

# Project 0

Jacob Kebe

## Introduction

This analysis concerns a dataset of cortisol and DHEA levels measured over the course of the day using a saliva capturing method. Participants took samples over the course of 3 days, and were instructed to collect samples at waking, before lunch, and 30 and 600 minutes after waking. We have data on Subject IDs, collection date and sample number, cortisol and DHEA (nmol/L), sample time and wake time recorded by participant in booklet, and sample time recorded by an electronic cap in the saliva collection interface. We are interested in testing 3 hypotheses. First we are interested in the agreement between times recorded in the booklet and times recorded by the electronic cap, and whether there is a bias in the booklet times. Statistically, we are interested in whether the slope of the relationship between the times is significantly different than 1, and whether the mean difference is significantly different than 0. We are also interested in what proportion of samples fall within a 15 minute and 30 minute window around the 30 minutes and 600 minutes since waking marks. Finally, we want to study the change in cortisol and DHEA over time. The clinical hypothesis is that cortisol and DHEA levels spike and peak 30 minutes after waking and decrease for the rest of

the day. We are interested in whether the slopes of the relationships of cortisol vs. time and DHEA vs. time are significantly different before and after the 30 minute mark.

## Methods

Data was processed in the following ways. Variables for minutes since waking recorded by booklet and cap were calculated using the provided times ('minWakeBook' and 'minWakeMEM', respectively). Data input as an empty string, and cortisol and DHEA outliers (anything greater than or equal to 80 for cortisol or 5.205 for DHEA in nmol/L) were all replaced by NAs in order to properly call functions necessary for the analysis. Missing values were excluded from analysis, assuming missing completely at random. One subject was removed for having multiple measured outliers. A second dataset only including observations for collection samples 2 and 4 (30 min and 600 min since waking) was created to answer the second research question.

To analyze the agreement between booklet times and cap times, we conducted linear mixed effect regression with cap times as a function of booklet times, accounting for a random intercept for each subject. Using a mixed effect model accounted for the within subject variation due to each subject having measurements over several days. We tested both the intercept and slope of the model as evidence of bias and agreement (or lack thereof) in the model.

We calculated the proportion of samples whose recorded cap time was within the 15 minute window and 30 minute window around the ideal 30 and 600 minutes since waking. We also

calculated this proportion for the booklet times.

For the third research question, linear models of cortisol and DHEA as a function of minutes since waking were proposed. Violation of linear regression assumptions was assessed since cortisol and DHEA are nonnegative right-skewed variables. We found that both residual plots and quantile plots were more indicative of adherence to assumptions when cortisol was log-transformed. Similarly, we log transformed DHEA. For cortisol, we used a piecewise linear mixed effect regression model with a knot at 30 minutes, allowing for the slope of the relationship between log cortisol and time to differ before and after the 30 minute mark, and adjusting for the repeated measures for each subject with a random intercept. We calculated minutes since waking for the model using booklet times. Less booklet times were missing than cap times, and since wake time is human measured, we thought it more consistent to calculate minutes awake using another human measured variable. We were interested in the coefficients for the time variable before and after 30 minutes. We created a similar model for DHEA using log DHEA as the outcome, but keeping the same knot and including random intercepts. For the linear mixed effect models we tested individual coefficients using the t-test from the lmerTest package using Satterthwaite's approximation for degrees of freedom. We assumed an  $\alpha$  level of .05 for all statistical tests.

## Results

For the model answering the first research question, the average minutes since waking recorded by the cap is 10.17 (95% CI: [4.53,15.81]) when the minutes recorded by booklet is 0, which is significantly different from 0 ( $p=.0008$ ), implying that booklet times tend to be earlier than

cap times, evidence of bias in the booklet measures. For every 1 minute increase recorded by book, the minutes recorded by cap increase by .9910 (95% CI:[.9771,1.004]). These measures are significantly positively correlated ( $p < 2E-16$ ) and the slope of the relationship is not significantly different from 1, based on Wald-type CIs, implying there is not evidence to reject agreement between the cap times and book times.

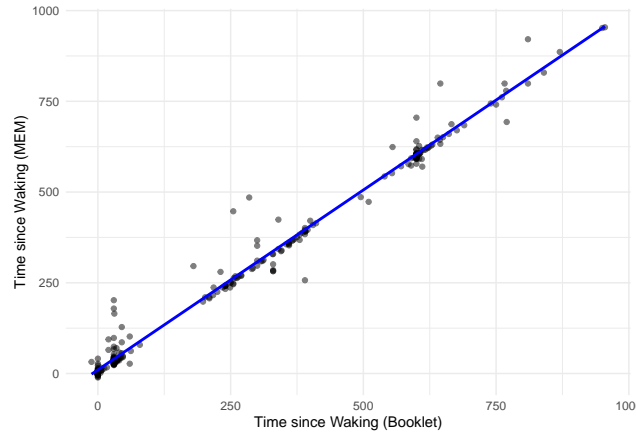


Figure 1: Regression of booklet minutes vs. MEM minutes

For the second research question addressing proportions, 62.9% of times recorded in the booklet were within 7.5 minutes of the ideal time, and 73.7% were within 15. For cap times, 41.8% of recorded times were within 7.5 minutes of the ideal time, and 54.2% were within 15.

Table 1: Adherence Proportions For Recorded Observations

	Proportion within 7.5 minutes	Proportion within 15 minutes
Booklet Recorded Times	0.629	0.737
Cap Recorded Times	0.418	0.542

For the third research question, our model showed differing trends for cortisol before and after 30 minutes post-waking. For the first waking 30 minutes, average cortisol increased by

.61% (95% CI:[-.10%,1.34%]) for each additional waking minute, which was not a significant relationship ( $p=.0936$ ) so we cannot reject the null hypothesis that time does not affect cortisol levels in the first 30 minutes. After the 30 minute mark, average cortisol levels were observed to decrease by .22% (95% CI: [.19%,.26%]) which was a significant decrease ( $p=.0262$ ), providing evidence that after the 30 minute mark, cortisol levels decrease throughout the day.

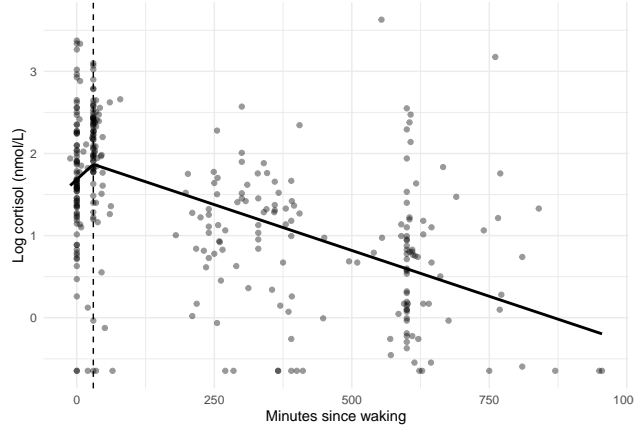


Figure 2: Log Cortisol as a function of time, with knot at 30 minutes

For average DHEA, levels decreased by 1.76% (95% CI:[1.17%,2.35%]) for each additional waking minute until minute 30, which was a significant trend ( $p=1.58E-8$ ). After waking minute 30, for each additional minute, average DHEA levels decreased by .15% (95% CI: [.12%,.18%]) which was a significant trend ( $p=4.19E-7$ ), which provides evidence that DHEA levels decrease throughout the day.

## Conclusion

We provided evidence that booklet times may be biased, but are significantly positively associated with with MEM times. We also showed that adherence of booklet times was generally higher than MEM times, showing a discrepancy between human-made and electronic

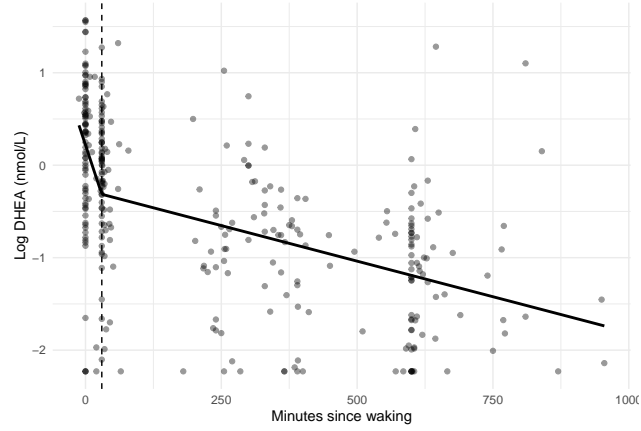


Figure 3: Log DHEA as a function of time, with knot at 30 minutes

measures of recording time. We failed to provide statistical evidence of the proposed cortisol trend before and after 30 minutes, and gave evidence that DHEA levels decrease throughout the day, more sharply in the first 30 minutes.

Table 2: Table of descriptive statistics by collection sample.

	1	2	3	4	Overall
	(N=93)	(N=93)	(N=93)	(N=93)	(N=372)
<b>Cortisol (nmol/L)</b>					
Mean (SD)	7.38 (6.05)	9.00 (4.94)	3.21 (2.29)	3.20 (4.98)	5.72 (5.40)
Median [Min, Max]	5.35 [0.524, 29.2]	8.84 [0.524, 25.5]	2.81 [0.524, 13.1]	1.89 [0.524, 37.7]	4.14 [0.524, 37.7]
Missing	0 (0%)	1 (1.1%)	2 (2.2%)	3 (3.2%)	6 (1.6%)
<b>DHEA (nmol/L)</b>					
Mean (SD)	1.63 (1.06)	1.02 (0.687)	0.505 (0.461)	0.503 (0.649)	0.911 (0.875)
Median [Min, Max]	1.54 [0.108, 4.80]	0.930 [0.108, 3.75]	0.404 [0.108, 2.78]	0.325 [0.108, 3.60]	0.594 [0.108, 4.80]
Missing	4 (4.3%)	3 (3.2%)	2 (2.2%)	2 (2.2%)	11 (3.0%)
<b>Missing Booklet Time</b>					
Yes	5 (5.4%)	6 (6.5%)	11 (11.8%)	13 (14.0%)	35 (9.4%)
No	88 (94.6%)	87 (93.5%)	82 (88.2%)	80 (86.0%)	337 (90.6%)
<b>Missing MEM Time</b>					
Yes	16 (17.2%)	23 (24.7%)	12 (12.9%)	10 (10.8%)	61 (16.4%)
No	77 (82.8%)	70 (75.3%)	81 (87.1%)	83 (89.2%)	311 (83.6%)
<b>Booklet vs. MEM discrepancy</b>					
Mean (SD)	-4.00 (8.75)	-16.0 (35.9)	-6.18 (42.7)	-5.40 (29.9)	-7.71 (32.1)
Median [Min, Max]	-2.00 [-44.0, 11.0]	-1.00 [-172, 33.0]	-1.00 [-200, 133]	0 [-154, 77.0]	-1.00 [-200, 133]
Missing	21 (22.6%)	27 (29.0%)	19 (20.4%)	20 (21.5%)	87 (23.4%)