Assessment of the cortisol awakening response: Real-time analysis and curvilinear effects of sample timing inaccuracy

Smyth, N., Thorn, L., Hucklebridge, F., Clow, A. and Evans, P.

https://dx.doi.org/10.1016/j.psyneuen.2016.09.026

Accepted Manuscript

Title: Assessment of the cortisol awakening response: Real-time analysis and curvilinear effects of sample timing inaccuracy

Author: Nina Smyth Lisa Thorn Frank Hucklebridge Angela Clow Phil Evans

PII: S0306-4530(16)30489-9
DOI: http://dx.doi.org/doi:10.1016/j.psyneuen.2016.09.026
Reference: PNEC 3408

To appear in:

Received date: 26-7-2016
Revised date: 26-9-2016
Accepted date: 28-9-2016

Please cite this article as: {http://dx.doi.org/

Assessment of the cortisol awakening response: real-time analysis and curvilinear effects of sample timing inaccuracy

Nina Smyth, Lisa Thorn, Frank Hucklebridge, Angela Clow and Phil Evans*

Psychophysiology and Stress Research Group, University of Westminster, 115 New Cavendish Street, London W1W 6UW, UK

*Corresponding author

Highlights

- All cortisol data objectively verified for awakening and saliva sampling times
- Cortisol growth curve plot against real-time not affected by protocol deviance
- In healthy young adults the mean CAR was a 100% increase from awakening cortisol
- Curvilinear delay effect on CAR size if protocol times wrongly assumed accurate
- Electronic-monitoring vital for CAR measurement and meaningful interpretation

Abstract

The cortisol awakening response (CAR) is typically measured in the domestic setting. Moderate sample timing inaccuracy has been shown to result in erroneous CAR estimates and such inaccuracy has been shown partially to explain inconsistency in the CAR literature. The need for more reliable measurement of the CAR has recently been highlighted in expert consensus guidelines where it was pointed out that less than 6% of published studies provided electronic-monitoring of saliva sampling time in the post-awakening period.

Analyses of a merged data-set of published studies from our laboratory are presented. To qualify for selection, both time of awakening and collection of the first sample must have been verified by electronic-monitoring and sampling commenced within 15 min of awakening. Participants (n=128) were young (median age of 20 years) and healthy. Cortisol values were determined in the 45 min post-awakening period on 215 sampling days. On 127 days, delay between verified awakening and collection of the first sample was less than 3 minutes ('no delay' group); on 45 days there was a delay of 4-6 min ('short delay' group); on 43 days the delay was 7-15 min ('moderate delay' group).

Cortisol values for verified sampling times accurately mapped on to the typical post-awakening cortisol growth curve, regardless of whether sampling deviated from desired protocol timings. This provides support for incorporating rather than excluding delayed data (up to 15 min) in CAR analyses. For this population the fitted cortisol growth curve equation predicted a mean cortisol awakening level of 6 nmols/l (+/-1 for 95% CI) and a mean CAR rise of 6nmols/l (+/- 2 for 95% CI). We also modelled the relationship between real delay and CAR magnitude, when the CAR is calculated erroneously by incorrectly assuming adherence to protocol time. Findings supported a curvilinear hypothesis in relation to effects of sample delay on the CAR. Short delays of 4-6 min between awakening and commencement of saliva sampling resulted an overestimated CAR. Moderate delays of 7-15 min were associated with an underestimated CAR. Findings emphasize the need to employ electronic-monitoring of sampling accuracy when measuring the CAR in the domestic setting.

Kewords: Cortisol; saliva; cortisol awakening response; CAR; growth cortisol curve; sample timing inaccuracy

1. Introduction

Typically, awakening triggers a marked rise in cortisol secretion, normally peaking within 45 min post-awakening at around 30 min, followed by a normal diurnal decline. This rise has by convention been termed the 'Cortisol Awakening Response' (CAR). Since its discovery, researchers have become interested in exploring this phenomenon in relation to possible trait and state correlates, especially in the domains of cognition, affect, health and well-being (Ennis et al., 2016; Evans et al., 2012; Evans et al., 2007; Juster et al., 2011; Lovell et al., 2011; Stalder et al., 2010a, b; Stalder et al., 2009; Steptoe et al., 2007; Steptoe et al., 2008). Assessment of the CAR is typically in the domestic setting, with self-collection of saliva samples on awakening and at fixed intervals up to 45 min post-awakening. Given the brief time-window of the post-awakening cortisol rise, accurate sampling times relative to awakening are imperative for assessment of the CAR (Smyth et al., 2013a), an issue that has been highlighted in a recent expert consensus guidelines paper (Stalder et al., 2016). Awakening and sampling times are typically based on participants' self-reports and only a small proportion (5.7%) of published studies conducted in the domestic setting (between 2013-2014) provided electronic-monitoring of sampling time relative to awakening. This is alarming given that delays between awakening and collection of saliva samples in the post-awakening period result in erroneous CAR measures (Broderick et al., 2004; DeSantis et al., 2010; Dockray et al., 2008; Golden et al., 2014; Griefahn and Robens, 2011; Kudielka et al., 2003; Kudielka et al., 2007; Kupper et al., 2005; Okun et al., 2010; Smyth et al., 2013a). There is also growing evidence that incorporation of uncorroborated data in computing CAR measures potentially influences findings and may partially explain some noted inconsistencies in the CAR literature (Smyth et al., 2015b).

When time derived from electronic monitoring does not match time defined by protocol, data are typically determined "inaccurately timed" and then excluded from CAR calculations (Ramachandran et al., 2016; Smyth et al., 2015b). Management of such inaccurate data in this way is costly, and could be avoided if it were possible to incorporate cortisol values outside of the fixed sample protocol times. Maximal use of hard-won data is not an insignificant issue and therefore it is vital to examine whether data obtained from such inaccurately timed samples are fully useable when analyzed in real-time. Such an approach was recommended in the expert consensus guidelines

(Stalder et al., 2016). However, no study has conducted an analysis of post-awakening cortisol data in electronically verified real-time. Real-time analysis would also allow the cortisol growth curve to be plotted effectively using something closer to near continuous sampling. Little is known about cortisol levels between the common fixed sampling points of most study protocols. Areas under the curve are typically presented as illustrations, with lines joining adjacent time points, with an unspoken assumption of interpolated linearity. However, where an assumption of linearity has been investigated with repeated time-verified sampling in the first 15 min interval following awakening (Smyth et al., 2013a), linearity was not supported. Rather there was a brief latency period immediately after awakening and ending sometime between 5 and 10-min later when cortisol rise is clearly evident with growth curve modelling estimating the point of up-swing in cortisol starting at 8-min.

If delayed cortisol values of known timing, plotted in real-time, fit the normal cortisol growth curve, the corollary has to be that they would perforce give rise to distorted CAR values if plotted in protocol time, significantly reinforcing concerns about imprecise measurement. Existing electronic monitoring studies tentatively suggest a curvilinear relationship between delay and CAR magnitude. Short delays between awakening and collection of the first sample (Md =7-min with a modal 28% of sample delayed by only 5-min) have yielded erroneously *larger* CARs if the first sample is wrongly assumed to be undelayed (Smyth et al., 2013a). Intensive 5 min sampling suggest this is a consequence of the 'latency' period which as we have just noted typically ends at or just after this time point relative to real awakening time (Smyth et al., 2013a; Smyth et al., 2015a). Thus the erroneous awakening value from which CAR rise is calculated is in reality delayed sufficiently to be just still in the latency period and likely therefore to be similar to what the hypothetical real awakening time sample would have been. However the short delay when carried forward to later samples will be associated with a higher average value in terms of their positioning under the real-time cortisol growth curve. The net effect is to yield erroneously higher CAR rise measures if protocol time is wrongly assumed. By contrast, a longer established evidence base has consistently found that longer delays are associated with smaller CARs (Dockray et al., 2008; Griefahn and Robens, 2011; Kupper et al., 2005; Okun et al., 2010), since the erroneously assumed base starting value, beyond the end of the brief latency period, will now most likely and rapidly become much higher than it should be and the potential for further CAR rise thus constrained.

In the present study, we present analyses of what we believe is the largest merged data-set yet assembled, where timing of awakening and collection of the first sample were verified using electronic-monitoring. Data were derived from studies of healthy young adults in our laboratory

using the saliva collection protocol of an awakening sample and three further samples at 15 min intervals over a 45 min period. Findings from each contributing data-set, in respect of diverse aims, have already been published (Ramachandran et al., 2016; Smyth et al., 2013a; Smyth et al., 2015a), but their merger permits exploitation of cortisol data of known collection timing previously excluded as not sufficiently protocol accurate. Inclusion of such data permits analysis, for the first time, of cortisol data in verified real-time, to clarify whether the typical growth curve of post-awakening salivary cortisol is evident, regardless of sampling accuracy. Data will also provide the most accurate parameters yet published of the typical CAR period growth curve, using real-time data, not just at fixed 15 min time intervals. Such data, allows us to compute predicted values for post-awakening cortisol and CAR measures as reference values for a young healthy sample.

This merged database has also enabled investigation of the impact of curvilinear effects of sample timing accuracy on composite measures of the CAR commonly used in the literature: the simple cortisol rise from awakening to 30 min and average rise across the whole 45 min period. The curvilinear hypothesis is currently based on plausible inference from a limited number of small studies; it has not been tested in a single large data-set. Using data in this pooled data-set, we modelled the impact and effect size of short and moderate sampling delay on measures of the CAR, with the expectation of finding a significant curvilinear effect. Specifically, if adherence to fixed protocol times is wrongly assumed for delayed data, the following curvilinear relationship should pertain: CAR magnitude will increase as sampling delay increases from a minimal range of 0-3-min to short delays of between 4-7-min, followed by a rapid decline in CAR magnitude as delay increases further (from 8-15min).

2. Method

2.1 Database

Data originating from four previously published studies by our group were merged into a single database. Data were drawn from two studies presented in a single paper by Smyth et al. (2013a), from Smyth et al. (2015a) and Ramachandran et al. (2016). In total, data derived from 128 healthy participants, recruited from the academic community at the University of Westminster (median age 20 years and inter-quartile range 18-24 years). The study samples were predominantly female (N = 102, 79%), non-smokers (N = 98, 77%) and none were suffering from any medical or psychiatric illness. Mean electronically-monitored wake time for all participants across two study days was 07:25 (±01:43).

In all four studies the protocols included provision for the collection of saliva samples immediately

on awakening (S1) and at 15 (S2), 30 (S3) and 45 (S4) min thereafter on two days. In all studies participants attended an individual induction session with the lead researcher, during which they were given full verbal and written instructions on saliva sampling and electronic-monitoring procedures, and were able to practise collecting saliva samples using either an Eppendorf tube (with straw to aid passive drool) or a salivette device (Sarstedt Ltd., Leicester, England). Participants were instructed to awake in their usual way and to refrain from smoking, brushing their teeth, exercising and to remain nil by mouth bar water during the saliva collection period. Participants completed record sheets on each day which included information regarding their awakening times, their protocol-required saliva sampling times based on their awakening time that day, and their actual saliva sampling times. In each study participants were informed about the need to adhere to the strict sampling regime relative to awakening time and were informed that the electronic devices would be used to verify their self-reported awakening and saliva sampling times, a strategy shown to increase sampling accuracy (Broderick et al., 2004; Kudielka et al., 2003). In a further effort to maximise protocol adherence participants received text-messages the evening prior to each study day, reminding them to wear the actiwatch to bed and store saliva sampling kit next to their bed. Samples were initially stored in a domestic freezer until they were returned to the laboratory to be stored at -20°C until assayed.

Cortisol concentrations were determined by enzyme linked immunosorbent assay developed by Salimetrics LLC (USA) at the Psychophysiology and Stress Research Group's laboratory at the University of Westminster. Standards, controls and all samples were assayed in duplicate and intra and inter-assay variations were both below 10%.

2.2 Electronic monitoring of sampling accuracy

In all data-providing studies there was electronic-monitoring of protocol adherence to ensure accurate (*non-delayed)* data were used to investigate their original research questions. Full details of how saliva sampling accuracy was determined can be found in the source publications. In summary participants were provided with wrist-worn activity-recording device (Cambridge Neurotechnology, Cambridge or Philips Respironics, UK) to monitor awakening times. These devices are a piezo-electric motion sensor recording physical activity. Awakening times were estimated using the actiwatch software that distinguishes sleep and awakening periods by reduced and increased activity respectively. In line with recommendations from (Boyne et al., 2013), actigraph awakening times were scored by the human eye rather than the computer algorithm. The lead researcher scored awakening times and other authors verified at least 10% of actigraph timings.

Track caps (Medication Event Monitoring Caps) were used to record the date and time of the opening of the bottle (containing straws or Salivette cotton swabs for sampling) and participants were instructed to open this device only at the time of saliva sampling. The timing of the track cap openings indicated the collection timing of samples.

To qualify for selection into the merged database both the time of awakening and collection of the first sample were verified by electronic devices. Smyth et al. (2013a) showed that sample timing inaccuracy was overwhelmingly attributed to delay in collection of the first sample, with later sampling times usually well synchronised to that of the first sample, i.e. at 15 min intervals thereafter, simply carrying forward delay between awakening and S1 as a constant.

For the pilot study reported alongside the main study in Smyth et al (2013a) and also Smyth et al. (2015a), MEMs monitoring after S1 was neither necessary nor feasible since sampling was carried out virtually continuously (every 5 min). However, in the two larger sample studies of Smyth et al., (2013a) and Ramachandran et al., (2016) it was possible to use MEMs in the vast majority of cases (86%) to monitor electronically collection times of samples 2-4. Missing sampling times in the 14% minority of cases were due either to participants opening and closing the track cap bottle too quickly so that the times were not recorded and/or leaving the bottle open for the sampling period despite being instructed to open and close the bottle at each sampling time. In the 86% of cases were MEMs data existed these were highly concordant with participants' self-reported timings, which is not surprising since participants were informed that sampling times were being electronically-monitored and this is known to increase the accuracy of self-reported timing (Broderick et al., 2004; Kudielka et al., 2003). Accordingly all data (EM or self-report) were used to check the assumption of approximately 15-min interval accuracy in providing samples 2-4. The assumption was overwhelmingly supported with both median and modal values of exactly 15-min obtained for each evaluated interval on each day, and IQRs averaging a single minute. Finally the very few cases (<3%) where checks suggested the possibility of a deviation of greater than 7.5-min were coded so that modelling could be carried out with or without these data. Subsidiary analyses confirmed they had no material impact on the reported outcome of this study.

In summary, in all cases known sampling times were based on electronic-monitoring of collection of S1 relative to participants' awakening-time. Sampling times for S2-4 were based on electronically-monitored S1 timing plus 15, 30 or 45 min.

Treatment of Data and Statistical Methods

All four studies providing data in the merged dataset involved both within and between participant variables. Cortisol values constituted repeated measures over days and sample-points within days. Given the emphasis placed on sample timing accuracy in these studies, the amount of delayed data (41%) was smaller than may otherwise have been the case, and the overall distribution of delay time proportionately more asymmetric. To mitigate this, we sought to maximise sample size of ordered interval ranges in comparison of delay conditions, while optimising approximation to expectations in regard to interval cut-off points for curvilinear predictions. Accordingly, SPSS visual binning routine was used to derive three interval ranges up to a maximum delay of 15 min which are labelled: (1) No significant delay (≈ None); (2) Short delay; and (3) Moderate delay. As can be seen from Table 1, the total N for the less populated short and moderate delay intervals divided very nearly equally between them (45 vs 43 days). At the same time the interval cut-off points reflected well the *a priori* expectation of where discrimination between contrasting delay effects (*increased* versus *decreased* CAR magnitude) would be optimized, based on existing studies reviewed in the introduction to this paper. Median average delay best represented typical delay within interval and these values were used to create a scale covariate.

Insert table 1 about here

Raw cortisol values were positively skewed and therefore root-transformed for inferential analyses so as to minimize the skew statistic relative to its standard error. Composite CAR measures and components were also treated similarly where skewness ratio to standard error exceeded two. Summary measures for descriptive purposes are reported in original units (nmols/l) for the overall CAR period growth curve, and standardized scores for composite CAR measures where comparison of delay conditions relative to the mean for the whole data-set is of cardinal interest.

A mixed regression approach was used to model the growth curve of cortisol values over known sampling times from awakening. Models were initially run to optimize the covariance structure for repeated measures expressed by day and sample-point order. A first-order auto-regressive structure was adopted in the final model based on tests of covariance parameters and minimizing of Schwarz's Bayesian Criterion (BIC). Estimates for cortisol values at 0, 15, 30 and 45 min post-awakening as well as composite cortisol measures were derived from the equation for the predicted quadratic fit.

Mixed regression modelling was used to examine polynomial curvilinear trend for delay effects on composite measures of the CAR. The first was the mean increase of cortisol from the first sample (MnInc: Mean [S2, S3, S4] - S1). When sampling intervals are approximately equal, the MnInc measure is virtually identical to an alternative CAR measure, viz., the area under the curve with

respect to increase from S1 (AUCi). Both constitute the typical composite measure of CAR magnitude in studies which collect multiple post-awakening saliva sample points. The MnInc comprises three components: delta 0-15min, delta 0-30min, and delta 0-45min. Increases in cortisol level from S1 to each subsequent sample points i.e. S2—S1, S3—S1, and S4—S1 were computed. The delta 0-30 min component (assuming accurate timing) represents the simple rise in cortisol in the first 30 min following awakening. It is pertinent to point out that delta 0-30 min has been used extensively (and delta 0-45 min occasionally) as a very simple CAR measure in the literature where limited resources have restricted saliva collection to two samples. These components were modelled separately to examine comparability of their estimated effect sizes in relation to delay, and further analyses were conducted controlling for potential confounding of reported effects by measured covariates (wake-time, sex, age, and smoking status). Given the only repeated measures variable in these analyses was study day with only two levels, sphericity was not an issue and reported effects are based on an intercepts only covariance structure (equivalent to compound symmetry). Simple regression was then performed entering a single delay vector weighted by coefficients of the quadratic equation (predicted scores), yielding estimates of effect size (pseudo-R^2) and associated probability based on a numerator df of 1.

3. Results

3.1 Modelling the growth curve of cortisol in verified real time

The first analysis addressed the question: do verified known sampling times, regardless of accuracy in relation to desired protocol sampling times, result in the typical post-awakening cortisol growth curve when plotted against verified real-time from awakening? The obtained growth curve of cortisol over 60 min post-awakening revealed the usual significant linear (F = 267.857; df = 1, 730.831; $p < 0.001$) and quadratic components (F = 46.813; df = 1, 653.823; $p < 0.001$), underpinned by the typical steep rise in cortisol over the first half hour followed by a flattening of the curve thereafter. There was no main effect of protocol delay category (F = 0.186; df = 1, 399.708; p = .666) and no evidence of any significant interaction between delay and linear (F = 0.023, df = 1, 670.531, p = .879) or quadratic (F = 0.245; df = *1*, 661.437; p = .621) time vector terms. This is very evident in Figure 1, where the quadratic lines of fit for protocol-delayed data, for protocol-adherent data, and for the full combined data set are overlaid. The growth curves are very closely overlapping, and yield similar estimates of starting values for the CAR (i.e. cortisol level at awakening) and two widely used composite CAR measures (MnInc and rise from 0-30min post-awakening).

Derived from the equation for quadratic fit, Table 2 provides estimates, standard errors, and 95% confidence intervals of cortisol values for 0, 15, 30 and 45 min post-awakening as well as the two

commonly used composite measures of CAR (MnInc, and Delta 0-30 min). To the nearest nmol/l, average cortisol level at awakening for this population was around 6 with a 95% confidence range between 5 and 7. Averaging across the two widely used CAR measures, mean rise is also around 6, amounting to a doubling of cortisol in the post-awakening period. Since CARs are fundamentally 'difference-scores', clearly their standard errors are proportionately greater than their counterparts for simple cortisol values. Thus the 95% confidence range of the CAR is wider, between 4 and 8 nmols/l of rise.

Insert Figure 1 about here

Insert Table 2 about here

3.2 Modelling the impact of delay on CAR measures under the false assumption of adherence to protocol required sampling times

We first examined how delay affected the composite MnInc measure of the CAR. This was repeated for the three components of MnInc, i.e. the simple delta scores from awakening to each of the three subsequent sample times. Results for all of these analyses are shown in Table 3. For the primary MnInc measure the means are very much in line with predicted curvilinear trend when delay scores are entered as a covariate (F (1, 209) = 5.589, p = .019). Predicted values are plotted in Figure 2a. Assuming virtually no delay, we can see that participants' CARs were more or less average (mean z = -.004) in terms of the whole data-set, i.e. a z-value close to zero. With a short delay of typically around 5 min, participants exhibited a larger mean CAR (.280), i.e. between a quarter and a third of a standard deviation above the data-set average. By contrast, with delays of typically 9 min, the mean CAR (-.263) fell to over a quarter of a standard deviation below the data-set average. While effects are clearly significant, the overall effect size associated with the significant curvilinear fit is small with delay accounting for only 2.6% of variation in MnInc.

A similar pattern to MnInc was observed for delta 0-15 min and delta 0-30 min composites (F (1, 209) = 6.807, p = .010 and F (1, 209) = 6.450, p = .012 respectively). Effect sizes are a little higher than for MnInc but with delay still accounting for only about 3% of variance in CAR magnitude. No significant effects were found for the delta 0-45 min measure (F (1, 209) = 2.156, p = .143). Predicted values are plotted in Figure 2b, 2c, and 2d.

Finally, a series of additional analyses were undertaken to explore whether simultaneous entry of potentially relevant covariates (wake-time, sex, age, and smoking status) into the models might

significantly change the reported effect sizes of delay on the magnitude of the CAR. Findings were robust to these tests of extraneous covariate influence.

Insert Figure 2 about here

4. Discussion

For the first time, using a merged data-set of electronically-monitored awakening and collection of the awakening sample, we have shown that a typical post-awakening growth curve is evident when cortisol values are plotted against verified real-sampling times, regardless of protocol sampling time accuracy. This means that data, which are delayed by less than 15 min between awakening and collection of the first sample, do not need to be excluded from analyses if real-time and not fixed protocol-time is modelled. This investigation provided an opportunity to report predicted post-awakening cortisol and CAR values in a young healthy sample, using a data-set with considerably more spread across the post-awakening 60 min period. Mean awakening level of cortisol was around 6 nmols/l with a mean rise of around 100%. In addition, the findings support a curvilinear hypothesis in relation to the effects of sample timing inaccuracy on summary measures of the CAR. When CAR magnitude was plotted against fixed protocol timings, short delays of between 4-6 min between awakening and commencement of saliva sampling resulted in overestimated CAR measures. In contrast, moderate delays of 7-15 min were associated with underestimated CAR estimates.

The goal of increased measurement precision of the CAR cannot be achieved without increased burden for the researcher and participant. Consequently, the question of how to deal with data derived from saliva samples of known timing which are termed inaccurate in so far as they have not been collected when the researcher asked them to be collected becomes imperative. In the past it has been recommended (Stalder et al., 2016) and practised (Ramachandran et al., 2016; Smyth et al., 2015b) to exclude such data from analyses. However, this is a wasteful and costly approach. The use of such data in analyses necessarily entails abandoning traditional repeated-measures ANOVA approaches, in favour of multi-level linear modelling techniques (see Smyth et al., 2013b; Stalder et al., 2016). However, there is considerable evidence that the latter techniques are more appropriate, more flexible, and more powerful than the former in regard to mixed effects designs (Blackwell et al., 2006; Smyth et al., 2013b; Stalder et al., 2016), and modelling software for the latter is now widely available. These novel analyses provide evidence that if sampling deviates from desired fixed protocol timings (up to 15 min delay) the cortisol growth curve can still be mapped on to real verified sampling times. We have shown that, controlling for sample timing using modelling

techniques, makes negligible difference to growth curve fit in real time and legitimises this approach.

This means that CAR researchers can become more inventive in terms of saliva collection protocols, sampling much more evenly and widely across the whole post-awakening period in order to produce authoritative parameters for cortisol levels at all time points on the curve. Sampling (ideally random after the initial post awakening sample) along the relevant length of the time vector should be the essence of a genuine growth curve. One consequence of more certain knowledge and confidence in growth curve parameters is the opportunity to explore the potential of alternative CAR measures based on the use of alternative time points (e.g. 35 or 40 min post-awakening) on the growth curve. Such alternatives may prove more robust at least to the small sampling timing errors which are indubitably currently present and continue to cast a shadow over the CAR literature by dint of the unquantifiable effects they may have had on reported findings.

Although the protocol-deviant data fitted alongside protocol-accurate data into an expected growth curve pattern in real-time there is a necessary corollary of that. It must entail that CAR measures, erroneously assuming accuracy to protocol timings, have to differ in magnitude as a function of their real delayed status. Direction of differences will in turn depend on amount of delay. For all four measures of CAR rise examined, the predicted curvilinear relationship was observed and was significant for all measures except the rise from awakening to 45 min. In terms of short delays (4-6 min) CAR magnitude was greater than with no delay, replicating a study by (Smyth et al., 2013a). However, longer delays (7-15 min) were associated with CARs of smaller magnitude than those with no delay, which replicates findings from a number of studies (Dockray et al., 2008; Griefahn and Robens, 2011; Kupper et al., 2005; Okun et al., 2010). This, however, is the first study, which has formally tested this curvilinear hypothesis. The hypothesis arose not only from considering separate existing studies together but more theoretically from 'time-shifting' the detailed cortisol curve between 0 and 15 min reported by (Smyth et al., 2013a; Smyth et al., 2015a) in order to predict the consequences of various delay intervals.

The utility of the CAR literature is limited by inconsistent associations (see Chida and Steptoe, 2009; Clow et al., 2004; Fries et al., 2009). The typical growth curve of cortisol in the post-awakening period is a clear and well-accepted phenomenon. However, it is only typical as applied to the plotting of averaged data from participants. The impact on a study's findings of a subset of unidentified delayed data is clearly impossible to assess, but should not be under-estimated, given the effects we have demonstrated of sample timing inaccuracy on the CAR magnitude. It is the reliability and validity of individual scores, which count when investigating associations between the

CAR and health, psychosocial or cognitive variables. The same goes for investigating temporal co-variability of within-participant changes in the CAR and changes in other state variables.

The results presented here strongly reinforce the recommendations for optimal measurement of the CAR presented by the recent expert consensus guidelines paper (Stalder et al., 2016), in particular, the recommendation wherever possible of using electronic-monitoring of both awakening time and the first (awakening) saliva collection time. Such recommendations have not been made lightly. However, the evidence presented here suggest they are essential in order to resolve current inconsistencies in the literature, and allow future research to shed much needed and clearer light on just how psychologically and neurologically important are the dramatic post-awakening changes in this endocrine measure. Of course, as research with more accurate measures hopefully accumulates, some flexibility and relaxation of stringency in CAR measurement may emerge, but that will only happen if meaningful empirical explorations are undertaken with access to a sufficiently large corpus of data drawn from samples of known timing.

Effects sizes alongside descriptive statistics for all delay effects have been presented. It is noticeable that they are relatively small effect sizes. Delay explains approximately 3 percent of variance in the magnitude of CAR measures. That compares with an approximately 5-10 percent range which is more typical of associations in domains such as psychophysiology, including associations between psychosocial variables and the CAR. However, that in no way diminishes the potential threat of erroneous measurement to the future of CAR research. Indeed, in the light of the extraordinarily small proportion of studies with electronic-monitoring identified by expert consensus guidelines authors (Stalder et al., 2016), it is somewhat difficult currently to estimate what average or range of effect size to assign to the general CAR literature in regard to psychological and neurological domains.

There are some potential limitations of the study that need to be mentioned. The results from this predominantly healthy young adult sample may not be generalizable to other age groups or clinical populations, where CAR dynamics may differ. The majority of the sample was female and although there were no obvious changes in reported effect sizes when gender was added as a covariate in the analyses, the possibility of modulating effects with a more equal gender balance should not be dismissed. It was not possible to verify electronic-monitoring of collection times for a minority of samples 2-4. However, to our knowledge this is the fullest data-set with electronic-monitoring in the post-awakening sampling period that exists to date, and we did examine and establish the general reliability across the data set of self-report in this minority of cases. One final limitation is worth emphasising since it sends a clear message about the need for wider sampling of cortisol values

across the whole post-awakening period. As observed, overwhelmingly CAR study protocols have locked the research community into a few fixed points within the CAR period upon which the cortisol growth curve has been routinely constructed. Thus the data included a highly asymmetric distribution of delay, since as originally intended most data were not significantly delayed in terms of the interval between awakening and the collection of the first sample (=< 4-min). We have done our best to take this asymmetry into account in arriving at predicted values for the CAR and underpinning time points on the cortisol growth curve. However, we would expect better estimates and confidence intervals to be forthcoming from fuller within-participant cortisol sampling across the range of the whole CAR period.

4.1 Conclusion

The novel analyses reported here make an important additional contribution to the evidence reviewed in the recent Experts' Guidelines on assessment of the CAR. Real-time analyses of all samples with less than 15 min delay between verified awakening and initiation of saliva sampling are presented. The expected cortisol curve is evident regardless of accuracy to the desired sample protocol timings, meaning that protocol inaccurate data need not be excluded from analyses. It needs emphasis that we refer to cortisol samples in a time-series where the first post-awakening sample is delayed by no more than 15-min from awakening. Longer delays will not only miss much of the peak CAR rise period but later samples will also partly represent the second wave in the continuous underlying ultradian rhythm of cortisol secretion. For these reasons, as stated in the Methods, we excluded those very few cases where first sample delay exceeded 15-min from awakening. Using this data-set, predicted cortisol values and CAR measures for healthy young participants were generated. The predicted curvilinear effects of delay on CAR measures if protocol times are wrongly assumed accurate, were tested and confirmed, and estimates of the size of such effects are presented. The results highlight again the importance of electronic-monitoring of awakening time and collection of the awakening sample (S1) in order to obtain known sampling timings in the post-awakening period. Finally, we have presented recommendations to help resolve current inconsistencies in the CAR literature.

References

Blackwell, E., de Leon, C.F., Miller, G.E., 2006. Applying mixed regression models to the analysis of repeated-measures data in psychosomatic medicine. Psychosom. Med. 68, 870-878.

Boyne, K., Sherry, D.D., Gallagher, P.R., Olsen, M., Brooks, L.J., 2013. Accuracy of computer algorithms and the human eye in scoring actigraphy. Sleep Breath 17, 411-417.

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

Chida, Y., Steptoe, A., 2009. Cortisol awakening response and psychosocial factors: A systematic review and meta-analysis. Biol. Psychol. 80, 265-278.

Clow, A., Thorn, L., Evans, P., Hucklebridge, F., 2004. The awakening cortisol response: Methodological issues and significance. Stress 7, 29-37.

DeSantis, A.S., Adam, E.K., Mendelsohn, K.A., Doane, L.D., 2010. Concordance between self-reported and objective wakeup times in ambulatory salivary cortisol research. Int. J. Behav. Med. 17, 74-78.

Dockray, S., Bhattacharyya, M.R., Molloy, G.J., Steptoe, A., 2008. The cortisol awakening response in relation to objective and subjective measures of waking in the morning. Psychoneuroendocrinology 33, 77-82.

Ennis, G.E., Moffat, S.D., Hertzog, C., 2016. The cortisol awakening response and cognition across the adult lifespan. Brain Cogn. 105, 66-77.

Evans, P., Clow, A., Hucklebridge, F., Loveday, C., 2012. The cortisol awakening response is related to executive function in older age. Int. J. Psychophysiol. 84, 201-204.

Evans, P., Forte, D., Jacobs, C., Fredhoi, C., Aitchison, E., Hucklebridge, F., Clow, A., 2007. Cortisol secretory activity in older people in relation to positive and negative well-being. Psychoneuroendocrinology 32, 922-930.

Fries, E., Dettenborn, L., Kirschbaum, C., 2009. The cortisol awakening response (CAR): facts and future directions. Int. J. Psychophysiol. 72, 67-73.

Golden, S.H., Sanchez, B.N., DeSantis, A.S., Wu, M.H., Castro, C., Seeman, T.E., Tadros, S., Shrager, S., Roux, A.V.D., 2014. Salivary cortisol protocol adherence and reliability by socio-demographic features: The Multi-Ethnic Study of Atherosclerosis. Psychoneuroendocrinology 43, 30-40.

Griefahn, B., Robens, S., 2011. Cortisol awakening response--Sampling delays of 15 minutes are not acceptable. Int. J. Psychophysiol. 82, 202-205.

Juster, R.-P., Sindi, S., Marin, M.-F., Perna, A., Hashemi, A., Pruessner, J.C., Lupien, S.J., 2011. A clinical allostatic load index is associated with burnout symptoms and hypocortisolemic profiles in healthy workers. Psychoneuroendocrinology 36, 797-805.

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

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

Kupper, N., de Geus, E.J.C., van den Berg, M., Kirschbaum, C., Boomsma, D.I., Willemsen, G., 2005. Familial influences on basal salivary cortisol in an adult population. Psychoneuroendocrinology 30, 857-868.

Lovell, B., Moss, M., Wetherell, M.A., 2011. Perceived stress, common health complaints and diurnal patterns of cortisol secretion in young, otherwise healthy individuals. Horm. Behav. 60, 301-305.

Okun, M.L., Krafty, R.T., Buysse, D.J., Monk, T.H., Reynolds III, C.F., Begley, A., Hall, M., 2010. What constitutes too long of a delay? Determining the cortisol awakening response (CAR) using self-report and PSG-assessed wake time. Psychoneuroendocrinology 35, 460-468.

Ramachandran, N., Smyth, N., Thorn, L., Eardley, A., Evans, P., Clow, A., 2016. Relationship between post-awakening salivary cortisol and melatonin secretion in healthy participants. Stress 19, 260-263.

Smyth, N., Clow, A., Thorn, L., Hucklebridge, F., Evans, P., 2013a. Delays of 5–15min between awakening and the start of saliva sampling matter in assessment of the cortisol awakening response. Psychoneuroendocrinology 38, 1476–1483.
Smyth, N., Hucklebridge, F., Thorn, L., Evans, P., Clow, A., 2013b. Salivary cortisol as a biomarker in social science research. Soc. Personal. Psychol. Compass 7, 605-625.
Smyth, N., Thorn, L., Hucklebridge, F., Evans, P., Clow, A., 2015a. Detailed time course of the cortisol awakening response in healthy participants. Psychoneuroendocrinology 62, 200-203.
Smyth, N., Thorn, L., Hucklebridge, F., Evans, P., Clow, A., 2015b. Post awakening salivary cortisol secretion and trait well-being: the importance of sample timing accuracy. Psychoneuroendocrinology 58, 141-151.
Stalder, T., Evans, P., Hucklebridge, F., Clow, A., 2010a. Associations between psychosocial state variables and the cortisol awakening response in a single case study. Psychoneuroendocrinology 35, 209-214.
Stalder, T., Evans, P., Hucklebridge, F., Clow, A., 2010b. State associations with the cortisol awakening response in healthy females. Psychoneuroendocrinology 35, 1245-1252.
Stalder, T., Hucklebridge, F., Evans, P., Clow, A., 2009. Use of a single case study design to examine state variation in the cortisol awakening response: relationship with time of awakening. Psychoneuroendocrinology 34, 607-614.
Stalder, T., Kirschbaum, C., Kudielka, B.M., Adam, E.K., Pruessner, J.C., Wuest, S., Dockray, S., Smyth, N., Evans, P., Hellhammer, D.H., Miller, R., Wetherell, M.A., Lupien, S.J., Clow, A., 2016. Assessment of the cortisol awakening response: Expert consensus guidelines. Psychoneuroendocrinology 63, 414-432.
Steptoe, A., Gibson, E.L., Hamer, M., Wardle, J., 2007. Neuroendocrine and cardiovascular correlates of positive affect measured by ecological momentary assessment and by questionnaire. Psychoneuroendocrinology 32, 56-64.
Steptoe, A., O'Donnell, K., Badrick, E., Kumari, M., Marmot, M., 2008. Neuroendocrine and inflammatory factors associated with positive affect in healthy men and women: the Whitehall II study. Am. J. Epidemiol. 167, 96-102.

Figures

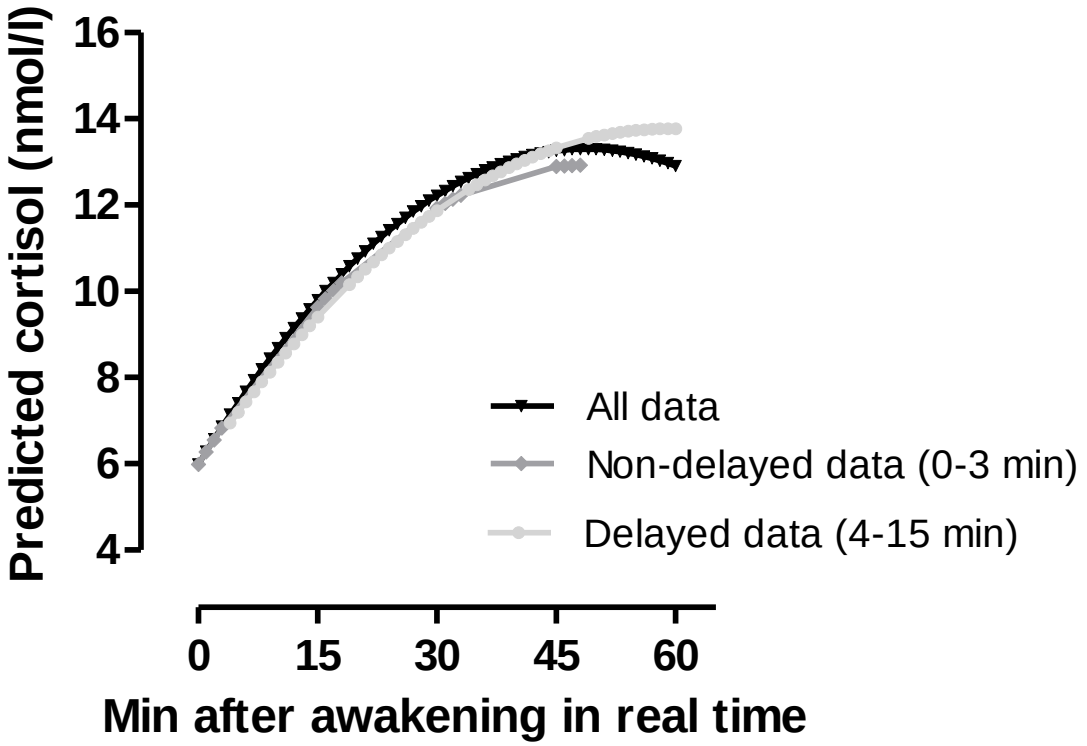

Figure 1. **Pattern of predicted cortisol over the post awakening period, showing close overlap of quadratic fit for non-delayed, delayed, and combined data.**

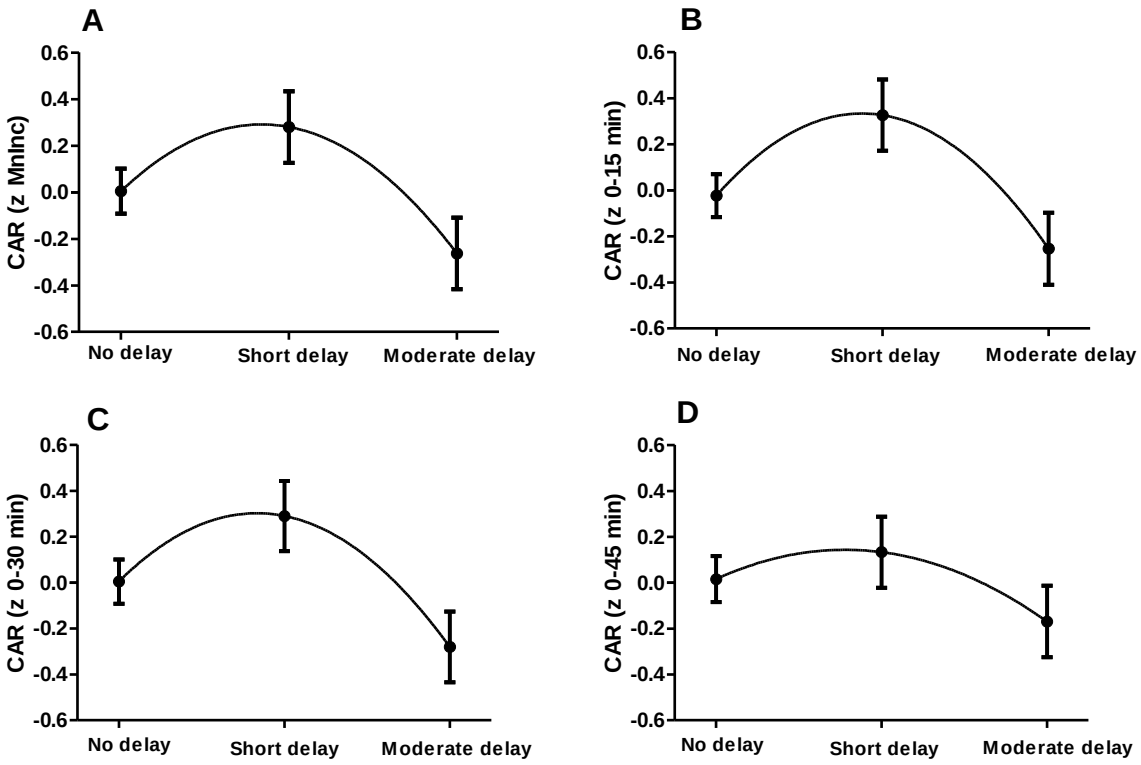

Figure 2. **Predicted values for CAR measures (a) MnInc, (b) delta 0-15 min (c) delta 0-30 min and (d) delta 0-45 min post-awakening for delay groups.**

Tables

Table 1. Descriptives for Delay Categories

Delay Intervals	N Days	Median average delay (min. from awakening)	Cut-off-range (minutes from awakening)
≈ None	127	1	0-3
Short	45	5	4-6
Moderate	43	9	7-15

Table 2. Predicted (S.E) values and 95% confidence intervals of cortisol values (0, 15, 30 and 45-min post-awakening) and CAR composites (MnInc, and Delta 0-30-min).

	Cortisol (nmols/l)			
	Predicted		95% CIs	
	Value	S.E	Lower	Upper
Awakening Level	5.98	.48	5.03	6.94
15-min	9.80	.42	8.97	10.63
30-min	12.22	.43	11.37	13.08
45-min	13.26	.43	12.42	14.09
CAR (MnInc)	5.78	.95	3.92	7.64
CAR (0-30min)	6.24	.95	4.38	8.10

Table 3. Modelling delay effects on CAR MnInc and components.

CAR measures	*Delay*	Mean Z	Standard Error	Curvilinear Trend Effect (% explained variance)
MnInc	≈ None	-.004	.097	2.6
	Short	.280	.154	
	Moderate	-.263	.154	
0-15min	≈ None	-.024	.093	3.2
	Short	.326	.154	
	Moderate	-.254	.156	
0-30min	≈ None	-.004	.096	3.0
	Short	.290	.153	
	Moderate	-.281	.154	
0-45min	≈ None	.015	.100	1.0
	Short	.133	.155	
	Moderate	-.170	.156	

Delay categories: ≈ None = 0-3 min, short = 4-6 min, moderate = 7-15 min delay between awakening and collection of sample 1.