Psychological states and physical fates

van Ockenburg, Sonja

Chapter 7

Cortisol dynamics in healthy individuals

Sonja L. van Ockenburg, Adriene Beltz, Ido Kema, Elisabeth H. Bos,
Hannah R. de Boer, Peter de Jonge, Reinold O.B. Gans,
Peter C. Molenaar, Judith G.M. Rosmalen

Submitted

ABSTRACT

Studies investigating the hypothalamic-pituitary-adrenal axis (HPA-axis) often focus on between-individual differences. Yet, for findings at the group level to be generalizable to the level of the individual two conditions have to be met. First, the population has to be homogeneous for the process under study (i.e. all subjects need to obey the same statistical model). Second, the process that is investigated needs to have stable statistical characteristics over time (e.g. should not contain trends). These conditions can only be validly studied by obtaining and analyzing time series of individual subjects. In the current study 10 healthy adults collected saliva samples thrice daily, and daily 24-h urine samples for 63 consecutive days. Cortisol in saliva and urine was measured by means of liquid chromatography tandem mass spectrometry. Furthermore, participants filled out an electronic diary twice a day to report about sleep and lifestyle factors. Time series were analyzed by means of unified structural equation modelling to assess the within-individual stability of cortisol over time and the effect of lifestyle variables on cortisol levels. Lifestyle factors did not show the same importance nor direction of effect for every individual. Furthermore, many time series displayed trends and very little stability over time. In conclusion, the two conditions necessary to be able to generalize group-based findings to the levels of the individual and vice versa have not been met. Thus, It is important for future studies to consider within-individual processes when studying the HPA-axis.

INTRODUCTION

It is often thought that associations between psychosocial stress and somatic disease occur because exposure to stressors leads to repeated activation of major stress responsive systems such as the hypothalamic pituitary adrenal axis (HPA-axis)[1,2]. Indeed, increased cortisol levels are associated with for instance cardiovascular mortality [3], diabetes mellitus [4], and major depressive disorder [5]. This has strengthened the belief that the HPA-axis plays a central role in tying psychosocial stress to various disease states. The question that arises is: what do these findings mean for the individual patient? This question can only be answered if we know *when* we are allowed to generalize findings from cohort studies to the individual.

Two conditions have to be met in order to be able to generalize findings from the level of the group to the level of the individual and vice versa. In mathematical statistics these conditions are specified in the classical ergodic theorems [6]. The first condition is that the study sample has to be relatively homogeneous for the process under study (i.e. subjects must be fairly similar and the same statistical model should be able to describe the process for each individual). This assumption is unlikely to hold in large cohort studies, as large differences between individuals in both genetics and environment set individuals out on different life course trajectories [7]. The second condition is that the process under study needs to have stable statistical properties over time. For Gaussian processes this means that if a person were to be measured repeatedly over time to create a time series, that the mean, variance, and autocorrelation should be stable over time. If for example a time series, such as day-to-day fluctuations in cortisol levels, contains trends (e.g. a steady increase over time or higher levels at weekdays than on weekends) the second condition has not been met [8]. Only processes meeting both conditions are called ergodic. The consequence of nonergodicity is that there is no a priori relationship between group-based findings and the trajectories that the individuals making up this group display over time [6].

Almost all studies investigating the effects of psychosocial stress on the HPA-axis [9,10] or of HPA-axis functioning on disease risk [3–5] aggregate data at a group level. Yet, it is currently unknown whether day-to-day fluctuations in cortisol levels can be classified as ergodic, as these have rarely been studied at the level of the individual. Three studies that investigated day-to-day stability of salivary cortisol (measuring cortisol for a few days) demonstrated that the differences within individuals over time were much larger than those between individuals [11–13]. Salivary cortisol levels are, however, heavily influenced by time of sampling, due to the circadian rhythm. Urinary cortisol may in that sense provide a more integrative measure of HPA-axis functioning and is frequently used as a clinical diagnostic tool for both hypo- and hypercortisolism [14]. The one study that we are aware of that studied cortisol dynamics over time within one healthy individual, does not describe the characteristics of the time series [15]. From the graphical display, however, it can be cautiously concluded that the series is not stationary. Naturally, the assumption of homogeneity of the group cannot be tested within one person. Hence, it remains uncertain whether findings from group studies generate any knowledge of within-individual processes.

Only a study that measures several individuals for a prolonged period of time can fill this gap in knowledge. To this end we studied 10 healthy individuals for 63 consecutive days

measuring daily cortisol levels in both saliva and urine. We used time series analysis to investigate the within-person dynamics of salivary and urinary cortisol levels. Additionally, we assessed if certain within-person factors such as time of awakening, the amount of exercise, caffeine, cigarettes, and alcohol consumption can help to explain day-to-day variations in cortisol levels.

METHODS

Cohort and study design

The study was a longitudinal prospective observational study generating time series data of 10 healthy participants who kept an electronic diary and collected urine and saliva samples for 63 consecutive days. The study took place in the city of Groningen, the Netherlands from 9th of July 2012 through 10th of March 2013. Participants were recruited by means of poster adverts that were displayed in university buildings, hospitals, and supermarkets in the city of Groningen. They were paid €5 per day of study participation, thus a total of €315 after completion of the entire study period. Inclusion criteria were being a healthy adult between the ages of 18 and 65 years and being available for 63 consecutive days. Exclusion criteria were any current somatic and/or mental illnesses and medication use other than oral contraceptives or occasional acetaminophen. The aim was to include 10 participants. A total of 11 participants were included in the study. One person discontinued participation in the study due to a major life event after two days. The 10 other participants successfully completed the entire study period. The study protocol was approved by the Medical Ethics Committee of the University Medical Center Groningen in the Netherlands. All participants were given extensive written and oral information about the study's purposes and protocol and had the option to consult an independent physician for additional information. Before enrollment participants gave written informed consent.

Electronic diary

Participants were asked to fill out an electronic diary in the evening before going to bed and in the morning directly after waking up. The diary was web based (Qualtrics, Provo, UT) and could be accessed with any type of browser, including those installed on smartphones. In the morning, participants filled out the Pittsburgh sleep diary (wake time). This questionnaire provides information about the time participants went to bed and the time of awakening. Not every daily assessment consisted of exactly 24 hours due to differences in bed and wake time. Information from the sleep diary was used to calculate of how many hours a daily assessment actually consisted (e.g. 23 hours). In the bed time diary, participants were asked how many caffeine containing drinks, alcoholic beverages, and cigarettes they had used that day. Physical activity was assessed by a simple activity inventory adapted from the Godin Leisure-Time exercise Questionnaire (Godin et al.). Participants were asked whether they were physically active and if they did mild (e.g. casual stroll), medium (e.g. tennis), and/or very strenuous exercise (e.g. sprinting). Mild exercise received a score of 3, medium exercise a score of 5, and very strenuous exercise a score of 9. The scores of the categories were summed and multiplied by the amount of minutes the exercise had taken place, creating an exercise score for each day. Furthermore, in the diary participants were asked to indicate

and describe any difficulties with collecting the samples that day or to describe any stressful events that had occurred.

Urinary cortisol

Participants collected all urine in two separate containers each day for 63 consecutive days. They were instructed to use the "night container" from the moment they went to bed until the first morning void. The "day container" was for all urine produced after the first morning void until the last void before going to bed. Containers were stored at room temperature until they were collected on every Monday, Wednesday, and Friday. Participants had the option of dropping of the containers at the University Medical Center Groningen or could have them collected at their home address at any convenient time. Before processing, the urine containers were weighted on a scale with a precision of 1 gram. After that, the urine was pipetted in Sarstedt 2 ml cups and sealed with a Sarstedt screwcap. Urine samples from the "day container" and the "night container" were stored separately at a temperature of -80 degrees Celcius in centrally monitored freezers. Cortisol was determined by means of liquid chromatography tandem mass spectrometry (LC-MS/MS). The day urine and night urine were pipetted in separate wells. To minimize the effects of interrun variability on the data, all urine samples of one participant were analyzed in one lot. Intra- and interrun coefficients of variation for the lower range of cortisol were 2.4% and 7.8%, and for the higher range of cortisol 1.4% and 3.8% respectively. 24-hour urinary free cortisol (24-h UFC) was computed in the following way: (cortisol concentration of morning urine X volume of morning urine) + (cortisol concentration of day urine X volume of day urine)= 24-h UFC output. The upper limit of cortisol output considered normal in our laboratory was 133 nmol/24-h, which is based on data from the Life Lines cohort.

Urinary creatinine

Completeness of the 24-h urine samples was assessed by means of 24-h urinary creatinine output. Urinary creatinine was measured with the creatinine plus enzymatic assay performed on the Roche Modular. Again, to minimize the effects of interrun variability on the results, all urine samples of each individual were analyzed in one lot. The interrun coefficient of variation was 2.4% and the intrarun coefficient of variation was 0.9%. Considerable within-individual variability in day-to-day creatinine excretion is known to exist. A review article mentions a within-subject coefficient of variation (CV) ranging from 3-20% with an average of 10% [17]. This is also the case for studies checking completeness with para-aminobenzoic acid administration who find an average within-person CV in urinary creatinine of 10% [18]. Also in our study each participant showed considerable day-to-day fluctuations in total creatinine output. As this variation was either normally distributed around the mean or slightly positively skewed, we assume, in accordance with other studies [17–19], that this concerns for the most part natural physiological variation. In line with the within-individual nature of the study, we decided to exclude samples from statistical analysis if the 24-hour creatinine output was lower than 2 standard deviations from the person's own mean, instead of looking at between-individual variation as an estimate for completeness. Based on this method we had to exclude day 23, 24, 28, 29, and 32 for participant 1 (P1), leaving 58 complete days for the analysis. For P2 we had to exclude the urine sample of day 41. Day 58 also contained too little urinary creatinine, but it was decided to keep this sample in the

study, as the participant's cortisol levels were exceptionally high this day and she indicated to have missed only one void due to a stressful life event that day. For P5, P6, P7, P8, and P10 day 63, day 22, days 1 and 34, day 56, and day 40 had to be excluded respectively. For P3, P4, and P9 all urine samples were judged complete based on their urinary creatinine output, and thus 63 days were available for the analysis.

Salivary cortisol

Participants collected saliva samples three times a day for 63 consecutive days. They were provided with 'Salivettes® Cortisol' (Sarstedt), synthetic swabs specifically designed to collect saliva for the measurement of cortisol. The salivettes were prelabeled with sampling moment (T1, T2, T3) and day number, and packed together in a plastic bag indicating the day for their intended use. Participants received supplies for a maximum number of three days. They were instructed to chew on the cotton for two minutes, and to take the first saliva sample (T1) immediately after awaking, the second sample 30 minutes later (T2), and the third sample at 7 pm (T3). Tooth brushing, eating, and drinking anything other than water were not allowed 30 minutes prior to sampling. Saliva samples were stored at room temperature until they were collected by the researcher on Monday, Wednesday, and Friday. The salivettes were centrifuged for 2 minutes at 1x1000 g. Saliva was then pipetted in Sarstadt 2 ml cups and sealed by a screwcap. The samples were stored at a temperature of -80 degrees Celcius in centrally monitored freezers. Cortisol was determined by means of liquid chromatography tandem mass spectrometry (LC-MS/MS). To minimize the effects of interrun variability on the within-person data, all saliva samples of one participant were analyzed in one lot. Intra- and interrun coefficients of variation for the lower range of cortisol were 8% and 9.4%, and for the higher range of cortisol 3.5% and 7.9% respectively. The cortisol awakening response (CAR) was calculated as the area under the curve with respect to the ground, based on the T1 and T2 saliva samples.

Statistical analyses

First the cortisol data was checked for stable statistical properties over time to test one of the ergodicity assumptions (stationarity) and to prepare data for the analysis. Time series were plotted and visually examined for the presence of trends (for example a gradual increase in the mean of the series). Linear and quadratic trends, and mean level shifts were removed from the data prior to further analysis. The distribution of the cortisol data was examined and turned out to be slightly positively skewed for most individuals. Data points larger than three box-lengths were considered outliers and modeled by adding a dummy variable to the time series model. To assess which within-person factors might be of influence on both current and lagged cortisol levels, we fitted multivariate unified structural equation (uSEM) models with a lag of order 1 using LISREL [20]. A lag of order 1 means that values of the previous day (t-1) are used to predict the values of the current day (t). USEM models consist of a structural equation model (SEM) part and a vector autoregressive (VAR) part. The SEM part estimates the contemporaneous relationships (i.e. variables from the same day predict each other), while the VAR part enables simultaneous estimation of the lagged relationships (i.e. previous values of the variables predict current values of the variables) [21,22]. Thus uSEM, unifies SEM and VAR in one model.

Three individual models were created for each participant separately with as outcome variable either 24-h UFC, salivary cortisol at 7 pm, or the CAR. The predictor variables for the 24-h UFC model were: duration/length of the day, time of awakening, and (if applicable) number of cigarettes used, number of caffeine containing drinks, number of alcohol containing beverages, and amount of exercise. The predictor variables for the model with salivary cortisol at 7 pm as outcome were: time of sampling, number of cigarettes, number of caffeine containing drinks, number of alcohol containing beverages, and amount of exercise. The CAR was predicted by time of awakening, number of cigarettes, number of caffeine containing drinks, number of alcohol containing beverages, and amount of exercise. This model differs from the previous models in that only lagged relationships were estimated for the lifestyle variables (cigarettes, caffeine, alcohol, and exercise), because we do not assume any contemporaneous relationships to exist (the CAR after all precedes the lifestyle factors in time on the same day). Thus only time of awakening is modelled as both a lagged and contemporaneous predictor in the CAR model.

All uSEM models were reversed fitted. This means that we first built a saturated model with zero degrees of freedom. Then the least significant predictor was removed and the difference in the chi-square square statistics between two consecutive models was compared. The winning model, that is, the one that was significantly better than the next with a p-value <0.05 is reported. The fit of the models was judged based on the non-normed fit index (NNFI) ≥ 0.95, the comparative fit index (CFI) ≥ 0.95, the root mean square error of approximation (RMSEA) <0.10, and the standardized root mean square residual (SRMR) ≤ 0.05, as these indices have found to be reliable in simulation studies [23]. We required a good fit according to at least two fit indices. A poor model fit is indicated in the results section. Additionally, fit indices for all uSEM models can be found in supplement 1. We used Stata [24] to assess if a lag of order 1 was sufficient, or whether a lag of two or three, etc. was required to adequately explain the data. This was tested by fitting VAR model containing the same variables that remained in the winning uSEM models. A Portmanteau test with a maximum of 6 lags was used to test for residual autocorrelation. Autocorrelation means that values of the previous day can be used to predict values of the following day(s). In addition to the Portmanteau test the autocorrelation function (ACF) of the residuals was plotted to visually assess whether all autocorrelation had been removed and only white noise remained.

RESULTS

Sample characteristics

An overview of the sample characteristics is given in table 1. Our sample consisted of three men and seven women. The youngest participant was 19 years and the oldest participant 58 years old. Two participants were related to each other and shared the same household (P 8 and P9). Only two participants smoked cigarettes; P5 smoked cigarettes at a daily basis and P1 smoked cigarettes only at social events. From table 1 it can be seen that participants, with respect to their median salivary and urinary cortisol levels, not only vary compared to each other, but that there is also a large within-individual variability (large interquartile range compared to the median).

Table 1: Sample characteristics

Participant	1	2	3	4	5	6	7	8	9	10
Sex	♂	♀	♀	♀	♂	♂	♀	♀	♀	♀
Age (years)	24	58	29	33	39	19	21	21	48	22
BMI kg/m²	23.2	26.5	20.0	17.2	20.0	21.6	21.3	20.1	25.3	23.1
Smoking yes/no	yes	no	no	no	yes	no	no	no	no	no
Amount of cigarettes Median (IQR)	0 0	n.a.	n.a.	n.a.	7 2.5	n.a.	n.a.	n.a.	n.a.	n.a.
Alcohol (units/day) Median (IQR)	2 2	0 0	0 1	1 2	1 2	0 0	0 0	n.a.	n.a.	2 4
Caffeine consumption (units/day) Median (IQR)	4 2	8 0	6 2	2 1	2 1	1 3	4 3	0.25 2	0 1	2 0
Exercise score Median (IQR)	0 203	180 810	0 90	60 120	600 540	135 135	135 230	0 0	0 0	0 180
Time of awakening Mean (SD)	8:50 (2:30)	8:30 (1:00)	7:10 (2:30)	7:33 (0:55)	7:57 (1:21)	7:39 (3:33)	9:03 2:35	10:30 (3:45)	7:45 (2:00)	7:25 (1:39)
Saliva T1 nmol Median (IQR)	5.21 (2.30)	12.86 (9.33)	7.26 (5.59)	8.18 (3.85)	8.20 (3.02)	9.00 (3.71)	7.20 (2.76)	7.99 (7.34)	6.37 (5.26)	12.47 (3.57)
Saliva T2 nmol Median (IQR)	9.33 (5.59)	11.86 (6.24)	13.89 (13.08)	7.97 (5.29)	10.93 (4.10)	13.50 (8.61)	12.93 (6.55)	13.06 (4.86)	12.37 (8.18)	13.64 (5.54)
Saliva T3 nmol Median (IQR)	1.08 (0.81)	1.16 (1.01)	1.51 (0.91)	1.67 (1.13)	2.30 (1.52)	2.28 (1.95)	1.98 (1.01)	2.78 (1.71)	2.61 (1.98)	2.04 (0.80)
CAR AUG nmol Median (IQR)	3.74 (1.75)	5.74 (3.08)	5.47 (3.26)	4.27 (1.74)	4.91 (1.47)	5.59 (1.94)	4.90 (2.86)	5.15 (2.22)	4.98 (2.41)	6.63 (2.32)
24-h urine volume (mL) Median (IQR)	1223 744	2094 472	2052 588	1590 443	1612 834	1386 772	1391 515	758 290	1222 549	1724 855
Urine 24-h nmol Median (IQR)	74.78 (24.99)	31.10 (8.19)	61.56 (24.41)	85.91 (37.70)	114.12 (54.12)	111.97 (36.01)	101.78 (35.06)	95.32 (46.81)	78.54 (33.98)	103.44 (34.78)
Number of days > 133 nmol	0	0	0	6	18	13	7	12	4	10

BMI=body mass index, IQR=interquartile range, n.a. not applicable`

Stability of cortisol levels over time

Figure 1 depicts the time series of the salivary and urinary cortisol levels for all 10 participants. To allow also for between-individual comparison, the scale was kept identical for all participants. The * in the graph of P2 in figure 1 indicates a stressful life event. At day 58 of the study her husband received some bad news at the hospital. It can be seen that the day time urinary cortisol, 24-h UFC, and the morning salivary cortisol levels are increased around this period and in particular on the day itself. Likewise, the * in the graph of P4 indicates a stressful life event. In the evening of day 42 of the study the participant had learned that her cat died after being run over by a car. She indicated that she experienced great grief that evening and also the next day. From the graph it can be seen that there is a rise in both salivary and urinary cortisol levels on day 43. The * in the graph of P6 is to indicate an outlier for which we cannot find a reason from the reported data in the electronic diary. On day 60 his 24-h UFC level was 645 nmol and the 7pm salivary cortisol sample was increased to 18 nmol.

It can be seen in figure 1 that there are large day-to-day fluctuations in cortisol levels, both in saliva and in urinary cortisol for all individuals. Many of the series have significant linear trends, quadratic trends or mean level shifts. For instance, the time series of the awakening salivary cortisol samples (T1) of P2 shows a quadratic trend, whereas those of P8, P9, and P10 have a positive linear trend. Two of the time series of the salivary cortisol samples 30 minutes post-awakening (T2) have trends: a negative linear trend in the series of P1 and a mean level shift in the series of P8. The time series of the 7pm salivary cortisol sample show fewer trends, and only the time series of P8 has a significant negative linear trend. The time series of the CAR, not depicted in figure 1, have a negative linear trend (P1), a quadratic trend (P2), and a mean level shift (P8 and P9). Moreover, the time series of 24-h UFC of P1 shows a mean level shift approximately halfway through the series, and the series of P2 a negative trend.

The influence of lifestyle factors on 24-h urinary cortisol levels

Results of the uSEM models with 24-h UFC as an outcome variable can be found in table 2. The exact regression estimates can be found in supplement 2. Of all participants only P8 has autocorrelation in the 24-h UFC, that is, for P8 cortisol levels of yesterday predict the cortisol levels of today. For all other participants every day is a new day and no apparent relationship exists between yesterday's levels of 24-h UFC and the levels of today. For six participants (P1, P2, P4, P7, P8, and P10) no relationship is found for either duration of the day, time of awakening, caffeine use, alcohol use, smoking behavior (applicable only to P1 and P5), or exercise. For P3 higher exercise levels significantly predict higher 24-h UFC levels on the same day. For P5 there is a lagged relationship between alcohol consumption and 24-h UFC; higher alcohol consumption yesterday significantly predicts higher 24-h UFC levels of today. Furthermore, there is a significant positive contemporaneous relationship between smoking and 24-h UFC levels. Surprisingly, for this participant higher caffeine consumption is significantly associated with lower 24-h UFC levels in the same day. P6, being a student, has great variability in the time of going to bed and the time of waking up, i.e. not every sampling day is equally long. This is reflected in the significant relationship between duration of the sampling day and 24-h UFC levels. The relationship is, however, not in the

expected direction and in his case the longer sampling days are associated with lower 24-h UFC levels. Furthermore, for P6 a significant positive contemporaneous relationship exists between alcohol consumption and 24-h UFC levels. Finally, for P9 a significant negative lagged relationship exists between exercise and 24-h UFC levels. Thus higher exercise levels of yesterday predict lower 24-h UFC levels today. For P8 the fit of the uSEM model was not optimal according to the two fit indices, but good according to two others.

Figure 1. Time series of salivary and urinary free cortisol of 10 healthy individuals

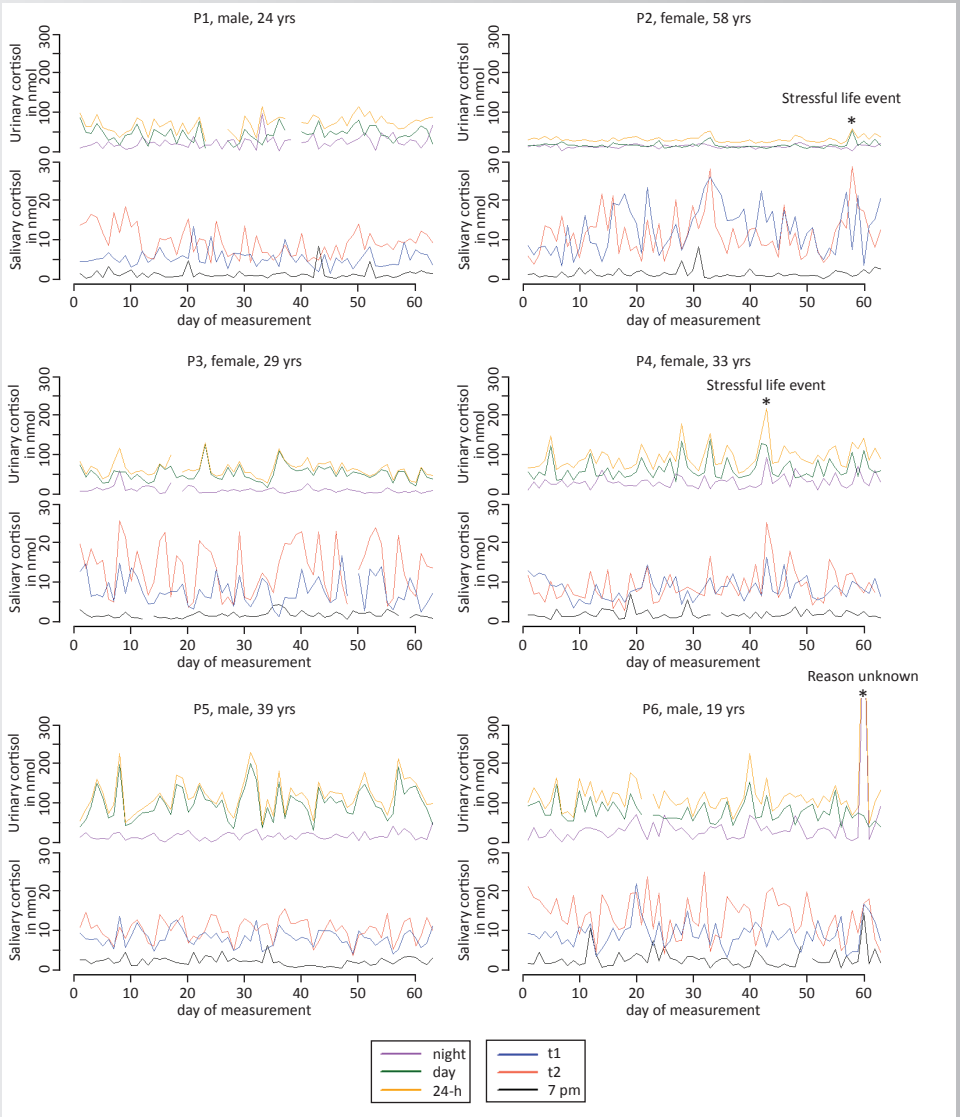
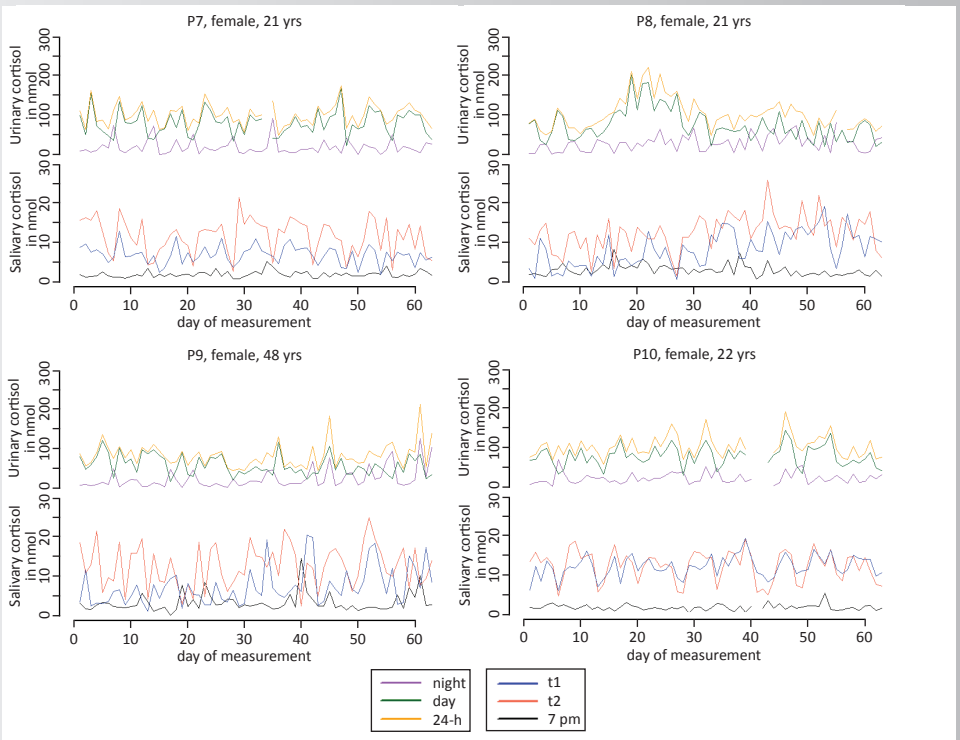


Figure 1 (continued). Time series of salivary and urinary free cortisol of 10 healthy individuals



The influence of lifestyle factors on evening salivary cortisol levels

Results of the uSEM models with evening salivary cortisol as an outcome variable can be found in table 2. The exact regression estimates can be found in supplement 2. None of the participants had autocorrelation in their series, that is, in no case did yesterday's salivary cortisol levels predict cortisol levels of today. For three people (P5, P8, and P9) no significant relationship exists between any of the lifestyle variables and evening salivary cortisol levels. For P1 there is a significant negative relationship between alcohol consumption and salivary cortisol levels on the same day. Some participants did not always take the saliva sample on the instructed time and this is reflected in the significant negative contemporaneous relationship between time of sampling and salivary cortisol levels (P2, P4, and P10). This is in accordance with what would be expected, as due to the circadian rhythm cortisol levels become progressively lower throughout the evening. For P3 a significant positive contemporaneous relationship exists between both caffeine and alcohol consumption and salivary cortisol levels. For P6 there is a significant positive lagged relationship between exercise and salivary cortisol levels, in that higher levels of exercise yesterday predict higher cortisol levels today. Opposite to P3, for P7 there is a significant negative contemporaneous relationship between both caffeine and alcohol consumption and salivary cortisol levels. Finally, also for P7, there is a significant positive contemporaneous relationship between exercise and salivary cortisol levels. For P8 the fit of the uSEM model was not optimal according to the two fit indices, but good according to two others.

Table 2. The results of the unified structural equation models with 24-h urinary free cortisol, the cortisol awakening response, and salivary cortisol at 7 pm as outcome variables

Predictor variables	participant									
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Cortisol (t-1)								24-h UFC+		CAR+
Duration of day (t-1)										
Duration of day (t)						24-h UFC-				
Time of awakening (t-1)			CAR-		CAR-					
Time of awakening (t)			CAR-	CAR-	CAR-					
Time of sampling (t-1)										
Time of sampling (t)		Sal 7pm-		Sal 7pm-						Sal 7pm-
Caffeine (t-1)										
Caffeine (t)			Sal 7pm+		24-h UFC-		Sal 7pm-			
Alcohol (t-1)					24-h UFC+					
Alcohol (t)	Sal 7pm-		Sal 7pm+			24-h UFC+	Sal 7pm-			
Smoking (t-1)					CAR-					
Smoking (t)					24-h UFC+					
Exercise (t-1)						Sal 7pm+			24-h UFC-	
Exercise (t)			24-h UFC+				Sal 7pm+			

Predictor(t-1)=the predictor variable with lag 1 (i.e. describing the relationship of yesterday's values of the predictor variable on the today's value of the outcome variable). Predictor(t)=the predictor variable with lag 0 (i.e. describing the contemporaneous relationship between the predictor variable and the outcome variable). 24-h UFC=24-h urinary free cortisol, CAR=cortisol awakening response, Sal 7pm=salivary cortisol at 7pm. Significant associations ($p<0.05$) between predictor variables and outcome variables are indicated by the outcome variable followed by a + or – sign. The direction of effects is indicated with a + for a significant positive relationship and a – for a significant negative relationship. In this table a blanc space indicates that either no significant effect exists or that it has not been modeled.

The influence of lifestyle factors on the cortisol awakening response

Results of the uSEM models with the CAR as an outcome variable can be found in table 2. The exact regression estimates can be found in supplement 2. For P10 a significant positive autocorrelation of cortisol levels exists. For all other participants no such relationship exists. For seven participants we did not find any relationship between lifestyle factors and the CAR. We want to remind the reader that except for time of awakening, only lagged relationships were modelled, as we cannot assume that lifestyle factors of the same day influenced the CAR. For P3 both a significant negative lagged and contemporaneous relationship exists between time of awakening and the CAR. Thus, the later P3 got out of bed both yesterday and today, the lower today's CAR was. For P4 a significant negative contemporaneous relationship exists between time of awakening and the CAR, meaning that the later she got up, the lower the CAR of the same day would be. For P5, the same as for P3, a significant negative lagged and contemporaneous relationship exists between time of awakening and the CAR. Finally, also for P5, a significant negative lagged relationship between smoking and the CAR exists. Thus the more cigarettes he used yesterday the lower the CAR today. For P8, and P9 the fit of the uSEM model was not optimal according to two out of four fit indices, and good according to two others.

Supranormal 24-h urinary cortisol levels

As can be seen in table 1, taking a cut-off of 133 nmol/24-h, seven participants had supranormal 24-h UFC levels at some point. For four participants this was the case on more than 15% of the days. One participant, P5, had supranormal levels for almost one third of the days (29%). For P6 supranormal levels were obtained on 20% of the days. P8 and P10 had supranormal levels on 19% and 16% of the days respectively.

DISCUSSION

This is the first study investigating the time series of urinary and salivary cortisol of several healthy individuals. We confirmed that the HPA-axis is a highly dynamical system. The within-individual variability was so large that it seemed as if every day was started with a clean slate, as most time series did not have any autocorrelation. Lack of autocorrelation means that previous values of cortisol levels were not useful for the prediction of future values. The lack of stability in day-to-day cortisol levels has also been demonstrated in a recent study by Ross and colleagues [11]. They used multilevel modeling on three independent multiwave studies to partition the day-to-day variance, visit to visit variance, and within-person variance in salivary cortisol measures. It turned out that most of the variance could be explained by day-to-day fluctuations and visit-to-visit fluctuations, indicating that cortisol has more state-like properties than trait-like properties. Over the time span of more than a year intra-class correlation coefficients reached an all-time low and were less than .13. This is in accordance with two other multiwave studies that reported a three-year stability estimate of the CAR of .17 [13], and .13 for the diurnal slope of cortisol over the course of six years [12]. Moreover, single case studies creating time series of the CAR [25] and urinary cortisol [26] also show great day-to-day variability. Ross and colleagues rightfully conclude that "focusing on short-term

cortisol fluctuations may be an especially fruitful research avenue" [11]. We would like to take it a few steps further, and will argue that the effects of time-varying processes such as stress or mood states on HPA-axis functioning should be studied at the level of the individual.

As discussed earlier, in order to generalize findings from the group (between-individual variation) to the level of the individual (within-individual variation) and vice versa two conditions have to be met [6]. We were able to show that patterns of cortisol over time, and the lifestyle variables influencing it, were distinctly different between individuals, and that our sample was quite heterogeneous. This entails that the first condition for ergodicity, homogeneity of the group for the process under study, has not been met (in the case of the mean, variance and autocorrelation of the series and the lifestyle factors). Furthermore, many time series contained trends, meaning that they did not have stable statistical properties over time, and thus the second condition for ergodicity has also not been met. From the above it can be concluded that the HPA-axis is a nonergodic process with highly dynamical properties. This has far-reaching consequences for the way in which we can interpret the results of previous group-based studies. Molenaar demonstrated for instance that for nonergodic processes factor analysis of between-individual covariation can yield a solution that has no relationship whatsoever to the factor structures characterizing the within-individual covariation of each individual subject [27]. This effectively means that for a dynamical system such as the HPA-axis, no a priori relationship between findings at the group level and findings at the individual level can be assumed. Thus, in order to gain insight into for example the effects of stressors on the HPA-axis, time series of both stressors and cortisol would need to be collected. These time series should then be linked together by means of time series analysis, to see if fluctuations in one variable can predict fluctuations in the other over time. Studying processes at the level of the individual does not have to preclude the generation of knowledge which is generalizable to larger groups. The statistical parameters estimated by means of time series analyses can for instance be subjected to cluster analysis techniques to find groups of people that are alike in the way that they respond to stressors over time. Between-individual factors can then be used to predict group membership.

The other important finding in our study is that several seemingly healthy individuals cross the upper threshold for normal population levels of 24-h UFC on multiple days. The diagnosis of Cushing's syndrome by means of 24-h urinary cortisol measurement requires a value of three times the upper normal limit. None of our participants passed this threshold at any time. Seven out of ten participants, however, had supranormal urinary cortisol excretion on several occasions. In a clinical setting, had such values been detected, this would have necessitated further evaluation. If these participants represent individuals with so-called cyclical hypercortisolism is unknown, nor if they suffer from subclinical Cushing's syndrome.

There are several limitations of the current study that need to be considered when interpreting our results. Firstly, we were not able to assess compliance of our participants with the saliva sampling protocol. Sampling time is important due to the circadian nature of cortisol excretion and noncompliance with the sampling protocol could have added noise to our time series (Kudielka et al.). We do not think that this is a sufficient explanation for the great variability of cortisol levels seen in our study, as another study which did

electronically monitor compliance essentially found the same results [11]. Moreover, we could show that our participants were compliant with the urinary sampling protocol, by checking urinary creatinine output, which was much more tedious than the saliva sampling protocol, indicating that they were highly motivated. Secondly, it is important to realize that in a naturalistic field study such as ours, 24-h UFC may not truly represent 24 hours. We have tried to adjust for this by adding the duration of each sampling day, as extracted from the electronic sleep diary, to our models. Unexpectedly, we showed that for one individual (P6) who was a student and had a very erratic sleep pattern, that longer duration of the day predicted lower cortisol levels. Third, as indicated in the results section, for a small minority of time series we were unable to find a uSEM model with a satisfactory fit according to all four fit indices. This limits the interpretation of the results for those particular series for which the fit was not good enough. Fourth, it is important to understand that for our models regarding the salivary cortisol sample at 7 pm the temporal order of events could not be established. Participants filled out the electronic diary on the lifestyle variables right before bedtime. It is thus impossible for us to know if a lifestyle variable preceded and thus could have caused the 7 pm cortisol levels or not. For some variables such as the amount of caffeine containing beverages this is likely the case, but for instance for alcohol, it might also have been the anticipation of a fun social event that lowered cortisol levels, instead of the alcohol itself. This brings us to the fifth limitation of the study. Due to the observational nature of our study residual confounding can never be excluded. It is thus possible that the effects of certain lifestyle variables were in reality driven by a third unmeasured variable. Our study also has several strengths. Firstly, unlike many other studies we used synthetic swabs specifically designed to collect saliva for cortisol measurement. Saliva cortisol levels of saliva collected with the 'Salivette® Cortisol' have been shown to correlate well with serum levels of cortisol [29], whereas usage of cotton swabs can lead to poor recoverability of cortisol [30]. Secondly, cortisol was measured by means of LC-MS/MS which, in contrast to antibody-based assays, is free of interferences from cortisol metabolites and conjugates, and also eliminates drug interferences [31]. Thirdly, we were able to verify completeness of the urine samples by measuring urinary creatinine excretion. With the exception of P1, for whom we had to exclude 5 days, most participants only had 1 or 2 days on which they were noncompliant.

In conclusion, this is the first study to analyze time series of salivary and urinary cortisol of several healthy individuals. We demonstrated that the HPA-axis is a highly dynamical system and that only very little stability within and homogeneity between individuals exists. Consequently, we advise future studies of the HPA-axis to focus on the study of within-individual change over time. Finally, seven out of ten subjects had supranormal 24-h UFC levels at any point. Larger time series studies are needed to clarify what this means for the way we define the normal range and if there are future clinical correlates in those individuals who frequently cross the threshold of the normal population values for 24-h UFC.

REFERENCES

1. McEwen BS. Stress, adaptation, and disease. Allostasis and allostatic load. *Ann N Y Acad Sci*. 1998;840:33–44.
2. Chrousos GP, Gold PW. The concepts of stress and stress system disorders. Overview of physical and behavioral homeostasis. *JAMA*. 1992;267:1244–1252.
3. Vogelzangs N, Beekman AT, Milaneschi Y, Bandinelli S, Ferrucci L, Penninx BW. Urinary cortisol and six-year risk of all-cause and cardiovascular mortality. *J Clin Endocrinol Metab*. 2010;95:4959–4964.
4. Champaneri S, Xu X, Carnethon MR, Bertoni AG, Seeman T, Diez Roux A, Golden SH. Diurnal salivary cortisol and urinary catecholamines are associated with diabetes mellitus: the Multi-Ethnic Study of Atherosclerosis. *Metabolism* . 2012;61:986–95.
5. Vrshek-Schallhorn S, Doane LD, Mineka S, Zinbarg RE, Craske MG, Adam EK. The cortisol awakening response predicts major depression: predictive stability over a 4-year follow-up and effect of depression history. *Psychol Med* . 2013;43:483–93.
6. Molenaar PCM. A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Meas Interdiscip Res Perspect*. 2004;2:201–218.
7. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol*. 2002;31:285–293.
8. Molenaar PCM, Campbell CG. The New Person-Specific Paradigm in Psychology. *Curr Dir Psychol Sci* . 2009;18:112–117.
9. Chida Y, Hamer M. Chronic psychosocial factors and acute physiological responses to laboratory-induced stress in healthy populations: a quantitative review of 30 years of investigations. *Psychol Bull*. 2008;134:829–885.
10. Miller GE, Chen E, Zhou ES. If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychol Bull*. 2007;133:25–45.
11. Ross KM, Murphy MLM, Adam EK, Chen E, Miller GE. How stable are diurnal cortisol activity indices in healthy individuals? Evidence from three multi-wave studies. *Psychoneuroendocrinology*. 2013; 39:184-93.
12. Shirtcliff EA, Allison AL, Armstrong JM, Slattery MJ, Kalin NH, Essex MJ. Longitudinal stability and developmental properties of salivary cortisol levels and circadian rhythms from childhood to adolescence. *Dev Psychobiol* . 2012;54:493–502.
13. Platje E, Vermeiren RRJM, Branje SJT, Doreleijers TAH, Meeus WHJ, Koot HM, Frijns T, van Lier PAC, Jansen LMC. Long-term stability of the cortisol awakening response over adolescence. *Psychoneuroendocrinology* . 2013;38:271–80.
14. Elamin MB, Murad MH, Mullan R, Erickson D, Harris K, Nadeem S, Ennis R, Erwin PJ, Montori VM. Accuracy of diagnostic tests for Cushing's syndrome: a systematic review and metaanalyses. *J Clin Endocrinol Metab* . 2008;93:1553–62.
15. Schubert C, Geser W, Noisternig B, Fuchs D, Welzenbach N, Konig P, Schussler G, Ocana-Peinado FM, Lampe A. Stress system dynamics during "life as it is lived": an integrative single-case study on a healthy woman. *PLoS One*. 2012;7:e29415.
16. Godin G, Jobin J, Bouillon J. Assessment of leisure time exercise behavior by self-report: a concurrent validity study. *Can J Public Health*. ;77:359–62.
17. Boeniger MF, Lowry LK, Rosenberg J. Interpretation of urine results used to assess chemical exposure with emphasis on creatinine adjustments: a review. *Am Ind Hyg Assoc J* . 1993;54:615–27.
18. Bingham SA, Williams R, Cole TJ, Price CP, Cummings JH. Reference values for analytes of 24-h urine collections known to be complete. *Ann Clin Biochem* . 1988;25 (Pt 6):610–9.
19. Perrone RD, Madias NE, Levey AS. Serum creatinine as an index of renal function: new insights into old concepts. *Clin Chem* . 1992;38:1933–53.
20. Jöreskog K, Sörbom D. LISREL. 1992;

21. Kim J, Zhu W, Chang L, Bentler PM, Ernst T. Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Hum Brain Mapp* . 2007;28:85–93.

22. Gates KM, Molenaar PCM, Hillary FG, Ram N, Rovine MJ. Automatic search for fMRI connectivity mapping: an alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. *Neuroimage* . 2010;50:1118–25.

23. Brown T. Confirmatory factor analysis for applied research. 1 edition. New York: Guilford Press; 2006.

24. StataCorp. Stata Statistical Software. 2011;

25. Stalder T, Hucklebridge F, Evans P, Clow A. Use of a single case study design to examine state variation in the cortisol awakening response: relationship with time of awakening. *Psychoneuroendocrinology* . 2009;34:607–14.

26. Schubert C, Lampe A, Geser W, Noisternig B, Fuchs D, König P, Chamson E, Schüssler G. Daily psychosocial stressors and cyclic response patterns in urine cortisol and neopterin in a patient with systemic lupus erythematosus. *Psychoneuroendocrinology* . 2003;28:459–73.

27. Molenaar PC. Time series analysis and its relationship with longitudinal analysis. *Int J Sports Med* . 1997;18 Suppl 3:S232–7.

28. Kudielka BM, Broderick JE, Kirschbaum C. Compliance with saliva sampling protocols: electronic monitoring reveals invalid cortisol daytime profiles in noncompliant subjects. *Psychosom Med*. ;65:313–9.

29. Poll E-M, Kreitschmann-Andermahr I, Langejuergen Y, Stanzel S, Gilsbach JM, Gressner A, Yagmur E. Saliva collection method affects predictability of serum cortisol. *Clin Chim Acta* . 2007;382:15–9.

30. Gröschl M, Rauh M. Influence of commercial collection devices for saliva on the reliability of salivary steroids analysis. *Steroids* . 2006;71:1097–100.

31. Taylor RL, Machacek D, Singh RJ. Validation of a high-throughput liquid chromatography-tandem mass spectrometry method for urinary cortisol and cortisone. *Clin Chem* . 2002;48:1511–9.

SUPPLEMENT 1: FIT INDICES OF USEM MODELS

Table 1. The fit indices of the unified structural equation models with 24-h urinary cortisol as outcome variable

Fit index	Participant									
	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10
NNFI	0.91	1.13	0.73	1.52	1.12	1.18	0.97	-0.61	1.61	1.08
CFI	0.99	1.00	0.96	1.00	1.00	1.00	0.99	n.a.	1.00	1.00
RMSEA	0.055	0.0	0.079	0.0	0.0	0.0	0.034	0.0	0.0	0.0
SRMSR	0.045	0.038	0.043	0.033	0.029	0.017	0.040	0.038	0.030	0.037

NNFI=Non-Normed Fit Index, CFI=Comparative fit index, RMSEA=Root mean square error of approximation, SRMSR=Standardized root mean square residual, n.a.=not available, index could not be computed

Table 2. The fit indices of the unified structural equation models with the salivary cortisol at 7 pm as outcome variable

Fit index	Participant									
	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10
NNFI	1.15	1.00	0.99	0.99	1.11	6.23	1.13	-24.54	9.06	1.71
CFI	1.00	1.00	1.00	1.00	1.00	1.00	1.00	n.a.	0.0	1.00
RMSEA	0.0	0.0	0.045	0.0	0.0	0.0	0.0	0.0	0.068	0.0
SRMSR	0.019	0.043	0.029	0.025	0.046	0.039	0.028	0.044	0.052	0.033

NNFI=Non-Normed Fit Index, CFI=Comparative fit index, RMSEA=Root mean square error of approximation, SRMSR=Standardized root mean square residual, n.a.=not available, index could not be computed

Table 3. The fit indices of the unified structural equation models with the cortisol awakening **r**esponse as outcome variable

Fit index	Participant									
	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10
NNFI	1.84	0.78	1.53	1.24	0.61	1.73	1.23	-18.23	-10.20	2.45
CFI	1.00	0.97	1.00	1.00	0.98	1.00	1.00	n.a.	n.a.	1.00
RMSEA	0.0	0.057	0.0	0.0	0.088	0.0	0.0	0.0	0.0	0.0
SRMSR	0.025	0.049	0.013	0.040	0.032	0.020	0.036	0.020	0.032	0.030

NNFI=Non-Normed Fit Index, CFI=Comparative fit index, RMSEA=Root mean square error of approximation, SRMSR=Standardized root mean square residual, n.a.=not available, index could not be computed

SUPPLEMENT 2: ESTIMATES OF THE USEM MODELS

Table 1. The results of the unified structural equation models with 24-h urinary cortisol as outcome variable

Variable—estimate (standard error) t-value	**participant**									
	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10
Cortisol (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.27 (0.12) 2.21	n.s.	n.s.
Duration of day (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Time of awakening (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Caffeine (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Alcohol (t-1)	n.s.	n.s.	n.s.	n.s.	0.26 (0.12) 2.10	n.s.	n.s.	n.a.	n.a.	n.s.
Smoking (t-1)	n.s.	n.a.	n.a.	n.a.	n.s.	n.a.	n.a.	n.a.	n.a.	n.s.
Exercise (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	-0.28 (0.12) -2.27	n.s.
Duration of day (t)	n.s.	n.s.	n.s.	n.s.	n.s.	-0.29 (0.12) -2.33	n.s.	n.s.	n.s.	n.s.
Time of awakening (t)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Caffeine (t)	n.s.	n.s.	n.s.	n.s.	-0.37 (0.17) -2.20	n.s.	n.s.	n.s.	n.s.	n.s.
Alcohol (t)	n.s.	n.s.	n.s.	n.s.	n.s.	0.32 (0.12) 2.59	n.s.	n.a.	n.a.	n.s.
Smoking (t)	n.s.	n.a.	n.a.	n.a.	0.47 (0.17) 2.76	n.a.	n.a.	n.a.	n.a.	n.a.
Exercise (t)	n.s.	n.s.	0.28 (0.12) 2.31	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

n.s. = not significant, n.a. = not applicable, Predictor(t-1)=the predictor variable with lag 1 (i.e. describing the relationship of yesterday's values of the predictor variable on the today's value of the outcome variable). Predictor(t)=the predictor variable with lag 0 (i.e. describing the contemporaneous relationship between the predictor variable and the outcome variable). Significant associations p<0.05.

Table 2. The results of the unified structural equation models with the salivary cortisol at 7 pm as outcome variable

Variable—estimate (standard error) t-value	**participant**									
	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10
Cortisol (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Time of sampling (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Caffeine (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Alcohol (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.	n.s.
Smoking (t-1)	n.s.	n.a.	n.a.	n.a.	n.s.	n.a.	n.a.	n.a.	n.a.	n.s.
Exercise (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	0.29 (0.12) 2.42	n.s.	n.s.	n.s.	n.s.
Time of sampling (t)	n.s.	-0.33 (0.12) -2.71	n.s.	-0.34 (0.12) -2.83	n.s.	n.s.	n.s.	n.s.	n.s.	-0.26 (0.12) -2.14
Caffeine (t)	n.s.	n.s.	1.84 (0.52) 3.54	n.s.	n.s.	n.s.	-0.44 (0.18) -2.44	n.s.	n.s.	n.s.
Alcohol (t)	-0.28 (0.12) -2.30	n.s.	1.80 (0.52) 3.47	n.s.	n.s.	n.s.	-0.63 (0.20) -3.20	n.a.	n.a.	n.s.
Smoking (t)	n.s.	n.a.	n.a.	n.a.	n.s.	n.a.	n.a.	n.a.	n.a.	n.a.
Exercise (t)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.33 (0.15) 2.12	n.s.	n.s.	n.s.

n.s. = not significant, n.a. = not applicable, Predictor(t-1)=the predictor variable with lag 1 (i.e. describing the relationship of yesterday's values of the predictor variable on the today's value of the outcome variable). Predictor(t)=the predictor variable with lag 0 (i.e. describing the contemporaneous relationship between the predictor variable and the outcome variable). Significant associations p<0.05.

Table 3. The results of the unified structural equation models with the cortisol awakening response as outcome variable

Variable—estimate (standard error) t-value	**participant**									
	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10
Cortisol (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.28 (0.12) 2.30
Time of awakening (t-1)	n.s.	n.s.	-0.28 (0.12) -2.39	n.s.	-0.31 (0.12) -2.61	n.s.	n.s.	n.s.	n.s.	n.s.
Caffeine (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Alcohol (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.	n.s.
Smoking (t-1)	n.s.	n.a.	n.a.	n.a.	-0.28 (0.12) -2.35	n.a.	n.a.	n.a.	n.a.	n.s.
Exercise (t-1)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Time of awakening (t)	n.s.	n.s.	-0.30 (0.12) -2.61	-0.29 (0.12) -2.34	-0.29 (0.11) -2.54	n.s.	n.s.	n.s.	n.s.	n.s.

n.s. = not significant, n.a. = not applicable, Predictor(t-1)=the predictor variable with lag 1 (i.e. describing the relationship of yesterday's values of the predictor variable on the today's value of the outcome variable). Predictor(t)=the predictor variable with lag 0 (i.e. describing the contemporaneous relationship between the predictor variable and the outcome variable). Significant associations p<0.05.