Project 0 Report

Analyst: Max McGrath

Investigator: Nichole Carlson

Report generated: 09/15/2021

Introduction

This analysis evaluates the viability of a novel, take-home collection device used to measure salivary cortisol and DHEA levels throughout the day and examines DHEA and cortisol level time trends. The device was provided to 31 healthy control subjects who were asked to collect saliva samples immediately upon waking, 30 minutes after waking, before lunch, and 600 minutes after waking. The time of waking and the time of lunch were left to the discretion of each subject. Each subject was asked to collect four samples per day for three days, providing a total of 372 observations.

The samples were collected using a filter paper device termed the Saliva Procurement and Integrated Testing (SPIT) booklet which was kept in a bottle with an electronic monitoring cap (MEM). Subjects were asked to manually record the time the sample was taken, and the electronic monitoring cap also recorded the timing of the sample. From each sample, cortisol and DHEA levels were measured in nanomoles per liter.

This analysis has three aims: (1) evaluate agreement between the subject's recordings of sampling times and those recorded by the MEM, (2) evaluate subject adherence to specified sample timing, and (3) quantify changes of cortisol and DHEA over time.

As such, two hypotheses will be tested: (1) Subject's booklet recordings and the times recorded by the MEM are significantly correlated, and (2) DHEA and cortisol levels are positively correlated with time until 30 minutes after waking and then negatively correlated with time for the remainder of the day. Further, conclusions will be drawn regarding subject adherence based on the findings of the analyses.

Methods

Data Preperation

In general, the data provided was filtered for missing and erroneous values in accordance with the needs of each section of analysis. When evaluating correlation between subjects'

booklet recorded times and MEM times, 87 missing time records were removed (26 with missing booklet records, 52 with missing MEM records, and 9 with both). When comparing strict differences between record types, only records with both values present were examined. Record missingness is further discussed when examining subject protocol compliance.

When examining DHEA levels and cortisol levels, all observation with missing booklet time records were removed. Further, 1 observation with cortesol levels greater than 40 nanomoles per liter was removed, and 6 observations with DHEA levels equal to 5.205 nanomoles per liter were removed, as those measurements were deemed to be errors caused instrumentation limitations.

Analysis

To evaluate agreement between the subjects' booklet records sample times and those recorded by the MEM, a linear model was fit with MEM time as the response variable and booklet time as the predictor. Raw times were used for this section, as minutes since waking (which is used for the other sections of this analysis), yielded a fit with obviously correlated residuals. The slope estimate of this model was then examined to quantify correlation between the times, and a t-test was conducted to evaluate the estimate's statistical significance. The mean difference between times recorded by MEM and those recorded in the booklets was also calculated. A plot comparing booklet times and MEM times is also provided.

To analyze protocol adherence, booklet records and MEM records were examined seperately through summary statistics identifying proportions of missingness and proportions of observations within 7.5 and 15 minutes of the 30 and 600 minute observation protocols. Average differences between protocol times and recorded times were also calculated for both record types.

Knowing the data consists of repeated observations from each subject, it is foreseeable that intrasubject DHEA and cortisol levels are correlated. So, to quantify changes in DHEA and cortisol over time, two linear mixed-effects models (one for each hormone) were fit

with intercepts being allowed to vary by subject, thereby accounting for this correlation. Further, given the known diurnal trend in which DHEA and cortisol levels peak 30 minutes after waking and then taper throughout the day, the fixed effects of the linear models were segmented into piecewise components wherein the linear trends were allowed to differ before and after 30 minutes after waking. For each model, time after waking was used as the sole predictor, and the natural log of the measurement of DHEA and cortisol levels were used as responses for each hormone's respective model.

Results

The linear model used to test the hypothesis that subject's booklet recordings and the times recorded by the MEM are significantly correlated had an estimated slope coefficient of 0.9938 with a p-value less than .001. A plot of the MEM recorded times versus the subject recorded times is included in the appendix in Figure 1. On average, the MEM recorded times are 7.71 minutes later than those recorded in the book. A plot of the difference between MEM and booklet records is provided in Figure 2.

Of the 93 observations for 30 minutes after waking, 33 MEM records and 6 booklet records were missing. Of those that were not missing, 75.0% of MEM records and 79.3% of booklet records were within 7.5 minutes of subjects' wake-time records. Further, 86.7% of MEM records and 95% of booklet records are within 15 minutes of the 30 minute protocol. On average, MEM records were 11.3 minutes after the 30 minute protocol, and booklet records were 3.49 minutes after the 30 minute protocol.

Of the 93 observations for the 600 minute protocol, 21 MEM records and 16 booklet records were missing. Of those that were not missing, 29.2% of MEM records and 42.9% of booklet records were within 7.5 minutes of subjects' wake-time records. 45.8% of MEM records and 53.2% of booklet records were within 15 minutes of the 600 minute protocol. On average, MEM records and book records differed from the 600 minute protocol by 54.1 and 42.1 minutes respectively. A box-plot for protocol adherence by record type is provided in

Figure 3.

Given greater number of complete observations and the seemingly superior protocol-adherence of booklet records, booklet records were used for the models analyzing DHEA and cortisol. Specifically, booklet recorded minutes after waking were used as the predictor. Records with missing booklet times were discarded. Additionally, to meet error assumptions for linear models, the natural log of DHEA and cortisol levels in nanomoles per Liter were used as the responses for their respective models.

For the DHEA model, random intercept estimates for the subjects ranged from -0.019 to 0.107 with a standard deviation of 0.530. The piecewise regression model used to analyze DHEA involves the estimation of two regression coefficients. The first estimates the slope of the data prior to the break at 30 minutes, and it was estimated to be -0.193 with a 95% confidence interval of (-0.026, -0.128). The second coefficient estimates the change in slope following the break and was estimated to be 0.018 with a 95% confidence interval of (0.011, 0.0244).

For the cortisol model, random intercept estimates for the subjects ranged from -0.019 to 0.107 with a standard deviation of 0.336. The estimate of the slope coefficient prior to 30 minutes after waking was 0.003829 with a 95% confidence interval of (-0.004, 0.011), and the estimate of the change in slope from before to after 30 minutes was -0.005913 with a 95% confidence interval of (-0.014, 0.001). Plots of both models for all estimated intercepts are provided in Figures 4 and 5.

Conclusions

Regarding, concordance between between booklet and MEM sample time records, we conclude that there is strong agreement between the two based on the magnitude of the regression slope estimate and its significance. However, on average MEM records were later than booklet records. This is seen again when looking at MEM and booklet record adherence to study protocol, where in both the 30 and 600 minutes after waking protocols booklet records were,

on average and in proportion, closer to the specified times.

Regarding time patterns of the hormone measurements, DHEA was shown to be negatively associated with time before 30 minutes after waking, with a reduction in the magnitude of negative relationship after 30 minutes, the former of which is contradictory to known biological rhythms. The model for cortisol did not have any statistically significant coefficients, so, once again, the known biological rhythms of the hormone were not significantly captured by the model with the provided data.

While many of the measurements of DHEA and cortisol were within the pre-specified time boundaries, there exist limitations to that analysis which, coupled with the findings of the linear models, question subject adherence. Specifically, all sample timing adherences were based on subjects reported wake times and the corresponding first daily sample. Should a subject fail to adhere to the first-sample protocol by failing to record a measurement immediately upon waking, it would bias any attempts to measure adherence for the remaining samples. Further, the measured hormone levels' failed to reflect known biological patterns in the linear models. While that failure may be caused by lack of power or poor choice of model, it is reasonable to suspect that the results may have been biased by lack of subject adherence to study protocol, thereby questioning the viability of the sample collection method used in this study.

Limitations

As discussed above, there is notable missingness throughout the data with regard to booklet and MEM sample time records, which may impact results. No missingness was imputed for any section, but, given the nature of the study in which there was an attempt to record each sample time twice using two different methods, an imputation strategy to increase the amount of data on sample timing seems viable.

Reproducibility

The code used to generate this analysis is available on GitHub at https://github.com/BIOS6624-UCD/bios6624-MaxMcGrath/tree/main/Project0 . The `Background` folder contains information pertinent to understanding the analysis but unnecessary for reproducing it. The `Code` folder contains five files: `1-DataPrep.R`, `2-EDA.R`, `3-Analysis-Aim1.R`, `4-Analysis-Aim2.R`, and `5-Analysis-Aim3.R`. These R scripts are dependent upon a data file `Data/Project0_Clean_v2.csv` which is not available on GitHub, but may be requested by emailing `max.mcgrath@ucdenver.edu`. To run the complete analysis, each script should be run in the order of the number preceding its filename. The last directory, `Report`, contains the RMarkdown file `report.Rmd` which may be used to generate this report (note that it also depends on the aforementioned data).

This analysis was performed using R version 4.0.4. The following non-base R packages were used: lme4 (1.1.26), dplyr (1.0.5), ggplot2 (3.3.3), stringr (1.4.0), hms (1.0.0), car (3.0-10), and tidyverse (1.3.0).

Appendix

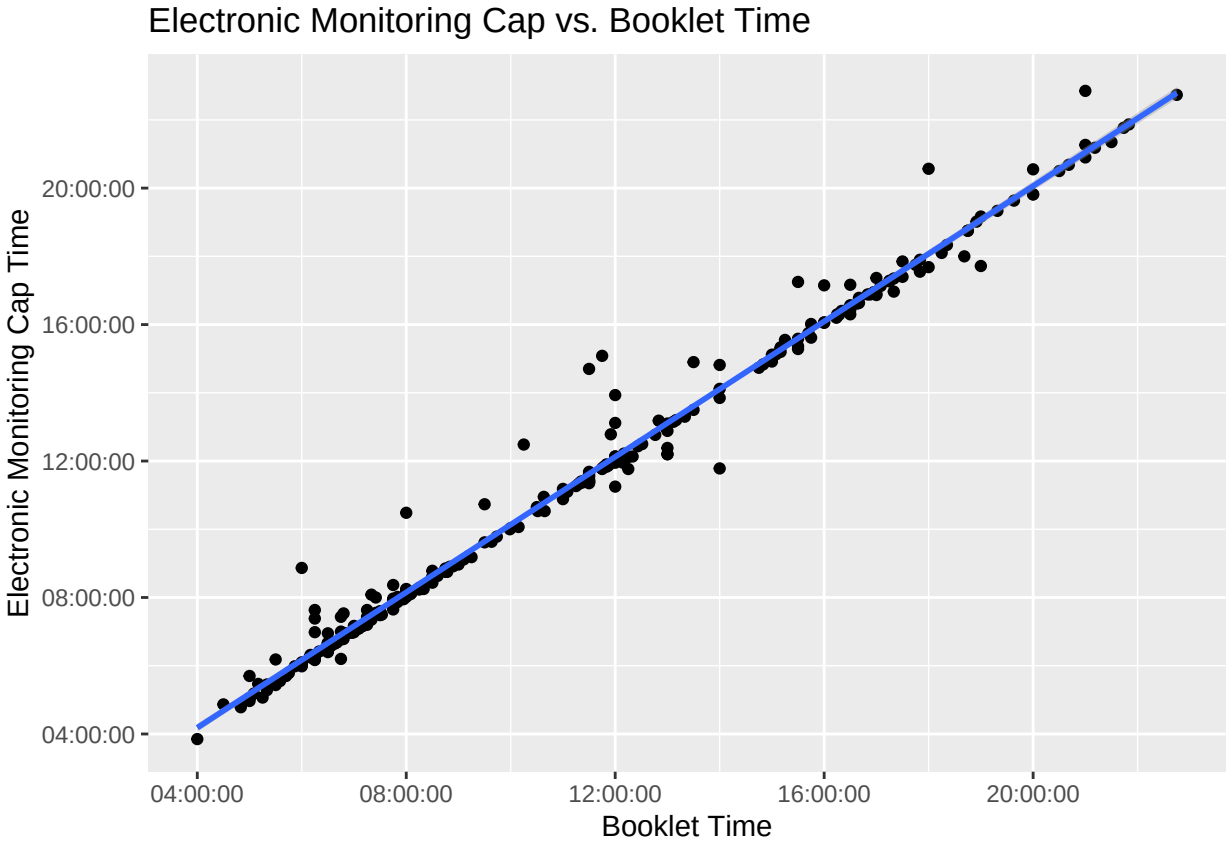

Figure 1: Plot comparing recorded times from electronic monitoring cap to those recorded by subjects in booklet. Blue line corresponds to the regression estimates described in Results section.

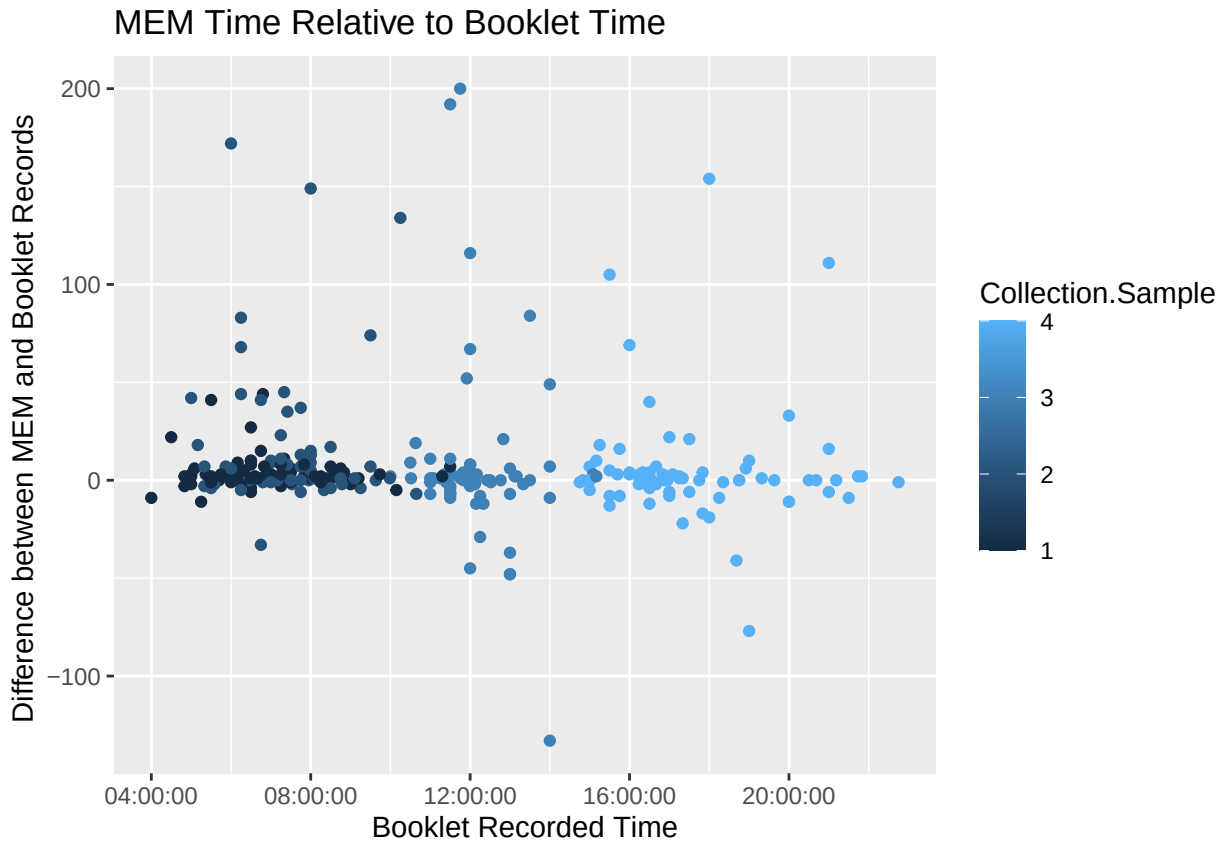

Figure 2: Plot comparing MEM time relative to booklet time. Observations above 0 on the y-axis represent observations where the MEM time was later than the booklet time, with observations below 0 on the y-axis are observations where the recorded MEM time preceded the recorded booklet time.

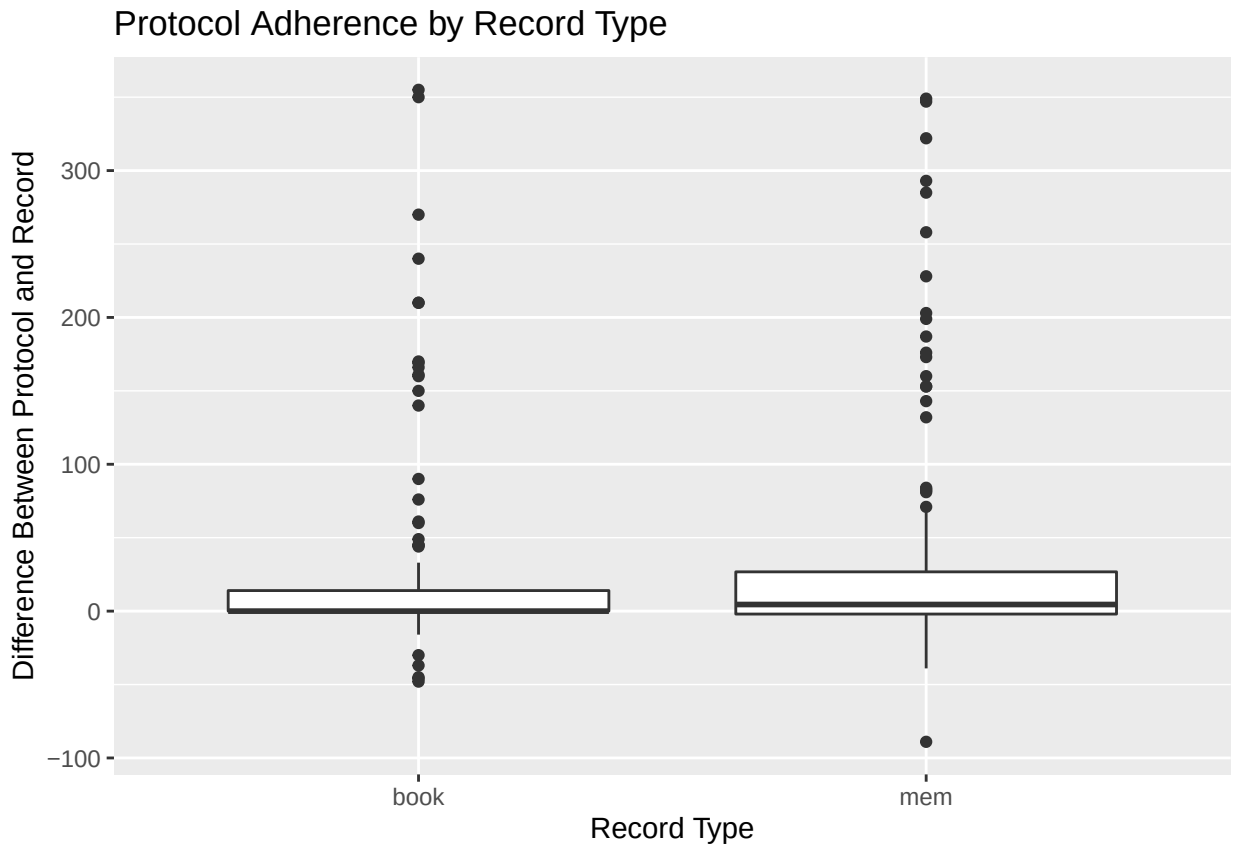

Figure 3: Plot comparing protocol adherence by record type. The left box-plot shows booklet records, with observations above 0 being records that were recorded after the time specified by study protocol. The right box-plot is the same but for MEM records.

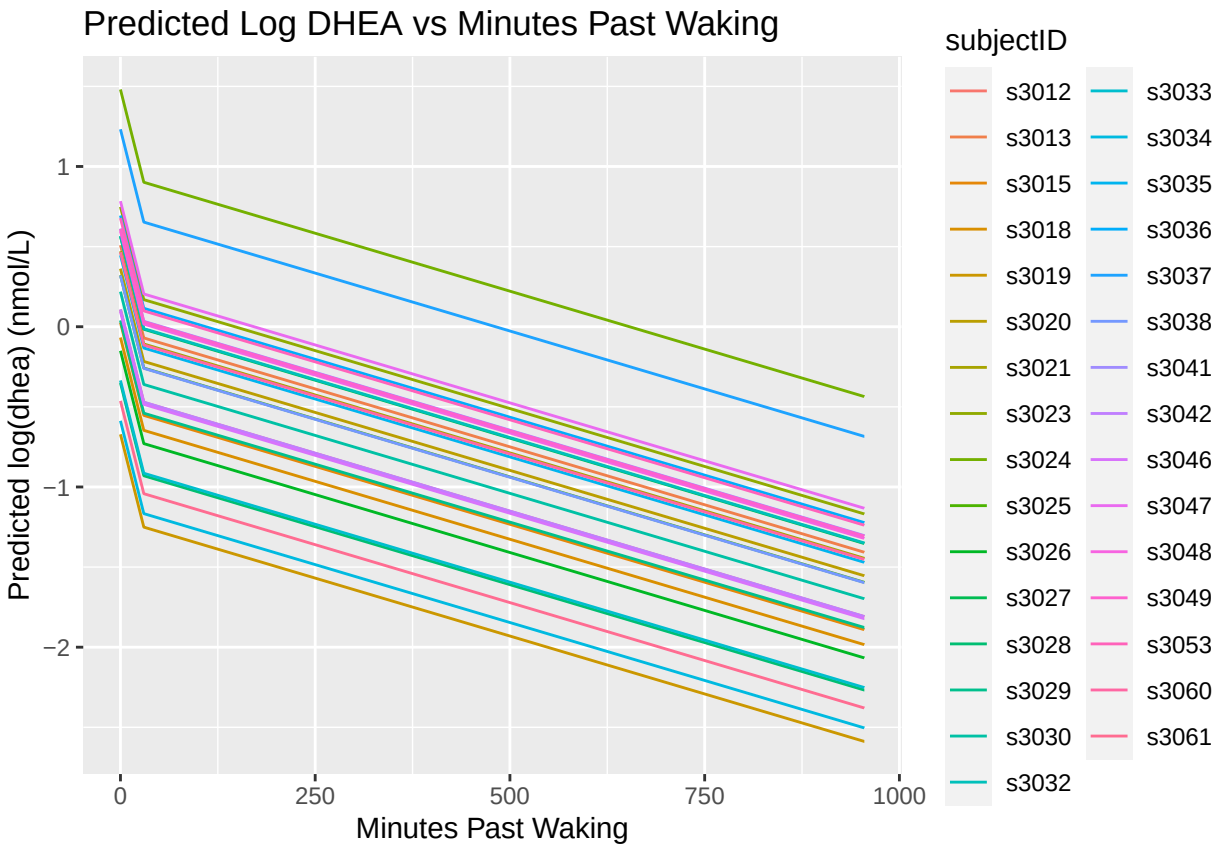

Figure 4: Predicted values for log(DHEA) for each subject's random intercept.

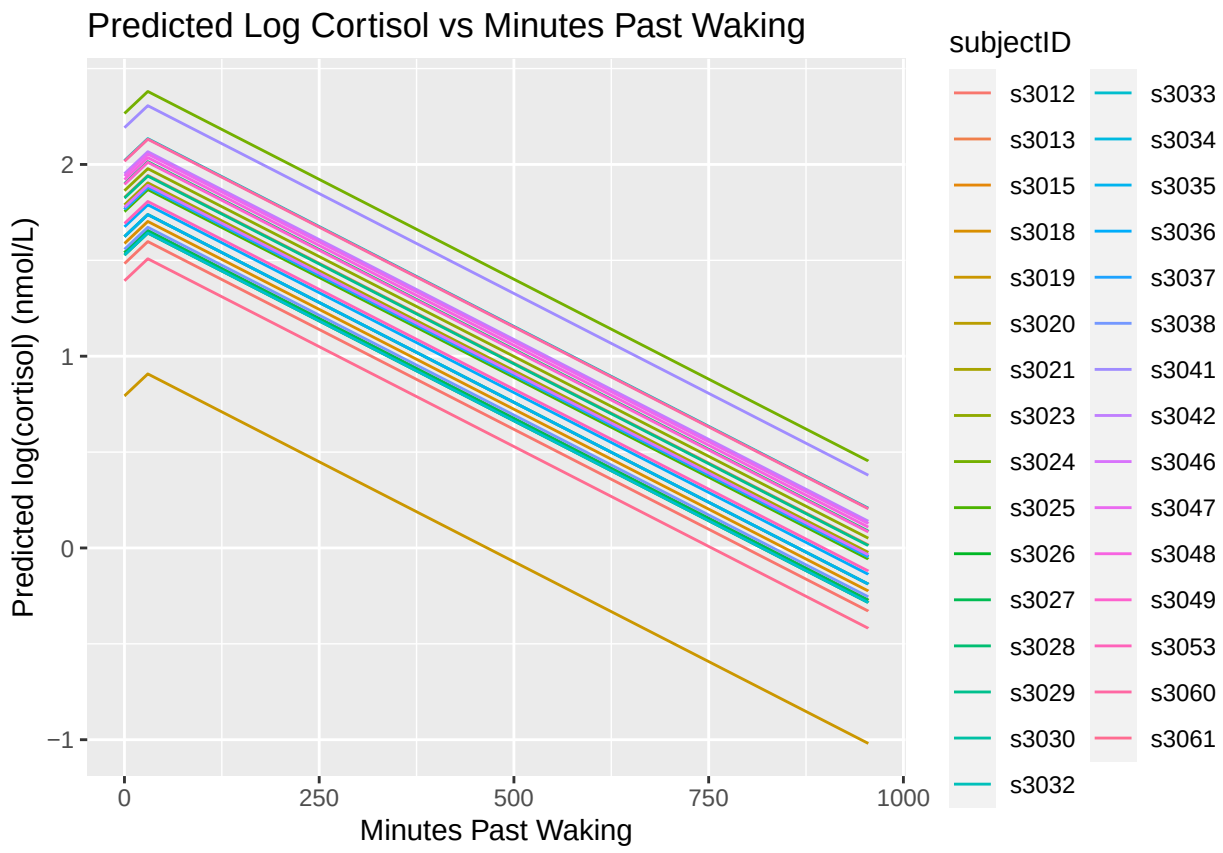

Figure 5: Predicted values for log(cortisol) for each subject's random intercept.