model-derived intraclass correlation coefficients (*ICC*s).
Results: Across studies, approximately 50% of the variance in cortisol indices was attributable to day-to-day fluctuations, suggesting state-like properties. Of the indices, total daily output emerged as the most stable over time, followed by diurnal slope and CAR, but stability estimates were generally quite modest regardless of index and sample. Over time spans of >1 year, *ICC*s were $\leq .13$.
Conclusions: Most of the variance in CAR, diurnal slope and total daily output reflects day-to-day fluctuation; there was little evidence for more stable trait-like influences. These findings suggest that future research should focus on short-term fluctuations in stress, cortisol and health, as opposed to lengthy disease processes.
© 2013 Elsevier Ltd. All rights reserved.

* Corresponding author at: Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208-2710, USA. Tel.: +1 847 491 5190; fax: +1 847 491 7859.
E-mail address: kmross@u.northwestern.edu (K.M. Ross).

0306-4530/$ — see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.psyneuen.2013.09.016

1. Introduction

Cortisol, a steroid hormone and glucocorticoid, is a key-end product of hypothalamic-pituitary-adrenal (HPA) axis activation. It is essential for life, and is best known in health psychology and related disciplines for its role in regulating the stress response cascade (Sapolsky, 2000). Much of the interest in cortisol revolves around its role as a proposed intermediary linking chronic stress with health problems (Miller et al., 2007; Adam and Kumari, 2009), including diabetes mellitus and the metabolic syndrome (Anagnostis et al., 2009; Champaneri et al., 2012), affective difficulties (Stetler and Miller, 2005; Adam et al., 2010; Stetler and Miller, 2011), clinical and sub-clinical heart disease (Matthews et al., 2006; Dekker et al., 2008; Hajat et al., 2013), cardiovascular disease mortality (Kumari et al., 2011), chronic fatigue syndrome (Strickland et al., 1998), arthritis (Chikanza et al., 1992; Catley et al., 2000), and asthma (Chen et al., 2003; Fei et al., 2004).

Cortisol is secreted in pulses over the course of the day. Typically, there is a steep rise in cortisol output during the first 30—45 min following awakening, followed by a steady decline across the morning, afternoon, and evening hours, with the daily nadir typically occurring around midnight (Kirschbaum and Hellhammer, 1989; Pruessner et al., 1997). To capture this variability in daily life most researchers have participants collect saliva samples 2—6 times during the waking hours. After cortisol has been measured in saliva, various indices can be extracted from different portions of the diurnal curve. Three of the most commonly used are the cortisol awakening response (CAR), diurnal cortisol slope (diurnal slope), and total daily cortisol output. The CAR is defined as the increase in cortisol concentrations during the first 30 min post-awakening, relative to waking cortisol values. It is believed to represent a physiological boost needed to meet the expected demands of the day, and evincing either high or blunted CARs has been linked to maladaptive outcomes (Stetler and Miller, 2005; Adam et al., 2010; Champaneri et al., 2012). The diurnal slope attempts to capture cortisol circadian fluctuations; it is usually operationally defined as the line resulting from regression of cortisol values collected across the day onto hours since awakening, excluding the morning awakening response. A negative diurnal slope is generally considered indicative of healthy HPA axis function (although some exceptions exist, see Smyth et al., 1997; Stone et al., 2001), with flattened or positive diurnal slopes suggestive of potential HPA axis dysfunction. Finally, total daily cortisol output, or total area-under-the-curve (total daily output), is defined as the area between ground and cortisol values taken across the day. Often, but not always, the morning awakening response is excluded from these calculations, to prevent morning awakening responses from having undue influence on calculated values. Total daily output is thought to reflect cumulative tissue exposure to cortisol across the day: Persistently high total daily output may create "wear and tear" on various bodily tissues, resulting in structural or functional changes that could affect disease vulnerability (McEwen, 1998).

Despite extensive research on these cortisol output indices, important questions about their determinants and characteristics remain unanswered. Of particular relevance to stress and health research is whether these indices possess relatively state-like vs. trait-like properties. If diurnal cortisol indices fluctuate widely across days, or are more state-like, they are probably best suited to explaining phenomenon that operate along similarly brief timeframes, like why arthritis symptoms wax and wane in concert with daily mood states (e.g., Schanberg et al., 2000). By contrast, if the cortisol indices are relatively stable over time, or are more trait-like, they could be well poised to explain processes that evolve over more lengthy time periods, for example why some enduring stressors, such as an abrasive marriage, bolster risk for heart disease (e.g., Orth-Gomer, 2000), a condition that develops over the course of multiple decades.

Surprisingly little is known about the stability of various cortisol indices. To the best of our knowledge, only two studies have examined stability over periods longer than 1 month. Both of these focused on adolescents and assessed a single domain of cortisol activity (i.e., cortisol following awakening or the diurnal rhythm of output). In a sample of 410 adolescents, CARs were assessed annually over three consecutive years. Analyses yielded a standard tracking coefficient of $\beta = .17$, which can be interpreted as indicating low stability (Platje et al., 2013). In another study, diurnal slopes were assessed biannually over 6 years in 357 children. Results indicated that 13% of the total variance in diurnal slopes was trait-like or between persons. The remainder was attributable to day-to-day or within-day fluctuations (Shirtcliff et al., 2012).

Although they provide initial evidence regarding cortisol's temporal stability, these studies leave open several important questions. First, to what extent does stability vary across the indices (CAR, diurnal slope, total daily output)? If it does, there could be significant implications for theory and methods in this area, with researchers attempting to match the temporal characteristics of cortisol indices with those of the phenomenon being studied. Second, to what extent does the stability of cortisol indices vary across the lifespan? All of the stability research to date has focused on adolescents, so it remains unclear whether patterns differ at other ages. Again, if they did, it would have significant implications for theory and methods in this literature, potentially leading researchers to focus on specific cortisol indices for specific populations or age groups. The purpose of this article is to begin addressing these questions, using cortisol datasets that were collected in three distinct multi-wave longitudinal studies spanning from childhood into the adult years.

2. Methods

We used data from three independent studies to assess cortisol index stability over various timeframes (see Table 1 for a summary of protocols and samples). Participants across the studies were recruited from the Vancouver, BC, Canada area through a combination of print and online advertisements. In order to be included, all participants had to be currently healthy and free of any history of major physical or psychiatric disorders. All projects were approved by the University of British Columbia Research Ethics Board. Written consent was obtained from all participants, and a parent or guardian also provided consent for participants under 18.

Table 1 Study sample characteristics and cortisol sampling schedules.

	Study 1	Study 2	Study 3
Description	Stress, SES and asthma (**controls**)	Risk for depression and atherosclerosis	Caregivers of brain cancer patients (**controls**)
Inclusion criteria	8–18 yo medically healthy	Female 15–19 yo, medically healthy Increased risk of having a depressive episode	>18 yo, medically healthy Free of major stressors
n	130	147	47
Age at entry	12.9 ± 2.4 years	17.0 ± 1.3 years	50.0 ± 13 years
# Female (%)	64 (49.2)	138 (100)	29 (61.7)
Sampling times Cortisol indices	1, 4, 9, 11 h Slope; AUC	0, ½, 1, 4, 9, 14 h CAR; Slope; AUC	0, ½, 1, 4, 9, 14 h CAR; Slope; AUC
# Sampling days per visit # Visits across study	2 5	2 3	3 4
Total sampling days	10	6	12
Interval between visits Total follow-up length	6 months 24 months	1 year 24 months	2–4 months 8 months
Cortisol indices			
CAR (log(ng/mL) h)	N/A	$8.40 \times 10^{-2} \pm 0.35$	$3.7 \times 10^{-3} \pm 0.22$
Diurnal slope (log(ng/mL)/h)	$-3.86 \times 10^{-2} \pm 3.9 \times 10^{-2}$	$-3.83 \times 10^{-2} \pm 3.1 \times 10^{-2}$	$-3.47 \times 10^{-2} \pm 2.12 \times 10^{-2}$
Total daily output (log(ng/mL) h)	6.52 ± -2.7	9.76 ± 3.7	7.47 ± 2.4

2.1. The studies

2.1.1. Study 1

A sample of 130 children and adolescents were originally recruited as the control group for a larger longitudinal study on stress, socioeconomic status, and asthma in children and adolescents (Chen et al., 2006). To qualify for participation, individuals had to be between the ages of 8 and 18. Participants were seen every 6 months for 2 years, for a total of five possible visits. After each visit, participants were asked to engage in two consecutive days of at-home salivary cortisol sample collection. Over the days of assessment, participants recorded their wake times and collected saliva samples at 1, 4, 9, and 11 h after waking. Compliance was evaluated using MEMS 6 TrackCap Monitors (Aardex Ltd., Switzerland). Because waking and 30-min after wake samples were not collected, only diurnal slope and total daily output were calculated for this sample. Across participants who collected saliva throughout the study, diurnal slope and total daily output could not be calculated on 6.1% and 6.3% of days respectively due to technical issues with collection or measurement, primarily the failure to provide enough saliva.

2.1.2. Study 2

Data from this sample are from a longitudinal study of depression and atherosclerosis among healthy adolescent girls at risk for developing a first episode of major depression (Miller and Cole, 2012). Individuals were eligible for the study if they were female between the ages of 15 and 19, and at high risk for developing a first episode of major depression. The latter was defined as having either a first-degree relative with a history of major depression, or scoring highly on one of the two indices of cognitive vulnerability. A total of 147 participants were recruited.

Participants collected saliva for cortisol annually, at baseline and 1- and 2-year follow-ups, for a total of three assessments. Following each visit, participants were asked to take part in two consecutive days of home salivary cortisol collection. For each day of assessment, participants recorded the time they awoke and collected samples at waking, ½, 1, 4, 9, and 14 h after waking. Participants were given a handheld computer (Palm Zire 21) to aid in saliva collection by sounding alarms at the necessary intervals based on a participant's wake time. To enforce compliance, the computer displayed a three-digit code with each alarm that participants were required to write on the corresponding Salivette container. These codes were then matched to computer codes in our laboratory, and samples without the correct code were excluded from analyses. CAR, total daily output and diurnal slope values were calculated for each day of cortisol sampling. Across participants who collected saliva throughout the study, values for CAR, total daily output, and diurnal slope could not be calculated on 15.9%, 11.7%, and 4.8% of days respectively due to issues with compliance or measurement.

2.1.3. Study 3

A sample of 47 healthy adults was originally recruited as a healthy control group for a longitudinal study on family caregivers of patients undergoing treatment for brain cancer (Rohleder et al., 2009). To be eligible to participate, individuals had to be at least 18 years old and free of major stressors, such as bereavement and family illness. Due to efforts to match the control group to the caregiver group, the healthy control group consisted of mostly middle-age adults

(see Table 1, baseline age range 23–78 years). Participants visited the laboratory at baseline and had follow-up appointments two, six, and eight months later, for a total of four possible visits. After each visit, participants underwent three consecutive days of home monitoring. Every day during the assessment period, participants recorded the time they awoke and collected salivary cortisol samples at wake, ½, 1, 4, 9, and 14 h after waking. CAR, total daily output and diurnal slope values were calculated for each day of cortisol sampling. Across participants who collected saliva throughout the study, values for CAR, total daily output, and diurnal slope could not be calculated on 5.3%, 2.4%, and 1.8% of days respectively due to issues with collection or measurement.

2.2. Salivary cortisol

Saliva was collected using cotton dental rolls (Salivettes; Sarstedt, Nuembrecht, Germany). Participants were asked to refrain from brushing their teeth, smoking, eating, and drinking (except water) for at least 30 min prior to collection time. To collect saliva, participants were instructed to chew on the roll and move it around their mouth for 1 min until fully saturated. After providing a saliva sample, participants were told to place the cotton dental roll in the plastic Salivette container and store samples in a refrigerator until the home monitoring session was completed, at which point packages were mailed back to our laboratory.

For Studies 1 and 2, completed saliva samples were shipped to the Technical University of Dresden in Germany, where levels of free cortisol were assessed in duplicate via a commercially available chemiluminescence assay (IBL, Hamburg, Germany). This assay had a sensitivity of 0.16 ng/mL, and intra- and interassay coefficients of variation of less than 12%. For Study 3, completed saliva samples were shipped to Brandeis University, where free cortisol was also assessed in duplicate using a commercial chemiluminescence assay (IBL-Hamburg, Hamburg, Germany). This assay had a sensitivity of 0.43 nmol/L, and intra- and interassay coefficients of less than 10%.

2.3. Cortisol indices

Three cortisol indices were calculated for every day of cortisol assessment: CAR, diurnal slope and total daily output. Prior to calculation, cortisol values were log-transformed to correct for non-normality, following standard practices (Stone et al., 2001).

The CAR is defined as the increase in cortisol concentration that occurs during the first 30 min after awakening, relative to wake values. CAR was quantified using an area-under-the-curve technique (Pruessner et al., 2003). Briefly, a right-angle triangle is formed by the time between the wake and ½ h post-wake cortisol values (horizontal) and the log-transformed ½ h post-wake cortisol value (vertical). The CAR is calculated by taking the area of the triangle.

The diurnal slope is defined as the linear degree of change in cortisol levels across the day, from morning to evening, excluding the CAR. It is calculated by linearly regressing the log-transformed cortisol values against hours of the day, excluding the ½-h post-waking sample (Stewart and Seeman, 2000).

The total daily output is defined as the total cortisol output across a day, or total area under the cortisol curve, excluding the CAR. It was calculated via a trapezoidal function using techniques described by Pruessner et al. (2003). Briefly, a trapezoid is formed by two adjacent log-transformed cortisol values (vertical) and between zero or ground and the line drawn connecting the two cortisol values (horizontal). Trapezoids, and their corresponding areas, are calculated for wake–1 h, 1–4 h, 4–9 h and 9–14 h. Total daily output is then calculated by summing the areas.

Cortisol sample sufficiency criteria varied by study. For Study 1, if the 1 h post-wake sample was not collected within 20 min of the expected time it was discarded. The 1 h post-wake sample had to be present, as well as two of the 4, 9 and 11 h samples, in order to calculate the total daily output or diurnal slope. If these requirements were not met, neither the total daily output nor diurnal slope was calculated. If only one of the 4, 9 and 11 h samples were missing, then the missing data was imputed using the Excel trend function. For Study 1, all cortisol indices were calculated using the actual time of sample collection, as electronically recorded by the MEMS cap.

For Studies 2 and 3, if the 1-h post wake sample was not collected within 20 min of the expected time, or if the 4, 9 or 14 h post-wake samples were not collected within 60 min of the expected time, they were discarded. If three or more of the 1, 4, 9, and 14 h samples were missing then neither total daily output nor diurnal slope were calculated for that day. Otherwise, missing data were imputed using the Excel trend function. If the waking cortisol sample was missing, this value was imputed based on the same-day 1 h sample and the percent difference between the waking and 1 h sample on a consecutive day of assessment. If the 30 min post-wake sample was not collected within 20 min of the expected time or was missing entirely, then the CAR was not calculated. Again, all cortisol indices were calculated based on the actual time of sample collection, as electronically recorded on the Palm Zire.

3. Analytical strategy

To assess the stability of cortisol indices over time, a series of multi-level models were estimated using HLM 6.08 software (Raudenbush et al., 2004). Separate three-level models were estimated for each available cortisol index for each study. In all models each level represents a different unit of time. The lowest unit was one of the cortisol indices calculated for each day of sampling. As such, Level-1 represents the base-unit of "days," Level-2 the intermediate-unit of "visits," and Level-3 the highest-unit of "persons." Three sets of models were run for each cortisol index for each study: intercept-only models and time and between-person covariate models (see below). For all models, estimated diurnal slopes and error terms were allowed to vary freely.

3.1. Intercept-only models and partitioning cortisol index variance

Three-level intercept-only models were constructed for each cortisol index for each study, as shown below:

$$\begin{aligned}
&\text{Level 1}: \quad \text{Cortisol index}_{ijk} = \pi_0 + e_{ijk}\\
&\text{Level 2}: \quad \pi_0 = \beta_{00} + R_{0jk}\\
&\text{Level 3}: \quad \beta_{00} = \gamma_{000} + U_{00}
\end{aligned}$$

We modeled each cortisol endpoint as the value for that particular index on day *i*, visit *j*, for person *k*. Intercept-only multi-level models were used to partition the total variance according to the nested time structure imposed by the model, as represented by the intercept random effects for each level. For each index, the Level-2 intercept random effect, σ^2_{R0}, represents variance that is systematically accounted at the level of "visits." Similarly, the Level-3 intercept random effect, σ^2_{U00}, represents the variance that is systematically accounted for at the level of "persons." The Level-1 intercept random effect, σ^2_e, quantifies the residual cortisol index variance not accounted for by stability between visits (Level-2) or between persons (Level-3). It can be thought of as an error term, but when standard practices for cortisol sample collection and analysis are observed, the amount of variance in this random effect due to unsystematic measurement error is negligible (Kirschbaum et al., 1990; Hellhammer et al., 2007). As such, the Level-1 intercept random effect was interpreted as the proportion of cortisol index variance attributable to short-term or day-to day factors, and was calculated as a percentage (Shirtcliff et al., 2012):

$$\% = \frac{\sigma^2_e}{\sigma^2_e + \sigma^2_{R0} + \sigma^2_{U00}} \times 100$$

The stability of cortisol indices at the higher model levels are quantified by intraclass correlation coefficients (*ICC*s), defined as the proportion of between-unit variance to total variance. At Level-2, the *ICC* represents cortisol index stability on a visit-to-visit basis, and is calculated as (Siddiqui et al., 1996; Hox, 2010):

$$ICC_{visit} = \frac{\sigma^2_{R0} + \sigma^2_{U00}}{\sigma^2_e + \sigma^2_{R0} + \sigma^2_{U00}}$$

Calculating the Level-2 *ICC* in this fashion produces an estimate of "real-world" cortisol index stability at the visit-to-visit level (Hox, 2010).

An *ICC* can also be calculated at Level-3, which represents cortisol index stability between-persons, the degree to which a cortisol index is trait-like across the total follow-up period of the given study, or on average how alike two cortisol index values are within the same person, excluding visit-to-visit stability (Snijders and Bosker, 2012). It is calculated as follows (Hox, 2010):

$$ICC_{person} = \frac{\sigma^2_{U00}}{\sigma^2_e + \sigma^2_{R0} + \sigma^2_{U00}}$$

4. Results

Table 1 presents descriptive information, including each study's demographics and sampling procedures, and values for the derived cortisol indices.

4.1. Intercept-only results

Figs. 1–3 display the primary results. In them we show estimates of how much variance in each index is attributable to day-to-day (i.e., Level-1; see Fig. 1), visit-to-visit (i.e., Level-2; see Fig. 2), and between-person factors (i.e., Level-3; Fig. 3).

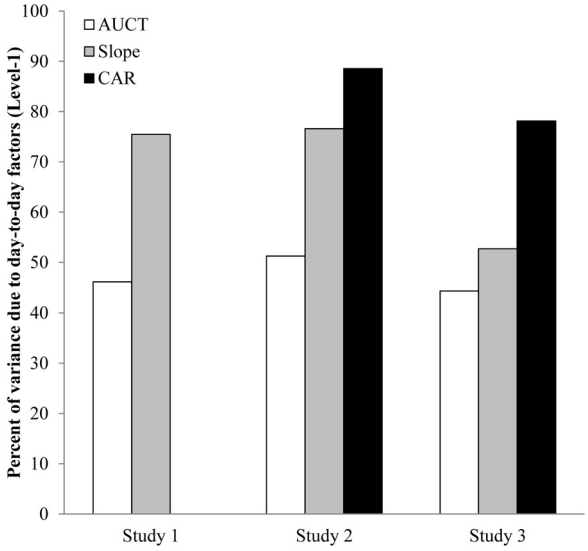

Figure 1 Proportion of total cortisol index variance explained by day-to-day factors (Level-1). Almost half or more of the total cortisol index variance across indices and studies is accounted for at this level.

4.1.1. Study 1

Partitioning analyses revealed that half or more of the variance in Study 1 cortisol indices was attributable to day-to-day fluctuations. For total daily output, the percent of variance at the Day level was 75.5%, and for diurnal slopes

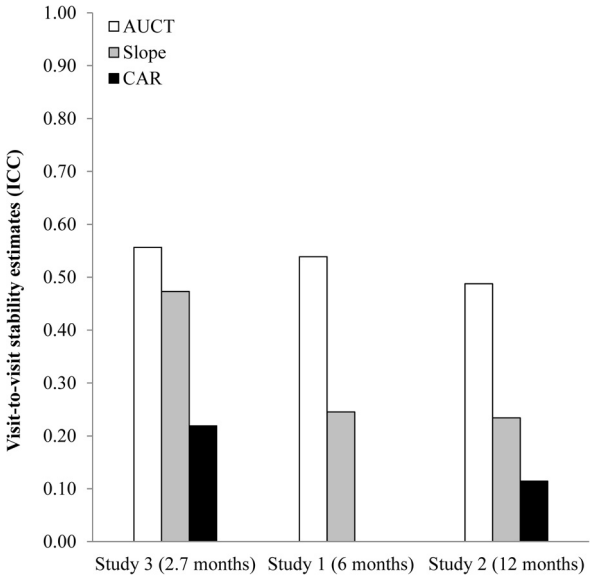

Figure 2 Cortisol index stability visit-to-visit or between visits (Level-2), arranged in order from shortest visit-to-visit follow-up (Study 3) to longest (Study 2). Larger ICCs denote greater stability over the time periods between visits. Across studies, total daily output appears the most stable, followed by diurnal slope and CAR. However, magnitude of stability estimates varies from sample to sample.

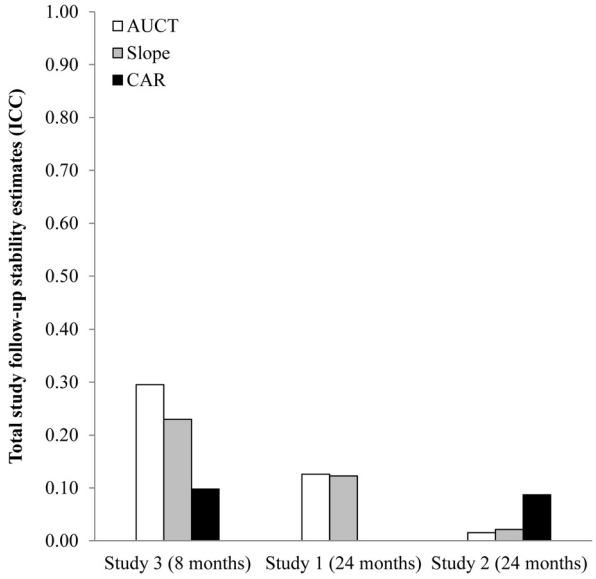

Figure 3 Cortisol index variance that is stable between-persons, or what is trait-like over the course of the follow-up, arranged in order from shortest total study follow-up (Study 3) to longest (Study 2). Larger *ICC*s denote greater stability over the course of the study. For Study 3, with total follow-up of less than a year, cortisol index stability reflects Level-2 patterns, with total daily output emerging as the most stable, followed by diurnal slope and CAR. For Studies 1 and 2, with total follow-up of more than a year, cortisol index stability is *ICC* = .13 or less. Again, even over studies with similar follow-up times, cortisol index stability estimates vary from study to study.

it was 46.1%. Study 1 did not assess the CAR so it is not reported here. Overall, these results suggest that the majority of the variance in cortisol indices is due to short-term, day-to-day or within-visit factors.

Analyses of the remaining variance indicated that statistically significant portions could be attributed to both the Visit and Person levels of the models. (For total daily output: Visit $\chi^2(282) = 737$, $p < .001$ and Person $\chi^2(114) = 195$, $p < .001$. For diurnal slope: Visit $\chi^2(283) = 379$, $p < .001$ and Person $\chi^2(114) = 208$, $p < .001$.) However, when ICCs were calculated it was apparent that for total daily output, most of this residual (non day-to-day) variance was explained at the level of Visits (*ICC* = .54) rather than Persons (*ICC* = .13). This was also true, but to a lesser extent, for diurnal slope, where values were modest at both the Visit (*ICC* = .25) and Person (*ICC* = .12) levels. As is apparent from these data, over the 6-month epoch between visits, total daily output values tend to be more stable than diurnal slope values. However, these indices seem to have fairly similar, and modest, levels of trait-like stability over the entire 2-year study.

4.1.2. Study 2

Similar to Study 1, partitioning analyses revealed that half or more of the cortisol index variance in Study 2 was attributable to day-to-day fluctuations: 51.3% for total daily output, 76.6% for diurnal slope and 88.6% for CAR. These results mirror those from Study 1, indicating that the majority of variance in cortisol indices is due to day-to-day or within-visit factors.

Again, most of the remaining variance was attributable to the level of Visits. Indeed, statistically significant portions of residual variance in each index could be attributed to Level-2, or Visit: Total daily output, $\chi^2(155) = 440$, $p < .001$, diurnal slope, $\chi^2(160) = 263$, $p < .001$, and CAR, $\chi^2(149) = 226$, $p < .001$. By contrast, the amount of residual variance explained at Level-3, or Person, was not significantly different from zero for either total daily output, $\chi^2(123) = 120$, $p > .50$, or diurnal slope, $\chi^2(123) = 123$, $p > .50$. These results indicate that, at least in this sample, total daily output and diurnal slope demonstrated no discernable trait-like tendencies over the 2 years of total follow-up. By contrast, a statistically significant portion of the residual variance in the CAR was attributable to the Persons level, $\chi^2(123) = 154$, $p = .032$.

Follow-up calculations revealed that for the 12-month epochs between visits, *ICCs* were .487, .234, and .114 for total daily output, diurnal slope, and CAR, respectively. As in Study 1, total daily output displayed the most between-visit stability of the indices. Because the Person Level did not explain significant residual variance in total daily output or diurnal slope, we did not estimate Level-3 *ICCs* for these indices. For CAR, this value was .087.

4.1.3. Study 3

As in Studies 1 and 2, partitioning analyses indicate that almost half or more of cortisol index variance in Study 3 was due to day-to-day fluctuations: 44.4% for total daily output, 52.7% for diurnal slope, and 78.1% for the CAR. Again, this mirrors the pattern observed in Studies 1 and 2, with the majority of the variance in cortisol indices is due to short-term, day-to-day factors.

Of the remaining variance, significant portions were attributed to both the Visit level (total daily output, $\chi^2(114) = 321$, $p < .001$, diurnal slope, $\chi^2(114) = 270$, $p < .001$, and CAR, $\chi^2(114) = 164$, $p = .002$) and Person level (total daily output, $\chi^2(43) = 154$, $p < .001$, diurnal slope, $\chi^2(43) = 128$, $p < .001$, and the CAR, $\chi^2(43) = 84.2$, $p < .001$). Again, calculated *ICCs* revealed that most of the residual, non day-to-day was explained at the level of Visits (total daily output: *ICC* = .556, diurnal slope: *ICC* = .473, CAR: *ICC* = .219) rather than at the level of Persons (total daily output: *ICC* = .295, diurnal slope: *ICC* = .230, CAR: *ICC* = .098). As in Study 1, over the average 3 months between visits, total daily output and diurnal slope values tended to be more stable than CAR. Over the average 8 months of study length, stability of all indices was modest.

4.1.4. Summary

Across studies and cortisol indices, the majority of the variance was attributable to short-term, day-to-day fluctuations. Of the indices, total daily output tended to be the most stable and trait-like, followed by diurnal slope, and then CAR. For illustrative purposes, we compared these stability estimates to those of two classic trait-like variables. Using the same partitioning analyses, we estimated stability coefficients in Study 2 for the personality traits Extraversion and Conscientiousness, which had been assessed annually with the Big Five Inventory (John et al., 1991), a widely used and extensively validated self-report measure of personality

(John et al., 2008). The *ICC*s at the highest level of the models were .77 and .75, respectively, suggesting a high degree of stability over 2 years of follow-up. None of the cortisol indices assessed here had stability coefficients that approached these values, even in the middle-aged sample (Study 3) or over the shortest time span between assessments (2.3 months, Study 3). We did the same analyses with depressed mood from Study 2, measured annually with the Beck Depression Inventory (Beck et al., 1961). The *ICC* at the highest level of the model was .48, a value greater than most of the cortisol indices stability estimates calculated here.

Also notable here are the differences in stability across studies. Across cortisol indices, estimates of stability tended to be higher in Study 3, which consisted of adults with a mean age of 50 years and shorter time between consecutive visits, compared to the other studies, which were comprised of children and adolescents with longer follow-ups.

4.2. Covariate models

It is possible that variations in sample timing, between visit lags, and follow-up lengths could be introducing systematic "noise" that constrains stability estimates. To test this possibility, we added variables to models reflecting weekend vs. weekday assessment, waking times, and between-visit length (see Supplemental material for more information). Stability estimates were re-calculated and compared to those from the simpler intercept-only models. Although including these covariates did change some of the stability estimates, the effect was neither systematic nor powerful: Most of the stability estimates were not greatly affected (Supplemental Tables 1 and 2). As such, we concluded that systematic variance in the timing of cortisol assessments does not appear to have substantially augmented our stability estimates.

It is also possible that the stability of cortisol indices is being obscured by individual difference variables, such as age, gender or body mass index (BMI). However, when we introduced these variables at Level-3, the stability estimates were only minimally affected (Supplemental Tables 3 and 4).

5. Discussion

The purpose of these analyses was to assess how state-like or trait-like common cortisol indices are, over both short- and long-term intervals, and in multiple healthy samples that span childhood through to middle-age. Across the studies and cortisol indices, half or more of the variance was partitioned at the day level, indicating that all of the indices exhibit state-like properties, characterized by substantial day-to-day fluctuations. The degree of stability varied somewhat across indices, with total cortisol output generally emerging as the most stable, followed by the diurnal slope and the CAR. However, none of the cortisol indices showed the degree of stability characteristic of classic traits like Extroversion and Conscientiousness, even under our most intensive assessment schedule (i.e., Study 3: saliva was collected 3 days per visit with 3 months between visits). These findings are consistent with previous reports on cortisol stability from Shirtcliff et al. (2012), who reported a 6-year stability estimate for the diurnal slope of .13, and Platje et al. (2013), who reported a 3-year CAR stability estimate of .17, as well as case studies demonstrating considerable day-to-day variation in the CAR (Stalder et al., 2009, 2010). Collectively, these findings suggest that, at least in healthy populations, the HPA axis is best viewed as a dynamic system, whose activity changes substantially from day-to-day.

These findings have implications for theory and research on cortisol in health psychology and related disciplines. In particular, focusing on short-term cortisol fluctuations may be an especially fruitful research avenue. Indeed, daily variations in negative mood (Adam et al., 2006; Giesbrecht et al., 2012), being alone or with other people (Adam, 2006; Matias et al., 2011), and interactions with other individuals (Papp et al., 2009; Slatcher et al., 2010) have been shown to co-vary with cortisol indices. There is also initial evidence that daily variations in cortisol indices co-vary with health indicators measured over the same time scale, such as self-reports of physical well-being (Adam et al., 2006) and joint pain severity (Savla and Almeida, 2008). However, more research of this nature is needed to clarify what role daily variations in cortisol plays on the pathway from stress to health. This work needs to focus on how daily swings in cortisol activity relate to downstream biological processes relevant to disease (e.g., daily variations in blood pressure, glucose regulation, etc.) and subjective reports of patient well-being (e.g., waxing and waning of symptoms).

Considered alongside the studies by Shirtcliff et al. (2012) and Platje et al. (2013), our findings suggest more limited prospects for research that approaches cortisol activity as a stable, trait-like feature of individuals. Across the studies reported here, cortisol indices displayed what could be called modest stability between visits, and even more limited durability over follow-up periods exceeding 1 year. To put these estimates into context, we used the Spearman–Brown Prophecy formula[1] to estimate how many study visits would be needed to achieve visit-to-visit stability of .60 for each index (Kingston and Tiemann, 2010). An *ICC* of .60 would suggest that 60% of the total variance reflects stable between-person differences, a value that is less than what we observed for personality characteristics like Extroversion and Conscientiousness, but still within the realm of what many psychologists would view as a reasonably durable individual difference. The most optimistic scenario was for achieving a stable total daily output measure in subjects like those who participated in Study 3 (Supplemental Table 5). To do that, one would need five visits, spaced on average 3 months apart, each of which entailed 3 days of sampling, with six collections per day. The least optimistic scenario entailed obtaining a .60 stability for CAR in subjects like those from Study 2. To do that, one would require 35 annual visits, with 2 days of sampling each, six collections per day. These estimates suggest that the temporal properties of cortisol indices vary a great deal, and that the feasibility of obtaining an estimate of stable, between-person activity does as well.

[1] The Spearman–Brown Prophecy formula is derived from Classic Test Theory, which assumes that values consist of a "true score" and measurement error. In other words, it applies to passive systems with a degree of inaccuracy. It is probably not appropriate to apply this or any Classic Test Theory formula to values from an adaptive, changing system. As such, the values given are meant to be rough estimates.

Unfortunately, stability estimates are not likely to be substantially improved by simply increasing the number of consecutive sampling days. When calculating stability estimates at Levels 2 and 3, our multilevel models took sampling days into account (Snijders and Bosker, 2012), essentially aggregating across the days subsumed within a visit. It is the case that including more sampling days within a visit will produce a more stable within-visit or day-to-day cortisol index aggregate (Hruschka et al., 2005; Hellhammer et al., 2007), but that will not necessarily translate into differences in stability over longer time intervals, such as months and years.

If as our data suggest, cortisol indices behave in a state-like manner, why might they have predicted longer-term health outcomes in some research? For example, in a 7-year study of patients with breast cancer (Sephton et al., 2000) reported flatter diurnal slopes presaged early mortality. And in study of adolescents and young adults, a high CAR predicted major depressive episodes 2.5 years later (Vrshek-Schallhorn et al., 2013). This is a difficult question to answer. One possibility is that people may vary in terms of cortisol index temporal stability. That is, although cortisol indices may have low stability generally or in healthy groups, there may be subgroups of healthy individuals or specific populations, such as medical patients or the elderly, who do not demonstrate plastic cortisol activity and/or consistently display a high total daily output, flat diurnal slope, or blunted CAR. If cortisol output is stably aberrant in these groups, it may confer increased susceptibility to disease-related processes that develop over protracted periods of time. Further research is required to assess cortisol index stability in medical patients or the elderly. For currently healthy subgroups, however, multi-wave datasets like ours could be used in conjunction with growth-mixture modeling to identify subgroups of highly stable individuals (if they exist) and evaluate their health outcomes. Indeed, it is possible that subgroups defined by age, gender or lifestyle factors, such as obesity, may vary in cortisol index stability.

There were several limitations to our analyses. First, the cortisol assessment schedule across studies produced a degree of freedom limitation in our models, restricting the number of covariates we could model and limiting power to detect significant random effects. Issues with power, however, should not have affected the size of our stability estimates. Another limitation related to restricted degrees of freedom is that we were unable to consider all variables that may have been introducing systematic ''noise'' or variability, and potentially masking ''true stability,'' such as daily stress, negative affect, sleep quality, puberty status, and health behaviors (Adam and Kumari, 2009). It is possible that if these variables were included in the model, stability estimates would have improved. However, controlling for puberty status did not greatly change long-term CAR and diurnal slope stability estimates in the Shirtcliff et al. (2012) and Platje et al. (2013) models. Data sets with more frequent cortisol sampling would be necessary to assess the influence of other moment-to-moment variables. However, if including these variables in models did boost stability estimates, it would only provide further evidence of the state-like tendencies of cortisol indices. It should also be acknowledged that the diurnal saliva profiles are only one technique for measuring cortisol secretion. Other techniques, like analysis of hair cortisol, may prove to be superior indicators of trait-like cortisol production and total exposure over long periods of time (Stalder et al., 2012), and cortisol reactivity to challenges may better tap state-related changes in HPA axis activity. Lastly, although our samples consisted of healthy community participants, they were not selected to be representative of the population at large. They also do not represent the full range of the lifespan. Further analyses are required to establish healthy and developmentally representative population norms.

In summary, most of the variance in cortisol indices in healthy samples is due to short-term, day-to-day fluctuations, reflecting the adaptive, dynamic activity of the HPA axis. These features need to be considered in theoretical frameworks that emphasize cortisol as a stress-health mediator, and in research designs that call for capturing ''trait-like'' differences in this hormone's activity.

Role of the funding source

The funding agency did not play any role in the analysis of data, interpretation of results, or drafting of manuscript.

Conflict of interest statement

None declared.

Acknowledgments

This research was supported by grants from the National Institute of Child Health and Human Development (R01 HD058052), the National Institutes of Health (HL073975), the Michael Smith Foundations for Health Research, and the Canadian Institutes of Health Research (#89763).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.psyneuen.2013.09.016.

References

Adam, E.K., 2006. Transactions among adolescent trait and state emotion and diurnal and momentary cortisol activity in naturalistic settings. Psychoneuroendocrinology 31, 664—679.

Adam, E.K., Doane, L.D., Zinbarg, R.E., Mineka, S., Craske, M.G., Griffith, J.W., 2010. Prospective prediction of major depressive disorder from cortisol awakening responses in adolescence. Psychoneuroendocrinology 35, 921—931.

Adam, E.K., Hawkley, L.C., Kudielka, B.M., Cacioppo, J.T., 2006. Day-to-day dynamics of experience-cortisol associations in a population-based sample of older adults. Proc. Natl. Acad. Sci. U. S. A. 103, 17058—17063.

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

Anagnostis, P., Athyros, V.G., Tziomalos, K., Karagiannis, A., Mikhailidis, D.P., 2009. The pathogenetic role of cortisol in the metabolic syndrome: a hypothesis. J. Clin. Endocrinol. Metab. 94, 2692—2701.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression. Arch. Gen. Psychiatry 4, 561—571.

Catley, D., Kaell, A.T., Kirschbaum, C., Stone, A.A., 2000. A naturalistic evaluation of cortisol secretion in persons with fibromyalgia and rheumatoid arthritis. Arthritis Care Res. 13, 51—61.

Champaneri, S., Xu, X., Carnethon, M.R., Bertoni, A.G., Seeman, T., Diez Roux, A., Golden, S.H., 2012. Diurnal salivary cortisol and urinary catecholamines are associated with diabetes mellitus: the Multi-Ethnic Study of Atherosclerosis. Metabolism. 61, 986—995.

Chen, E., Fisher, E.B., Bacharier, L.B., Strunk, R.C., 2003. Socioeconomic status, stress, and immune markers in adolescents with asthma. Psychosom. Med. 65, 984—992.

Chen, E., Hanson, M.D., Paterson, L.Q., Griffin, M.J., Walker, H.A., Miller, G.E., 2006. Socioeconomic status and inflammatory processes in childhood asthma: the role of psychological stress. J. Allergy Clin. Immunol. 117, 1014—1020.

Chikanza, I.C., Petrou, P., Kingsley, G., Chrousos, G., Panayi, G.S., 1992. Defective hypothalamic response to immune and inflammatory stimuli in patients with rheumatoid arthritis. Arthritis Rheum. 35, 1281—1288.

Dekker, M.J., Koper, J.W., van Aken, M.O., Pols, H.A., Hofman, A., de Jong, F.H., Kirschbaum, C., Witteman, J.C., Lamberts, S.W., Tiemeier, H., 2008. Salivary cortisol is related to atherosclerosis of carotid arteries. J. Clin. Endocrinol. Metab. 93, 3741—3747.

Fei, G.H., Liu, R.Y., Zhang, Z.H., Zhou, J.N., 2004. Alterations in circadian rhythms of melatonin and cortisol in patients with bronchial asthma. Acta Pharmacol. Sin. 25, 651—656.

Giesbrecht, G.F., Campbell, T., Letourneau, N., Kooistra, L., Kaplan, B., Team, A.P.S., 2012. Psychological distress and salivary cortisol covary within persons during pregnancy. Psychoneuroendocrinology 37, 270—279.

Hajat, A., Diez-Roux, A.V., Sanchez, B.N., Holvoet, P., Lima, J.A., Merkin, S.S., Polak, J.F., Seeman, T.E., Wu, M., 2013. Examining the association between salivary cortisol levels and subclinical measures of atherosclerosis: the Multi-Ethnic Study of Atherosclerosis. Psychoneuroendocrinology 38, 1036—1046.

Hellhammer, J., Fries, E., Schweisthal, O.W., Schlotz, W., Stone, A.A., Hagemann, D., 2007. Several daily measurements are necessary to reliably assess the cortisol rise after awakening: state- and trait components. Psychoneuroendocrinology 32, 80—86.

Hox, J.J., 2010. Multilevel Analysis: Techniques and Applications, 2nd ed. Routledge, Great Britain.

Hruschka, D.J., Kohrt, B.A., Worthman, C.M., 2005. Estimating between- and within-individual variation in cortisol levels using multilevel models. Psychoneuroendocrinology 30, 698—714.

John, O.P., Donahue, E.M., Kentle, R.L., 1991. The Big Five Inventory — Versions 4a and 54. University of California, Berkeley, Institute of Personality and Social Research, Berkeley.

John, O.P., Naumann, L.P., Soto, S.J., 2008. Paradigm shift to the integrative big-five trait taxonomy: history, measurement and conceptual issues. In: John, O.P., Robins, W.J. (Eds.), Handbook of Personality: Theory and Research. Guilford Press, New York.

Kingston, N., Tiemann, G., 2010. Spearman—Brown Prophecy formula. In: Salkind, N.J. (Ed.), Encyclopedia of Research Design. Sage Reference, Thousand Oaks, CA, pp. 1402—1404.

Kirschbaum, C., Hellhammer, D.H., 1989. Salivary cortisol in psychobiological research: an overview. Neuropsychobiology 22, 150—169.

Kirschbaum, C., Steyer, R., Eid, M., Patalla, U., Schwenkmezger, P., Hellhammer, D.H., 1990. Cortisol and behavior: 2. Application of a latent state-trait model to salivary cortisol. Psychoneuroendocrinology 15, 297—307.

Kumari, M., Shipley, M., Stafford, M., Kivimaki, M., 2011. Association of diurnal patterns in salivary cortisol with all-cause and cardiovascular mortality: findings from the Whitehall II study. J. Clin. Endocrinol. Metab. 96, 1478—1485.

Matias, G.P., Nicolson, N.A., Freire, T., 2011. Solitude and cortisol: associations with state and trait affect in daily life. Biol. Psychol. 86, 314—319.

Matthews, K., Schwartz, J., Cohen, S., Seeman, T., 2006. Diurnal cortisol decline is related to coronary calcification: CARDIA study. Psychosom. Med. 68, 657—661.

McEwen, B.S., 1998. Protective and damaging effects of stress mediators. N. Engl. J. Med. 338, 171—179.

Miller, G.E., Chen, E., Zhou, E.S., 2007. If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. Psychol. Bull. 133, 25—45.

Miller, G.E., Cole, S.W., 2012. Clustering of depression and inflammation in adolescents previously exposed to childhood adversity. Biol. Psychiatry 72, 34—40.

Orth-Gomer, K., 2000. Marital stress worsens prognosis in women with coronary Heart Disease: The Stockholm Female Coronary Risk Study. JAMA 284, 3008—3014.

Papp, L.M., Pendry, P., Adam, E.K., 2009. Mother—adolescent physiological synchrony in naturalistic settings: within-family cortisol associations and moderators. J. Fam. Psychol. 23, 882—894.

Platje, E., Vermeiren, R.R., Branje, S.J., Doreleijers, T.A., Meeus, W.H., Koot, H.M., Frijns, T., van Lier, P.A., Jansen, L.M., 2013. Long-term stability of the cortisol awakening response over adolescence. Psychoneuroendocrinology 38, 271—280.

Pruessner, J.C., Kirschbaum, C., Meinlschmid, G., Hellhammer, D.H., 2003. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. Psychoneuroendocrinology 28, 916—931.

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.

Raudenbush, S.W., Bryk, A.S., Congdon, R., 2004. HLM 6 for Windows, 6.08 ed. , Lincolnwood, IL.

Rohleder, N., Marin, T.J., Ma, R., Miller, G.E., 2009. Biologic cost of caring for a cancer patient: dysregulation of pro- and anti-inflammatory signaling pathways. JCO 27, 2909—2915.

Sapolsky, R.M., 2000. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. Endocr. Rev. 21, 55—89.

Savla, T., Almeida, D.M., 2008. Daily Stress, Affect and Fluctuations in Pain Symptoms and Cortisol Rhythm, Gerontologist. Gerontological Society of America, National Harbor, MD394.

Schanberg, L.E., Sandstrom, M.J., Starr, K., Gil, K.M., Lefebvre, J.C., Keefe, F.J., Affleck, G., Tennen, H., 2000. The relationship of daily mood and stressful events to symptoms in juvenile rheumatic disease. Arthritis Care Res. 13, 33—41.

Sephton, S., Sapolsky, R.M., Kraemer, H.C., Spiegel, D., 2000. Diurnal cortisol rhythm as a predictor of breast cancer survival. J. Natl. Cancer Inst. 92, 994—1000.

Shirtcliff, E.A., Allison, A.L., Armstrong, J.M., Slattery, M.J., Kalin, N.H., Essex, M.J., 2012. Longitudinal stability and developmental properties of salivary cortisol levels and circadian rhythms from childhood to adolescence. Dev. Psychobiol. 54, 493—502.

Siddiqui, O., Hedeker, D., Flay, B.R., Hu, F.B., 1996. Intraclass correlation estimates in a school-based smoking prevention study. Outcome and mediating variables, by sex and ethnicity. Am. J. Epidemiol. 144, 425—433.

Slatcher, R.B., Robles, T.F., Repetti, R.L., Fellows, M.D., 2010. Momentary work worries, marital disclosure, and salivary cortisol among parents of young children. Psychosom. Med. 72, 887—896.

Smyth, J.M., Ockenfels, M.C., Gorin, A.A., Catley, D., Porter, L.S., Kirschbaum, C., Hellhammer, D.H., Stone, A.A., 1997. Individual differences in the diurnal cycle of cortisol. Psychoneuroendocrinology 22, 89—105.

Snijders, T.A.B., Bosker, R.J., 2012. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, 2nd ed. Sage, Thousand Oaks, CA.

Stalder, T., Evans, P., Hucklebridge, F., Clow, A., 2010. Associations between psychosocial state variables and the cortisol awakening response in a single case study. Psychoneuroendocrinology 35, 209—214.

Stalder, T., Hucklebridge, F., Evans, P., Clow, A., 2009. Use of a single case study design to examine state variation in the cortisol awakening response: relationship with time of awakening. Psychoneuroendocrinology 34, 607—614.

Stalder, T., Steudte, S., Miller, R., Skoluda, N., Dettenborn, L., Kirschbaum, C., 2012. Intraindividual stability of hair cortisol concentrations. Psychoneuroendocrinology 37, 602—610.

Stetler, C., Miller, G.E., 2005. Blunted cortisol response to awakening in mild to moderate depression: regulatory influences of sleep patterns and social contacts. J. Abnorm. Psychol. 114, 697—705.

Stetler, C., Miller, G.E., 2011. Depression and hypothalamic-pituitary-adrenal activation: a quantitative summary of four decades of research. Psychosom. Med. 73, 114—126.

Stewart, J.C., Seeman, T., 2000. Salivary Cortisol Measurement. MacArthur Research Network on SES & Health, Allostatic Load Notebook.

Stone, A., Schwartz, J.E., Smyth, J., Kirschbaum, C., Cohen, S., Hellhammer, D.H., Grossman, S., 2001. Individual differences in the diurnal cycle of salivary free cortisol: a replication of flattened cycles for some individuals. Psychoneuroendocrinology 26, 295—306.

Strickland, P., Morriss, R., Wearden, A., Deakin, B., 1998. A comparison of salivary cortisol in chronic fatigue syndrome, community depression and healthy controls. J. Affect. Disord. 47, 191—194.

Vrshek-Schallhorn, S., Doane, L.D., Mineka, S., Zinbarg, R.E., Craske, M.G., Adam, E.K., 2013. The cortisol awakening response predicts major depression: predictive stability over a 4-year follow-up and effect of depression history. Psychol. Med. 43, 483—493.