EB_data

Ariel Mundo

3/16/2021

GAM analysis of EB's DRS data. The first part is to load all the required packages, and create another column in the dataframe that labels each mouse as "Control" or "Treatment", for plotting purposes.

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
## v purrr   0.3.4

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

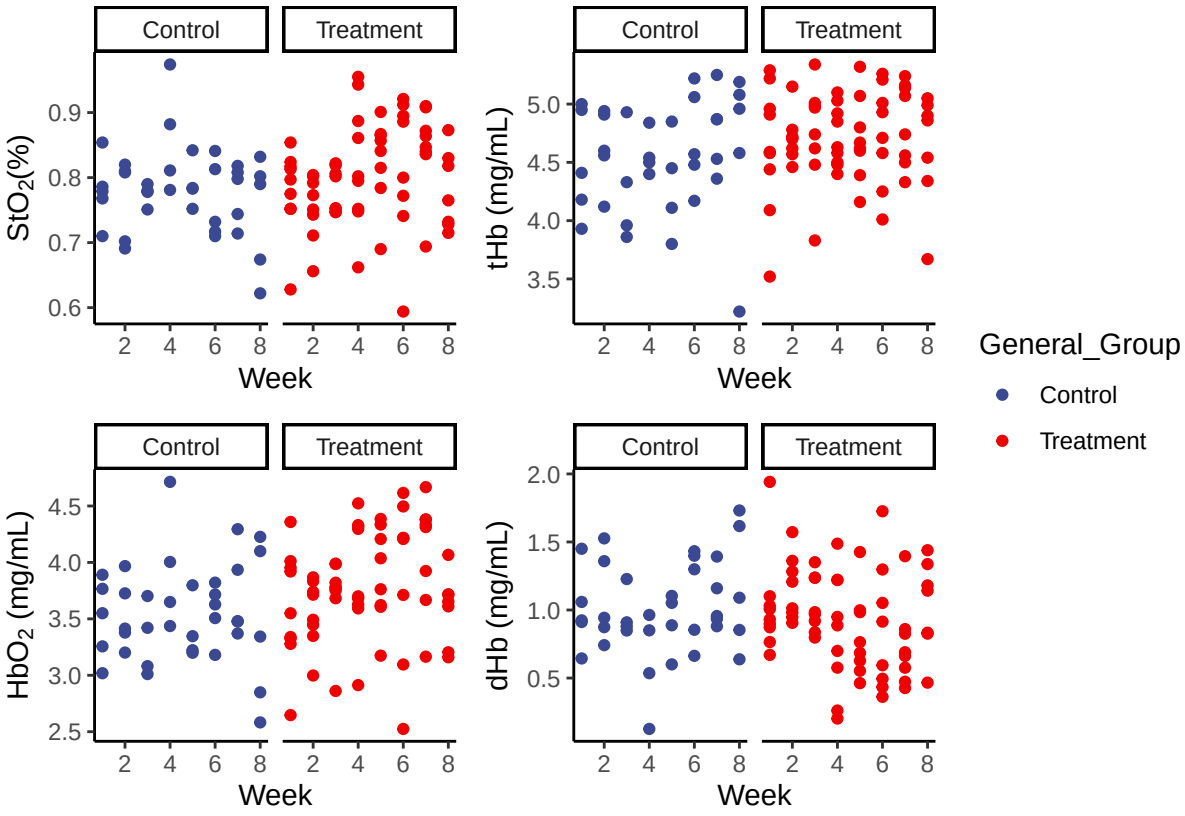
```
library(mgcv)
```

```
## Loading required package: nlme

##
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':
##
##     collapse

## This is mgcv 1.8-33. For overview type 'help("mgcv-package")'.
```

```
library(ggsci)
library(gratia)
library(patchwork)
dat<-read.csv("GAM data_weeks.csv")

#create new column for the Treatment and Control Groups (for plotting purposes)


dat<-dat %>%
    mutate(General_Group=factor(ifelse(grepl("C",Group),"Control","Treatment")))
```

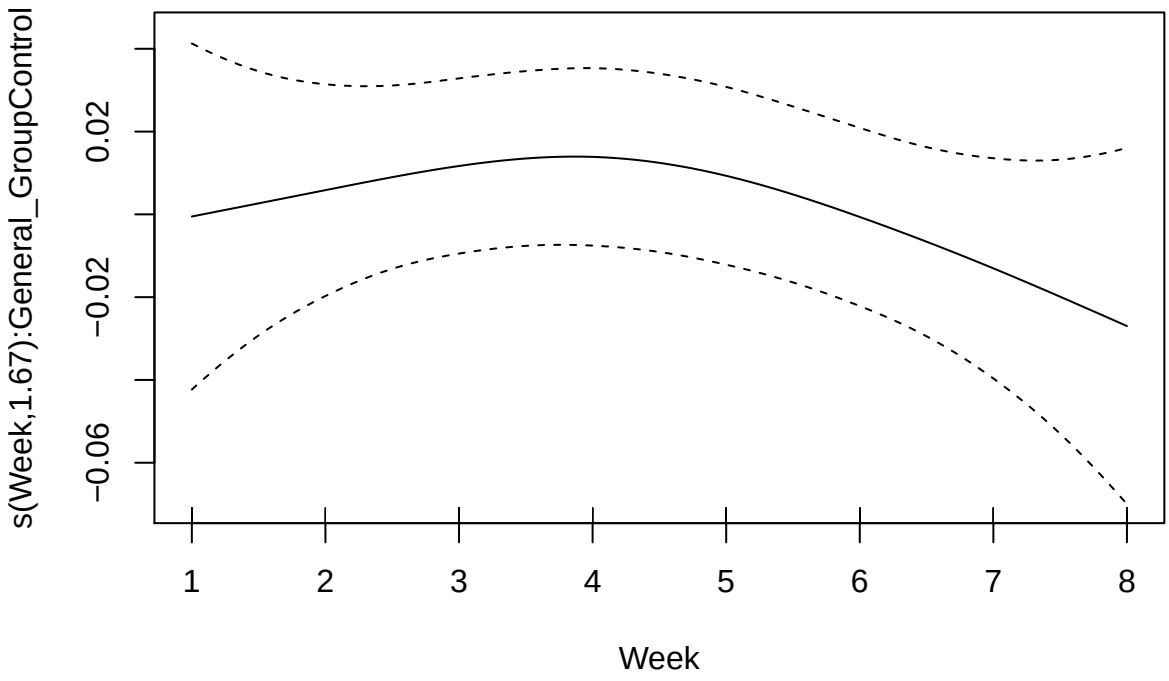
Next, the trends for each group are plotted for StO_2, tHb, HbO_2 and dHb.

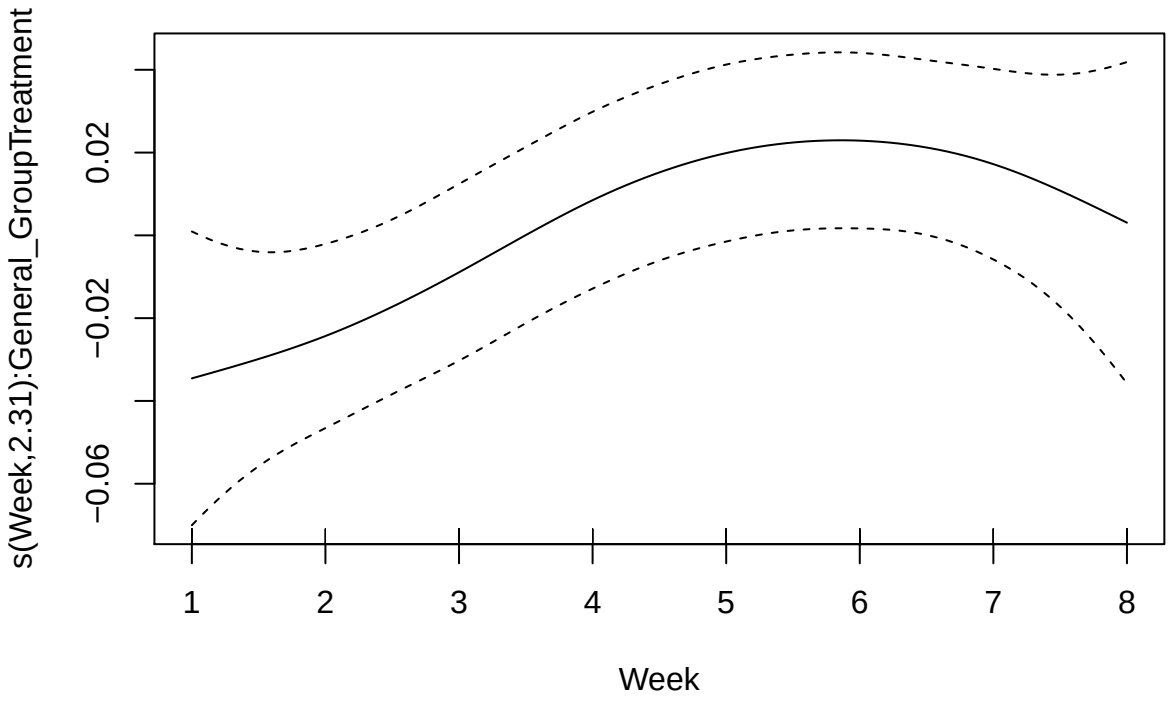
```
a1<-ggplot(data=dat,aes(x=Week, y=StO2))+
    geom_point(aes(color=General_Group))+
    labs(y=expression(paste(StO[2],'(%)')))+
    scale_color_aaas()+
    theme_classic()+
    facet_wrap(~General_Group)



a2<-ggplot(data=dat,aes(x=Week, y=tHb))+
    geom_point(aes(color=General_Group))+
    labs(y=expression(paste(tHb,' (mg/mL)')))+
    scale_color_aaas()+
    theme_classic()+
    facet_wrap(~General_Group)

a3<-ggplot(data=dat,aes(x=Week, y=HbO2))+
     geom_point(aes(color=General_Group))+
    labs(y=expression(paste(HbO[2],' (mg/mL)')))+
    scale_color_aaas()+
    theme_classic()+
    facet_wrap(~General_Group)

a4<-ggplot(data=dat,aes(x=Week, y=dHb))+
     geom_point(aes(color=General_Group))+
    labs(y=expression(paste(dHb,' (mg/mL)')))+
    scale_color_aaas()+
    theme_classic()+
    facet_wrap(~General_Group)

(a1+a2)/(a3+a4)+plot_layout(guides="collect") #plotting all together
```



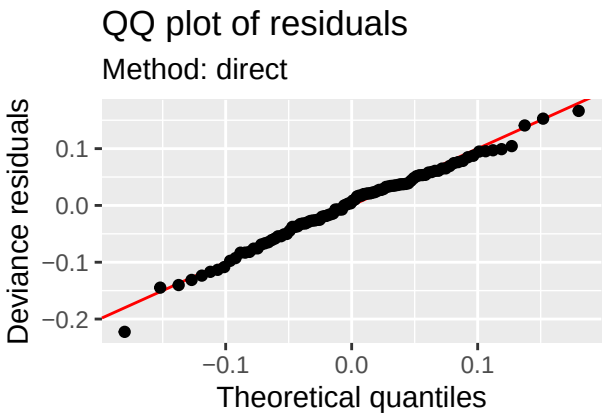
We are interested in determining how the DRS-derived parameters differ between the "Control" and "Treatment" Groups over time. Because we don't have a huge amount of datapoints the model needs to be straightforward. We will fit a model where we want a smooth term by *time*, which can be different for each treatment group.

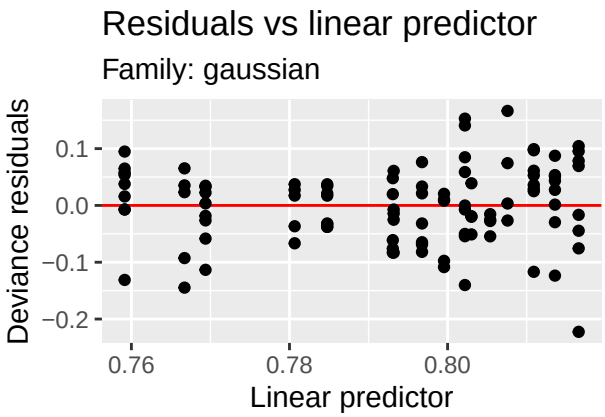
```
StO2_GAM<-gam(StO2~s(Week,by=General_Group,k=7),data=dat)
plot(StO2_GAM)
```

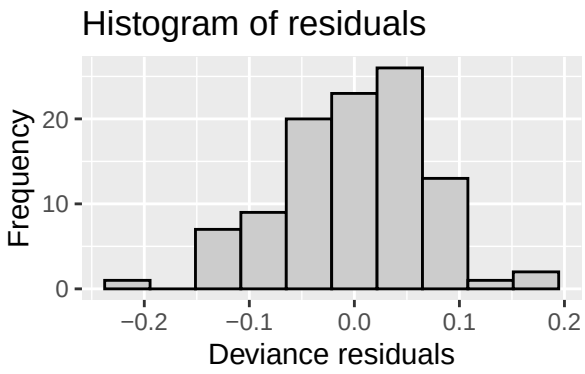

s(Week,1.67):General_GroupControl
0.02
−0.02
−0.06
1
2
3
4
5
6
7
8
Week

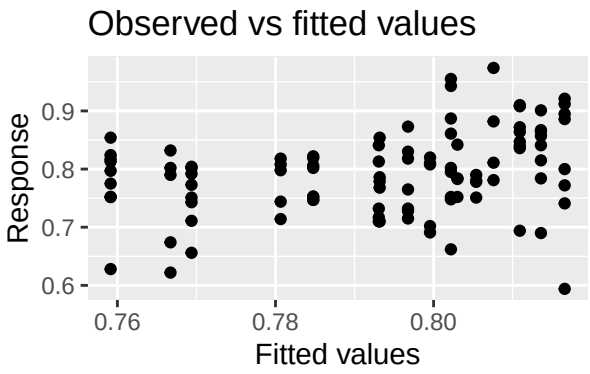



```
appraise(StO2_GAM)
```

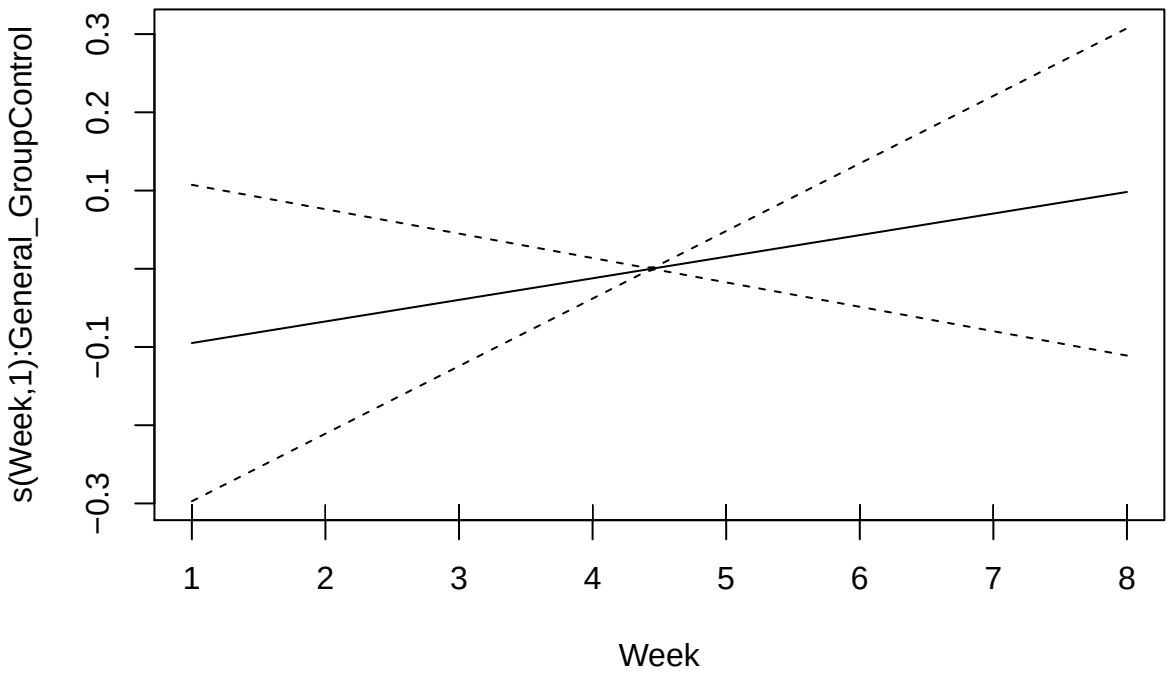


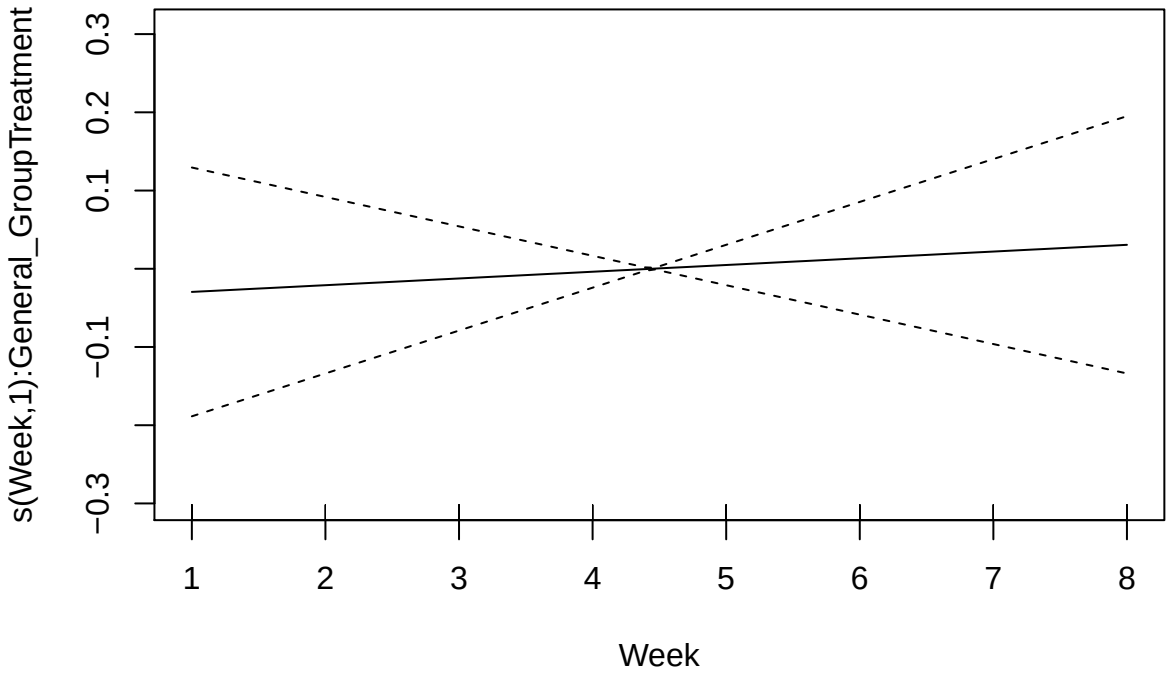




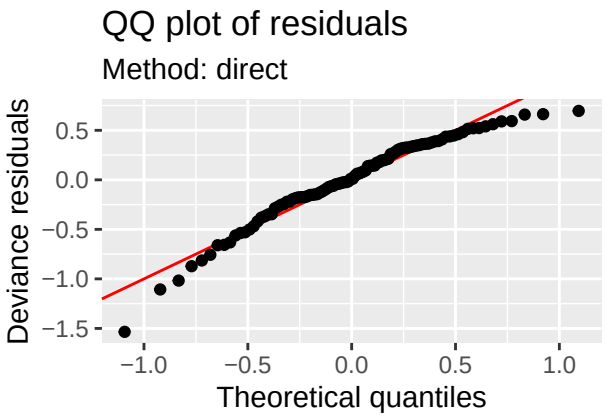


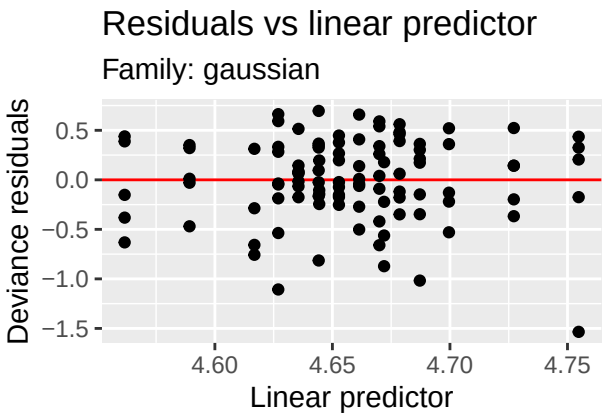

```
tHb_GAM<-gam(tHb~s(Week,by=General_Group,k=7),data=dat)
plot(tHb_GAM)
```

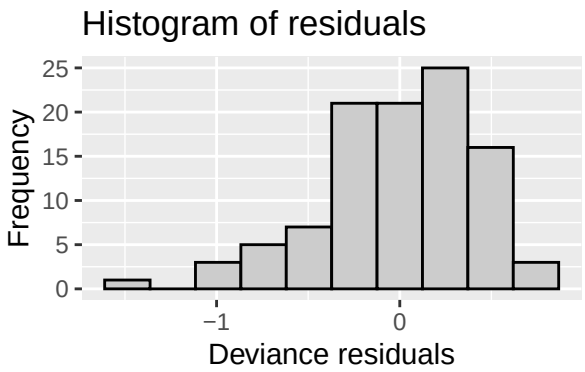

s(Week,1):General_GroupControl
−0.3
−0.1
0.1
0.2
0.3
1
2
3
4
5
6
7
8
Week

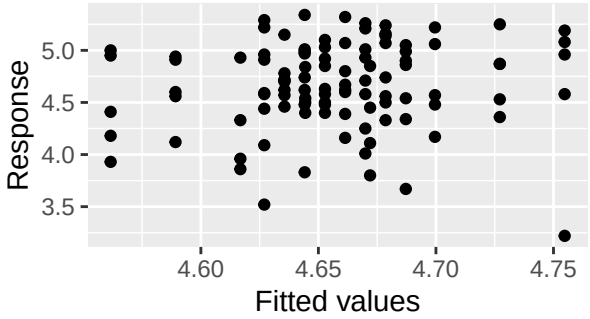



```
appraise(tHb_GAM)
```

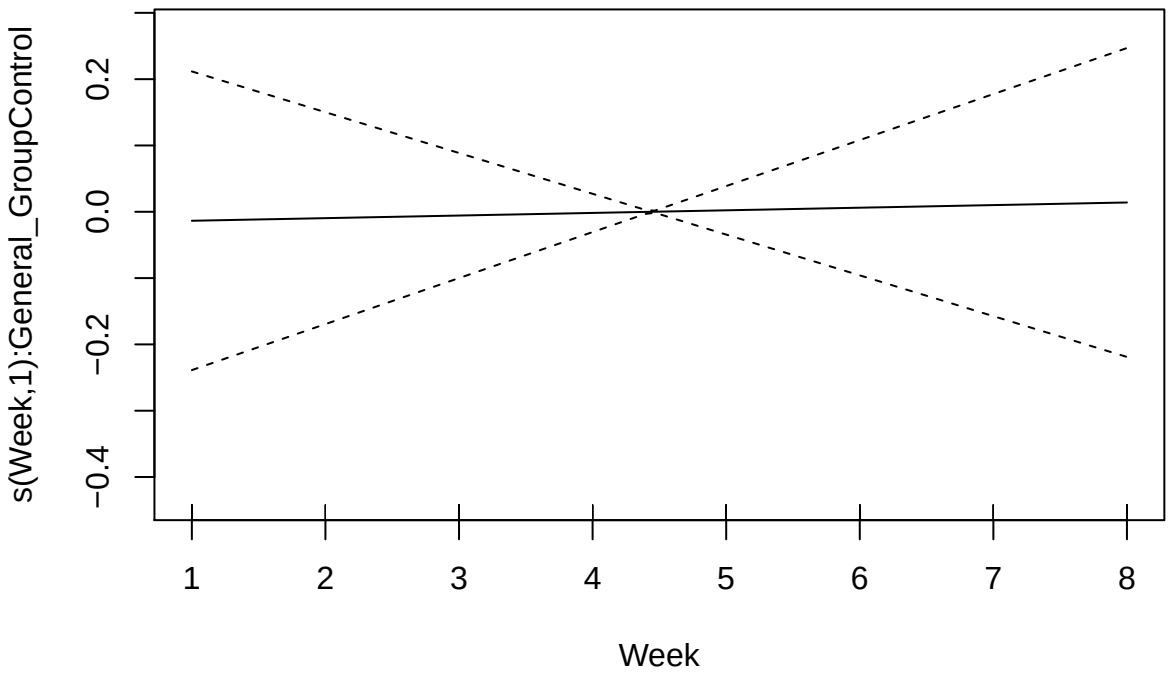


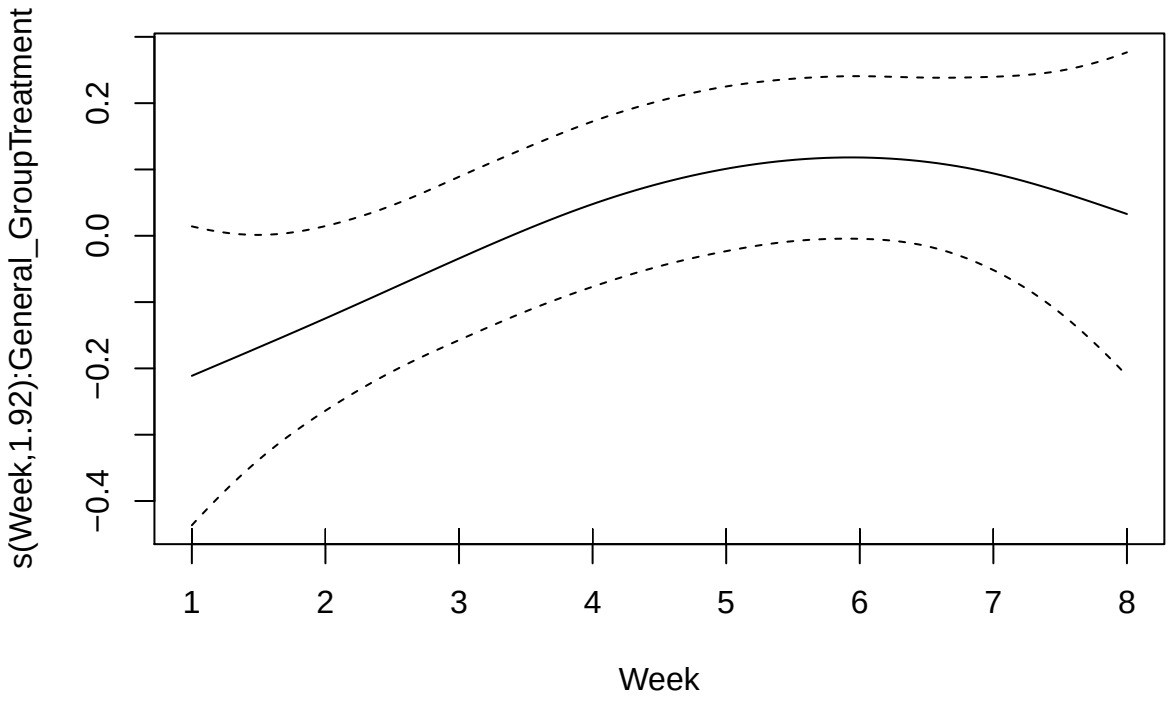




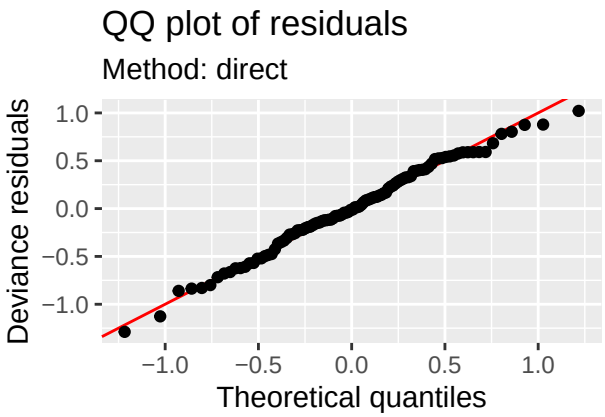


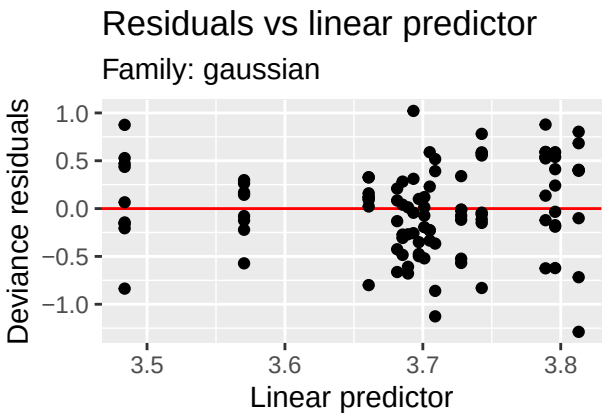

```
HbO2_GAM<-gam(HbO2~s(Week,by=General_Group,k=7),data=dat)
plot(HbO2_GAM)
```

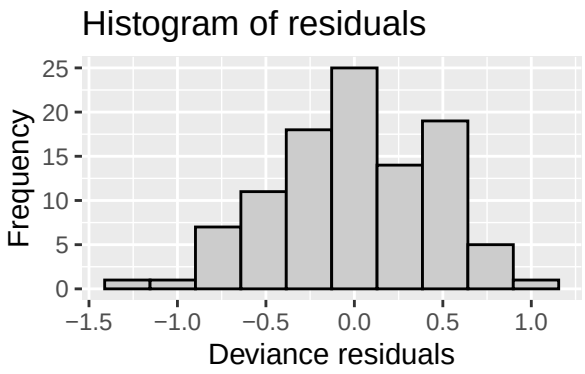

s(Week,1):General_GroupControl
0.2
0.0
−0.2
−0.4
1
2
3
4
5
6
7
8
Week

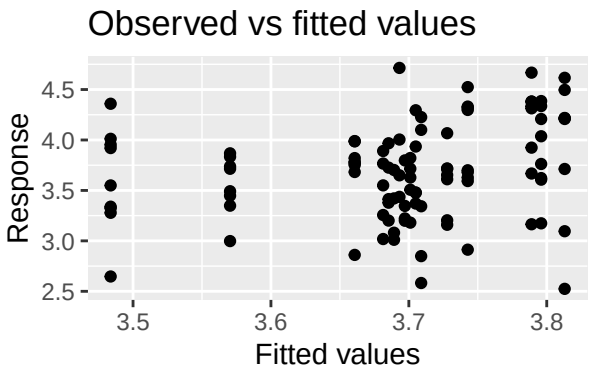



```
appraise(HbO2_GAM)
```

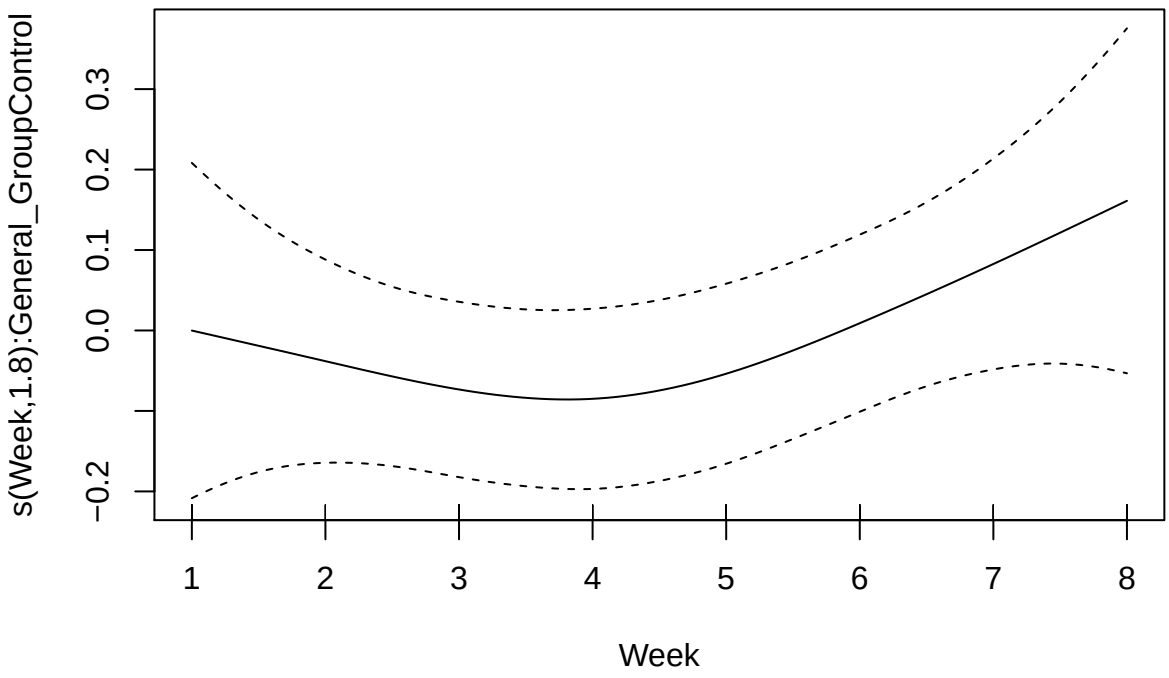


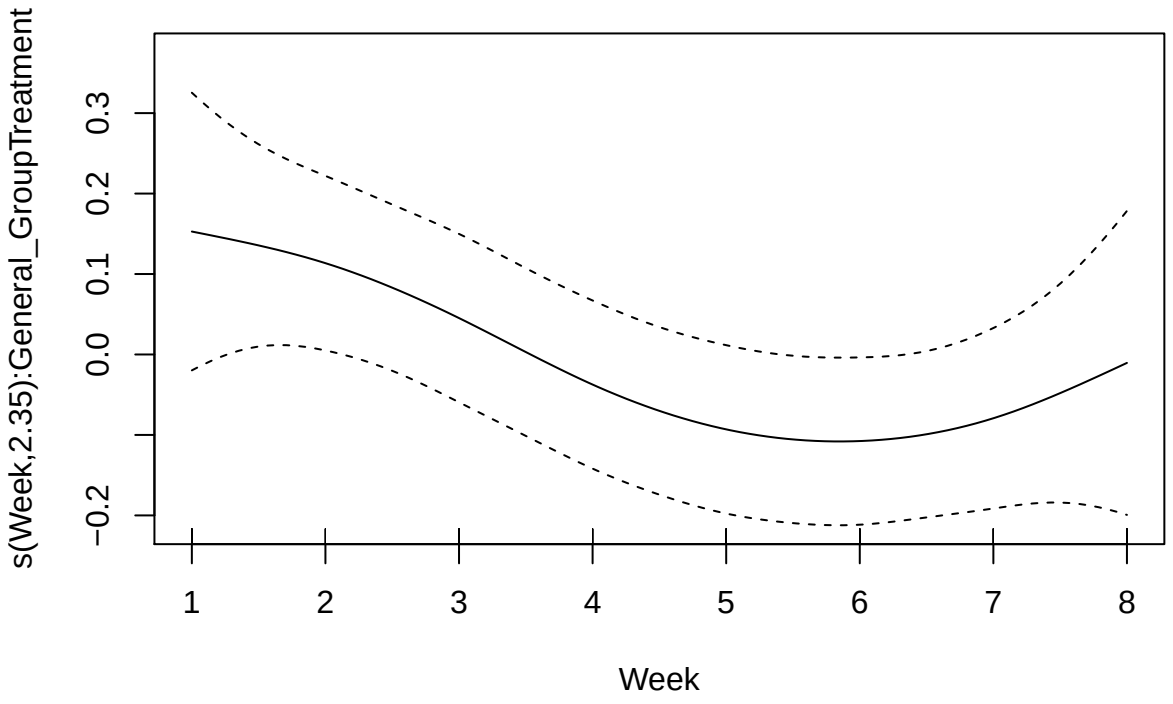




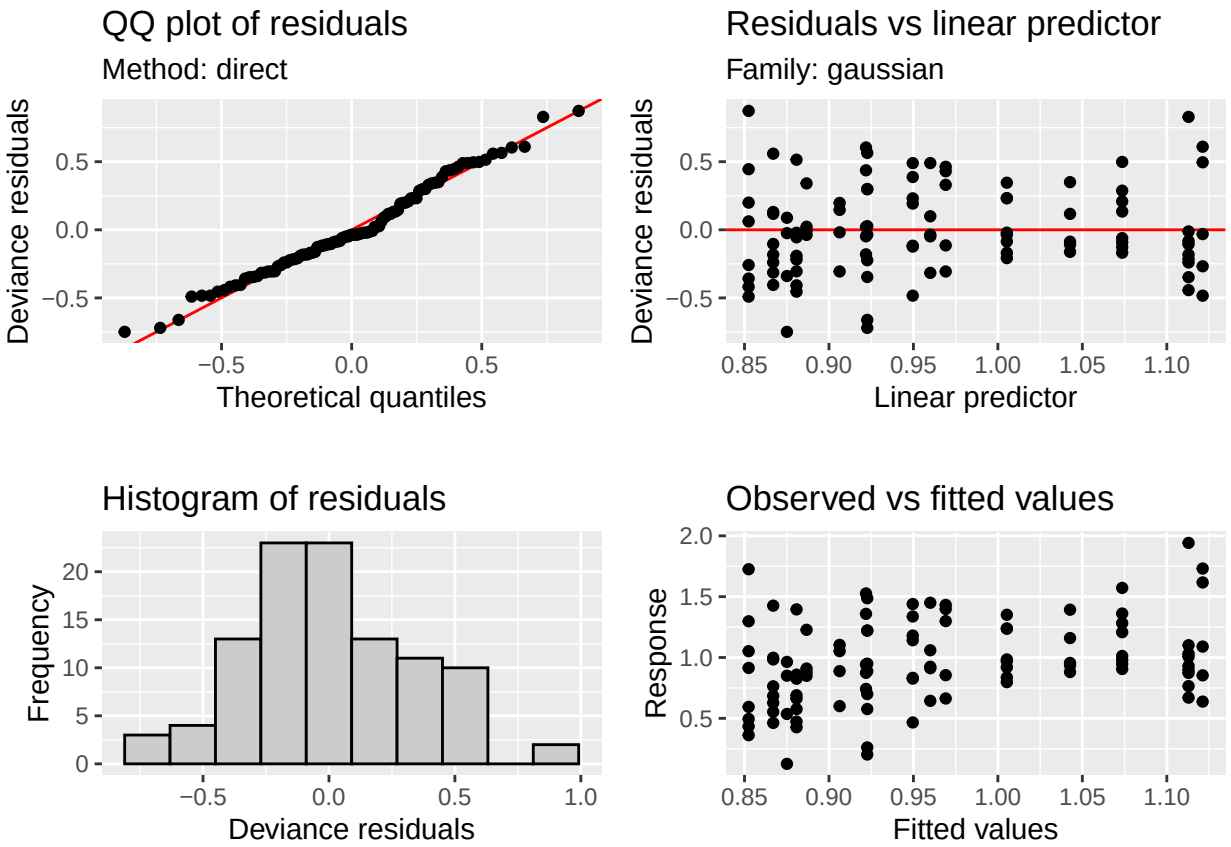



```
dHb_GAM<-gam(dHb~s(Week,by=General_Group,k=7),data=dat)
plot(dHb_GAM)
```


s(Week,1.8):General_GroupControl
−0.2
0.0
0.1
0.2
0.3
1
2
3
4
5
6
7
8
Week




```
appraise(dHb_GAM)
```



From the plots above (specifically, the plot for each model) it seems there are different smooths per treatment for StO2, HbO2 and dHb. However, there seems not to be a clear difference for the tHb model. This is somewhat expected based on what we have seen in the tumor model, but also from the initial data scatterplots, there is not a clear difference in the trends of this value over time. The diagnostics for the models of StO2, HbO2 and dHB also look good: the QQ plot shows that the residuals are following a normal distribution and the the residuals vs linear predictor plot show no trend (which is what we want because the residuals should be randomly distributed).

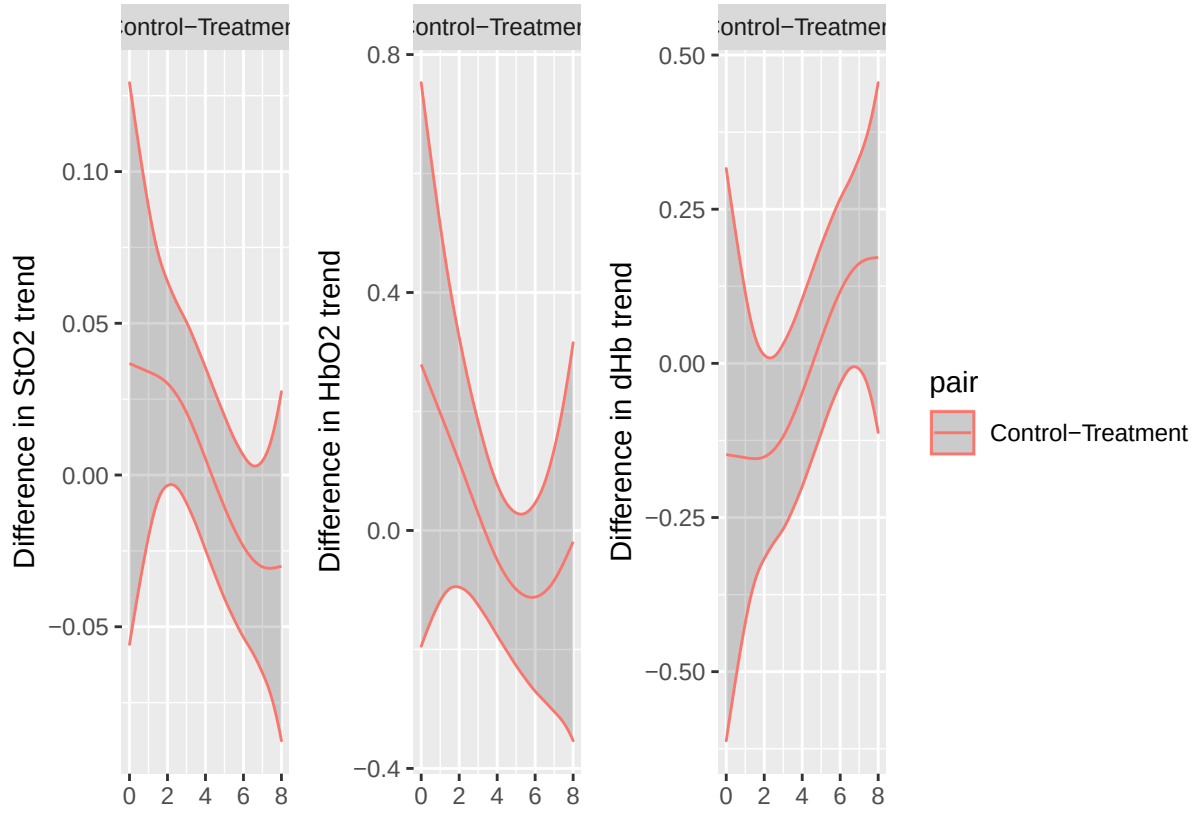
The final piece of this analysis is to do a comparison between the different smooths for each group for each model. Basically, we will not get a *p-value* as we would do for an ANOVA analysis, but insted we are going to compare the smooth and its confidence interval for each treatment group for each DRS-derived parameter. We can think about this comparison in the following way: Say you have two treatments and the trend/smooth across time in each treatment is following the same pattern. If you calculate the *difference* between the smooths it should be near or equal to zero, because the trend is the same. That would mean that the treatments are having a similar effect (which is typically what we don't want). But if in turn we find that the difference is non-zero at some time point, that would mean that the treatment is causing a change in the response, and that is what we want. That is the main idea.

To do this comparisons we need to do some additional coding because of how the GAM models are computed in R. We basically create a grid of points that spans 8 weeks and get the information of the models over this grid to calculate the trends. We will also create 95% confidence intervals for the smooths (that is the line is doing in the code below).

```
#creating grid of points of Day to evaluate smooths
pdat <- expand.grid(Week = seq(0, 8, length = 400),
                    General_Group = c('Treatment', 'Control'))
```

```r
smooth_diff <- function(model, newdata, f1, f2, alpha = 0.05,
                        unconditional = FALSE) {
    xp <- predict(model, newdata = newdata, type = 'lpmatrix')
    c1 <- grepl(f1, colnames(xp))
    c2 <- grepl(f2, colnames(xp))
    #r1 <- newdata[[var]] == f1
    #r2 <- newdata[[var]] == f2
    r1 <- with(newdata, General_Group == f1)
    r2 <- with(newdata, General_Group == f2)

    ## difference rows of xp for data from comparison
    X <- xp[r1, ] - xp[r2, ]
    ## zero out cols of X related to splines for other lochs
    X[, ! (c1 | c2)] <- 0
    ## zero out the parametric cols
    X[, !grepl('^s\\(', colnames(xp))] <- 0
    dif <- X %*% coef(model)
    se <- sqrt(rowSums((X %*% vcov(model, unconditional = unconditional)) * X))
    crit <- qt(alpha/2, df.residual(model), lower.tail = FALSE)
    upr <- dif + (crit * se)
    lwr <- dif - (crit * se)
    data.frame(pair = paste(f1, f2, sep = '-'),
               diff = dif,
               se = se,
               upper = upr,
               lower = lwr)
}

comp1<-smooth_diff(StO2_GAM,pdat,'Control','Treatment')
comp_StO2 <- cbind(Week = seq(0, 8, length = 400),
              rbind(comp1))


c1<-ggplot(comp_StO2, aes(x = Week, y = diff, group = pair,color=pair)) +
    geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2) + geom_line() +
  facet_wrap(~ pair) +
     labs(x = NULL, y = 'Difference in StO2 trend')


comp2<-smooth_diff(HbO2_GAM,pdat,'Control','Treatment')
comp_HbO2 <- cbind(Week = seq(0, 8, length = 400),
              rbind(comp2))


c2<-ggplot(comp_HbO2, aes(x = Week, y = diff, group = pair,color=pair)) +
    geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2) + geom_line() +
  facet_wrap(~ pair) +
     labs(x = NULL, y = 'Difference in HbO2 trend')



comp3<-smooth_diff(dHb_GAM,pdat,'Control','Treatment')
```

```
comp_dHb <- cbind(Week = seq(0, 8, length = 400),
                  rbind(comp3))


c3<-ggplot(comp_dHb, aes(x = Week, y = diff, group = pair,color=pair)) +
    geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2) + geom_line() +
  facet_wrap(~ pair) +
     labs(x = NULL, y = 'Difference in dHb trend')


c1+c2+c3+plot_layout(guides="collect")
```



From the plots it seems that there are different trends in the response, but they don't get to the point of being significant (all the intervals pass at 0 at all timepoints). This makes sense because if you look at the original scatterplots, the values are close between treatment groups (for example, StO2 trends have maximum and minimum values around 0.9 and 0.6 in both groups). This would be a case where more samples are needed to make a better distinction between the groups (the famous issue of "power analysis"). I would say that your results are good, for the small scale of your study you are showing trends, but it is expected to not have a huge gap between the treatment groups at this point.