Balancing scientific accuracy and participant burden: testing the impact of sampling intensity on diurnal cortisol indices

Lindsay Till Hoyt[a], **Katherine B. Ehrlich**[b], **Heining Cham**[a], **Emma K. Adam**[b,c]

[a]Department of Psychology, Fordham University, Bronx, NY, USA

[b]Institute for Policy Research, Northwestern University, Evanston, IL, USA

[c]School of Education and Social Policy, Northwestern University, Evanston, IL, USA

Abstract

Despite the increasing popularity of incorporating salivary cortisol measurement into health and social science research, relatively little empirical work has been conducted on the number of saliva samples across the day required to capture key features of the diurnal cortisol rhythm, such as the diurnal cortisol slope, the area under the curve (AUC), and the cortisol awakening response (CAR). The primary purpose of this study is to compare slope, AUC, and CAR measures obtained from an intensive sampling protocol with estimates from less intensive protocols, to identify sampling protocols with minimal participant burden that still provide reasonably accurate assessment of each of these measures. Twenty-four healthy adults provided samples four times in the first hour awake, and then every hour throughout the rest of the day until bedtime ($M = 17.8$ samples/day; $SD = 2.0$), over two consecutive days ($N = 862$ total samples). We compared measures calculated from this maximum intensity protocol to measures calculated from two to six sampling points per day. Overall, results show that salivary cortisol protocols with two fixed samples (waking and bedtime) and three additional daily samples, closely approximates the full cortisol decline (slope). Abbreviated sampling protocols of total cortisol exposure across the day (AUC), however, were not well approximated by reduced sampling protocols. CAR measures based on only two samples, including waking cortisol and a second sample measured at a fixed time point between 30 and 60 min after waking, provided a measure of the CAR that closely approximated CAR measures obtained from 3 or 4 sampling points.

Keywords

HPA axis; salivary cortisol measurement; diurnal cortisol; cortisol slope; area under the curve; cortisol awakening response

---

CONTACT Lindsay Till Hoyt lhoyt1@fordham.edu Department of Psychology, Fordham University, Dealy Hall, 411 East Fordham Rd, Bronx, NY 10458, USA.

Introduction

Salivary cortisol is one of the most popular biomarkers employed in research on stress and stress-related disease. Healthy hypothalamic-pituitary-adrenocortical (HPA) axis function is characterized by a strong cortisol diurnal rhythm, and deviations from the typical diurnal cycle provide valuable information regarding environmental influences on the HPA axis and the role of the HPA axis in disease processes (Chrousos, 2009; Nader et al., 2010). Three key aspects of the diurnal cortisol rhythm are the diurnal cortisol slope, the area under the curve (AUC), and the cortisol awakening response (CAR).

Diurnal cortisol slope is measured as the rate of change in cortisol levels from waking to bedtime and a steeper decline (i.e. a more negative slope) is typically associated with better health and psychosocial functioning (Adam et al., 2006; Huppert, 2006). A flatter slope, on the other hand, has been linked with a history of stress exposure (Adam & Gunnar, 2001; Gunnar & Vazquez, 2001; Matthews et al., 2006; Miller et al., 2007; Suglia et al., 2010) and disease processes (Abercrombie et al., 2004; Heim et al., 2000; Matthews et al., 2006; Sephton et al., 2000). The AUC reflects the average level of cortisol across the day, which is not strongly associated with the cortisol slope (Adam & Kumari, 2009). Although associations among stress, health variables, and AUC cortisol are inconsistent, it is generally believed that both very large and very small AUCs (representing hyperactivity and hypoactivity, respectively) signify poor psychological and physiological functioning (Saxbe, 2008). Finally, the CAR (i.e. the increase in cortisol from waking to approximately 30–45 min after waking) is a distinct feature of the cortisol diurnal rhythm, thought to play a role in regaining arousal upon waking or helping people meet the anticipated demands of their day (Adam et al., 2006; Clow et al., 2010). However, the CAR is sensitive to chronic stress (Chida & Steptoe, 2009; Kunz-Ebrecht et al., 2004; Schlotz et al., 2004) and a higher CAR prospectively predicts the development of major depressive disorder (Adam et al., 2010; Vrshek-Schallhorn et al., 2013) and first onsets of anxiety disorder (Adam et al., 2014).

In order to capture these diurnal cortisol indices efficiently, and without overtaxing participants, many studies rely on the abbreviated sampling protocols utilizing 2–6 samples per day. However, no one has empirically tested whether such abbreviated protocols reasonably approximate the pattern that would be obtained if more frequent sampling were employed. To investigate this question, 24 healthy adults participated in an intensive two-day data collection protocol, providing samples four times in the first hour awake, and then every hour throughout the day until bedtime. Although this highly intensive protocol is not practical or feasible in most naturalistic research, it provides an idealized standard from which we can validate commonly used diurnal salivary cortisol protocols in health and social science research.

Methods

Participants

Students and community members from a large Midwestern city were recruited by word-of-mouth and flyers posted on campus. Eligible participants had to be: (a) between the ages of 18 and 49, (b) not currently taking corticosteroid medication, and (c) not pregnant. The first

eligible 25 individuals to respond were selected for the study and consented in person by a member of the study research team. One person had to withdraw for medical reasons. The final sample consisted of 24 healthy adults (17 female) between the ages of 21 and 42 years ($M = 27.5$; $SD = 5.2$). Most of the sample was White ($n = 20$) and the rest were Asian ($n = 4$). Participants received a $30 gift card upon completion of the study.

Procedures

For each of two typical weekdays, participants were asked to provide small samples of saliva in the morning immediately upon awakening (waking cortisol level), 30, 45, and 60 min after waking (three CAR samples), in the evening immediately before bedtime (bedtime cortisol level), and every hour on the hour during the day. Participants were instructed not to eat, drink, or brush their teeth in the first hour after waking, and avoid eating food in the half hour before other samples, if possible. Saliva sampling involved expelling the saliva through a small straw into a sterile cryogenic vial. They wrote the exact time of collection on a label attached to the vial. Samples were refrigerated by participants as soon as possible, and then returned (in person) to the lab when sampling was complete, where they were stored at −20 °C until they were shipped for processing. Salivary cortisol levels are robust to variations in temperature and motion similar to those experienced in a trip through the postal system (Clements & Parker, 1998).

Samples were sent on dry ice to the Biochemisches Labor at the University of Trier, Germany and were assayed in duplicate for cortisol using a time-resolved immunoassay with fluorometric detection (DELFIA). Duplicate cortisol results were averaged and mean values were used in analysis. Intra-assay coefficients of variation (CVs) were between 4.0% and 6.7%, and inter-assay CVs ranged from 7.1% to 9.0%. Raw cortisol values were winsorized at 1.8 μg/dl ($n = 2$) to reduce the effects of outliers on the analysis.

Participants also reported about health and lifestyle factors, such as medication use (e.g. birth control), consumption of caffeine and alcohol, use of nicotine, timing of menstrual cycle, pregnancy, presence of chronic illness, and their height and weight. When participants collected their hourly saliva samples, they also completed brief diary reports of their moods, activities, and health behaviors over the past hour. Participants also wore the Actiwatch Score (Phillips Respironics Inc., Bend, OR), a wrist-based accelerometer placed on the non-dominant hand that quantifies movement across the day and during sleep. All procedures were approved by the Institutional Review Board at Northwestern University.

Data analysis

All diurnal cortisol measures were created based on natural logarithmic transformed cortisol values. This transformation serves several important purposes: It reduces the positive skew of the distribution of cortisol, it reduces the impact of outlying values, and it serves to help linearize the association between cortisol levels and time of day (Adam & Kumari, 2009). For the AUC and CAR, indices were calculated separately for each day and averaged; for the diurnal cortisol slope, regression-based slopes were fit by regressing time of day on cortisol levels across the average of both days of data.

For the diurnal slope and AUC, the accuracy of the reduced cortisol sampling protocols was examined by comparing measures based on the *maximum* number of data points (i.e. waking, all the CAR samples, plus all the hourly samples; approximately 36 samples per person across both days) compared to those based on *medium* number of data points (i.e. waking, CAR, 3, 8, and 12 h after waking, and bedtime each day) or *minimum* number of data points (i.e. waking, CAR, and bedtime each day). Medium and minimum intensity indices were estimated by selecting subsets of the data points obtained during the maximum intensity protocol.

Slope calculations

Slopes were estimated using multiple regression techniques using all cortisol data samples across both study days. Because we wanted to estimate the same slope parameter across all three protocols (maximum, medium, minimum), we calculated the linear slope at waking (i.e. time of day is centered at waking). Natural logarithmic transformed cortisol levels were regressed on cortisol sampling times separately for each person, with the person-specific beta coefficient or the effect of time of day on cortisol representing the cortisol slope at waking: b_1 in Equation (1), below.

$$LnCORT = b_0 + b_1 \times Time + e \tag{1}$$

Although previous work often includes a quadratic term in regression-based slope calculations, we chose not to include this term in our Maximum and Medium equations because it was not possible to estimate a quadratic form in the Minimum procedures. As reported, we used a natural log transformation across all slope measures to help linearize the association between cortisol levels and time of day.

Most researchers exclude the CAR in the slope calculation (Adam et al., 2006; Cohen et al., 2006; Polk et al., 2005; Weissbecker et al., 2006) because of suggestions that the CAR may be regulated by different neurobiological mechanisms than the rest of the diurnal curve (Clow et al., 2004). Therefore, our primary slope analyses were those in which the CAR samples were removed from the dataset before running the regression (MaxSlope$_E$, MedSlope$_E$, MinSlope$_E$). However, because some researchers have calculated slopes from the peak of the CAR to bedtime, we also calculated a second set of slope analyses in which the CAR data points were included when estimating the slopes (MaxSlope$_I$, MedSlope$_I$, MinSlope$_I$). To create a consistent naming convention, prefixes refer to the level of intensity of the protocol (Max, Med, Min) and subscripts refer to whether the CAR was *excluded* (E) or *included* (I) in each calculation. (Based on the most common procedures from previous research, we selected the 30-min post-waking value to represent the CAR for the MedSlope$_I$ and MinSlope$_I$. The MaxSlope$_I$ includes all three CAR data points.)

Finally, we created multiple diurnal cortisol slope estimates calculated from just two data points – a simplified measure that has been reported in recent studies (e.g. Folkesson et al., 2014; Rotenberg et al., 2012). A slope from wakeup to bedtime and a series of slope measurements from each of the 30/45/60-min post-awakening CAR sample to bedtime were calculated by subtracting the value of the earlier cortisol sample from the value of the later

cortisol sample, and dividing by the total length of time between their respective sampling times.

$$\text{Slope}_{\text{WAKE}} = (\text{LnCORT}_{\text{bed}} - \text{LnCORT}_{\text{wake}})/(\text{time}_{\text{bed}} - \text{time}_{\text{wake}}) \quad (2)$$

$$\text{Slope}_{\text{CAR}} = (\text{LnCORT}_{\text{bed}} - \text{LnCORT}_{\text{CAR}})/(\text{time}_{\text{bed}} - \text{time}_{\text{CAR}}) \quad (3)$$

Area under the curve

AUC measures were calculated using the trapezoid method [AUC with respect to ground, AUC_G, as described by Pruessner and colleagues (2003)]. In summary, when a line graph is plotted for each individual across the day, with cortisol level on the Y-axis and time since waking for each sample (*n* total samples) on the X-axis, the result is ($n - 1$) polygons under that line, the areas of which can be combined to create a summary measure of total daily cortisol. An optional transformation is to divide the AUC_G value by each person's total time awake (i.e. wake time subtracted from bedtime), which can then be interpreted as average cortisol exposure per hour across the day. We examine both versions of the AUC_G, but focus on the latter because it adjusts for individual differences in total time awake.

The CAR value is sometimes, but not always, excluded in AUC measurements to prevent morning awakening responses from having an undue influence on total daily output values. Therefore, the first set of AUC measurements in the current study incorporates all data points, including the CAR data points (MaxAUC_I, MedAUC_I, MinAUC_I). (As with the slope estimates, we selected the 30-min post-waking value to represent the CAR for the MedAUC_I and MinAUC_I; the MaxAUC_I includes all three CAR data points.) The second set of AUC measurements excludes all of the CAR data points (MaxAUC_E, MedAUC_E, MinAUC_E).

Cortisol awakening response

Our protocol called for four saliva samples during the first hour after waking in order to compare the most common markers of the post-awakening cortisol rise. This includes three *simple difference* measures at 30-, 45-, and 60-min post-awakening (CAR30_D, CAR45_D, CAR60_D) with the general formula:

$$\text{CAR}_D = \left(\text{LnCORT}_{\text{post-awakening}} - \text{LnCORT}_{\text{wake}}\right) \quad (4)$$

We also measured total awakening cortisol output using *area under the curve* across the first 30 min ($\text{CAR30}_{\text{AUC}}$), 45 min ($\text{CAR45}_{\text{AUC}}$), or 60 min ($\text{CAR60}_{\text{AUC}}$) awake in the morning. Although the AUC measure in the previous section captures total hormonal output across the day (i.e. AUC_G), the CAR_{AUC} only measures the area above the waking value (i.e. AUC with respect to increase; AUC_I), which captures the amount of cortisol increase above the waking value. (AUC equations are described in detail elsewhere [Pruessner et al., 2003]). As reported in recent expert consensus guidelines, only the dynamic post-awakening cortisol secretion (i.e. cortisol change due to awakening) is accurately referred to as the

"CAR" (Stalder et al., 2016). The two types of CAR measures in this paper (i.e. CAR_D and CAR_{AUC}) are illustrated in Figure 1.

Importantly, in order to compare the CAR_{AUC} values calculated across different lengths of time (i.e. 30, 45, 60 min), we divided each of the CAR_{AUC} values by the time from waking to the time of the final CAR sample utilized (i.e. total time awake thus far, in minutes), which captures a measure of the average cortisol level per minute for each CAR_{AUC} measure.

Comparison analysis

In order to examine the extent to which the reduced sampling protocols resembled the maximum intensity protocol, we ran a series of intraclass correlations (ICCs) between the minimum, medium, and maximum versions of each set of cortisol parameters: slopes, AUCs, and CARs. ICCs take into account both association (i.e. covariation) and bias (i.e. whether levels are systematically higher or lower), and thus represent a more stringent test of the similarity of the maximum, medium, and minimum diurnal cortisol composites than Pearson correlations (McGraw & Wong, 1996; Shrout & Fleiss, 1979). ICCs were calculated comparing the various cortisol measures computed using the maximum, medium and minimum protocols for each person, averaged across the two days of testing. (ICCs were very similar when estimated with one day of data [Day 1 or Day 2] versus the average [Day 1 and Day 2 mean], therefore the results will focus on the average only.)

These comparisons speak to how much measurement of each cortisol index (slope, AUC, CAR) is affected by reducing the number of samples included in the calculations. Of course, more intensive sampling protocols are desirable, in that they provide more precise measurements of the diurnal cortisol measure of interest; however, if lower intensity protocols decrease participant burden, without resulting in dramatically lowered accuracy, these protocols should be considered as possible options for naturalistic diurnal cortisol research. We used the nonparametric bootstrapping method to compare the differences between a pair of ICCs (DiCiccio & Efron, 1996; Efron & Tibshirani, 1986). We obtained 2000 equally sized bootstrap samples by simple random sampling with replacement. For each bootstrap sample, we estimated the difference of the ICCs. The empirical sampling distribution of the difference of the ICCs is the distribution of these 2000 values from the bootstrap samples. The lower and upper limits of the 95% confidence interval of the difference of two ICCs equal to the 2.5th and 97.5th percentiles of the empirical sampling distribution. When the 95% confidence interval does not include the value of zero, the two ICCs are significantly different at $p = 0.05$.

Results

Overall, participants demonstrated very high compliance with the intensive study protocols, providing an average of 18 samples per day (range of 14–20 samples). There was no missing data for waking and bedtime cortisol, and only minimal missing samples from the incremental CAR sampling each day ($n = 1$ for CAR30, $n = 2$ for CAR45, $n = 2$ for CAR60). After data were aggregated across both study days, there were only 5 missing data

points for diurnal slope and AUC calculations (missing samples at hour 9 [$n = 1$], hour 12 [$n = 1$], hour 14 [$n = 2$], hour 15 [$n = 1$]) and no missing data points for CAR calculations.

Given the dynamic changes in cortisol that typically occur in the first hour after waking, correct measurement of waking time is critical. Therefore, we examined actigraphy data to check compliance for waking cortisol reporting (i.e. number of minutes between the waking cortisol sample and the actigraph recorded waking time). We identified 2 participants with one day of late waking cortisol (i.e. the sample was taken more than 10 min after the actigraph recorded exact time of waking), and 4 participants with both days of late waking cortisol. Therefore, we ran two sets of analyses, one set removing all 10 instances of low compliance ($n = 20$ participants; 38 total sampling days), and a second set with the full sample ($n = 24$ participants; 48 total sampling days). We will focus on the former (compliant) set of measures in the main text and tables.

We also examined the self-reported CAR times to ensure that all morning saliva samples were collected within close proximity of the targeted times (CAR30 $M = 30.20$ min, $SD = 1.66$; CAR45 $M = 45.32$ min, $SD = 1.68$; CAR60 $M = 60.43$ min, $SD = 2.04$). Specifically, all participants provided each of their CAR samples within 3 min of the targeted sampling times (i.e. 30/45/60-min post-waking), except two persons who were 3.6 and 7.2 min late for their CAR60 sample on one of the two sampling days. (Results do not change when removing the two individuals who were more than 3 min late.)

Overall, most participants demonstrated the typical pattern of cortisol across the day: cortisol levels were high upon waking ($M = 0.353$ μg/dl), increased strongly after waking ($M = 0.535$ μg/dl, 52% increase after 30 min; $M = 0.550$, 56% increase from waking through 45 min; $M = 0.496$, 41% increase from waking through 60 min), dropped rapidly over the first few hours after waking, and then declined more slowly throughout the remainder of the day. As expected given the typical diurnal cortisol rhythm, the average slopes of all type and calculation method were negative in value and bedtime cortisol values were low ($M = 0.072$ μg/dl). These descriptive statistics are summarized in Table 1, and cortisol patterns (showing all data points across the two days of testing by time of day) for a randomly selected subsample of four participants are visually represented in Figure 2.

Morning peaks occurred at 30 min in 32.5% of study days, and at 45 min in 32.5% of study days, as expected from previous research (Clow et al., 2004). However for 20% of study days, participants displayed a continuing morning rise until a full hour after waking, and on 5% of days, samples did not peak until approximately 2-h post-waking. Across the final 10% of study days, participants did *not* exhibit a morning rise at all, with peak morning cortisol levels upon waking.

Diurnal slope

There were relatively high ICCs between regression-based slope estimates for protocols excluding the CAR values and those including the CAR values (Max: ICC = 0.805; Med: ICC = 0.902; Min: ICC = 0.890; $ps < 0.001$). The medium intensity protocol slope excluding the CAR (i.e. five samples per day averaged across both study days) was highly correlated with the maximum intensity protocol slope excluding the CAR (i.e. hourly samples) with

an ICC of 0.867 ($p < 0.001$). The ICC between medium and maximum protocol slopes including the CAR was 0.873, $p < 0.001$. The difference between these two ICCs (excluding and including the CAR) was not significant ($p > 0.05$). Table 2 shows the full set of ICCs between the maximum, medium, and minimal sampling protocols excluding the CAR (the most common method in current research), and the four additional "simple slope" measures. (Sampling protocols excluding the CAR [i.e. MaxSlope$_E$] were very similar to those including CAR [i.e. MaxSlope$_I$], therefore the results in Table 2 are limited to MaxSlope$_E$ only. Slope analyses including the CAR are available upon request.) In the supplemental set of analyses with the full sample we found a very similar set of results (i.e. all ICC coefficients were within 0.09 of the original estimates).

The estimates for the minimum regression-based slope and 2-point simple slopes in relation to the maximum protocol ranged from 0.550 to 0.728 ($ps < 0.001$), with the highest ICC (0.728) for the wake to bedtime simple slope. The ICCs between the two-point slope estimates from the CAR (30/45/60-min post-waking) to bedtime were all under 0.600 ($p < 0.001$). These ICCs were significantly less than the ICC between the medium and maximum protocol ($ps < 0.05$). This reduced accuracy was even more pronounced when the less compliant data were used: the associations between the two-point slope estimates from the CAR (30/45/60-min post-waking) to bedtime in the full sample ($n = 24$) were all under 0.500 ($ps < 0.001$).

Area under the curve

Table 3 summarizes the relations between the six AUC measures under study. (The AUC values reported in the text and tables were divided by total time awake, to account for individual differences in waking hours. However, effects were almost identical without this transformation [all differences in ICCs ranged from 0.005 to 0.033]). Similar to the slope estimates, there were no significant differences between ICC between AUC estimates excluding versus including the CAR for the maximum intensity (ICC = 0.914, $p < 0.001$) and medium intensity (ICC = 0.811, $p < 0.001$) protocols, $p > 0.05$. However, the addition of the CAR (wake +30) in the minimal intensity procedure had a large effect on the average size (MinAUC$_E$ $M = 0.212$ versus MinAUC$_I$ $M = 0.308$, t(19) = −3.07, $p < 0.01$), and on the ICC between these two AUC measures (= 0.321, $p < 0.01$). The differences between this ICC and the corresponding ICCs for the maximum and medium intensity were significant ($ps < 0.05$). Overall, none of the medium or minimum intensity protocols showed a strong ICC with the hourly (maximum) sampling protocol, with especially low ICCs for the minimum protocols ranging from 0.177 ($ps < 0.01$) to 0.554 ($ps < 0.001$). In the supplemental set of analyses with the full sample we found a very similar set of results (i.e. all ICC coefficients were within 0.07 of the original estimates).

Cortisol awakening response

The CAR comparisons are shown in Table 4. First, we examined CAR$_{AUC}$ estimates derived from one, two, or three incremental CAR points (30/45/60-min post-waking) divided by the total time between waking and the final CAR sample (to adjust for difference in total sampling time between the three CAR measures). Results showed that these three measures were highly correlated (all ICCs >0.870, $ps < 0.001$). (ICCs for the CAR30$_{AUC}$ [AUC$_I$

with just two values: waking and 30-min post-waking] drop significantly if you do not divide by time. $CAR30_{AUC}$ was associated with $CAR45_{AUC}$ [ICC = 0.711, $p < 0.001$] and $CAR60_{AUC}$ [ICC = 0.481, $p < 0.01$]. Associations between the $CAR30_{AUC}$ and the simple difference CAR measures also drop if you do not divide by time [all ICCs range from 0.329 to 0.360, $ps < 0.05$]. These ICCs were significantly smaller than the corresponding ICCs of $CAR30_{AUC}$ divided by time [$ps < 0.05$].)

Simple difference measures, at either 30-, 45-, or 60-min post-waking were also highly interrelated: $CAR60_D$ was correlated with $CAR45_D$ (ICC = 0.941, $p < 0.001$) and $CAR30_D$ (ICC = 0.891, $p < 0.001$); $CAR45_D$ was very strongly correlated with $CAR30_D$ (ICC = 0.968, $p < 0.001$). Overall, the first row in Table 4 shows the similarities between CAR estimates calculated using all four data points ($CAR60_{AUC}$) compared to protocols using two or three sampling points – ICCs were high, ranging from 0.870 to 0.985 ($ps < 0.001$). In the supplemental set of analyses with the full sample we found a very similar set of results (i.e. all ICC coefficients were within 0.06 of the original estimates).

Discussion

The primary purpose of the current study was to understand the impact of various intensities of salivary sampling across the day on the accuracy of estimates of key diurnal cortisol measures – diurnal cortisol slopes, the CAR, and the AUC. Overall, we found that medium intensity protocols with two fixed samples (waking, bedtime) and three additional samples measured across the day, closely approximates the cortisol decline (slope) derived from an intensive protocol including about 18 data points per day. However, more data points may be necessary to adequately measure the total cortisol exposure across the day (AUC). Additionally, our CAR analyses suggest that two samples (waking cortisol and a second sample between 30 and 60 min after waking) provides a reasonable estimate of the CAR, with evidence showing that these two-point protocols are surprisingly highly associated (in both level and covariation) with each other and with protocols including three to four CAR data points. This finding is particularly noteworthy in light of the fact that we found considerable variability in *when* individuals experienced their morning peak cortisol levels, and considering that recent expert guidelines recommend a 3-sample protocol (at minimum) as a result (Stalder et al., 2016).

Accuracy of reduced protocols for diurnal cortisol exposure

The medium sampling approach for the diurnal slope yielded a high ICC with the maximum intensity approach (ICC = 0.867) with salivary cortisol samples provided hourly from waking to bedtime. Although further reduction would help to alleviate participant burden, minimal procedures that collect only two (waking and bedtime or CAR and bedtime) samples had relatively small associations with the maximum intensity protocol. As a result, investigators using minimum intensity protocols should be aware that their measures do not strongly approximate a cortisol slope based on a more intensive measurement protocol.

Furthermore, although a medium intensity protocol provided a relatively strong estimate of the slope, this reduced protocol was not a good estimate of the AUC. Therefore, studies with

limited sampling points may achieve more accuracy in their measure of the slope rather than the AUC.

Number of samples

Theoretically, if the diurnal rhythm were entirely linear, taking cortisol at any two points in the day would provide a good estimate. Although applying a logarithmic transformation of raw cortisol values helps to linearize the association between salivary cortisol and time of day, the association is still not entirely linear, and thus slopes and AUCs with just a few data points do not perfectly represent slopes and AUCs derived from hourly protocols, which reveal greater cortisol variability and curvilinearity across the day. Systematic differences in stress exposures, emotional state, activity levels, or behaviors such as napping, eating, and smoking at particular points in the day may impact specific cortisol sampling points and further reduce the similarities among slopes estimated with different protocols. Random error also adds additional variability to the differences among estimates.

Regardless of the explanation for the observed variability, it is clear that sampling intensity affects the calculation of diurnal measures of cortisol, and researchers must be cognizant of this fact when designing studies. Furthermore, in order for results to be comparable across studies, it will be important for the field to establish norms for the best number of samples used in the measurement of diurnal cortisol slopes and AUCs.

Estimating slope and AUC with and without the CAR

Another important comparison that deserves further discussion is the distinction between exclusion versus inclusion of CAR samples in estimating diurnal slopes and AUCs. Given that cortisol values are known to dramatically increase in the first 30- to 45-min post-awakening, previous research suggests that when CAR samples are included in the analysis, they have a strong influence on diurnal measures (Hruschka et al., 2005; Ranjit et al., 2005). This pattern is especially true with fewer overall data points. For example, the 2-point simple slope from wakeup to bedtime will provide different values than measuring the CAR to bedtime slope. Furthermore, the interpretation and meaning of the CAR to bedtime slope is different than the interpretation of wakeup to bedtime slope, with CAR slopes being strongly influenced by the size of the morning CAR peak. There may be reason to believe that the CAR is influenced by distinct psychobiological processes, being regulated by different psychosocial, and different neurobiological mechanisms, than the rest of the diurnal rhythm (Clow et al., 2010; Pruessner et al., 1999; Wüst et al., 2000). Similarly, the minimum AUC measure including the CAR (wake, wake +30, bedtime) had a dramatic effect on the ICCs, reducing all ICCs below 0.412. Thus, although there is some intuitive appeal to including CAR data in slope and AUC measures, there has been a trend toward using slope and AUC estimates that purposefully do not include the influence of CAR values (Adam, 2006; Polk et al., 2005; Weissbecker et al., 2006). Further research is needed, however, to establish whether excluding or including CAR sampling points in slope and AUC calculations represent the more psychologically and medically meaningful diurnal cortisol measure.

Accuracy of reduced protocols for CAR measures

In the current study, we found that measuring the difference between the waking cortisol sample, and a second sample between 30 and 60 min after waking, provides a strong approximation of the CAR compared to more intensive protocols with accumulating CAR measurement at 30-, 45-, and 60-min post-waking. The ICCs were around 0.90, suggesting that additional morning samples, although intuitively and theoretically helpful, seem to require additional participant burden without statistical or mathematically large returns. Importantly, the current study accounted for the amount of time between waking and each CAR sample (by dividing by the difference between sampling time and wake time). Without this adjustment, ICCs between these different protocols is reduced.

Because the first hour after waking can be a hectic time of day, reducing the number of samples during this period should greatly decrease participant burden. Based on the findings from this study, we conclude that measuring the CAR with just two samples (i.e. waking and 30 min after waking) is a reasonable approach, providing highly comparable measures to those obtained with a more intensive procedure involving 4 morning measurements, or a procedure involving 3 measurements. Although our sample was evenly split between days with the morning peak at 30 min versus 45 min after waking, choosing the earlier sampling time may increase compliance with sampling protocol (i.e. not eating, drinking, or brushing your teeth before CAR sampling).

Notably, the newly published CAR expert consensus guidelines (Stalder et al., 2016) report recommends using a 4- to 5-sample protocol (e.g. waking, 15-, 30-, 45-, and 60-min post-waking), or a minimum of three samples in case of financial restrictions (i.e. equivalent to our $CAR45_{AUC}$ measure: waking, 30- and 45-min post-waking). Stalder et al. (2016) argue that a two-sample protocol cannot be recommended because peak levels varies between people (e.g. gender differences) and within-person across days, based on situational factors (e.g. stressful events). Although the results of the current study suggest that little measurement accuracy in the size of the CAR_{AUC} was lost by relying on two samples, we agree that in order to (a) capture the shape of the CAR curve, (b) identify the exact CAR peak, or (c) measure the rate of recovery from the CAR peak, adding additional measures at 45 min, 60 min, or even 2 h after waking are important. Indeed, 25% of the study sample peaked at 60 min or later, but most current protocols only measure the CAR at 30- or 45-min post-waking. Future research with additional participants (and additional morning cortisol measures) is needed to evaluate whether these findings replicate across diverse samples.

Limitations

In the current study, we asked participants to provide hourly saliva samples for two days, a rigorous protocol with a very high participant burden. Participants provided, on average, 36 total saliva samples across the two days – a major disruption in everyday life. Therefore, one notable limitation of the current study is the small and homogenous nature of the sample: A convenience sample of 24 individuals (mostly white female young adults) who agreed to participate in this intensive protocol. Future research is needed to validate reduced sampling protocols in more diverse samples.

Another limitation of the current study is that it did not use electronic monitoring devices such as MEMS® track caps to monitor the timing of compliance with requested sampling times. Electronic monitoring of compliance, and awareness of that monitoring, is associated with improved compliance rates (Broderick et al., 2004). In the current study, because all comparisons were conducted with different configurations of the same data, we have less concern regarding the impact of compliance on our comparisons across measures, than other studies measuring the relation between cortisol measures and individual characteristics. Additionally, we monitored objective wake times using wrist actigraphy to verify awakening times, as recommended by the new CAR guidelines. We found that people's subjective reports of wake time, overall, closely matched the actigraph data. Comparing our two sets of models, with ($n = 20$) and without ($n = 24$) compliant wake-up times, we found slightly stronger associations across minimum, medium, and maximum diurnal cortisol measures for the high compliance sample. Overall, to the extent that post-waking sample timing can be electronically monitored in future research, or at least to the extent that participants believe such monitoring is occurring, it should improve the quality of estimates (Stalder et al., 2016). Future studies, incorporating larger and more diverse samples, and employing electronic monitoring, should further investigate the implications of various intensity sampling protocols in order to continue to inform protocol decisions for naturalistic diurnal cortisol research.

Conclusions

Overall, these results are encouraging for stress researchers who are interested in efficient, yet accurate, cortisol sampling protocols to measure diurnal cortisol (slope) and the CAR. There are already a number of large-scale datasets including Midlife in the United States (MIDUS), the National Study of Daily Experiences, Coronary Artery Risk Development in Young Adults (CARDIA), Whitehall II, and the Multi-Ethnic Study of Atherosclerosis (MESA) that incorporate medium intensity protocols (approximately 4–6 samples per day) in large groups of participants. Although increasing the number of samples will help to reduce error and improve reliability, in order to measure cortisol in large samples in naturalistic settings, investigators need to carefully select sampling points, using empirically informed judgments, in order to balance scientific accuracy with participant burden.

Funding

References

Abercrombie HC, Giese-Davis J, Sephton S, Epel ES, Turner-Cobb JM, Spiegel D. (2004). Flattened cortisol rhythms in metastatic breast cancer patients. Psychoneuroendocrinology 29:1082–92. [PubMed: 15219660]

Adam EK. (2006). Transactions among adolescent trait and state emotion and diurnal and momentary cortisol activity in naturalistic settings. Psychoneuroendocrinology 31:664–79. [PubMed: 16584847]

Adam EK, Doane LD, Zinbarg RE, Mineka S, Craske MG, Griffith JW. (2010). Prospective prediction of major depressive disorder from cortisol awakening responses in adolescence. Psychoneuroendocrinology 35:921–31. [PubMed: 20079576]

Adam EK, Gunnar MR. (2001). Relationship functioning and home and work demands predict individual differences in diurnal cortisol patterns in women. Psychoneuroendocrinology 26:189–208. [PubMed: 11087964]

Adam EK, Hawkley LC, Kudielka BM, Cacioppo JT. (2006). Day-to-day dynamics of experience-cortisol associations in a population-based sample of older adults. Proc Natl Acad Sci USA 103:17058–63. [PubMed: 17075058]

Adam EK, Kumari M. (2009). Assessing salivary cortisol in large-scale, epidemiological research. Psychoneuroendocrinology 34:1423–36. [PubMed: 19647372]

Adam EK, Vrshek-Schallhorn S, Kendall AD, Mineka S, Zinbarg RE, Craske MG. (2014). Prospective associations between the cortisol awakening response and first onsets of anxiety disorders over a six-year follow-up. Psychoneuroendocrinology 44:47–59. [PubMed: 24767619]

Broderick JE, Arnold D, Kudielka BM, Kirschbaum C. (2004). Salivary cortisol sampling compliance: comparison of patients and healthy volunteers. Psychoneuroendocrinology 29:636–50. [PubMed: 15041086]

Chida Y, Steptoe A. (2009). Cortisol awakening response and psychosocial factors: a systematic review and meta-analysis. Biol Psychol 80:265–78. [PubMed: 19022335]

Chrousos GP. (2009). Stress and disorders of the stress system. Nat Rev Endocrinol 5:374–81. [PubMed: 19488073]

Clements AD, Parker CR. (1998). The relationship between salivary cortisol concentrations in frozen versus mailed samples. Psychoneuroendocrinology 23:613–16. [PubMed: 9802131]

Clow A, Hucklebridge F, Stalder T, Evans P, Thorn L. (2010). The cortisol awakening response: more than a measure of HPA axis function. Neurosci Biobehav Rev 35:97–103. [PubMed: 20026350]

Clow A, Thorn L, Evans P, Hucklebridge F. (2004). The awakening cortisol response: methodological issues and significance. Stress 7:29–37. [PubMed: 15204030]

Cohen S, Schwartz JE, Epel E, Kirschbaum C, Sidney S, Seeman T. (2006). Socioeconomic status, race, and diurnal cortisol decline in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. Psychosom Med 68:41–50. [PubMed: 16449410]

DiCiccio TJ, Efron B. (1996). Bootstrap confidence intervals. Stat Sci 11:189–212.

Efron B, Tibshirani R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54–75.

Folkesson L, Riva R, Östberg V, Lindfors P. (2014). Single and aggregate salivary cortisol measures during two schooldays in midadolescent girls and boys. PsyCh J 3:121–31. [PubMed: 26271764]

Gunnar MR, Vazquez DM. (2001). Low cortisol and a flattening of expected daytime rhythm: potential indices of risk in human development. Dev Psychopathol 13:515–38. [PubMed: 11523846]

Heim C, Ehlert U, Hellhammer DH. (2000). The potential role of hypocortisolism in the pathophysiology of stress-related bodily disorders. Psychoneuroendocrinology 25:1–35. [PubMed: 10633533]

Hruschka DJ, Kohrt BA, Worthman CM. (2005). Estimating between-and within-individual variation in cortisol levels using multilevel models. Psychoneuroendocrinology 30:698–714. [PubMed: 15854786]

Huppert FA. Positive emotions and cognition: developmental, neuroscience and health perspectives. In: Forgas JP, editor. Affect in social thinking and behavior. Vol. 8. New York, NY: Psychology Press; 2006. p 235–52.

Kunz-Ebrecht SR, Kirschbaum C, Marmot M, Steptoe A. (2004). Differences in cortisol awakening response on work days and weekends in women and men from the Whitehall II cohort. Psychoneuroendocrinology 29:516–28. [PubMed: 14749096]

Matthews K, Schwartz J, Cohen S, Seeman T. (2006). Diurnal cortisol decline is related to coronary calcification: CARDIA study. Psychosom Med 68:657–61. [PubMed: 17012518]

McGraw KO, Wong SP. (1996). Forming inferences about some intraclass correlation coefficients. Psychol Methods 1:30–46.

Miller GE, Chen E, Zhou ES. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. Psychol Bull 133:25–45. [PubMed: 17201569]

Nader N, Chrousos GP, Kino T. (2010). Interactions of the circadian CLOCK system and the HPA axis. Trends Endocrinol Metab 21:277–86. [PubMed: 20106676]

Polk D, Cohen S, Doyle W, Skoner D, Kirschbaum C. (2005). State and trait affect as predictors of salivary cortisol in healthy adults. Psychoneuroendocrinology 30:261–72. [PubMed: 15511600]

Pruessner J, Hellhammer D, Kirschbaum C. (1999). Burnout, perceived stress, and cortisol responses to awakening. Psychosom Med 61:197–204. [PubMed: 10204973]

Pruessner J, Kirschbaum C, Meinlschmid G, Hellhammer DH. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. Psychoneuroendocrinology 28:916–31. [PubMed: 12892658]

Ranjit N, Young EA, Raghunathan TE, Kaplan GA. (2005). Modeling cortisol rhythms in a population-based study. Psychoneuroendocrinology 30:615–24. [PubMed: 15854778]

Rotenberg S, McGrath JJ, Roy-Gagnon M-H, Tu MT. (2012). Stability of the diurnal cortisol profile in children and adolescents. Psychoneuroendocrinology 37:1981–9. [PubMed: 22658393]

Saxbe DE. (2008). A field (researcher's) guide to cortisol: tracking HPA axis functioning in everyday life. Health Psychol Rev 2:163–90.

Schlotz W, Hellhammer J, Schulz P, Stone AA. (2004). Perceived work overload and chronic worrying predict weekend–weekday differences in the cortisol awakening response. Psychosom Med 66:207–14. [PubMed: 15039505]

Sephton SE, Sapolsky RM, Kraemer HC, Spiegel D. (2000). Diurnal cortisol rhythm as a predictor of breast cancer survival. J Natl Cancer Inst 92:994–1000. [PubMed: 10861311]

Shrout PE, Fleiss JL. (1979). Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–8. [PubMed: 18839484]

Stalder T, Kirschbaum C, Kudielka BM, Adam EK, Pruessner JC, Wüst S, Dockray S, et al. (2016). Assessment of the cortisol awakening response: expert consensus guidelines. Psychoneuroendocrinology 63:414–32. [PubMed: 26563991]

Suglia SF, Staudenmayer J, Cohen S, Enlow MB, Rich-Edwards JW, Wright RJ. (2010). Cumulative stress and cortisol disruption among Black and Hispanic pregnant women in an urban cohort. Psychol Trauma Theory Res Pract Policy 2:326–34.

Vrshek-Schallhorn S, Doane L, Mineka S, Zinbarg R, Craske M, Adam E. (2013). The cortisol awakening response predicts major depression: predictive stability over a 4-year follow-up and effect of depression history. Psychol Med 43:483–93. [PubMed: 22652338]

Weissbecker I, Floyd A, Dedert E, Salmon P, Sephton S. (2006). Childhood trauma and diurnal cortisol disruption in fibromyalgia syndrome. Psychoneuroendocrinology 31:312–24. [PubMed: 16274933]

Wüst S, Federenko I, Hellhammer DH, Kirschbaum C. (2000). Genetic factors, perceived chronic stress, and the free cortisol response to awakening. Psychoneuroendocrinology 25:707–20. [PubMed: 10938450]
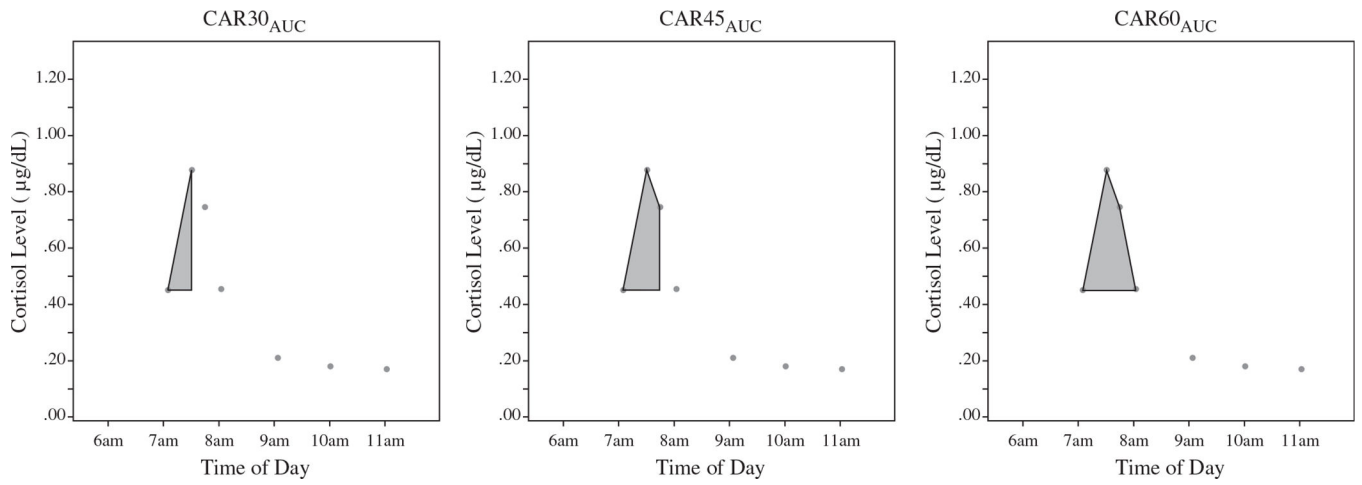
Figure 1.
Incremental CAR$_{AUC}$ measures including 30-, 45-, and 60-min post-waking.
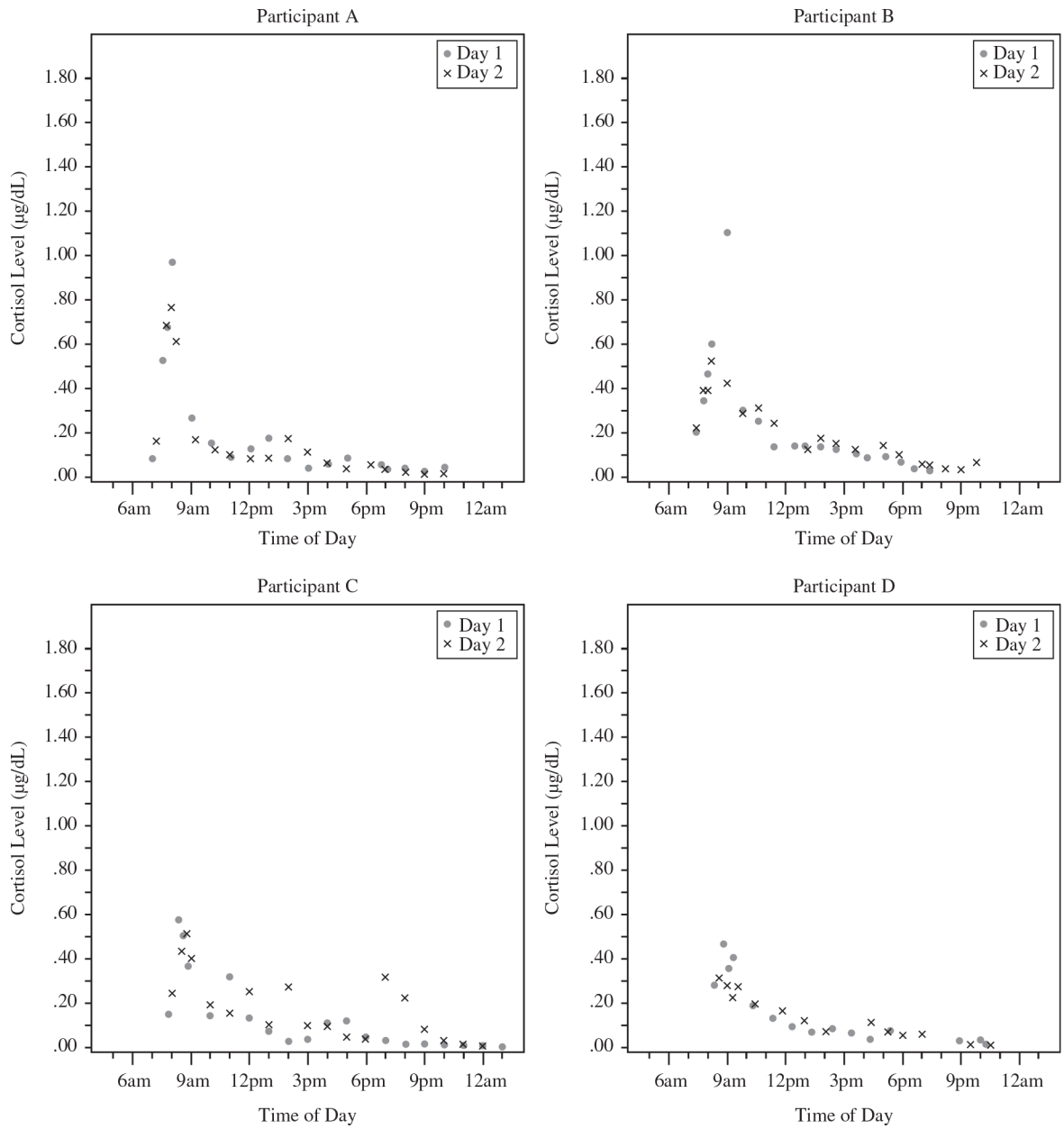
Figure 2.
Cortisol data across the day from four randomly selected study participants.

Table 1.

Descriptive statistics for sample (n = 20).

	Min	Max	Mean	SD
Number of samples	14	20	17.950	1.849
Total time awake (hours)	13.933	17.658	15.910	1.060
Time of waking sample	5.30 AM	9:00 AM	7:15 AM	56.8 m
Time of bedtime sample	9:00 PM	12:44 AM	11:11 PM	1 h 12.4 m
Cortisol at waking (µg/dl) [a]	0.124	1.235	0.353	0.275
Cortisol at bedtime (µg/dl) [a]	0.006	0.713	0.072	0.155
Diurnal cortisol slope				
MaxSlope$_E$	−0.199	−0.021	−0.144	0.042
MedSlope$_E$	−0.215	−0.038	−0.142	0.045
MinSlope$_E$	−0.224	−0.022	−0.141	0.050
Slope$_{WAKE}$	−0.222	−0.022	−0.142	0.050
Slope$_{CAR30}$	−0.288	0.061	−0.180	0.075
Slope$_{CAR45}$	−0.294	0.071	−0.186	0.075
Slope$_{CAR60}$	−0.281	0.071	−0.181	0.072
Cortisol area under the curve (AUC) average level per hour [a]				
MaxAUC$_E$	0.064	0.484	0.144	0.089
MedAUC$_E$	0.105	0.593	0.205	0.100
MinAUC$_E$	0.078	0.772	0.212	0.199
MaxAUC$_I$	0.072	0.486	0.159	0.085
MedAUC$_I$	0.088	0.470	0.166	0.078
MinAUC$_I$	0.173	0.632	0.308	0.127
Cortisol awakening response (CAR) average level per 15 min [a]				
CAR60$_{AUC}$	−0.133	0.113	0.033	0.051
CAR45$_{AUC}$	−0.117	0.116	0.032	0.047
CAR30$_{AUC}$	−0.087	0.089	0.023	0.035

	Min	Max	Mean	SD
CAR simple difference values[a]				
CAR60$_D$	−0.720	0.668	0.143	0.285
CAR45$_D$	−0.726	0.632	0.199	0.290
CAR30$_D$	−0.691	0.714	0.184	0.279
		Mean level in µg/dl (SD)		% Peak
Cortisol morning values[a]				
Wake-up		0.353 (0.275)		10.0
30 min after waking		0.535 (0.219)		32.5
45 min after waking		0.550 (0.240)		32.5
60 min after waking		0.496 (0.211)		20.0
2 h after waking		0.317 (0.223)		5.0

[a]Raw cortisol values are presented for descriptive purposes but log transformed values are used in all analyses.

Table 2.

Intraclass correlation table for different diurnal slope calculations.

	1	2	3	4	5	6	7
Regression-based slopes							
1. MaxSlope$_E$	–	0.867 ***	0.725 ***	0.728 ***	0.598 ***	0.550 ***	0.591 ***
2. MedSlope$_E$		–	0.919 ***	0.914 ***	0.625 ***	0.548 ***	0.571 ***
3. MinSlope$_E$			–	0.998 ***	0.661 ***	0.588 ***	0.607 ***
Simple slopes							
4. Slope$_{WAKE}$				–	0.662 ***	0.589 ***	0.610 ***
5. Slope$_{CAR30}$					–	0.986 ***	0.968 ***
6. Slope$_{CAR45}$						–	0.978 ***
7. Slope$_{CAR60}$							–

Cortisol was log transformed prior to slope calculations. Regression-based slopes were based on maximum (MaxSlope$_E$; hourly samples) medium (MedSlope$_E$; six samples per day), or minimum (MinSlope$_E$; two samples per day) sampling procedures, either excluding (E) or including (I) the cortisol awakening response (CAR). Slope$_{WAKE}$ is the "rise over run" slope from waking to bedtime; Slope$_{CAR30/45/60}$ are the three "rise over run" slopes from 30/45/60-min post-waking to bedtime.

***
$p < 0.001$.

Table 3.

Intraclass correlation table for AUC calculations.

	1	2	3	4	5	6
Excluding CAR						
1. MaxAUC$_E$	–	0.598 ***	0.554 **	0.914 ***	0.775 ***	0.177 **
2. MedAUC$_E$		–	0.534 **	0.727 ***	0.811 ***	0.411 ***
3. MinAUC$_E$			–	0.439 *	0.412 *	0.321 *
Including CAR						
4. MaxAUC$_I$				–	0.921 ***	0.237 ***
5. MedAUC$_I$					–	0.261 ***
6. MinAUC$_I$						–

Cortisol was log transformed after AUC calculations. AUC indices were based on maximum (MaxAUC; hourly samples) medium (MedAUC; six samples per day), or minimum (MinAUC; two samples per day) sampling procedures, either excluding (E) or including (I) the cortisol awakening response (CAR). All AUC scores were divided by total time awake to account for individual differences in waking hours.

*
$p < 0.05$.

**
$p < 0.01$.

***
$p < 0.001$.

Table 4.

Intraclass correlation table for CAR calculations.

	1	2	3	4	5	6
CAR area under the curve						
1. $CAR60_{AUC}$ (4 samples)	–	0.985^{***}	0.870^{***}	0.898^{***}	0.905^{***}	0.923^{***}
2. $CAR45_{AUC}$ (3 samples)		–	0.930^{***}	0.822^{***}	0.839^{***}	0.881^{***}
3. $CAR30_{AUC}$ (2 samples)			–	0.662^{***}	0.654^{***}	0.707^{***}
Simple difference (2 samples)						
4. $CAR60_D$				–	0.941^{***}	0.891^{***}
5. $CAR45_D$					–	0.968^{***}
6. $CAR30_D$						–

Cortisol was log transformed prior to CAR calculations. Area under the curve CAR indices (CAR_{AUC}) were calculated using two points ($CAR30_{AUC}$; wake and wake +30), three points ($CAR45_{AUC}$; wake, wake +30, and wake +45), or four points ($CAR60_{AUC}$; wake, wake +30, wake +45, and wake +60), and then divided by total time from waking to sample. Simple difference CAR indices (CAR_D) were calculated between two points: waking and 30-, 45-, or 60-min post-waking.

*** $p < 0.001$.