SUPPLEMENTARY MATERIALS for

Generalized additive models to analyze biomedical non-linear longitudinal data in R:

Beyond repeated measures ANOVA and Linear Mixed Models

APPENDIX B: CODE AND FUNCTIONS

Ariel I. Mundo, John R. Tipton, and Timothy J. Muldoon*

tmuldoon@uark.edu

This appendix shows the code for the functions used through the main manuscript, which can be found in the *scripts* folder in the GitHub repository. We provide a brief explanation of the purpose of each function.

B.1 Setup

First, we load all required libraries and set seed.

```
library(patchwork)
library(tidyverse)
library(mvnfast)
library(nlme)
library(mgcv)
library(gratia)
library(here)
library(scico)
set.seed(2021) #set seed for reproducibility

#alpha for ribbon in the smooth plots
al<-0.8

thm1<-scale_fill_scico_d(palette="tokyo",begin=0.3, end=0.8, direction =
   -1, aesthetics = c("colour","fill"))
```

B.2 Linear and quadratic longitudinal trends

B.2.1 Function for linear and quadratic trends, rm-ANOVA and LMEM fits

The first function is `example.R`, which allows to simulate linear and quadratic data in the same manner as in Section 3.5 in the main manuscript and estimates rm-ANOVA and LMEM with interaction to the data. The error for each simulated trend can be correlated or uncorrelated

```
##########Section for calculations###########

## Example with linear response

#This function simulates data using a linear or quadratic mean response
   and each with correlated
#or uncorrelated errors. Each group has a different slope/concavity.
example <- function(n_time = 6, #number of time points
                    fun_type = "linear", #type of response
                    error_type = "correlated") {

  if (!(fun_type %in% c("linear", "quadratic")))
    stop('fun_type must be either "linear", or "quadratic"')
  if (!(error_type %in% c("correlated", "independent")))
    stop('fun_type must be either "correlated", or "independent"')


  x <- seq(1,6, length.out = n_time)

  #Create mean response matrix: linear or quadratic
  mu <- matrix(0, length(x), 2)
  # linear response
  if (fun_type == "linear") {
    mu[, 1] <- - (0.25*x)+2
```

```r
    mu[, 2] <- 0.25*x+2
  } else {
    # quadratic response (non-linear)

    mu[, 1] <-  -(0.25 * x^2) +1.5*x-1.25
    mu[, 2] <- (0.25 * x^2) -1.5*x+1.25
  }


  #create an array where individual observations per each time point for
     each group are to be stored. Currently using 10 observations per
     timepoint
  y <- array(0, dim = c(length(x), 2, 10))

  #Create array to store the "errors" for each group at each timepoint.
     The "errors" are the
  #between-group variability in the response.
  errors <- array(0, dim = c(length(x), 2, 10))
  #create an array where 10 observations per each time point for each
     group are to be stored

  #The following loops create independent or correlated responses. To each
      value of mu (mean response per group) a randomly generated error (
     correlated or uncorrelated) is added and thus the individual response
      is created.
  if (error_type == "independent") {
    ## independent errors
    for (i in 1:2) {
      for (j in 1:10) {
        errors[, i, j] <- rnorm(6, 0, 0.25)
        y[, i, j] <- mu[, i] + errors[, i, j]
      }
    }
  } else {
    for (i in 1:2) {      # number of treatments
      for (j in 1:10) {  # number of subjects
        # compound symmetry errors: variance covariance matrix
        errors[, i, j] <- rmvn(1, rep(0, length(x)), 0.1 * diag(6) + 0.25
            * matrix(1, 6, 6))
        y[, i, j] <- mu[, i] + errors[, i, j]
      }
    }
  }



  ## subject random effects

  ## visualizing the difference between independent errors and compound
     symmetry
  ## why do we need to account for this -- overly confident inference

#labeling y and errors
  dimnames(y) <- list(time = x,
                      treatment = 1:2,
```

```
                      subject = 1:10)

  dimnames(errors) <- list(time = x,
                           treatment = 1:2,
                           subject = 1:10)

  #labeling the mean response
  dimnames(mu) <- list(time = x,
                       treatment = 1:2)

  #convert y, mu and errors to  dataframes with time, treatment and
     subject columns
  dat <- as.data.frame.table(y,
                             responseName = "y")
  dat_errors <- as.data.frame.table(errors,
                                    responseName = "errors")
  dat_mu <- as.data.frame.table(mu,
                                responseName = "mu")

  #join the dataframes to show mean response and errors per subject
  dat <- left_join(dat, dat_errors,
                   by = c("time", "treatment", "subject"))
  dat <- left_join(dat, dat_mu,
                   by = c("time", "treatment"))
  #add time
  dat$time <- as.numeric(as.character(dat$time))
  #label subjects per group
  dat <- dat %>%
    mutate(subject = factor(paste(subject,
                                  treatment,
                                  sep = "-")))


  ## repeated measures ANOVA

  fit_anova <- lm(y ~ time + treatment + time * treatment, data = dat)

#LMEM: time and treatment interaction model, compound symmetry
  fit_lme <- lme(y ~ treatment + time + treatment:time,
                 data = dat,
                 random = ~ 1 | subject,
                 correlation = corCompSymm(form = ~ 1 | subject)
  )

  #create a prediction frame where the model can be used for plotting
     purposes
  pred_dat <- expand.grid(
    treatment = factor(1:2),
    time = unique(dat$time)
  )

  #add model predictions to the dataframe that has the simulated data
  dat$pred_anova <- predict(fit_anova)
  dat$pred_lmem <- predict(fit_lme)
```

```
  #return everything in a list
  return(list(
    dat = dat,
    pred_dat = pred_dat,
    fit_anova=fit_anova,
    fit_lme = fit_lme
  ))
}
```

B.2.2 A composite plot for the trends

Function `plot_example.R` uses the output of `example.R` to show the fit of a rm-ANOVA and a LMEM. It can be used to show an expanded version of Figure 1 in the main manuscript, presenting simulated data with correlated and uncorrelated errors and the correspoding rm-ANOVA and LMEM fits, which we do in the next subsection.

```
## This function plots the rm-ANOVA and LMEM for the data simulated in
   example.R
plot_example <- function(sim_dat) {
    # Plot the simulated data (scatterplot)
    p1 <- sim_dat$dat %>%
        ggplot(aes(x = time,
                   y = y,
                   group = treatment,
                   color = treatment)
        ) +
        geom_point(alpha=0.5,
                   show.legend=FALSE) +
        labs(y='response')+
        geom_line(aes(x = time,
                      y = mu,
                      color = treatment),
                  size=3,
                  show.legend=FALSE) +
        theme_classic() +
        theme(plot.title = element_text(size = 20,
                                        face = "bold"),
              text=element_text(size=20))+
        thm1

    #plot the model predictions for rm-ANOVA
    p2 <- ggplot(sim_dat$dat,
                 aes(x = time,
                     y = y,
                     color = treatment)) +
        geom_point(alpha=0.5,
                   show.legend=FALSE)+
        labs(y='response')+
        geom_line(aes(y = predict(sim_dat$fit_anova),
                      group = subject, size = "Subjects"),
                  show.legend = FALSE) +
        geom_line(data = sim_dat$pred_dat,
                  aes(y = predict(sim_dat$fit_anova,
```

```
                                  level = 0,
                                  newdata = sim_dat$pred_dat),
                      size = "Population"),
                  show.legend=FALSE) +
        guides(color = guide_legend(override.aes = list(size = 2)))+
        scale_size_manual(name = "Predictions",
                          values=c("Subjects" = 0.5, "Population" = 3)) +
        theme_classic() +
        theme(plot.title = element_text(size = 20,
                                        face = "bold"),
              text=element_text(size=20))+
        thm1

    #plot the model predictions for LMEM
    p4 <- ggplot(sim_dat$dat,
                 aes(x = time,
                     y = y,
                     color = treatment)) +
        geom_point(alpha=0.5)+
        labs(y='response')+
        geom_line(aes(y = predict(sim_dat$fit_lme),
                      group = subject, size = "Subjects")) +
        geom_line(data = sim_dat$pred_dat,
                  aes(y = predict(sim_dat$fit_lme,
                                  level = 0,
                                  newdata = sim_dat$pred_dat),
                      size = "Population")) +
        guides(color = guide_legend(override.aes = list(size = 2)))+
        scale_size_manual(name = "Predictions",
                          values=c("Subjects" = 0.5, "Population" = 3)) +
        theme_classic() +
        theme(plot.title = element_text(size = 20,
                                        face = "bold"),
              text=element_text(size=20))+
        thm1

    return((p1+p3+p2+p4+p5)+plot_layout(nrow=1)+plot_annotation(tag_levels
        = 'A'))

}
```

B.2.3 Plotting rm-ANOVA and LMEM fits for linear and quadratic trends in data

In this subsection, we use `example.R` and a modified version `plot_example.R` (`plot_example_Appendix.R`) to create an expanded version of Figure 1 on the main manuscript. The only difference between `plot_example.R` and `plot_example_Appendix.R` is the inclusion of a `ggplot2` object (`p3`) that allows to plot the simulated errors.

Figure B.1 show in panels A and D the simulated mean responses and individual data points. Panels C and G show a visual interpretation of "correlation" in the responses: In panel C, subjects that have a value of the random error ε either above or below the mean group response are more likely to have other observations that follow the same trajectory, thereby demonstrating correlation in the response. In panel G,because the errors are independent, there is no expectation that responses are likely to follow a similar pattern. Panels D and H show the predictions from the rm-ANOVA model.

B.2.3.1 Fits for linear trends The chunk below calls both `example.R` and `plot_example_Appendix.R` to simulate data and create the composite plots.

```
source(here::here("Manuscripts/Manuscript_by_chapters-SIM_Revisions_final/
   scripts","example.R"))
source(here::here("Manuscripts/Manuscript_by_chapters-SIM_Revisions_final/
   scripts","plot_example_Appendix.R"))

A1<-plot_example_Appendix(example(fun_type = "linear", error_type = "
   correlated"))

B1<-plot_example_Appendix(example(fun_type = "linear", error_type = "
   independent"))

C1<-plot_example_Appendix(example(fun_type = "quadratic", error_type = "
   correlated"))

D1<-plot_example_Appendix(example(fun_type = "quadratic", error_type = "
   independent"))
```

B.2.3.2 Fits for quadratic trends For the quadratic response case, Figure B.2 shows the simulated responses using compound symmetry and independent errors.

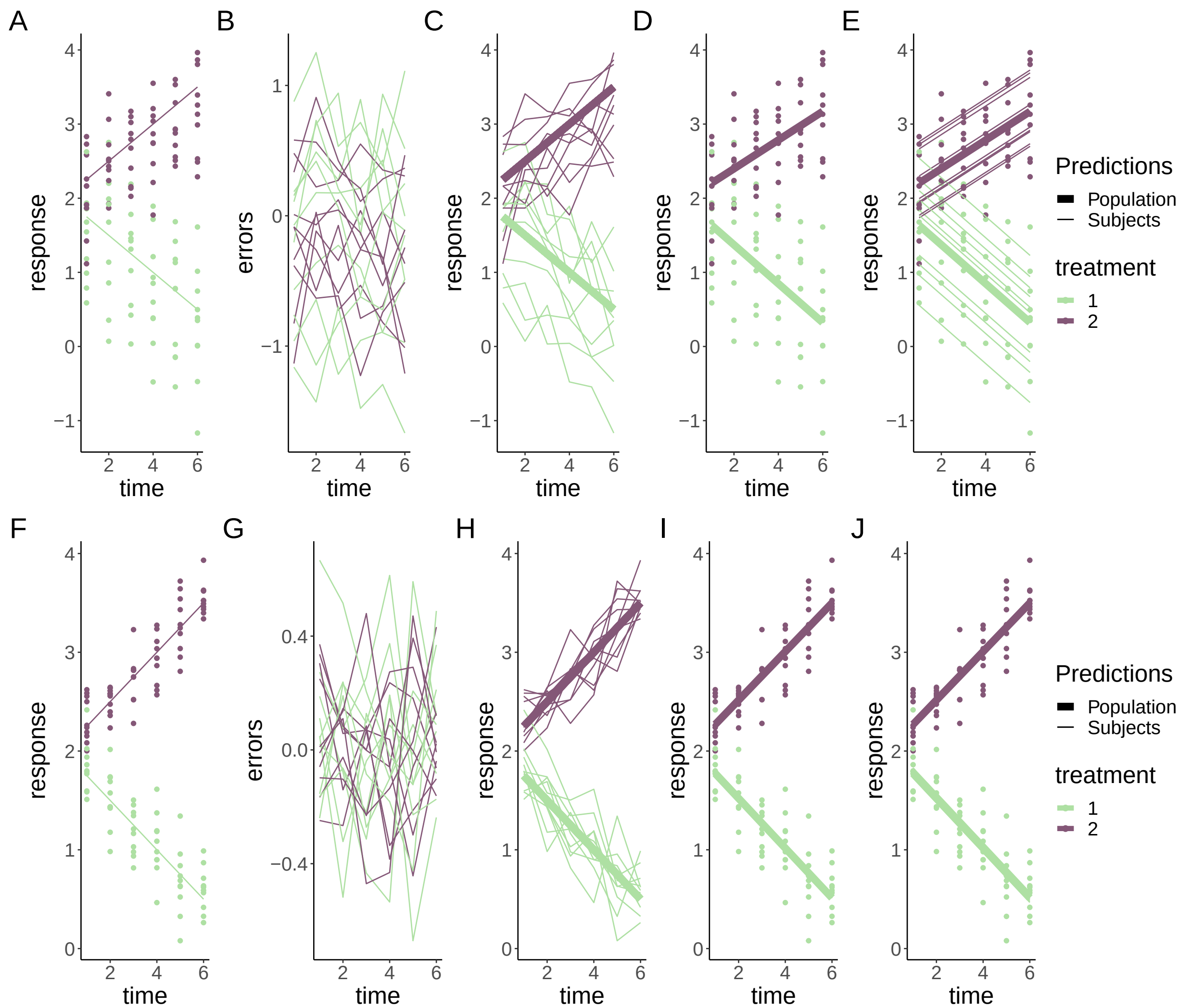

Figure B.1: Simulated linear responses from two groups with correlated (top row) or independent (bottom row) errors using a rm-ANOVA model and a LMEM. A, F:Simulated data with known mean response and individual responses (points) showing the dispersion of the data. B,G: Generated errors showing the difference in the behavior of correlated and independent errors. C,H: Simulated data with thin lines representing individual trajectories. D,I: Estimations from the rm-ANOVA model for the mean group response. E, J: Estimations from the LMEM for the mean group response and individual responses (thin lines). In all panels, thick lines are the predicted mean response per group, thin lines are the random effects for each subject and points represent the original raw data. Both rm-ANOVA and the LMEM are able to capture the trend of the data.

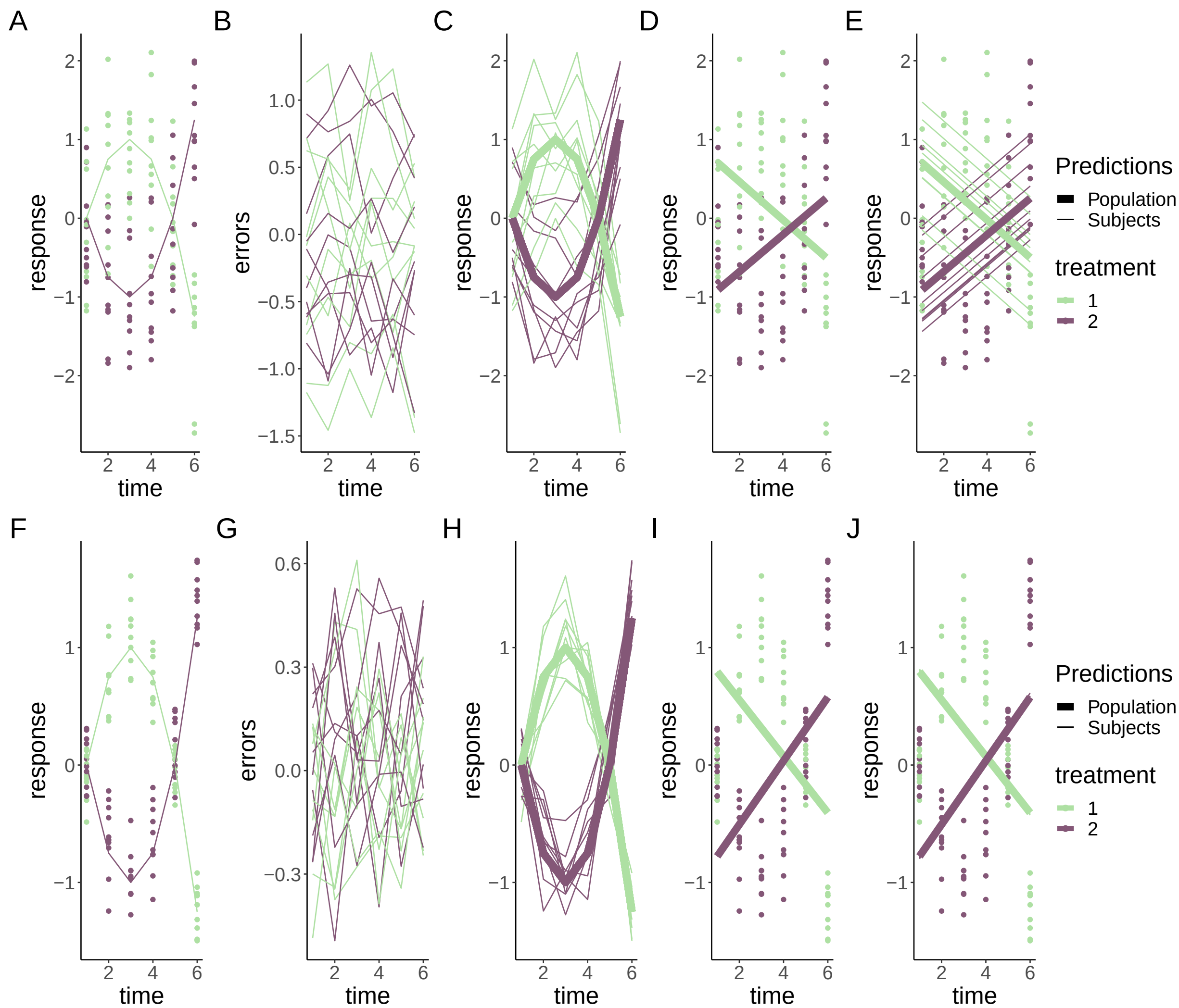

Figure B.2: Simulated quadratic responses from two groups with correlated (top row) or independent (bottom row) errors using a rm-ANOVA model and a LMEM. A, F:Simulated data with known mean response and individual responses (points) showing the dispersion of the data. B,G: Generated errors showing the difference in the behavior of correlated and independent errors. C,H: Simulated data with thin lines representing individual trajectories. D,I: Estimations from the rm-ANOVA model for the mean group response. E, J: Estimations from the LMEM for the mean group response and individual responses (thin lines). In all panels, thick lines are the predicted mean response per group, thin lines are the random effects for each subject and points represent the original raw data. Both rm-ANOVA and the LMEM are not able to capture the changes in each group over time.