

Feasibility and Timing Accuracy of a Novel Filter Paper–Based Saliva Collection Method for Assessing Diurnal Cortisol and DHEA

Mahfuza Haque Mahi¹

Introduction

This project analyzes data from a feasibility study evaluating a novel saliva collection device, the Saliva Procurement and Integrated Testing (SPIT) booklet, for measuring diurnal patterns of salivary cortisol and dehydroepiandrosterone (DHEA) in free-living conditions. The dataset provided by the investigator includes repeated saliva samples from 31 healthy adult participants collected over three consecutive days, with up to four samples per day per participant. Samples were scheduled at waking, approximately 30 minutes after waking, before lunch, and approximately 10 hours after waking. For each sample, timing information was recorded using three sources: participant-recorded clock time in the booklet, electronically recorded clock time from a MEMS monitoring cap, and self-reported wake time from a sleep diary. Salivary cortisol and DHEA concentrations were measured in nmol/L, along with indicators of collection day and sample order.

The scientific objective of the study is to evaluate whether the SPIT booklet provides accurate sampling times and hormone measurements suitable for characterizing diurnal hormone patterns. This report focuses on three related analytic aims. First, agreement between booklet-recorded and MEMS-recorded sampling times is assessed. Second, adherence to the protocol-specified sampling schedule, particularly the +30-minute and +10-hour post-waking collections, is evaluated. Third, changes in cortisol and DHEA concentrations over time since waking are examined.

These objectives are translated into testable statistical hypotheses. Agreement is assessed by testing whether booklet-recorded minutes since waking are linearly related to MEMS-recorded minutes since waking with an intercept of zero and a slope of one. Protocol adherence is evaluated by estimating the proportion of samples collected within predefined ± 7.5 minute and ± 15 minute windows around scheduled targets. Diurnal hormone patterns are examined by testing for changes in cortisol and DHEA from waking to 30 minutes post-waking and for subsequent rates of change after 30 minutes post-waking using mixed-effects regression models that account for repeated measures within participants.

Methods

Data Management and Cleaning Data management and cleaning were performed prior to analysis in consultation with the investigator and followed prespecified quality-control criteria. Sample interval variables derived from booklet and MEMS recorded times were excluded because their accuracy could not be verified. Time since waking was instead calculated by combining the collection date with recorded clock times and subtracting the sleep diary-reported wake time.

Hormone measurements reported in alternative units ($\mu\text{g/dL}$ for cortisol and pg/dL for DHEA) were excluded to avoid redundancy. Cortisol values greater than 80 nmol/L were treated as biologically implausible laboratory values and set to missing, while values greater than 26 nmol/L were retained as biologically possible. DHEA measurements equal to the assay upper detection limit (5.205 nmol/L) were excluded because these represent censored observations that could obscure diurnal patterns. In addition, participants with more than two DHEA measurements at the detection limit across all samples were excluded from analyses examining diurnal hormone change, as advised by the investigator.

Missing timing data were retained for descriptive summaries. Missing MEMS times could occur due to device malfunction or participant noncompliance, whereas missing booklet or wake times could occur due to incomplete recording. Derived variables included minutes since waking based on booklet times, minutes since waking based on MEMS times, and the difference between booklet and MEMS recorded sampling times.

¹Biostatistics PhD Student, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO

Preliminary Data Analysis Preliminary analyses included descriptive summaries of sample counts, participant counts, and missing data patterns across collection times. Graphical methods were used to explore agreement between booklet and MEMS recorded times and to visualize hormone concentrations over time since waking. These exploratory analyses were used to assess data quality and guide model specification but were not used for formal inference.

Statistical Analysis

Agreement Between Booklet and MEMS Sampling Times Agreement between booklet-recorded and MEMS-recorded sampling times was evaluated among samples with non-missing time since waking values for both measures. Graphical agreement was assessed using scatter plots and Bland–Altman plots. Formal statistical assessment was conducted using a linear mixed-effects model with booklet-recorded minutes since waking as the outcome and MEMS-recorded minutes since waking as the primary explanatory variable. A random intercept for participant was included to account for within-subject correlation due to repeated measurements.

Statistical hypotheses were evaluated by testing fixed-effect parameters from the mixed-effects model. Perfect agreement was defined as an intercept equal to zero and a slope equal to one. Wald-based hypothesis tests and confidence intervals were used to assess whether the intercept differed from zero and whether the slope differed from one. Statistical significance was evaluated at the 0.05 level.

Protocol Adherence to Scheduled Sampling Times Protocol adherence was evaluated for the +30 minute and +10 hour post-waking samples. Adherence was defined as collection within predefined ± 7.5 -minute and ± 15 -minute windows around the scheduled target times, using sleep diary reported wake time as the reference. Adherence was assessed separately using booklet recorded times and MEMS recorded times. Device agreement was defined as the absolute difference between booklet and MEMS recorded sampling times. Adherence and agreement metrics were summarized descriptively, as the primary goal was to characterize feasibility rather than conduct formal hypothesis testing.

Diurnal Changes in Cortisol and DHEA Diurnal changes in cortisol and DHEA were examined using linear mixed-effects regression models. Hormone concentrations were log-transformed prior to analysis to address right-skewness and to model proportional changes over time. Model diagnostics, including histograms of raw and log-transformed hormone concentrations and residual assessments, are provided in Appendix A. Time since waking was modeled using a piecewise linear specification with a knot at 30 minutes after waking, allowing separate estimation of early post-waking change and subsequent diurnal change. Fixed effects included time before 30 minutes and time after 30 minutes, and a random intercept for participant was included to account for repeated measurements.

Hypotheses regarding early post-waking change and post-30 minute rates of change were evaluated using Wald tests of the corresponding fixed-effect coefficients. Estimates and 95% confidence intervals from log-scale models were back-transformed to obtain percent changes. Statistical significance was assessed at the 0.05 level without adjustment for multiple comparisons, as analyses were prespecified and limited in scope.

Statistical Software All analyses were conducted using R version 4.3.1 (2023-06-16 ucrt).

Results

Descriptive statistics (Table 1)

Table 1 below summarizes the availability of timing data and the distribution of cortisol and DHEA measurements across scheduled collection time points and overall.

Table 1: Table 1. Hormone measurements and timing data availability by collection time point.

	Waking	+30 min	Lunch	+10 hours	Overall
	(N=90)	(N=90)	(N=90)	(N=90)	(N=360)
Booklet time recorded					
Yes	85 (94.4%)	84 (93.3%)	79 (87.8%)	77 (85.6%)	325 (90.6%)
No	5 (5.6%)	6 (6.7%)	11 (12.2%)	13 (14.4%)	35 (9.4%)
MEMS cap time recorded					
Yes	74 (82.2%)	68 (75.6%)	78 (86.7%)	80 (88.9%)	300 (83.3%)
No	16 (17.8%)	22 (24.4%)	12 (13.3%)	10 (11.1%)	60 (16.7%)
Sleep diary wake time recorded					
Yes	90 (100%)	0 (0%)	0 (0%)	0 (0%)	90 (25.0%)
No	0 (0%)	90 (100%)	90 (100%)	90 (100%)	270 (75.0%)
Cortisol (nmol/L)					
Mean (SD)	7.30 (6.04)	9.12 (4.93)	3.22 (2.29)	2.98 (4.54)	5.68 (5.15)
Median [Min, Max]	5.34 [0.524, 29.2]	8.91 [0.524, 25.5]	2.86 [0.524, 13.1]	1.79 [0.524, 37.7]	4.11 [0.524, 29.2]
Missing	0 (0%)	1 (1.1%)	2 (2.2%)	3 (3.3%)	6 (1.7%)
DHEA (nmol/L)					
Mean (SD)	1.63 (1.07)	1.03 (0.684)	0.481 (0.433)	0.445 (0.553)	0.897 (0.745)
Median [Min, Max]	1.49 [0.108, 4.80]	0.930 [0.108, 3.75]	0.398 [0.108, 2.78]	0.305 [0.108, 3.60]	0.577 [0.108, 4.80]
Missing	2 (2.2%)	1 (1.1%)	2 (2.2%)	2 (2.2%)	7 (1.9%)

Timing data were largely complete. Booklet recorded times were available for over 90% of samples overall, while MEMS recorded times were available for more than 80%, with slightly higher missingness at the +30 minute timepoint. Cortisol and DHEA concentrations varied across collection time points, with higher concentrations at waking and +30 minutes followed by lower levels at lunch and +10 hours, and hormone measurements were missing in fewer than 3% of samples across all timepoints.

Agreement Between Booklet-Recorded and MEMS-Recorded Sampling Times

Figure 1: Agreement between booklet and MEMS-recorded sampling times

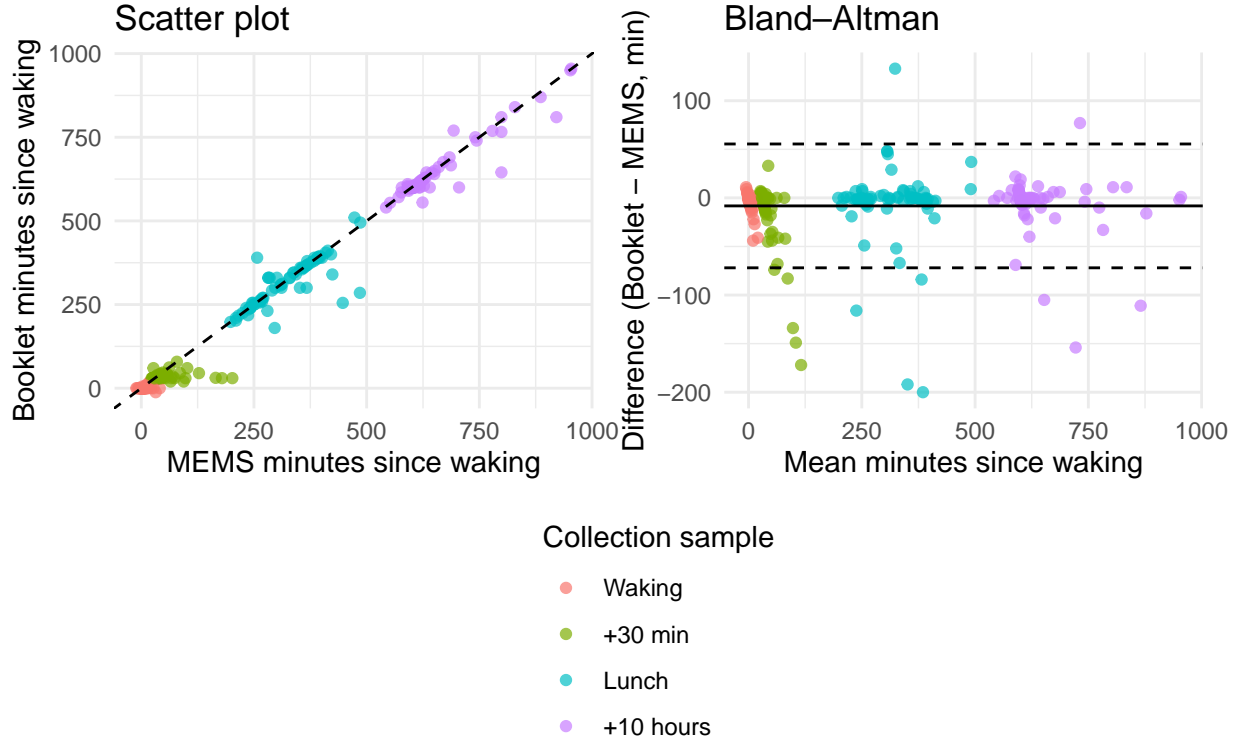


Figure 1 shows agreement between booklet recorded and MEMS recorded sampling times. In the scatter plot (Panel A), points lie close to the line of identity across all collection times, indicating strong proportional agreement between the two measures. Agreement appears tight near waking and the +30 minute sample, with slightly greater dispersion at later time points, particularly the +10 hour collection.

The Bland-Altman plot (Panel B) demonstrates minimal overall bias, with the mean difference close to zero, but increasing variability as time since waking increases. This pattern suggests that while booklet and MEMS times generally agree, discrepancies are more variable later in the day.

Together, these graphical assessments suggest strong proportional agreement between booklet and MEMS recorded times with minimal systematic bias, supporting the use of regression based methods to formally quantify agreement while accounting for repeated measurements within participants.

Guided by the strong linear relationship observed in Figure 1, agreement was formally assessed using a linear mixed-effects model with subject-specific random intercepts. Table 2 below presents results from a linear mixed-effects model evaluating the agreement between booklet recorded and MEMS recorded minutes since waking, accounting for repeated measurements within subjects.

Table 2: Table 2. Linear mixed-effects model assessing agreement between booklet and MEMS recorded sampling times.

Term	Estimate	Std. Error	95% CI (Lower)	95% CI (Upper)
Intercept	-6.703	2.972	-12.639	-0.768
MEMS minutes since waking	0.994	0.007	0.979	1.008

Note:

Outcome: booklet-recorded minutes since waking. Predictor: MEMS-recorded minutes since waking. Model includes a random intercept for each participant.

Booklet recorded sampling times were strongly associated with MEMS recorded times (slope = 0.99, 95% CI: 0.98-1.01; $p < 0.001$), indicating near proportional agreement between the two methods. The intercept differed modestly from zero (-6.7 minutes, 95% CI: -12.6 to -0.8), suggesting a small systematic tendency for booklet times to be recorded slightly earlier than MEMS times. Overall, these results support good agreement between devices with minimal bias at the population level.

Adherence to Scheduled Saliva Collection Times (+30 Minutes and +10 Hours)

Given the strong proportional agreement and minimal systematic bias between booklet and MEMS times, MEMS-recorded times were considered an appropriate reference for evaluating protocol adherence. This analysis focuses on whether participants collected samples at the required +30 minute and +10 hour time-points. Table 3 below summarizes device agreement between booklet and MEMS recorded sampling times and participant adherence to the scheduled +30 minute and +10 hour saliva collection windows.

Collection time	Device agreement (Booklet vs MEMS)				Booklet ± 7.5 (%)	Booklet ± 15 (%)	MEMS ± 7.5 (%)
	N	Missing	± 7.5 min (%)	± 15 min (%)			
Overall	274	86	67.5	82.1	—	—	—
+30 min	64	26	59.4	75.0	78.6	90.5	52.9
+10 hours	70	20	64.3	80.0	48.1	55.8	33.8

Note:

Device agreement is defined as the absolute difference between booklet- and MEMS-recorded sampling times. Protocol ad

Overall agreement between booklet- and MEMS-recorded sampling times was moderate, with 67.5% of non-missing samples within ± 7.5 minutes and 82.1% within ± 15 minutes.

Adherence to the study protocol varied by collection time and measurement source. For the +30-minute sample, adherence was generally high, with 78.6% of booklet-recorded times and 52.9% of MEMS-recorded times within ± 7.5 minutes of the scheduled target, increasing to 90.5% and 70.6%, respectively, within ± 15 minutes. In contrast, adherence declined substantially for the +10 hour sample, particularly when assessed using MEMS data, with only 33.8% of samples within ± 7.5 minutes and 41.2% within ± 15 minutes. These findings indicate that participants adhered more closely to the early post-waking collection than the late-day collection and that self-reported booklet times tended to overestimate adherence relative to electronically monitored times.

Diurnal Changes in Cortisol and DHEA Over Time Since Waking

Having established that booklet recorded times generally agree with MEMS times and that adherence to scheduled collections was reasonable, particularly in the early post-waking period, we next examined whether the SPIT device captured the expected diurnal patterns of cortisol and DHEA over the course of the day.

Figure 2: Diurnal patterns of cortisol and DHEA relative to waking

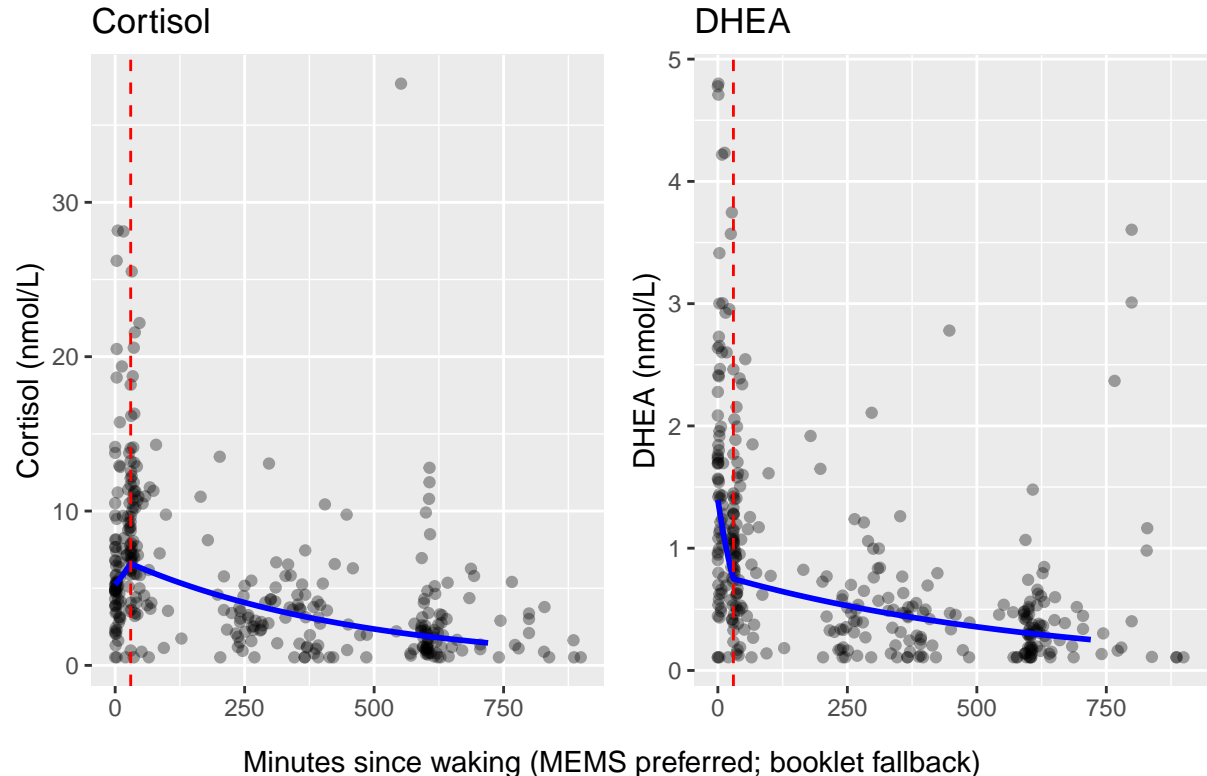


Figure 2 illustrates observed cortisol and DHEA concentrations over time since waking, with superimposed population-level model predictions. Cortisol concentrations show an early rise shortly after waking followed by a gradual decline throughout the day, whereas DHEA concentrations decline rapidly after waking and continue to decrease more slowly over time.

To formally quantify these observed patterns and estimate rates of change before and after 30 minutes post-waking, we fit piecewise linear mixed-effects models with a knot at 30 minutes after waking. Table 4 below presents the estimated percent changes in cortisol and DHEA levels over time since waking, including the early post-waking period and the subsequent diurnal decline.

Table 4: Table 4. Estimated percent changes in cortisol and DHEA levels over time since waking

Hormone	Time interval	Percent change (%)	95% CI
Cortisol	Waking → 30 min	25.9	(-2.3, 62.2)
Cortisol	After 30 min (per hour)	-12.3	(-14.1, -10.4)
DHEA	Waking → 30 min	-46.2	(-56.7, -33.3)
DHEA	After 30 min (per hour)	-9.0	(-10.6, -7.4)

Note:

Percent changes and confidence intervals were obtained by back-transforming log-scale mixed-effects model estimates. Post

Cortisol increased by an estimated 25.9% from waking to 30 minutes post-waking (95% CI: -2.3% to 62.2%; $p = 0.076$). After 30 minutes, cortisol declined significantly at a rate of 12.3% per hour (95% CI: -14.1% to -10.4%; $p < 0.001$).

In contrast, DHEA exhibited a marked decline immediately after waking, decreasing by 46.2% within the first 30 minutes (95% CI: -56.7% to -33.3%; $p < 0.001$), followed by a continued but slower decline of 9.0% per

hour thereafter (95% CI: -10.6% to -7.4%; $p < 0.001$). These results indicate distinct early-morning dynamics for cortisol and DHEA and support the ability of the SPIT method to capture biologically plausible diurnal hormone patterns.

Discussion

This analysis evaluated the feasibility and timing accuracy of the SPIT saliva collection method by assessing agreement between participant-recorded and electronically recorded sampling times, adherence to protocol-specified collection windows, and expected diurnal patterns of cortisol and DHEA. Together, these analyses provide insight into both the timing accuracy and biological validity of the SPIT method.

Booklet-recorded sampling times demonstrated strong proportional agreement with MEMS-recorded times, indicating that participants were generally able to record sampling times accurately using the SPIT booklet. Although a small negative intercept suggested a tendency for booklet times to be recorded slightly earlier than MEMS times, the magnitude of this bias was modest relative to the overall range of sampling times and is unlikely to be meaningful for characterizing diurnal hormone patterns.

Adherence to scheduled sampling windows was higher for the +30-minute post-waking collection than for the +10-hour collection, particularly when adherence was assessed using MEMS-recorded times. This pattern is consistent with greater challenges in maintaining protocol compliance later in the day under free-living conditions. The discrepancy between booklet-based and MEMS-based adherence estimates suggests that self-reported times may overestimate compliance, highlighting the value of electronic monitoring when protocol adherence is a key concern.

Despite modest variability in sampling times, the SPIT booklet captured biologically plausible diurnal patterns for both cortisol and DHEA. Cortisol exhibited an early post-waking rise followed by a sustained decline, while DHEA showed a sharp decrease shortly after waking with continued decline thereafter. These patterns are consistent with established diurnal profiles measured using standard saliva collection methods, supporting the ability of the SPIT device to capture meaningful temporal variation in hormone levels.

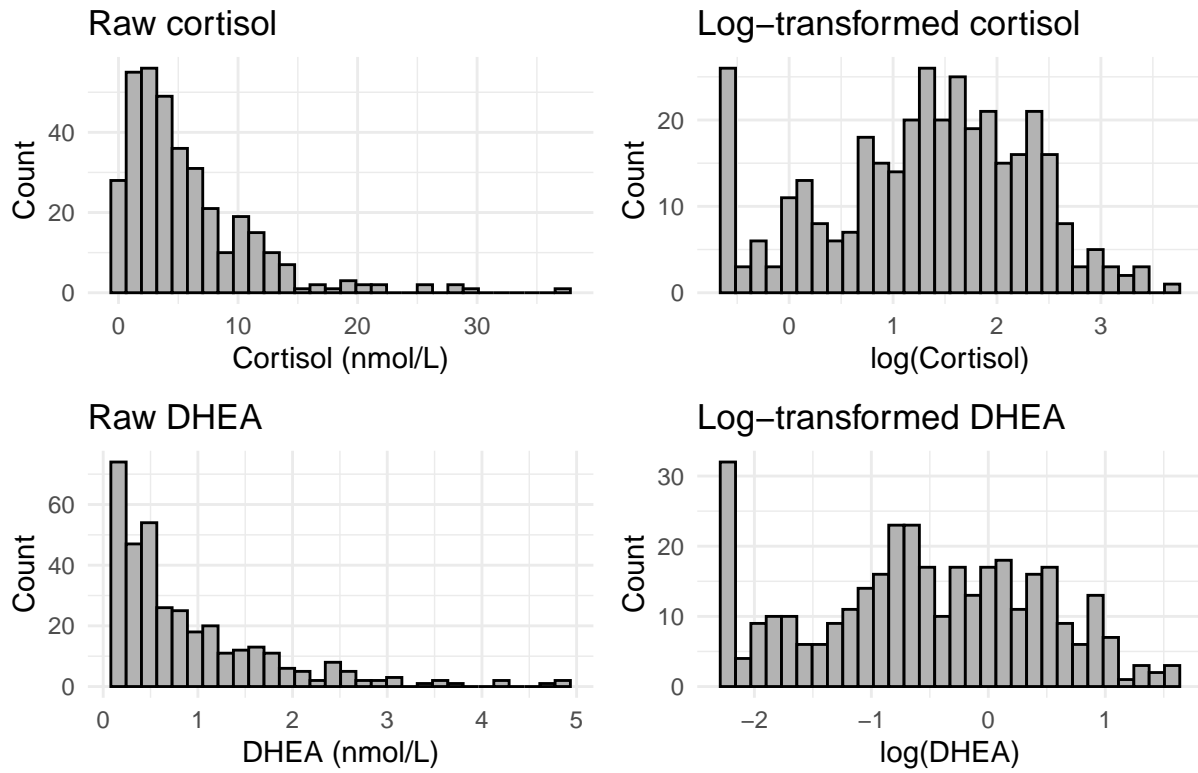
Overall, these findings suggest that the SPIT booklet is a feasible tool for assessment of cortisol and DHEA data. The combination of good agreement with electronic monitoring, reasonable protocol adherence and biologically plausible hormone patterns supports its potential utility in large scale or resource limited studies. However, these findings also emphasize the value of electronic monitoring for evaluating timing accuracy and protocol adherence, particularly for late-day sample collections.

Limitations

This analysis has several limitations that should be considered when interpreting the findings. The study was conducted in a relatively small sample of healthy adult participants, which may limit the generalizability of the results to more diverse or clinical populations where adherence and timing accuracy may differ. Estimates of time since waking relied on self-reported wake times from sleep diaries, which may introduce measurement error even when electronically recorded sampling times were available. Missing timing data occurred for both booklet and MEMS measures due to incomplete recording or device malfunction. Adherence and agreement metrics were summarized descriptively, and regression analyses were conducted without adjustment for multiple comparisons, consistent with the exploratory aims of the study. Additionally, hormone concentrations were measured over only three consecutive days which limits the ability to assess longer-term variability in diurnal patterns. Despite these limitations, the findings provide useful evidence supporting the feasibility of the SPIT method and highlight areas for refinement in future studies.

Appendix A. Model diagnostics for cortisol and DHEA mixed-effects models

Distributions of raw and log-transformed hormone concentrations



Raw hormone concentrations exhibit substantial right skewness, which is reduced after log transformation, supporting the use of log-scale outcomes in mixed-effects modeling.

Reproducible Research Information

GitHub Repository: