BIOS 6624 : Project 0

Ryan Summers

2026-01-22

Read in the csv

```
spit_raw <- read_csv("/Users/ryan_summers/GitHub/BIOS6624-Class/Project0/DataRaw/Project0_Clean_v2.csv")
```

```
## Rows: 372 Columns: 15
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): Collection Date
## dbl  (9): SubjectID, Collection Sample, Booklet: Sample interval Decimal Tim...
## time (5): Booket: Clock Time, MEMs: Clock Time, Sleep Diary reported wake ti...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# number of subjects
length(unique(spit_raw$SubjectID))
```

```
## [1] 31
```

```
# number of missing measurements per collection sample
spit_raw %>%
  #filter(`Collection Sample` != 1) %>%
  group_by(`Collection Sample`) %>%
  summarise(
    bk_miss = sum(is.na(`Booket: Clock Time`)),
    mem_miss = sum(is.na(`MEMs: Clock Time`)),
    n_obs = n(),
    `bk_%` = bk_miss/n_obs,
    `mem_%` = mem_miss/n_obs)
```

```
## # A tibble: 4 x 6
##   `Collection Sample` bk_miss mem_miss n_obs `bk_%` `mem_%`
##                 <dbl>   <int>    <int> <int>  <dbl>   <dbl>
## 1                   1       5       16    93 0.0538   0.172
## 2                   2       6       23    93 0.0645   0.247
## 3                   3      11       12    93 0.118    0.129
## 4                   4      13       10    93 0.140    0.108
```

Data Management & Cleanup

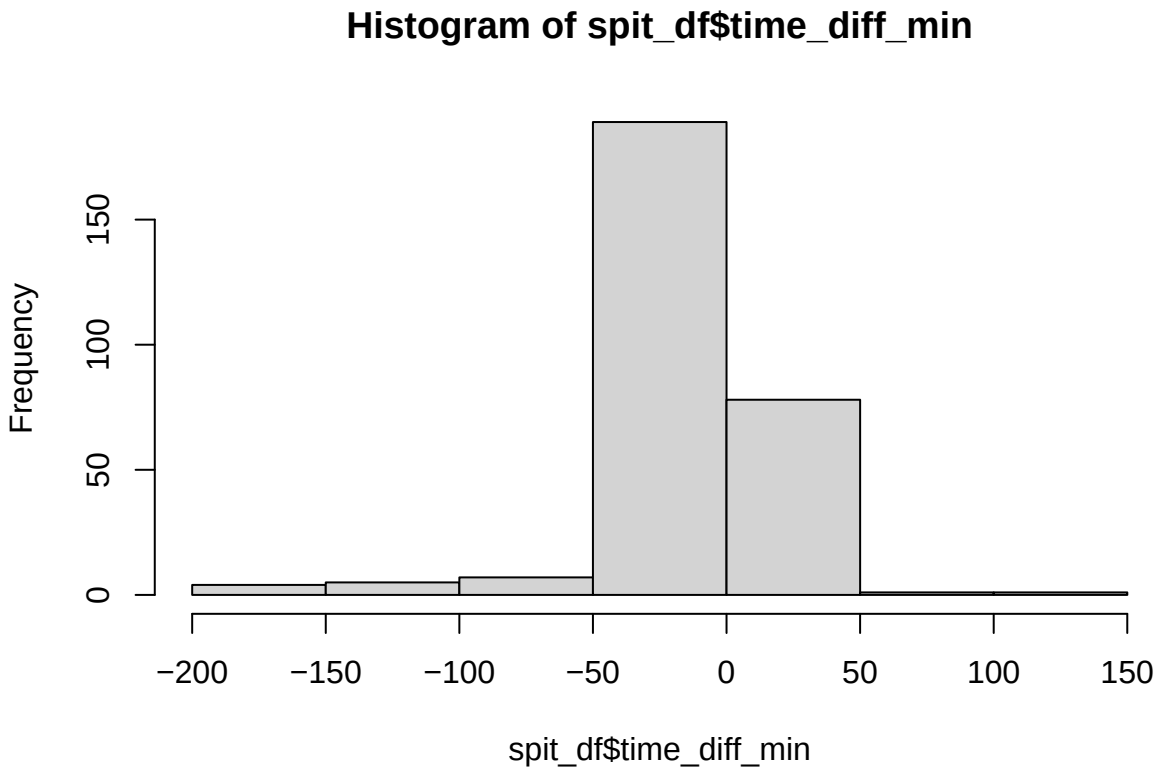
```
# clean up the column names
spit_df <- spit_raw %>%
  # lowercase columns & rows
  setNames(tolower(names(.))) %>%
  mutate_if(is.character, tolower) %>%
  # remove white-space in columns
  setNames(gsub("\\s+", "_", names(.)))
```

```
spit_df <- spit_df %>%
  mutate(
    # get weekday for collection dates
    collection_date = as.Date(collection_date, format = "%m/%d/%Y"),
    weekday = weekdays(collection_date),
    # determine if weekday is the weekend
    weekend = factor(if_else(weekday %in% c("Saturday","Sunday"),
                             "Weekend", "Weekday")),
    # add zeroes for MEMS collection on 1st sample
    `mems:_sample_interval_decimal_time_(mins)` =
      ifelse(collection_sample == 1, 0,
             `mems:_sample_interval_decimal_time_(mins)`),
    # log transform response
    log_cortisol = log(`cortisol_(nmol/l)`),
    # calculate time differences between book and MEM
    time_diff_min = as.numeric(difftime(`booket:_clock_time`,
                                        `mems:_clock_time`, units = "mins")),
    interval_diff_min = `booklet:_sample_interval_decimal_time_(mins)` -
      `mems:_sample_interval_decimal_time_(mins)`)
```

```
summary(spit_df$time_diff_min)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
## -200.000   -7.000   -1.000   -7.712    1.000  133.000       87
```

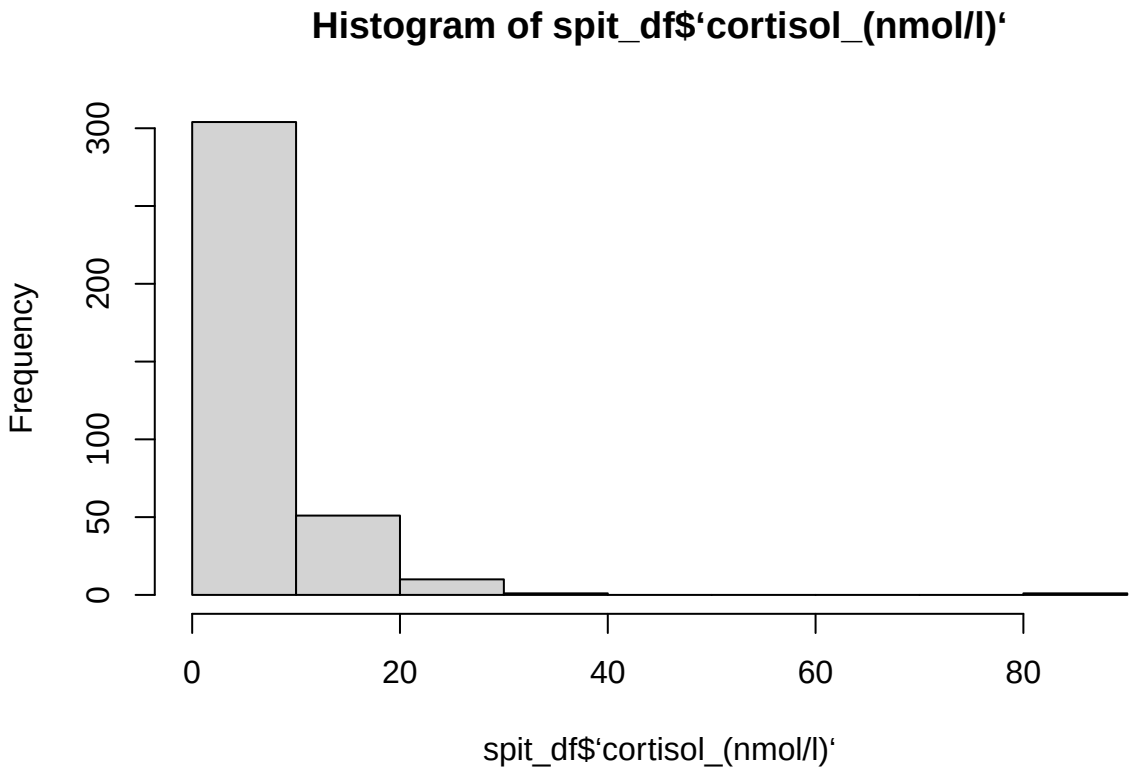
```
hist(spit_df$time_diff_min)
```




```
# distribution of cortisol and dhea values
summary(spit_df$`cortisol_(nmol/l)`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.5242  2.0555  4.1385  5.9509  7.6562 89.5571       5
```

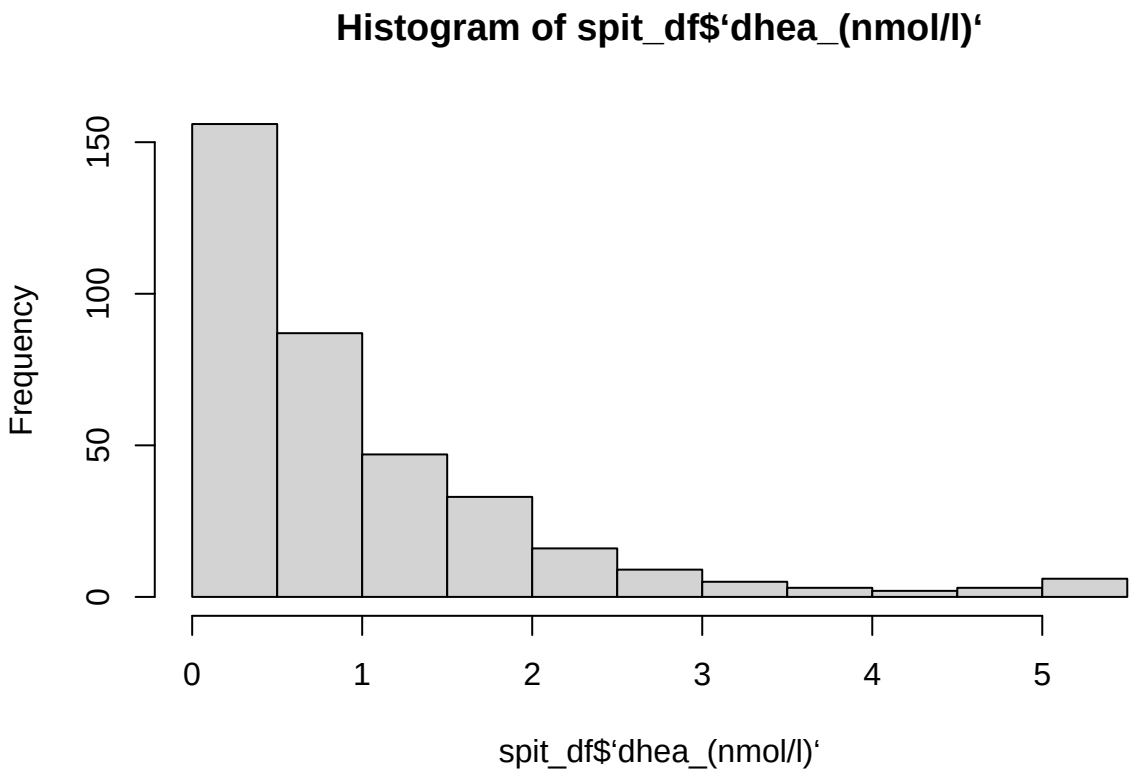
```
hist(spit_df$`cortisol_(nmol/l)`)
```



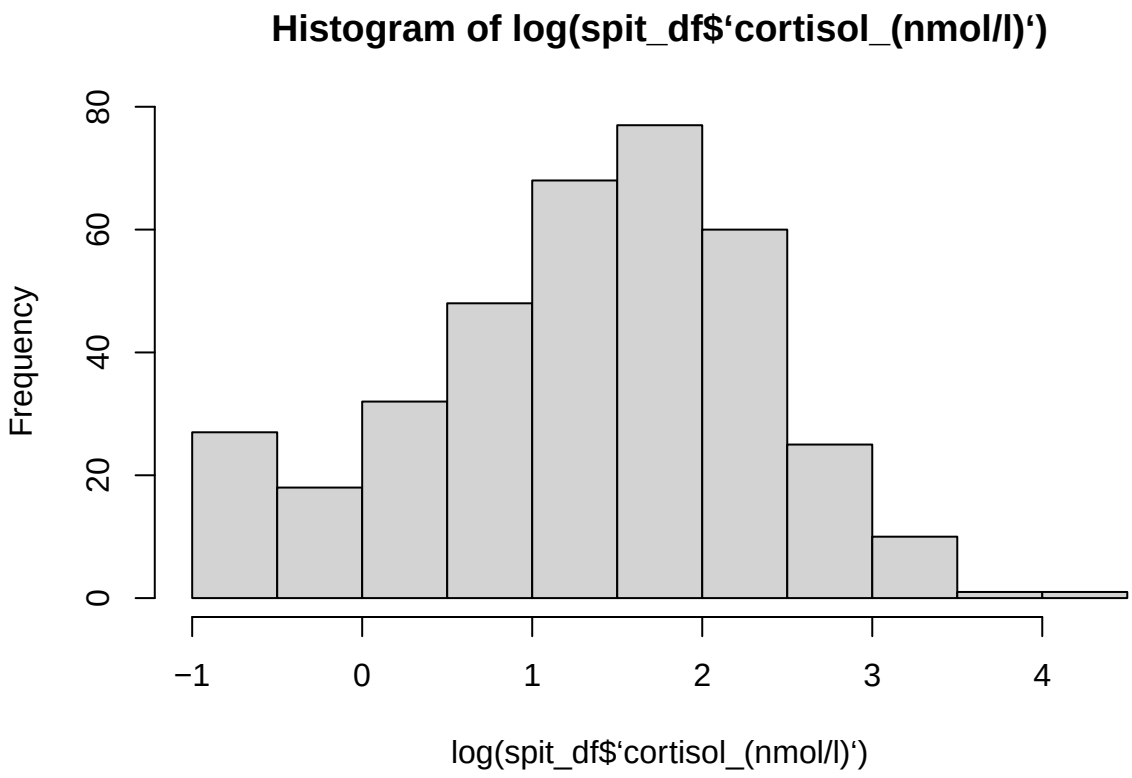

```
summary(spit_df$`dhea_(nmol/l)`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.1076  0.3093  0.6121  0.9811  1.2834  5.2050       5
```

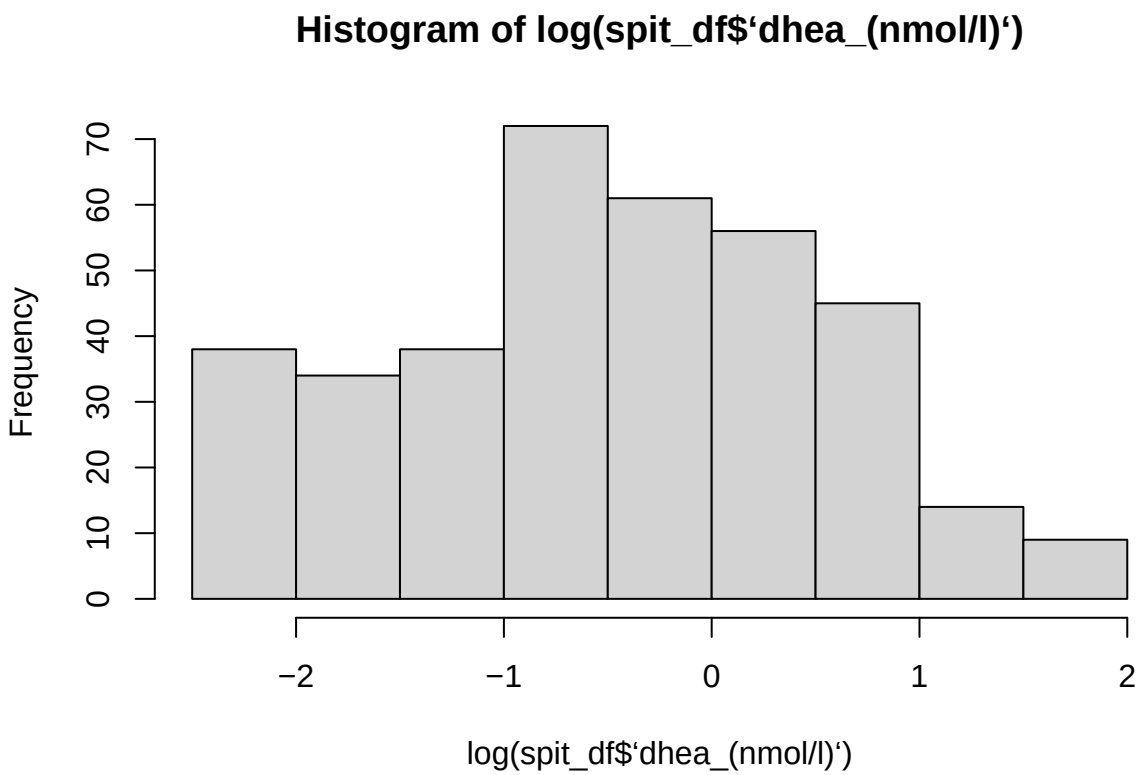
```
hist(spit_df$`dhea_(nmol/l)`)
```




```
# log-transformed
hist(log(spit_df$`cortisol_(nmol/l)`))
```



```r
hist(log(spit_df$`dhea_(nmol/l)`))
```



Create a datetime variable that merges the collection date and booklet clock time. * Use this variable as an anchor to determine minutest since wake-up * Additionally, create a column that stores either booklet or MEM clock time based on whether booklet contains a missing value.

```r
spit_df <- spit_df %>%
  mutate(
    # create a variable that includes the date and clock time for booklet and MEMs
    bk_dt = as.POSIXct(paste(collection_date, `booket:_clock_time`),
                       format = "%Y-%m-%d %H:%M", tz = "UTC"),
    mems_dt    = as.POSIXct(paste(collection_date, `mems:_clock_time`),
                       format = "%Y-%m-%d %H:%M", tz = "UTC")) %>%

group_by(subjectid, collection_date) %>%

  mutate(
    # use wake time determined by sleep diary as PI noted
    diary_wake_time = as.POSIXct(
      paste(collection_date, sleep_diary_reported_wake_time),
      format = "%Y-%m-%d %H:%M", tz = "UTC"),
    # fill down datetime values for each subject on each day
    diary_wake_time = fill(data.frame(diary_wake_time),
                   diary_wake_time,
                   .direction = "down")$diary_wake_time,

    # extract collection times from booklet, if NA grab MEMs
    sample_dt = coalesce(`booket:_clock_time`, `mems:_clock_time`),
    dt_source = if_else(!is.na(`booket:_clock_time`), "Booklet", "MEMS"),
```

```
    # calculate time since waking for both booklet and MEMs from diary start
    bk_min_since_wk = as.numeric(difftime(bk_dt, diary_wake_time,
                                          units = "mins")),
    mem_min_since_wk = as.numeric(difftime(mems_dt, diary_wake_time,
                                           units = "mins")),
    # create knot for piecewise regression - PI wants to kknow if theres
    # an increase in cortisol and/or DHEA from waking to 30min. Also, rate of
    # decline after 30min of waking.
    t30 = pmin(`mems:_sample_interval_decimal_time_(mins)`, 30),
    t_after30 = pmax(0, `mems:_sample_interval_decimal_time_(mins)`-30),) %>%
  ungroup()
```

Indicate whether Samples 2 and 4 are within 7.5min and 15min acceptance windows for adherence

```
targets <- c(`2` = 30, `4` = 600)

spit_df2 <- spit_df %>%
  #filter(collection_sample %in% c(2, 4)) %>%
  mutate(
    target_min = targets[as.character(collection_sample)],
    good_bk    = abs(bk_min_since_wk - target_min) <= 7.5,
    adeq_bkt = abs(bk_min_since_wk - target_min) <= 15,
    good_mem    = abs(mem_min_since_wk - target_min) <= 7.5,
    adeq_mem = abs(mem_min_since_wk - target_min) <= 15)

spit_df2 <- spit_df %>%
  mutate(
    target_min = case_when(
      collection_sample == 2 ~ 30,
      collection_sample == 4 ~ 600,
      TRUE ~ NA_real_),
    # booklet times
    good_bk = if_else(
      !is.na(target_min),
      abs(bk_min_since_wk - target_min) <= 7.5,
      NA),
    adeq_bkt = if_else(
      !is.na(target_min),
      abs(bk_min_since_wk - target_min) <= 15,
      NA),
    # MEM times
    good_mem = if_else(
      !is.na(target_min),
      abs(mem_min_since_wk - target_min) <= 7.5,
      NA),
    adeq_mem = if_else(
      !is.na(target_min),
      abs(mem_min_since_wk - target_min) <= 15,
      NA))
```

Missing Data Analysis

```
# number of NAs
lapply(spit_df, function(x) sum(is.na(x)))
```

```
## $subjectid
## [1] 0
##
## $collection_date
## [1] 0
##
## $collection_sample
## [1] 0
##
## $`booket:_clock_time`
## [1] 35
##
## $`mems:_clock_time`
## [1] 61
##
## $sleep_diary_reported_wake_time
## [1] 279
##
## $`booklet:_sample_interval`
## [1] 130
##
## $`booklet:_sample_interval_decimal_time_(mins)`
## [1] 55
##
## $`mems:_sample_interval`
## [1] 177
##
## $`mems:_sample_interval_decimal_time_(mins)`
## [1] 84
##
## $`cortisol_(ug/dl)`
## [1] 0
##
## $`dhea_(pg/dl)`
## [1] 5
##
## $`cortisol_(nmol/l)`
## [1] 5
##
## $`dhea_(nmol/l)`
## [1] 5
##
## $daynumb
## [1] 0
##
## $weekday
## [1] 0
##
```

```
## $weekend
## [1] 0
##
## $log_cortisol
## [1] 5
##
## $time_diff_min
## [1] 87
##
## $interval_diff_min
## [1] 122
##
## $bk_dt
## [1] 35
##
## $mems_dt
## [1] 61
##
## $diary_wake_time
## [1] 0
##
## $sample_dt
## [1] 9
##
## $dt_source
## [1] 0
##
## $bk_min_since_wk
## [1] 35
##
## $mem_min_since_wk
## [1] 61
##
## $t30
## [1] 84
##
## $t_after30
## [1] 84
```

```
spit_df %>%
  ungroup() %>%
  #group_by(collection_sample) %>%
  select(`booket:_clock_time`, `mems:_clock_time`, `cortisol_(ug/dl)`, `dhea_(pg/dl)`,
         `cortisol_(nmol/l)`, `dhea_(nmol/l)`) %>%
  miss_var_summary()
```

```
## # A tibble: 6 x 3
##   variable           n_miss pct_miss
##   <chr>               <int>    <num>
## 1 mems:_clock_time       61    16.4
## 2 booket:_clock_time     35     9.41
## 3 dhea_(pg/dl)            5     1.34
## 4 cortisol_(nmol/l)       5     1.34
## 5 dhea_(nmol/l)           5     1.34
```

```
## 6 cortisol_(ug/dl)        0     0
```

Address PI's question about agreement between the subject's recordings of sampling times compared to the times recorded by an electronic monitoring cap (1)

Scatter-plot of agreement (stratify by collection sample) for PI

```
cor_by_sample <- spit_df %>%
  filter(!is.na(mem_min_since_wk), !is.na(bk_min_since_wk)) %>%
  group_by(collection_sample) %>%
  summarise(
    r = cor(mem_min_since_wk, bk_min_since_wk),
    p = cor.test(mem_min_since_wk, bk_min_since_wk)$p.value)

cor_by_sample
```

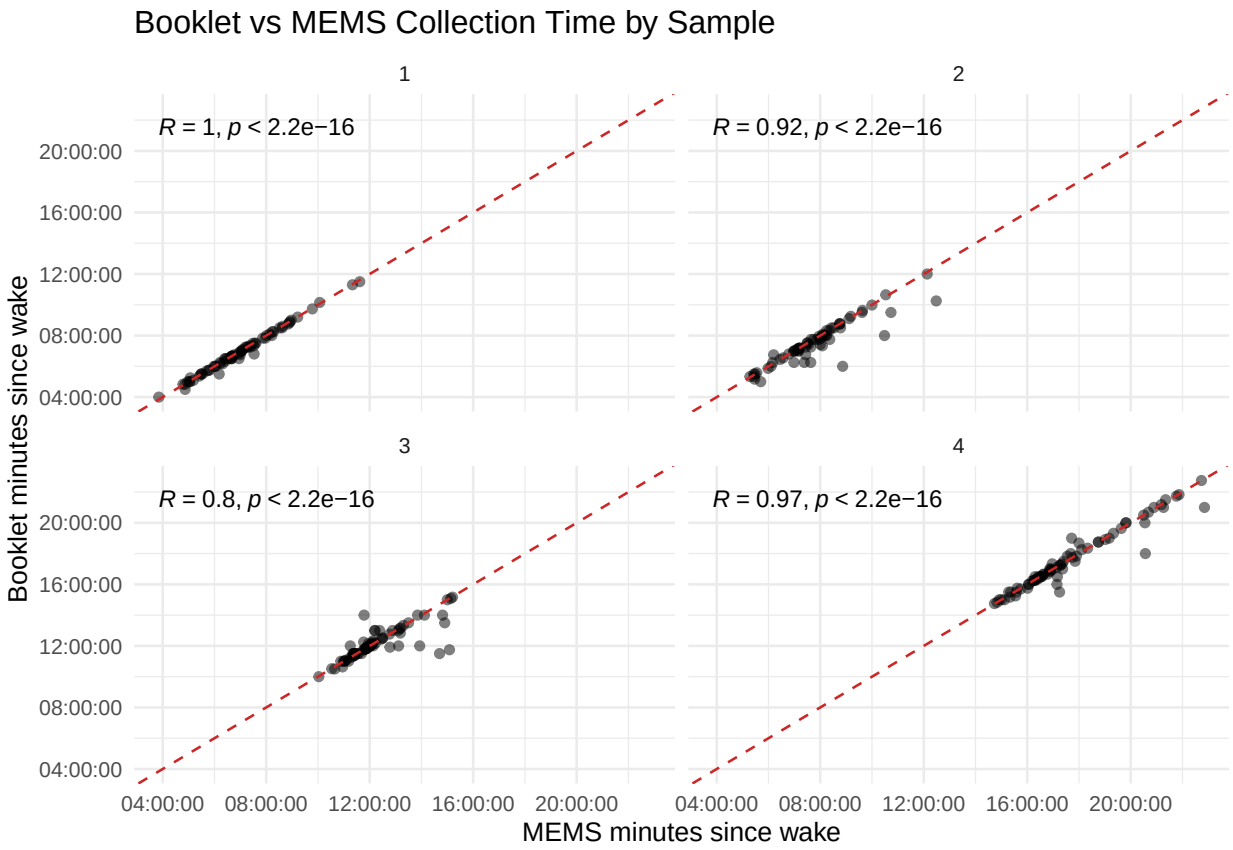
```
## # A tibble: 4 x 3
##   collection_sample      r        p
##               <dbl>  <dbl>    <dbl>
## 1                 1 0.0734 5.40e- 1
## 2                 2 0.125  3.18e- 1
## 3                 3 0.853  5.13e-22
## 4                 4 0.945  3.07e-36
```

```
library(ggpubr)

ggplot(spit_df, aes(x = `mems:_clock_time`, y = `booket:_clock_time`)) +
  geom_point(alpha = 0.5) +
  geom_abline(slope = 1, intercept = 0, linetype = 2, color = "firebrick3") +
  facet_wrap(~ collection_sample) +
  stat_cor(
    method = "pearson",
    label.x.npc = "left",
    label.y.npc = "top",
    size = 3
  ) +
  labs(
    title = "Booklet vs MEMS Collection Time by Sample",
    x = "MEMS minutes since wake",
    y = "Booklet minutes since wake"
  ) +
  theme_minimal(base_size = 10)
```

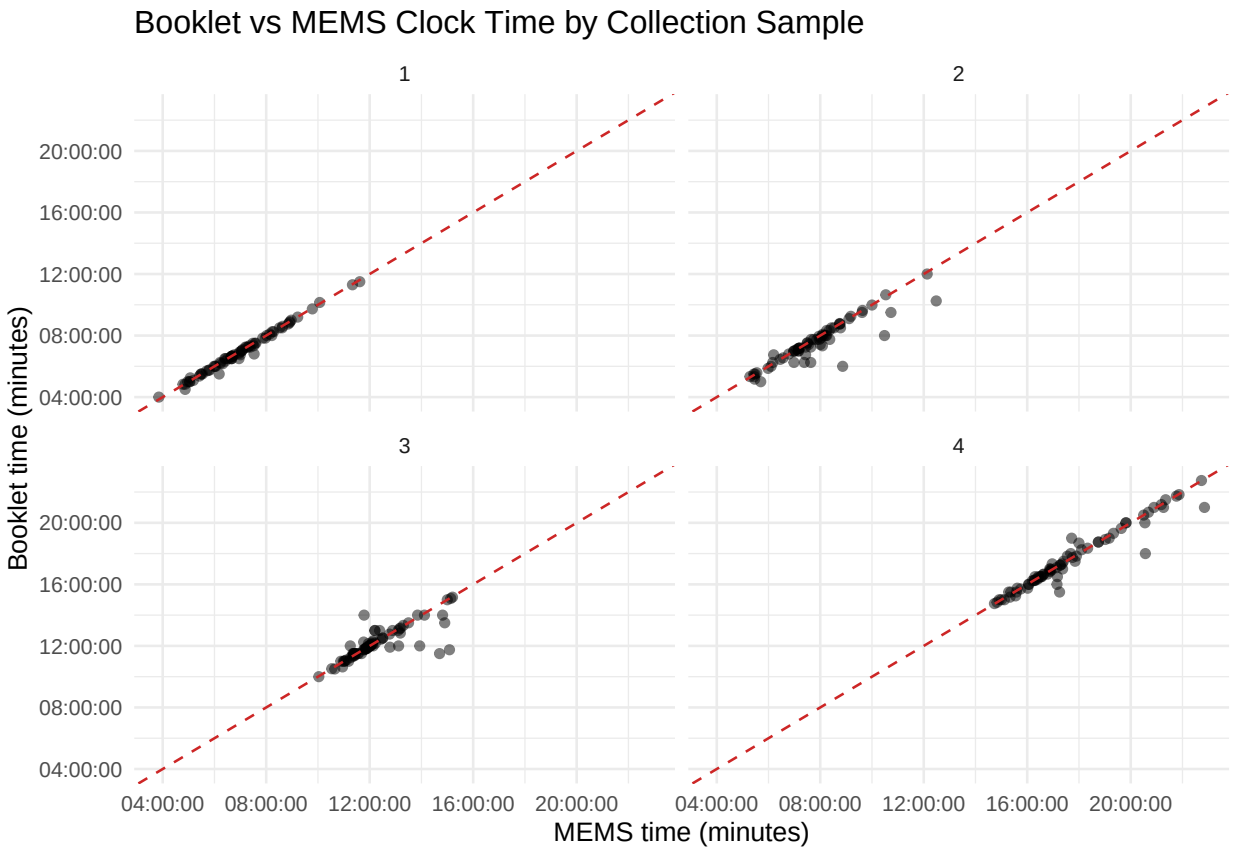
```
## Warning: Removed 87 rows containing non-finite outside the scale range
## (`stat_cor()`).
```

```
## Warning: Removed 87 rows containing missing values or values outside the scale range
## (`geom_point()`).
```




```
ggplot(spit_df, aes(x = `mems:_clock_time`, y = `booket:_clock_time`)) +
  geom_point(alpha = 0.5) +
  geom_abline(slope = 1, intercept = 0, linetype = 2, color="firebrick3") +
  facet_wrap(~ collection_sample) +
  labs(
    title = "Booklet vs MEMS Clock Time by Collection Sample",
    x = "MEMS time (minutes)",
    y = "Booklet time (minutes)"
  ) +
  theme_minimal(base_size = 10)
```

```
## Warning: Removed 87 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Mixed Model: Is there bias in self-reported times, and does it differ by sample, accounting for repeated measurements within subjects?

```
# convert collection sample to factor and create labels for it
spit_df$collection_sample <- factor(
  spit_df$collection_sample,
  labels = c(
    "Wake",
    "+30min",
    "Before lunch",
    "+600min"))

# booklet minus mem as outcome
agrmnt.model <- lmer(
  time_diff_min ~ collection_sample + (1 | subjectid),
  data = spit_df)

summary(agrmnt.model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: time_diff_min ~ collection_sample + (1 | subjectid)
##    Data: spit_df
##
## REML criterion at convergence: 2759
##
## Scaled residuals:
```

Characteristic	Beta	95% CI	p-value
Wake	-4.12	-11.8, 3.61	0.294
Collection Sample			
Wake	—	—	
+30min	-12.2	-22.7, -1.65	0.024
Before lunch	-2.12	-12.3, 8.07	0.682
+600min	-1.47	-11.7, 8.77	0.778

Abbreviation: CI = Confidence Interval

```
##     Min      1Q  Median      3Q     Max
## -6.0659  0.0084  0.1101  0.3667  4.3412
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  subjectid (Intercept)  47.96    6.925
##  Residual              973.36   31.199
## Number of obs: 285, groups:  subjectid, 31
##
## Fixed effects:
##                               Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)                     -4.118      3.912 158.733  -1.053   0.2941
## collection_sample+30min        -12.163      5.341 260.787  -2.277   0.0236 *
## collection_sampleBefore lunch   -2.122      5.175 255.232  -0.410   0.6821
## collection_sample+600min        -1.471      5.200 259.170  -0.283   0.7775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) cl_+30 cll_Bl
## cllctn_s+30 -0.652
## cllctn_smBl -0.673  0.492
## cllctn_+600 -0.672  0.488  0.505
```

```
tbl_regression(
  agrmnt.model,
  intercept = T,
  estimate_fun = label_style_sigfig(digits = 3),
  pvalue_fun = label_style_pvalue(digits = 3),
    label = list(
    `(Intercept)` ~ "Wake",
    collection_sample ~ "Collection Sample"))
```

```
# class recommened one - **debatable**
agrmnt.model2 <- lmer(
  bk_min_since_wk ~ mem_min_since_wk + (1 | subjectid),
  data = spit_df)

summary(agrmnt.model2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
```

```
## lmerModLmerTest]
## Formula: bk_min_since_wk ~ mem_min_since_wk + (1 | subjectid)
##    Data: spit_df
##
## REML criterion at convergence: 2787.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.9420  0.0024  0.1588  0.2635  4.3866
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  subjectid (Intercept)  44.11    6.642
##  Residual              989.60   31.458
## Number of obs: 285, groups:  subjectid, 31
##
## Fixed effects:
##                   Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)       -6.50779    2.91608  66.29918  -2.232    0.029 *
## mem_min_since_wk   0.99494    0.00706 262.91666 140.931   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## mm_mn_snc_w -0.641
```

Intercept should be zero and slope should be 1.

Interpretation: In lme model of time differences (Booklet - MEMS), agreement at waking did not show significant bias ($\beta_0 : t = -1.1; p = 0.29$); however, the +30-minute sample showed that participants recorded times approximately 16.3 minutes earlier (-4.1 + -12.2) than the MEMS timestamp, on average. There is no evidence of disagreement for samples before lunch ($p = 0.7$) and 600min after waking ($p = 0.8$).

- Between-subject differences in reporting bias are modest (~7 min)
- Within-subject variability is much larger (~31 min)

Most disagreement arises from occasion-to-occasion variability, not consistent subject-specific bias.

Correlation is invariant to shifts: * If Booklet times are always 10 minutes later than MEMS, correlation can still be ~1.0. * High correlation does not preclude bias; the mixed model identifies a consistent offset at Sample 2 after accounting for subject-level variability.

Address PI's question about adhering to +30min and 600min sampling times (2)

and create a table 1 (Descriptive Table)

```r
# variables used to remove NA calculation for samples 1 and 3
adherence_vars <- c("good_bk", "adeq_bkt", "good_mem", "adeq_mem")

tbl1 <- spit_df2 %>%
  mutate(
    collection_sample = factor(
```

```
    collection_sample,
    levels = c(1, 2, 3, 4),
    labels = c(
      "Wake",
      "+30 min",
      "Before lunch",
      "+600 min"))) %>%
ungroup() %>%
select(
  time_diff_min,
  collection_sample,
  `dhea_(nmol/l)`,
  `cortisol_(nmol/l)`,
  good_bk,
  adeq_bkt,
  good_mem,
  adeq_mem ) %>%

tbl_summary(
  by = collection_sample,
  type = all_continuous() ~ "continuous2",
  statistic = list(
    all_continuous() ~ c(
      "{mean} ({sd})",
      "{median} ({p25}, {p75})",
      "{min} - {max}"),
    all_categorical() ~ "{p}%"),

  digits = all_continuous() ~ 1,
  missing = "ifany",
  missing_text = "NA",
  missing_stat = "{p_miss}%",

  # change variable names
  label = list(
    time_diff_min ~ "Booklet vs MEM Time",
  `dhea_(nmol/l)` ~ "DHEA (nmol/L)"  ,
  `cortisol_(nmol/l)` ~ "Cortisol (nmol/L)",
  good_bk  ~ "Booklet (±7.5 min)",
  adeq_bkt  ~ "Booklet (±15 min)",
  good_mem ~ "MEMS (±7.5 min)",
  adeq_mem ~ "MEMS (±15 min)"
  )) %>%
add_overall() %>%
# modify_table_body(
#    ~ .x %>%
#     mutate(
#       across(
#         starts_with("stat_"),
#         ~ if_else(is.na(.) | str_detect(., "^NA"), "-", .)))) %>%

  modify_table_body(
  ~ .x %>%
```

```
      mutate(
        across(
          starts_with("stat_"),
          ~ if_else(
            # only replace NA-looking output for categorical rows / missing rows,
            # and NOT for continuous variables (continuous2 multi-line rows)
            var_type != "continuous" &
              row_type %in% c("level", "missing") &
              (is.na(.) | str_detect(., "^NA")), "-", .)))) %>%

    modify_table_body(
    ~ .x %>%
      mutate(
        # remove Overall NA calculation for adherence variables (misleading to have)
        stat_0 = if_else(row_type == "missing" & variable %in% adherence_vars,
                         "-", stat_0),
        # remove NA calculation for collection 1 and 3
        stat_1 = if_else(row_type == "missing" & variable %in% adherence_vars,
                         "-", stat_1),
        stat_3 = if_else(row_type == "missing" & variable %in% adherence_vars,
                         "-", stat_3))) %>%

    modify_table_body(
    ~ .x %>%
      mutate(
        across(
          starts_with("stat_"),
          ~ if_else(
            # Only touch adherence variables, any row type,
            # and replace anything that is NA / starts with NA (e.g., NA%, NA (NA%))
            variable %in% adherence_vars & (is.na(.) | str_detect(., "^NA")),
            "-", .)))) %>%
  bold_labels() %>%
  italicize_levels() %>%
  add_p(pvalue_fun = label_style_pvalue(digits = 2)) %>%
  modify_spanning_header(
    c("stat_1", "stat_2", "stat_3", "stat_4") ~ "**Collection Sample**") %>%

  as_kable_extra(format = "latex") %>%
  kableExtra::kable_styling(
    latex_options = c("hold_position", "scale_down", "repeat_header"),
    font_size = 8)
  # as_gt() %>%
  # tab_options(
  #   table.font.size = "small",
  #   heading.align = "left",
  #   data_row.padding = px(3),
  #   table.border.top.width = px(2),
  #   table.border.bottom.width = px(2))


tbl1
```

Characteristic	Overall N = 372	Collection Sample				p-value
		Wake N = 93	+30 min N = 93	Before lunch N = 93	+600 min N = 93	
Booklet vs MEM Time						0.015
Mean (SD)	-7.7 (32.1)	-4.0 (8.7)	-16.0 (35.9)	-6.2 (42.7)	-5.4 (29.9)	
Median (Q1, Q3)	-1.0 (-7.0, 1.0)	-2.0 (-6.0, -0.5)	-1.0 (-13.0, 0.0)	-1.0 (-3.0, 3.0)	0.0 (-4.0, 4.0)	
Min - Max	-200.0 - 133.0	-44.0 - 11.0	-172.0 - 33.0	-200.0 - 133.0	-154.0 - 77.0	
NA	23%	23%	29%	20%	22%	
DHEA (nmol/L)						<0.001
Mean (SD)	1.0 (1.0)	1.8 (1.3)	1.1 (0.9)	0.5 (0.5)	0.5 (0.6)	
Median (Q1, Q3)	0.6 (0.3, 1.3)	1.6 (0.8, 2.4)	0.9 (0.5, 1.4)	0.4 (0.2, 0.6)	0.3 (0.2, 0.5)	
Min - Max	0.1 - 5.2	0.1 - 5.2	0.1 - 5.2	0.1 - 2.8	0.1 - 3.6	
NA	1.3%	0%	1.1%	2.2%	2.2%	
Cortisol (nmol/L)						<0.001
Mean (SD)	6.0 (6.9)	7.4 (6.0)	9.0 (4.9)	3.2 (2.3)	4.2 (10.3)	
Median (Q1, Q3)	4.1 (2.0, 7.7)	5.4 (3.9, 8.2)	8.8 (5.9, 11.4)	2.8 (1.4, 4.3)	2.0 (1.0, 3.3)	
Min - Max	0.5 - 89.6	0.5 - 29.2	0.5 - 25.5	0.5 - 13.1	0.5 - 89.6	
NA	1.3%	0%	1.1%	2.2%	2.2%	
Booklet (±7.5 min)	63%	—	78%	—	46%	<0.001
NA	—	—	6.5%	—	14%	
Booklet (±15 min)	74%	—	90%	—	56%	<0.001
NA	—	—	6.5%	—	14%	
MEMS (±7.5 min)	42%	—	53%	—	33%	0.014
NA	—	—	25%	—	11%	
MEMS (±15 min)	54%	—	71%	—	40%	<0.001
NA	—	—	25%	—	11%	

[1] %

[2] Kruskal-Wallis rank sum test; Fisher's exact test

```
# instances where there is no booklet or mem time recorded but cortisol/dhea
# levels are recorded
spit_df2 %>%
  filter(is.na(`booket:_clock_time`) & is.na(`mems:_clock_time`))
```

```
## # A tibble: 9 x 34
##   subjectid collection_date collection_sample `booket:_clock_time`
##       <dbl> <date>                      <dbl> <time>
## 1      3019 2018-10-03                      3    NA
## 2      3021 2018-10-21                      4    NA
## 3      3028 2018-10-22                      3    NA
## 4      3029 2018-10-07                      4    NA
## 5      3030 2018-11-06                      3    NA
## 6      3038 2018-10-29                      2    NA
## 7      3038 2018-10-29                      3    NA
## 8      3038 2018-10-29                      4    NA
## 9      3049 2018-12-04                      2    NA
## # i 30 more variables: `mems:_clock_time` <time>,
## #   sleep_diary_reported_wake_time <time>, `booklet:_sample_interval` <time>,
## #   `booklet:_sample_interval_decimal_time_(mins)` <dbl>,
## #   `mems:_sample_interval` <time>,
## #   `mems:_sample_interval_decimal_time_(mins)` <dbl>,
## #   `cortisol_(ug/dl)` <dbl>, `dhea_(pg/dl)` <dbl>, `cortisol_(nmol/l)` <dbl>,
## #   `dhea_(nmol/l)` <dbl>, daynumb <dbl>, weekday <chr>, weekend <fct>, ...
```

Confirm the rates of adherence are accurate in tbl_summary

```r
# booklet = good adherence
spit_df2 %>%
  ungroup() %>%
  filter(collection_sample %in% c(2, 4)) %>%
  group_by(collection_sample) %>%
  summarise(
    n_total_rows = n(),
    n_nonmissing = sum(!is.na(good_bk)),
    n_good  = sum(good_bk == T, na.rm = T),
    pct_good = 100 * n_good / n_nonmissing) %>%
  ungroup()
```

```
## # A tibble: 2 x 5
##   collection_sample n_total_rows n_nonmissing n_good pct_good
##               <dbl>        <int>        <int>  <int>    <dbl>
## 1                 2           93           87     68     78.2
## 2                 4           93           80     37     46.2
```

```r
# mem = good adherence
spit_df2 %>%
  ungroup() %>%
  filter(collection_sample %in% c(2, 4)) %>%
  group_by(collection_sample) %>%
  summarise(
    n_total_rows = n(),
    n_nonmissing = sum(!is.na(good_mem)),
    n_good  = sum(good_mem == T, na.rm = T),
    pct_good = 100 * n_good / n_nonmissing) %>%
  ungroup()
```

```
## # A tibble: 2 x 5
##   collection_sample n_total_rows n_nonmissing n_good pct_good
##               <dbl>        <int>        <int>  <int>    <dbl>
## 1                 2           93           70     37     52.9
## 2                 4           93           83     27     32.5
```

Confirm proportions test is accurate in tbl_summary

```r
# booklet good test
tab <- spit_df2 %>%
  filter(collection_sample %in% c(2, 4)) %>%
  filter(!is.na(good_bk)) %>%
  with(table(collection_sample, good_bk))

addmargins(tab)
```

```
##                 good_bk
## collection_sample FALSE TRUE Sum
##               2      19   68  87
##               4      43   37  80
##               Sum    62  105 167
```

```r
chisq.test(tab)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tab
## X-squared = 16.84, df = 1, p-value = 4.067e-05
```

```r
fisher.test(tab)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  tab
## p-value = 2.638e-05
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.1155585 0.4945701
## sample estimates:
## odds ratio
##  0.2426543
```

Address PI's question about changes of cortisol and DHEA over time (3)

Identify subjects with cortisol levels over 80 nmol/L and DHEA upper limit of 5.205 nmol/L. These subjects could have underlying health problems and should be excluded from analysis

```r
# identify subjects with biologically implausible values and DHEA levels at 5.205
spit_df %>%
  filter(`cortisol_(nmol/l)` > 80 | `dhea_(nmol/l)`==5.205)
```

```
## # A tibble: 7 x 29
##   subjectid collection_date collection_sample `booket:_clock_time`
##       <dbl> <date>          <fct>             <time>
## 1      3021 2018-10-21      Wake              05:45
## 2      3024 2018-10-21      Wake              07:15
## 3      3025 2018-10-20      +600min           16:15
## 4      3037 2018-11-04      Wake              08:06
## 5      3037 2018-11-04      +30min            08:47
## 6      3037 2018-11-05      +30min            09:06
## 7      3037 2018-11-06      Wake              08:30
## # i 25 more variables: `mems:_clock_time` <time>,
## #   sleep_diary_reported_wake_time <time>, `booklet:_sample_interval` <time>,
## #   `booklet:_sample_interval_decimal_time_(mins)` <dbl>,
## #   `mems:_sample_interval` <time>,
## #   `mems:_sample_interval_decimal_time_(mins)` <dbl>,
## #   `cortisol_(ug/dl)` <dbl>, `dhea_(pg/dl)` <dbl>, `cortisol_(nmol/l)` <dbl>,
## #   `dhea_(nmol/l)` <dbl>, daynumb <dbl>, weekday <chr>, weekend <fct>, ...
```

```r
# exclude any observations for patiens from analysis
spit_df_filt <- spit_df %>%
  filter(`cortisol_(nmol/l)` <= 80 | `dhea_(nmol/l)`!=5.205) %>%
  mutate(
    # split time from wake to 30min and after 30min after waking (PI interest)
    # use booklet time as PI requested
    t30 = pmin(bk_min_since_wk, 30),
    t_after30 = pmax(0, bk_min_since_wk - 30))
```

```r
pw.model <- lmer(log_cortisol ~ t30 + t_after30 +
                   (1|subjectid),
                 data=spit_df_filt)

summary(pw.model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: log_cortisol ~ t30 + t_after30 + (1 | subjectid)
##    Data: spit_df_filt
##
## REML criterion at convergence: 851.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1803 -0.5049  0.0459  0.5482  4.6607
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  subjectid (Intercept) 0.1052   0.3244
##  Residual              0.6227   0.7891
## Number of obs: 336, groups:  subjectid, 31
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  1.698e+00  1.044e-01  1.067e+02  16.271   <2e-16 ***
## t30          4.054e-03  3.866e-03  3.031e+02   1.049    0.295
## t_after30   -2.020e-03  1.951e-04  3.064e+02 -10.353   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) t30
## t30       -0.620
## t_after30  0.005 -0.503
```

```r
exp(fixef(pw.model))
```

```
## (Intercept)         t30   t_after30
##   5.4632034   1.0040619   0.9979822
```

t30: * During first 30min of waking, cortisol increases 0.71% per minute * 1.0070562^30 1.2348 means cortisol increases to about 23.5% total between waking and 30min

t_after30: * after 30min, cortisol decreases by about 0.21% per minute * 0.9978616^60 0.8795 or after 30min, cortisol decreases by about 12.1% per hour

```
tr1 <- emtrends(pw.model, specs = ~ 1, var = "t30")
tr2 <- emtrends(pw.model, specs = ~ 1, var = "t_after30")

# Combine into one object and contrast
tr_both <- rbind(tr1, tr2)
contrast(tr_both, method = list("t30 - t_after30" = c(1, -1)))
```

```
##  contrast        estimate      SE  df t.ratio p.value
##  t30 - t_after30  0.00607 0.00397 305   1.531  0.1269
##
## Degrees-of-freedom method: kenward-roger
```

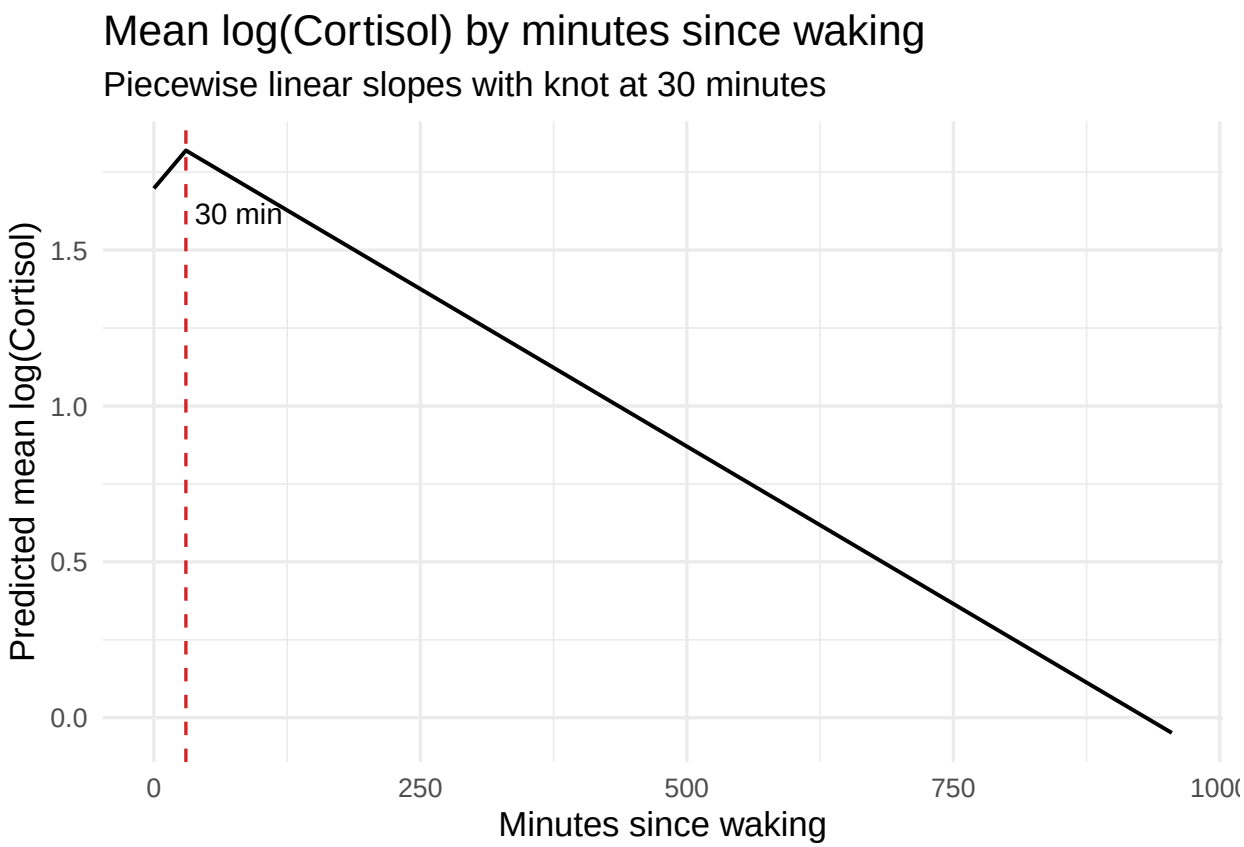
```
knot <- 30

newdat <- expand.grid(
  time_min = seq(0, max(spit_df_filt$bk_min_since_wk, na.rm = T), by = 1)) %>%
  mutate(
    t30       = pmin(time_min, knot),
    t_after30 = pmax(time_min - knot, 0))

newdat$fit <- predict(pw.model, newdata = newdat, re.form = NA)

# convert back to original units
newdat$exp_fit <- exp(newdat$fit)


ggplot(newdat, aes(x = time_min, y = fit)) +
  geom_line(linewidth = 0.7) +
  geom_vline(xintercept = knot, linetype = 2, color="firebrick3") +
  annotate("text", x = knot, y = max(newdat$fit, na.rm = TRUE),
           label = paste0(knot, " min"), vjust = 3.5, hjust = -0.1,
           size = 4, color = "black") +
  labs(
    title = "Mean log(Cortisol) by minutes since waking",
    subtitle = paste0("Piecewise linear slopes with knot at ", knot, " minutes"),
    x = "Minutes since waking",
    y = "Predicted mean log(Cortisol)") +
  theme_minimal(base_size = 13)
```


Mean log(Cortisol) by minutes since waking
Piecewise linear slopes with knot at 30 minutes
30 min
Predicted mean log(Cortisol)
1.5
1.0
0.5
0.0
0
250
500
750
Minutes since waking