

Report Pilot CFTR

“Dario Righelli”

“Institute for Applied Mathematics”M.Picone” - CNR“

“DISA-MIS - Univeristy of Salerno”

3/6/2018

Loading dataset

```
dataset <- read.xlsx("../data/CPET FC 2018 31.5.xlsx", sheetIndex=1, as.data.frame=TRUE, header=TRUE)
colnames(dataset) <- gsub("__", "_", gsub("\\.", "_", colnames(dataset)))
colnames(dataset) <- gsub("X_", "", colnames(dataset))
colnames(dataset) <- gsub("__", "_", colnames(dataset))
colnames(dataset) <- gsub("_$", "", colnames(dataset))
```

computing t-tests

Without multiple testing correction (so, without comparing all the variables together) we can notice a significance just in **FEV FVC PRED**. Using a threshold on the p.values of 0.05. If we raise this threshold to 0.1, we can include **FEV FVC** and **RV** variables. Anyway three observations are too few to give a real answer of the response of these treatment. So, what we can tell is that for sure we need to extend our observations (patients) to obtain a more robust response and give a better answer to our investigation.

```
t.p <- c()
tt <- c()
idx.pre <- grep(pattern="PRE", x=colnames(dataset))
for(i in idx.pre)
{
  t <- t.test(dataset[,i], dataset[,i+1])
  t.p <- c(t.p, t$p.value)
  tt <- c(tt, t$statistic)
}

rownames <- colnames(dataset)[idx.pre]
rownames <- gsub(pattern="_PRE", replacement="", x=rownames)
tt.p <- as.data.frame(cbind(tt, t.p), row.names=rownames)
tt.p <- tt.p[order(tt.p[,2]),,drop=FALSE]
tt.p$FDR <- p.adjust(p=tt.p$t.p, method="fdr")
colnames(tt.p) <- c("t-test", "p-value", "FDR")
print(tt.p)
```

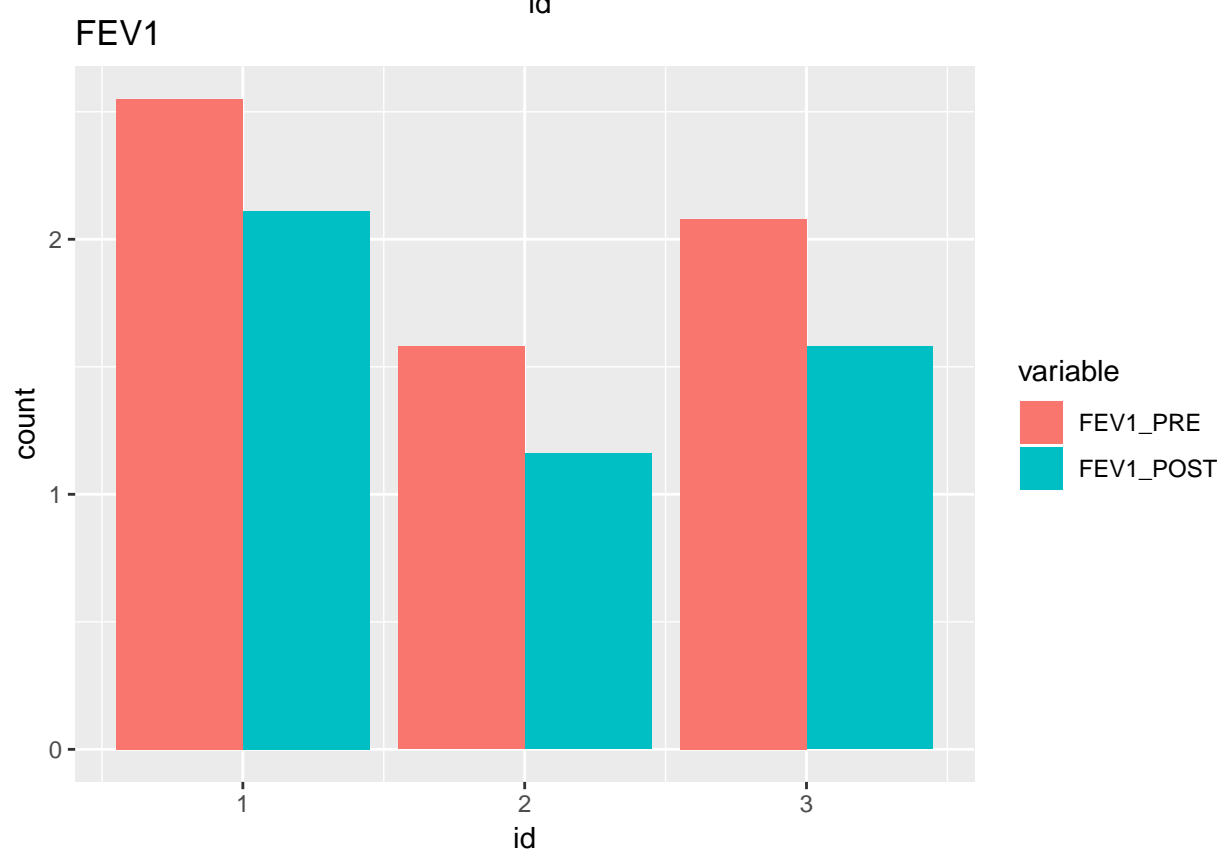
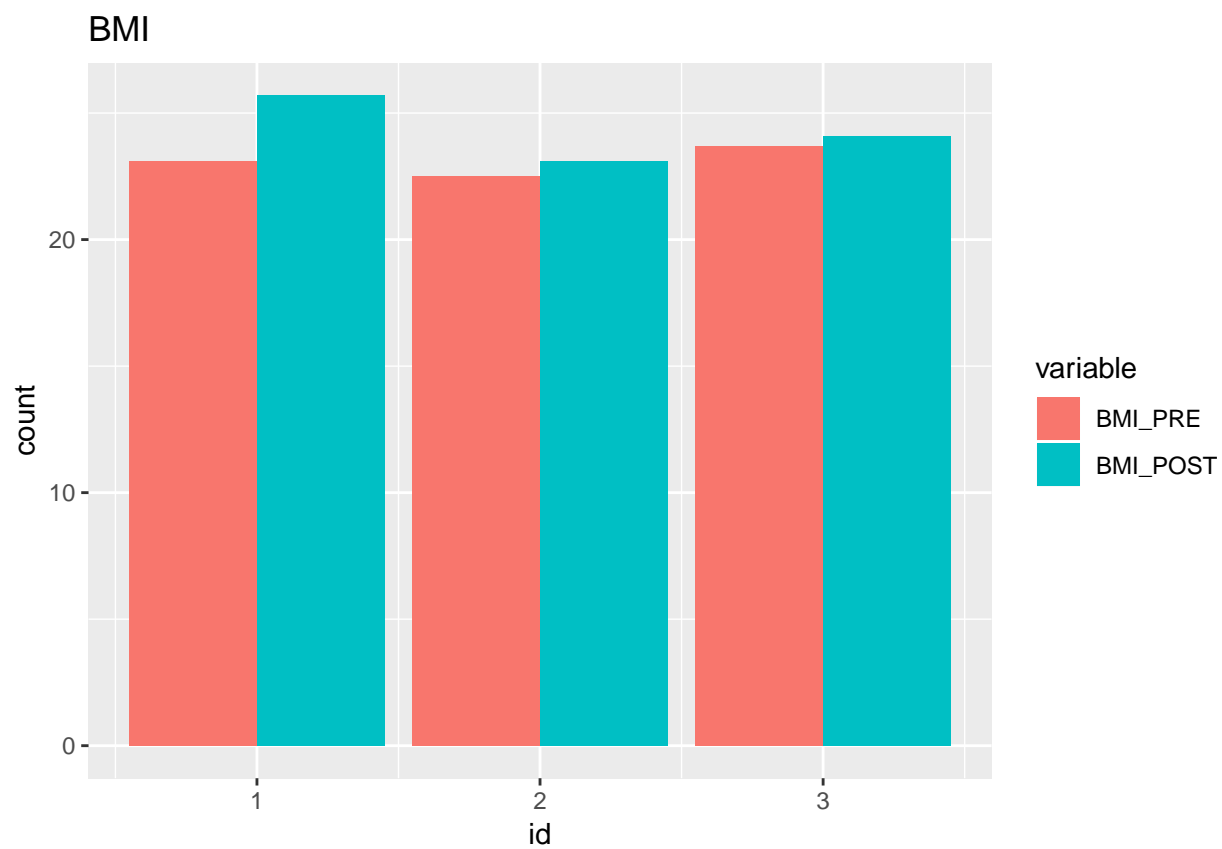
##	t-test	p-value	FDR
## FEV1_FVC_pred	3.4230125	0.03182687	0.7826012
## FEV1_FVC	2.6055562	0.05993569	0.7826012
## RV	-2.6635858	0.07586433	0.7826012
## N_total_Exacerbations	1.8898224	0.13323256	0.7826012
## RV_TLC	-1.8460735	0.14846951	0.7826012
## RV_Pred	-1.8498061	0.15970711	0.7826012
## FEF25_75	1.5499516	0.19733935	0.7826012

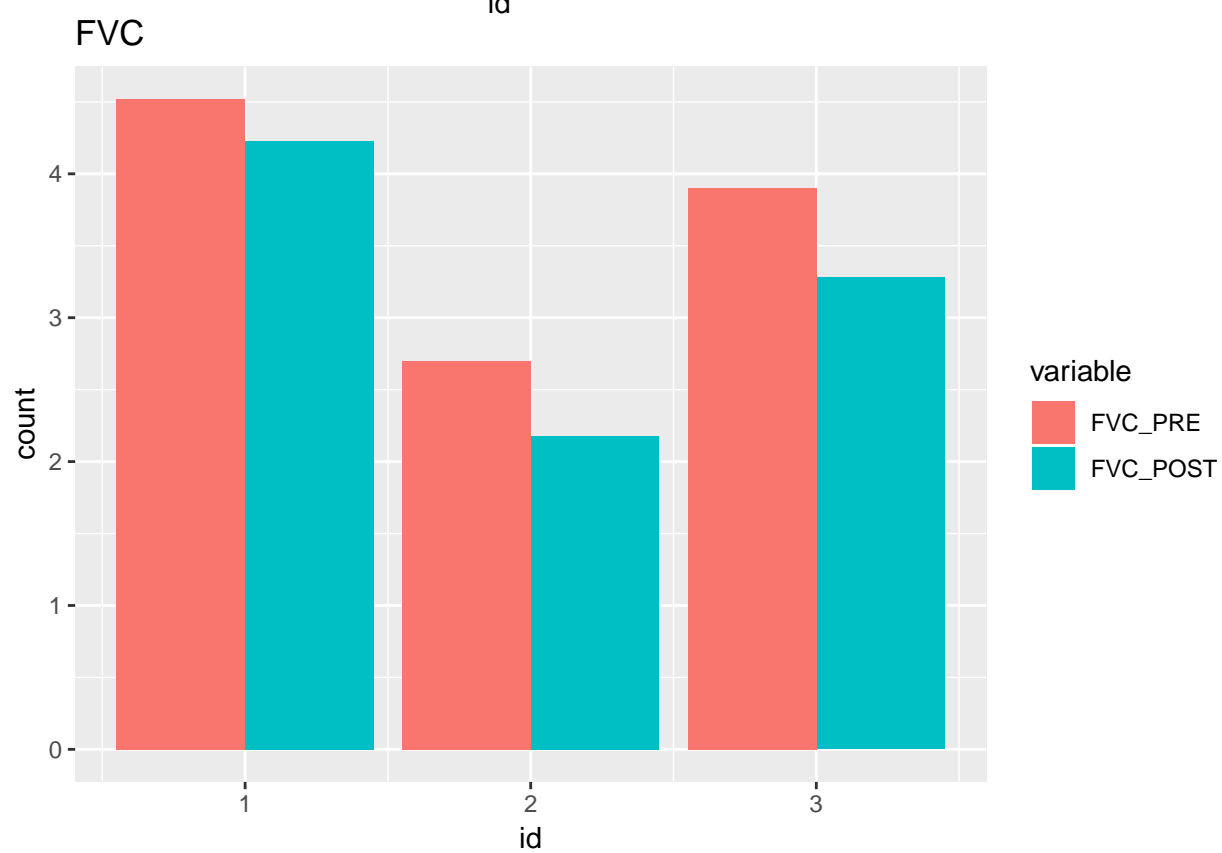
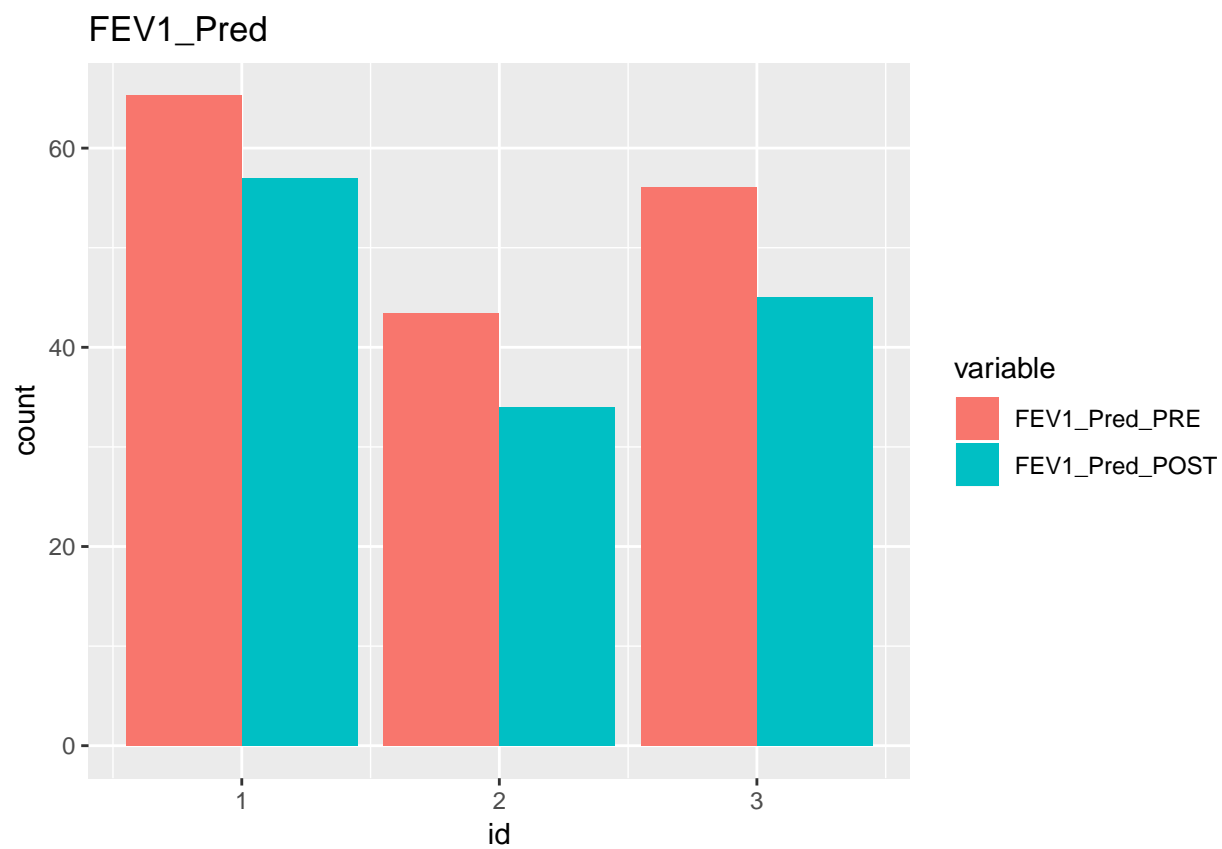
```
## Mild_activity_3_4_8_METs -1.4199011 0.22982993 0.7826012
## BMI -1.4411534 0.25126281 0.7826012
## FEF25_75_pred 1.2596913 0.27630028 0.7826012
## FEV1 1.1552837 0.31230800 0.7826012
## FRC -1.1392344 0.32239363 0.7826012
## FRC_TLC -1.1044161 0.33943000 0.7826012
## FEV1_Pred 1.0448439 0.35518849 0.7826012
## Duration_Phys_Act -1.0301143 0.37116591 0.7826012
## No_Steps 0.9827888 0.40656416 0.7826012
## FRC_Pred -1.0031249 0.41354081 0.7826012
## Vigorous_7_2METs 0.9933993 0.42286480 0.7826012
## Active_energy_exp_cal -0.9490503 0.43733597 0.7826012
## Sleep_duration -0.7709558 0.49444990 0.7923522
## Tot_energy_exp_cal -0.7241153 0.52941440 0.7923522
## RV_TLC_Pred -0.6795838 0.53533920 0.7923522
## FRC_TLC_Pred -0.7083737 0.53913896 0.7923522
## TLC_Pred -0.6503738 0.56070718 0.7923522
## FVC 0.5976049 0.58261194 0.7923522
## ModerateVigorous 0.4754234 0.66797164 0.8151006
## FVCPred 0.4598890 0.67001584 0.8151006
## TLC -0.4619758 0.67125930 0.8151006
## Moderate_4_8_7_2_METs 0.3177020 0.76991178 0.8747882
## Avg_mets -0.3123475 0.77962186 0.8747882
## Lying_Time -0.2798867 0.79760098 0.8747882
## Time_on_body 0.2242114 0.83376249 0.8858726
## DLCO_Pred -0.1571522 0.88305584 0.8940739
## DLCO 0.1439959 0.89407394 0.8940739
```

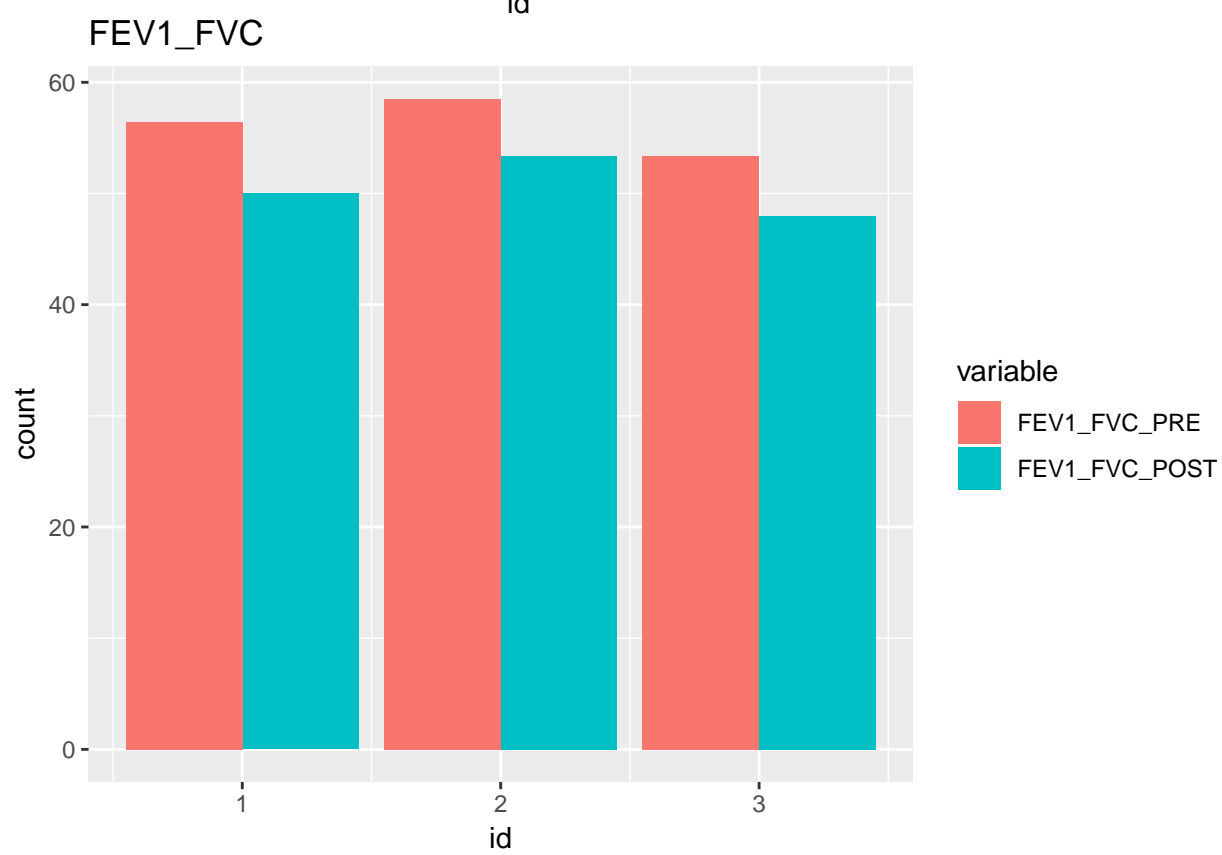
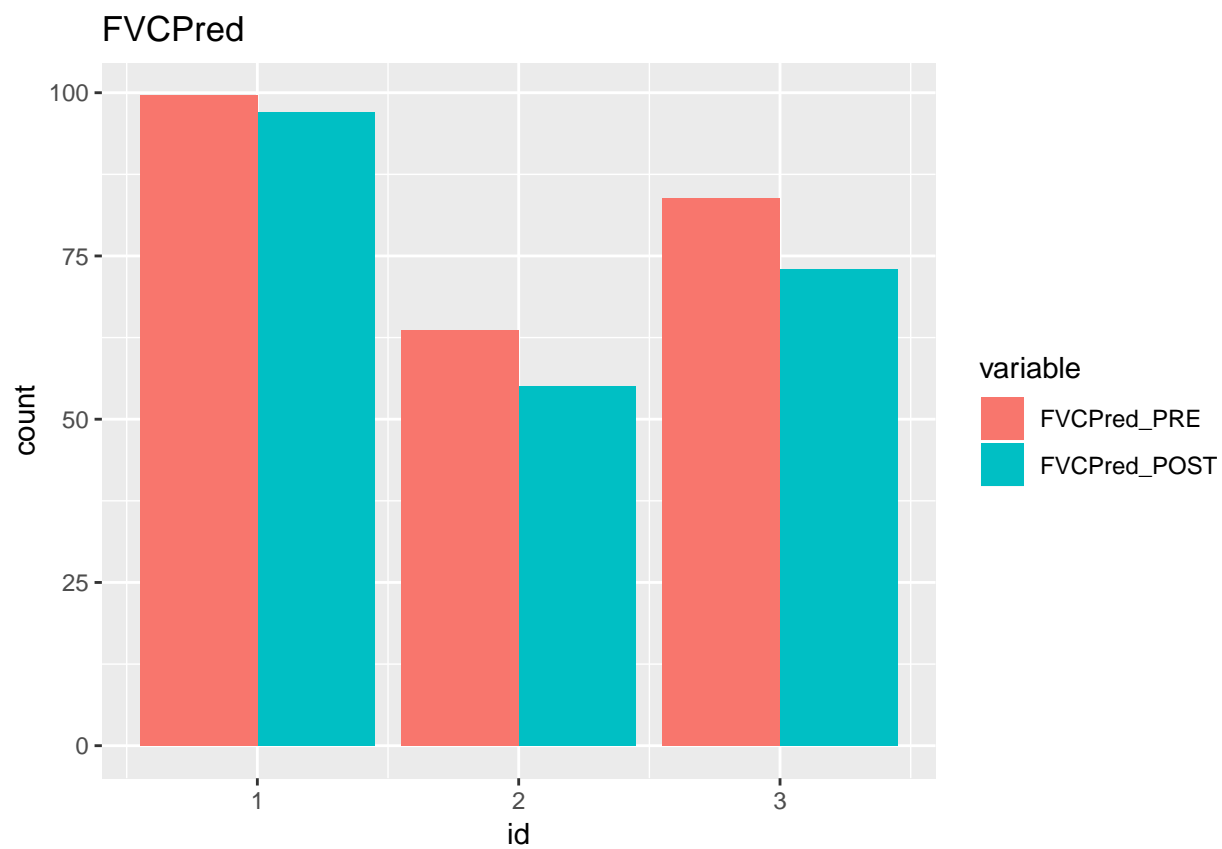
barplots

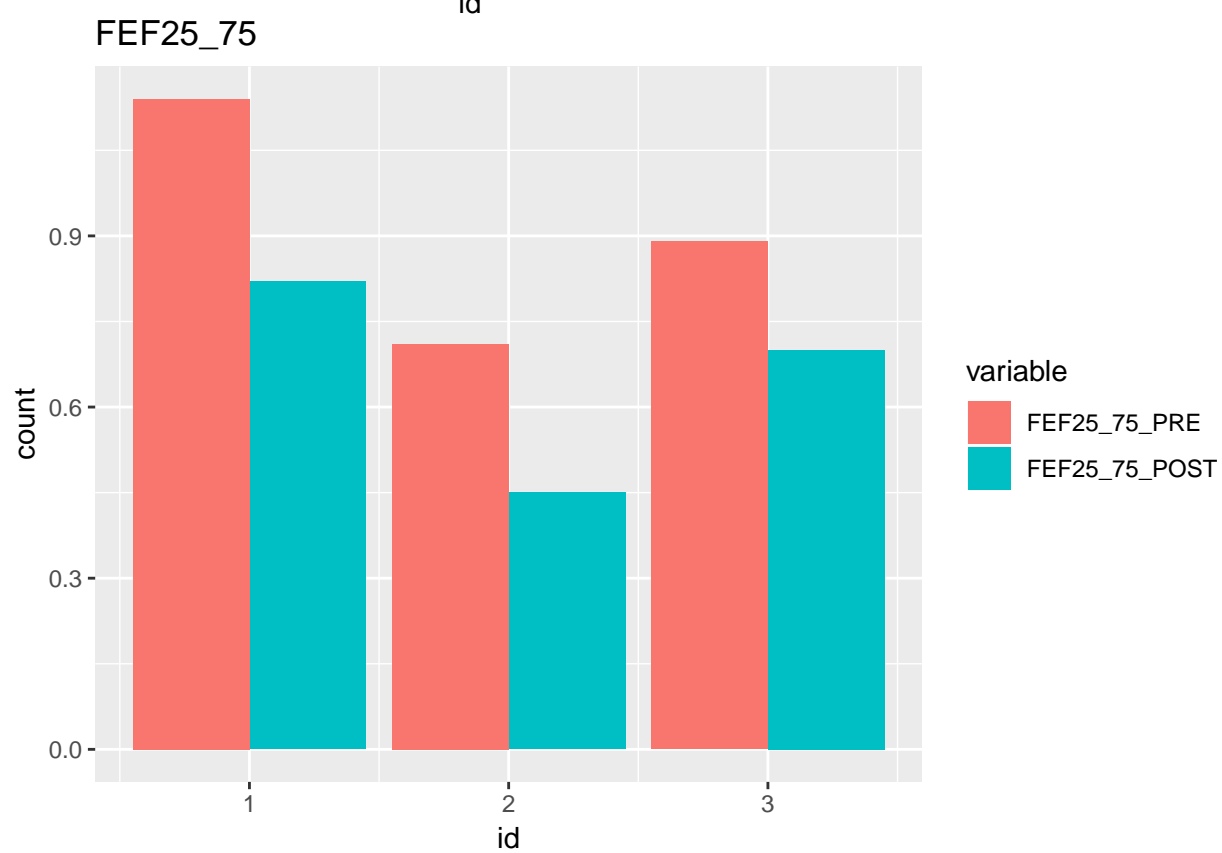
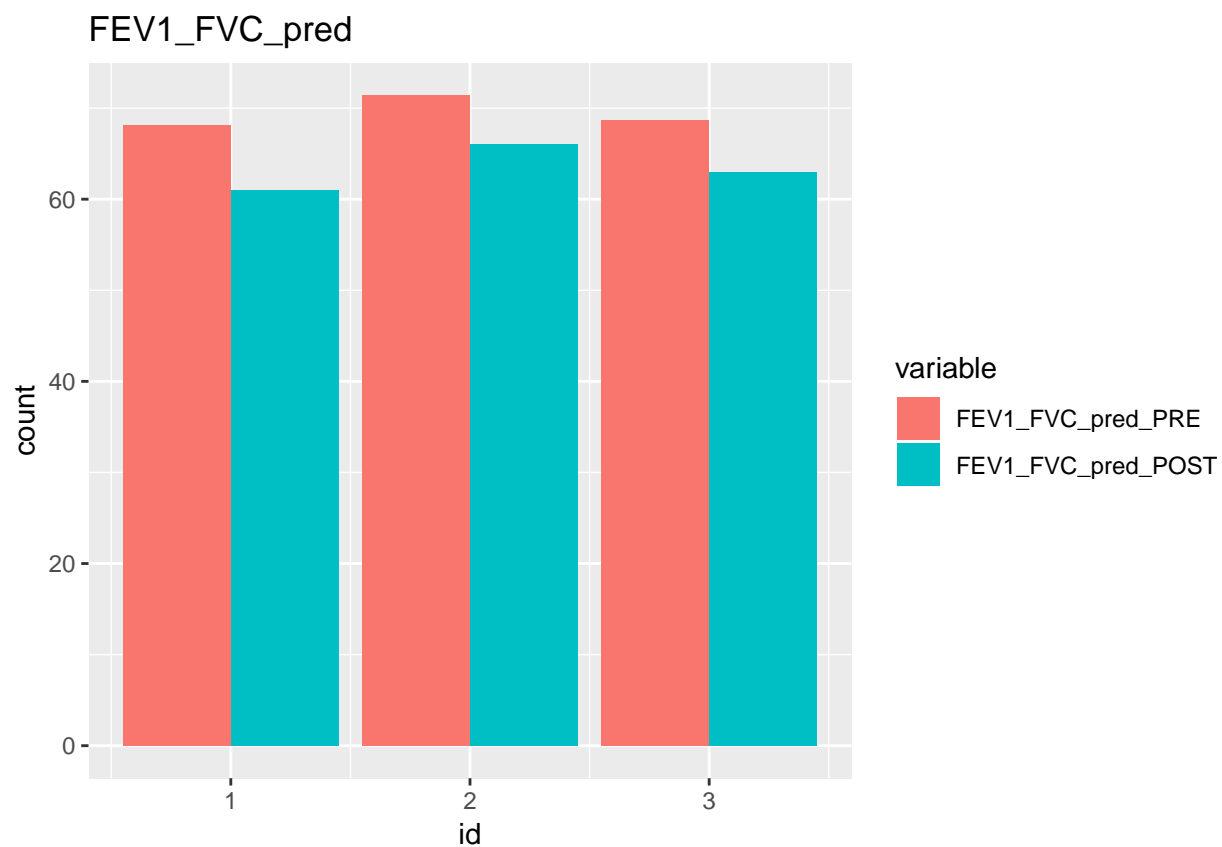
Here we plot for each patient his levels of each variable comparing pre and post values. When a color is not present in a bar, means that the value is 0.

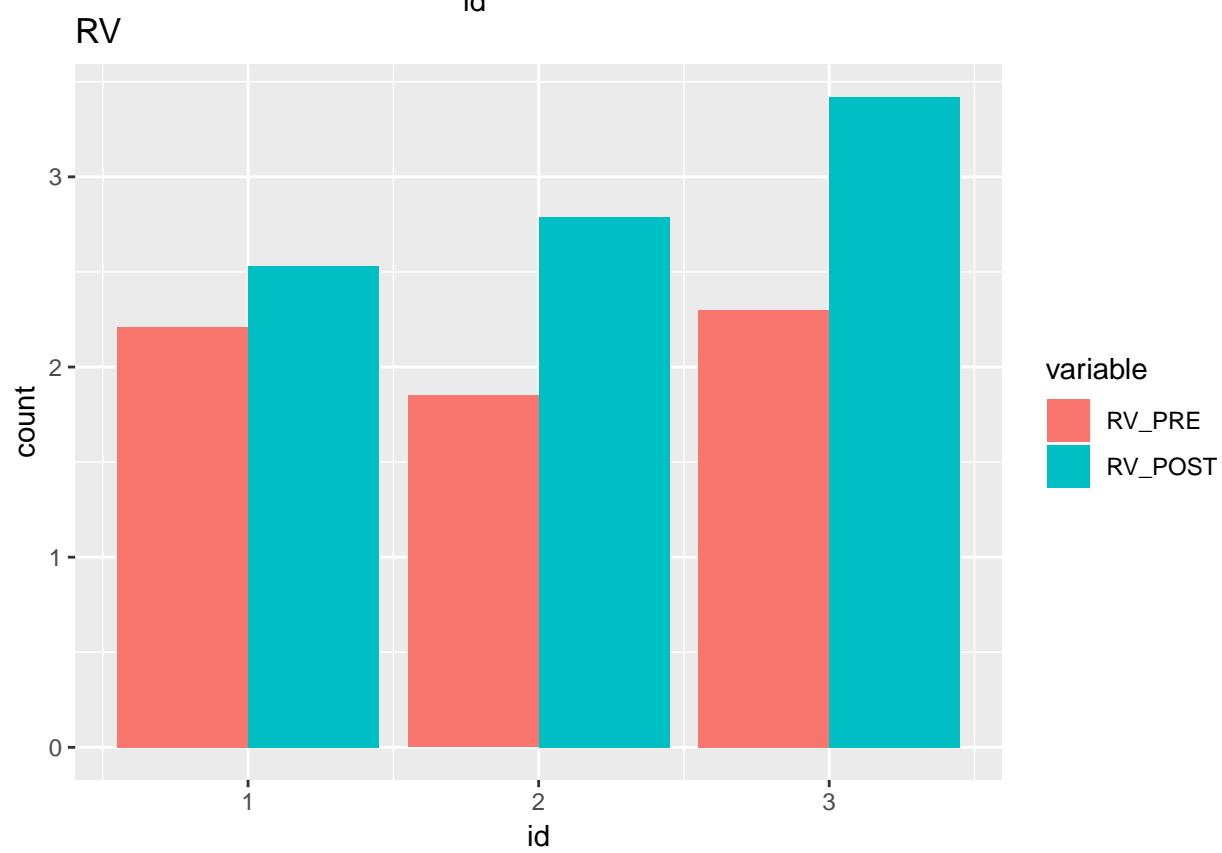
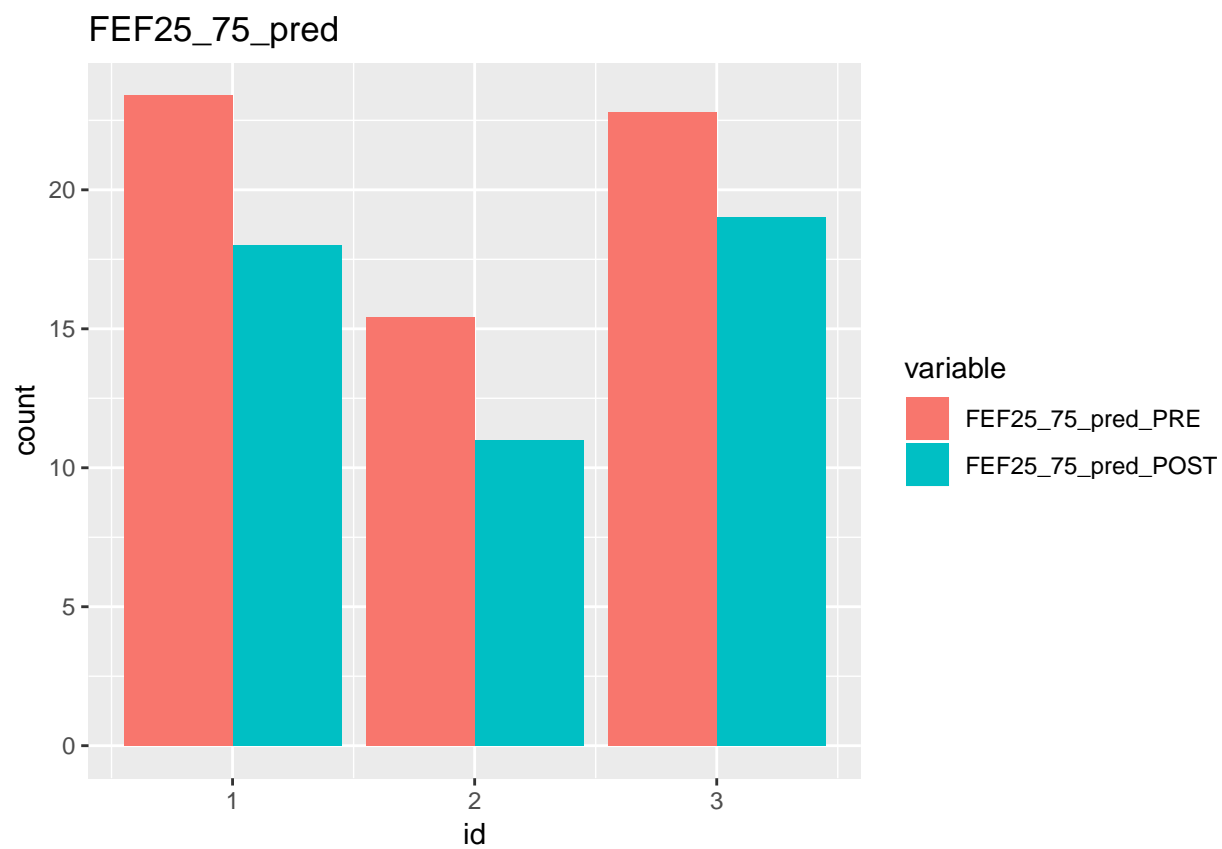
```
for(i in idx.pre)
{
  meltedi <- melt(dataset[,c(i, i+1)])
  meltedi$id <- rep(dataset$ID, 2)
  g <- ggplot(melted, aes(id)) + geom_bar(mapping=aes(weight=value, fill=variable), data=melted, position="dodge")
  print(g)
}
```

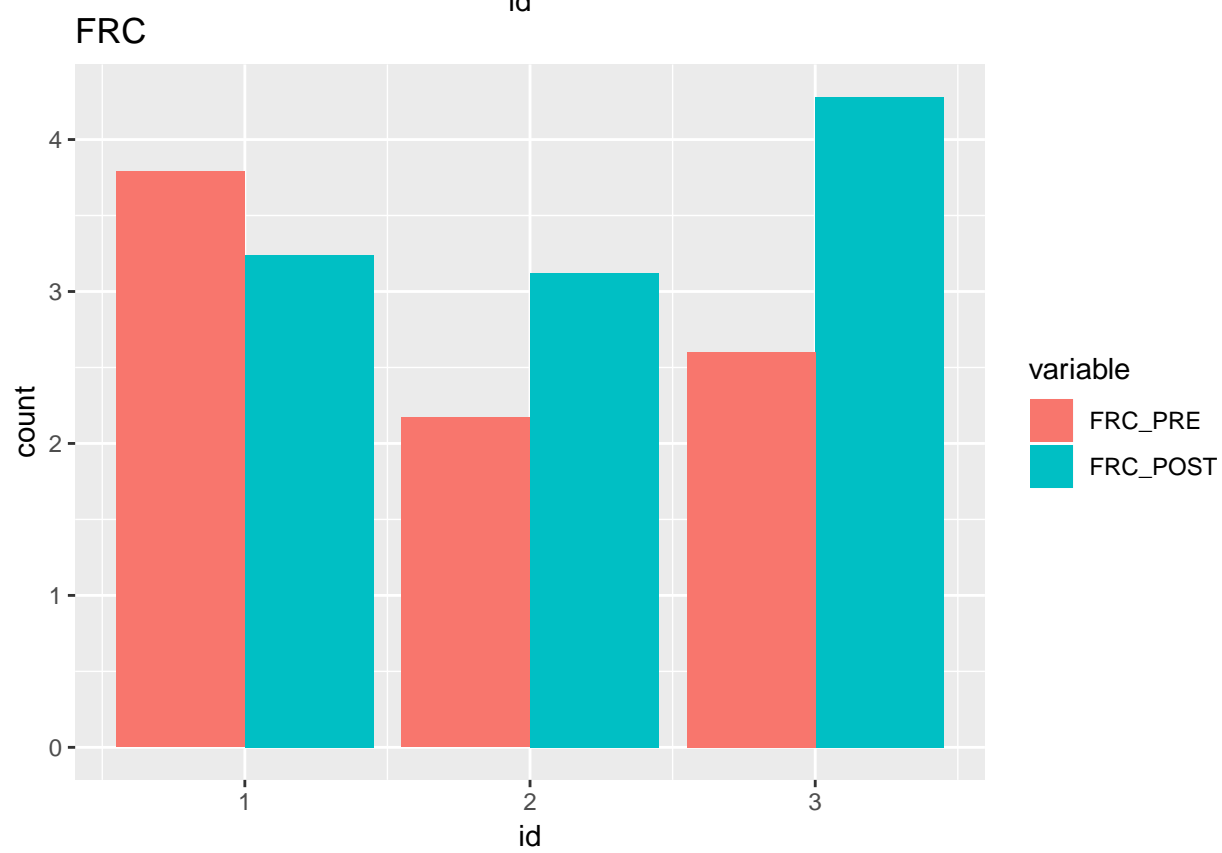
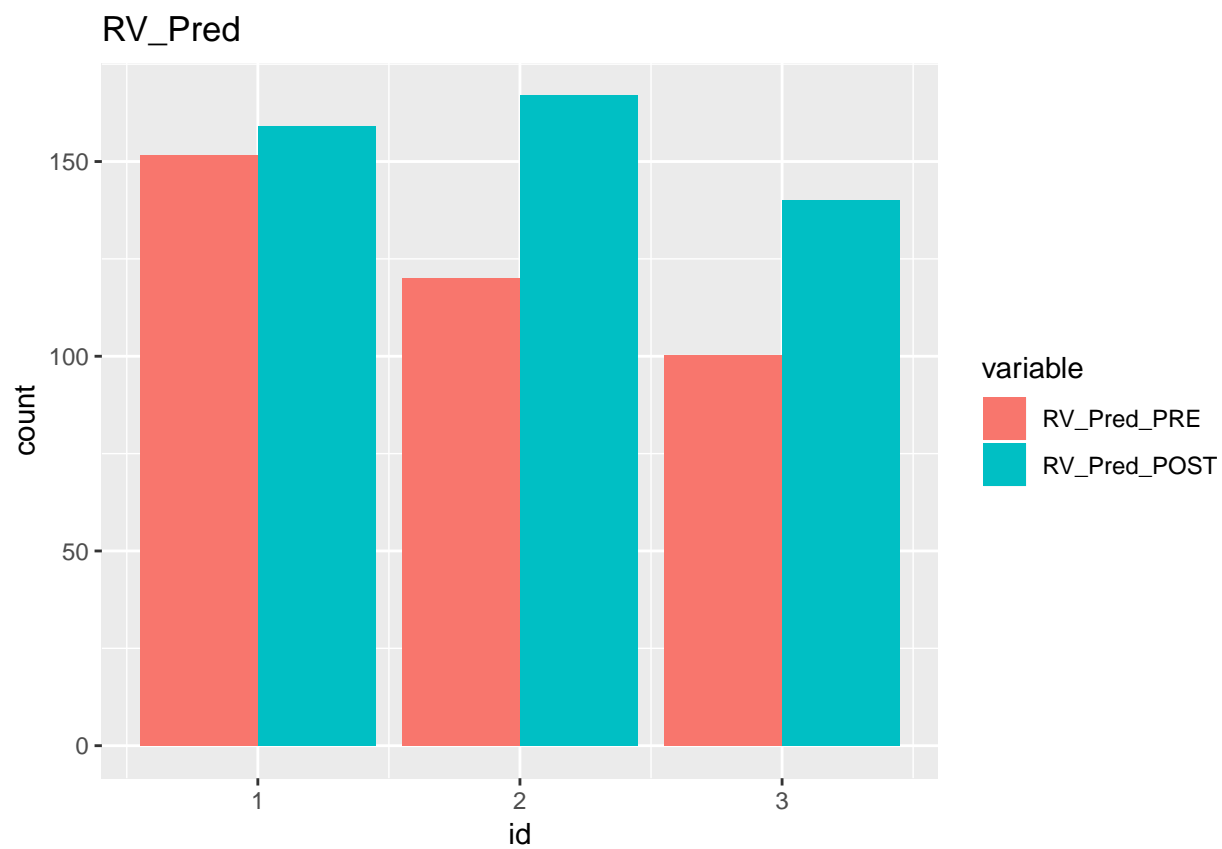


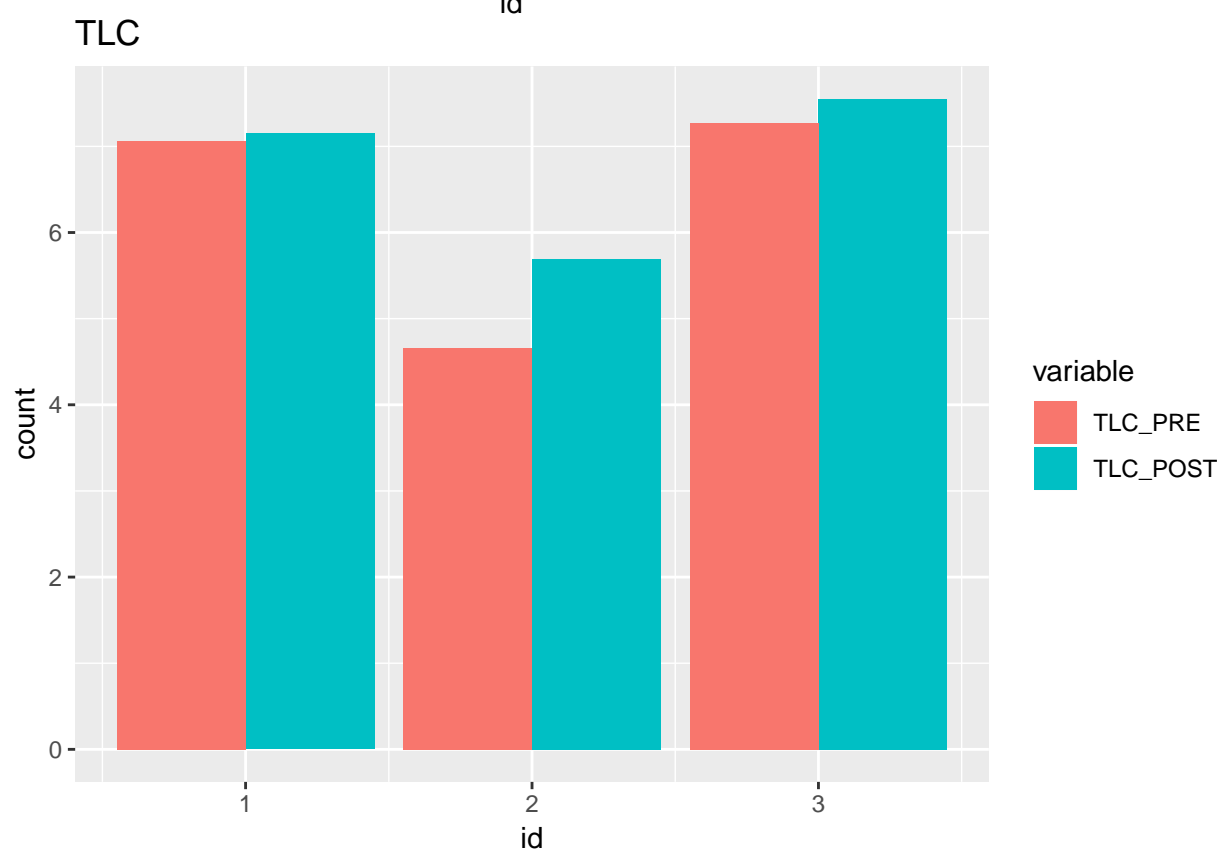
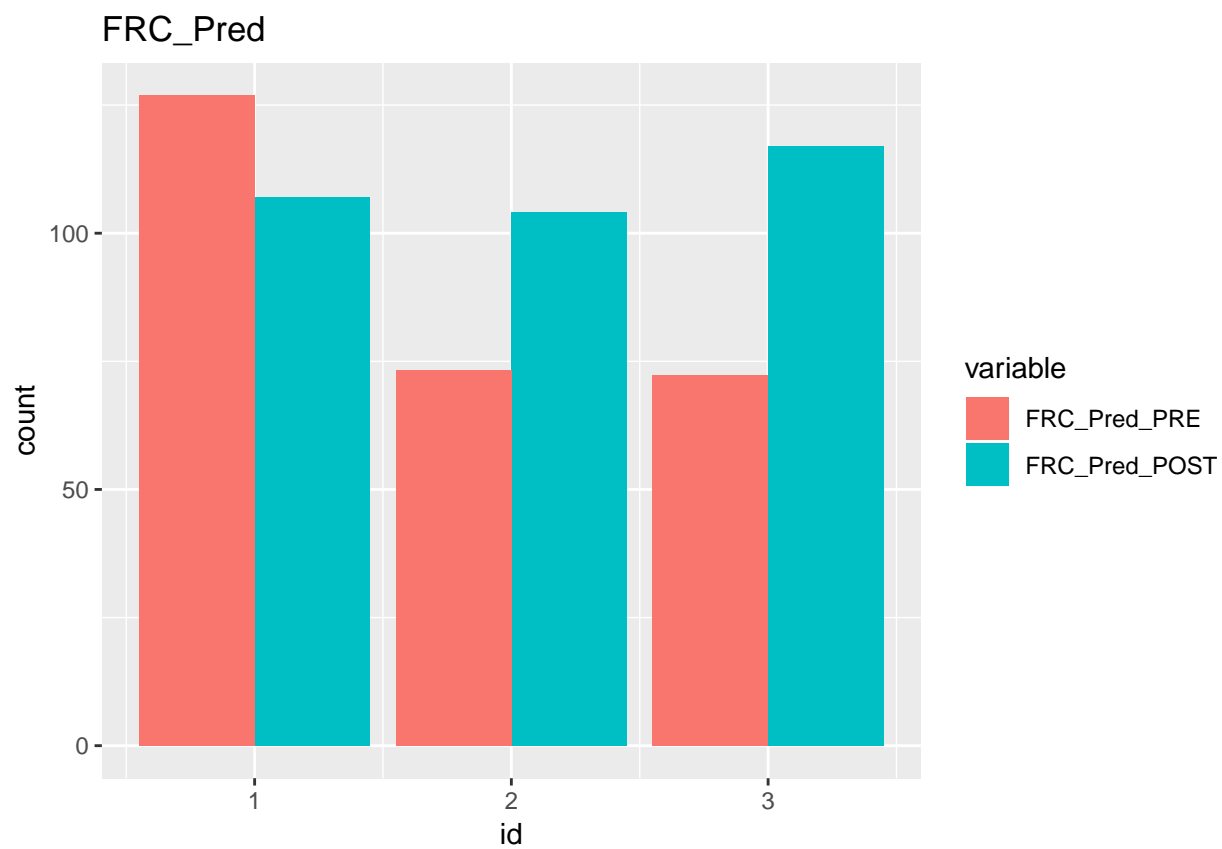


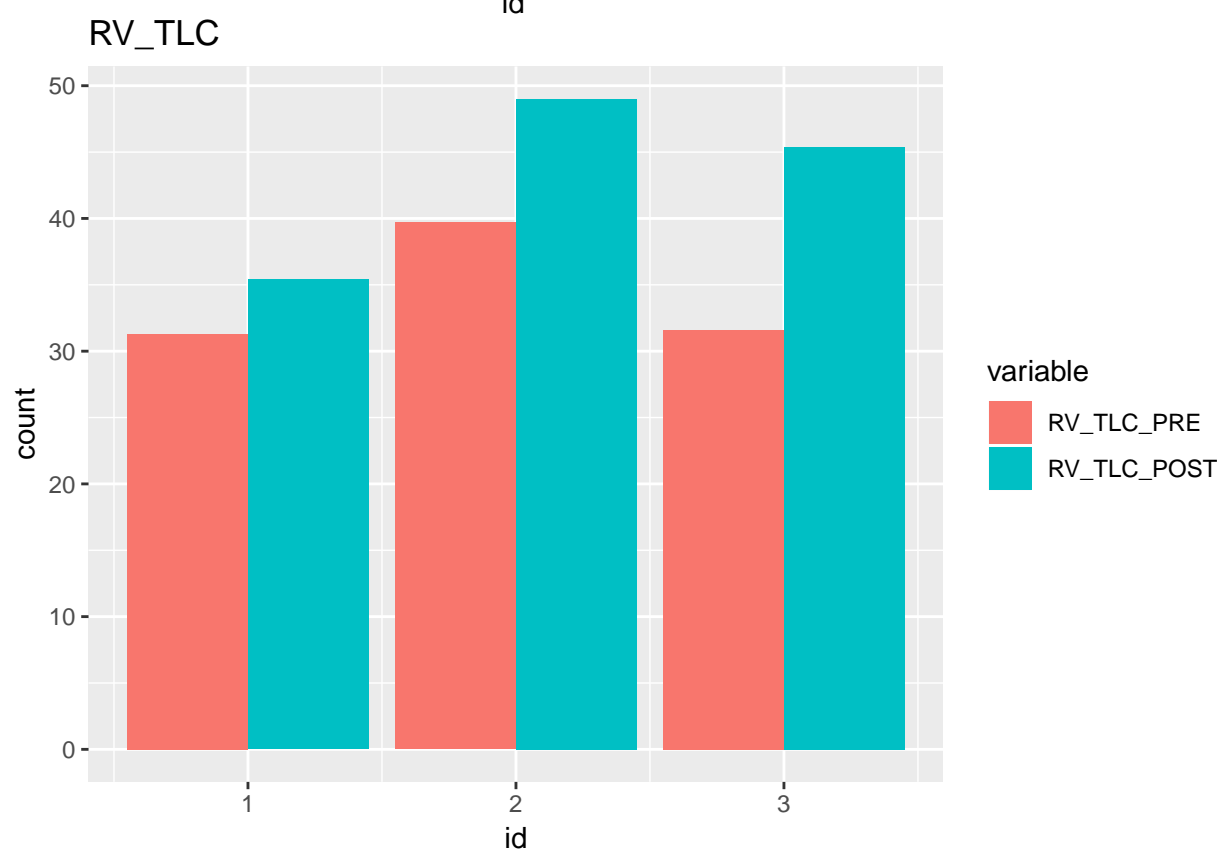
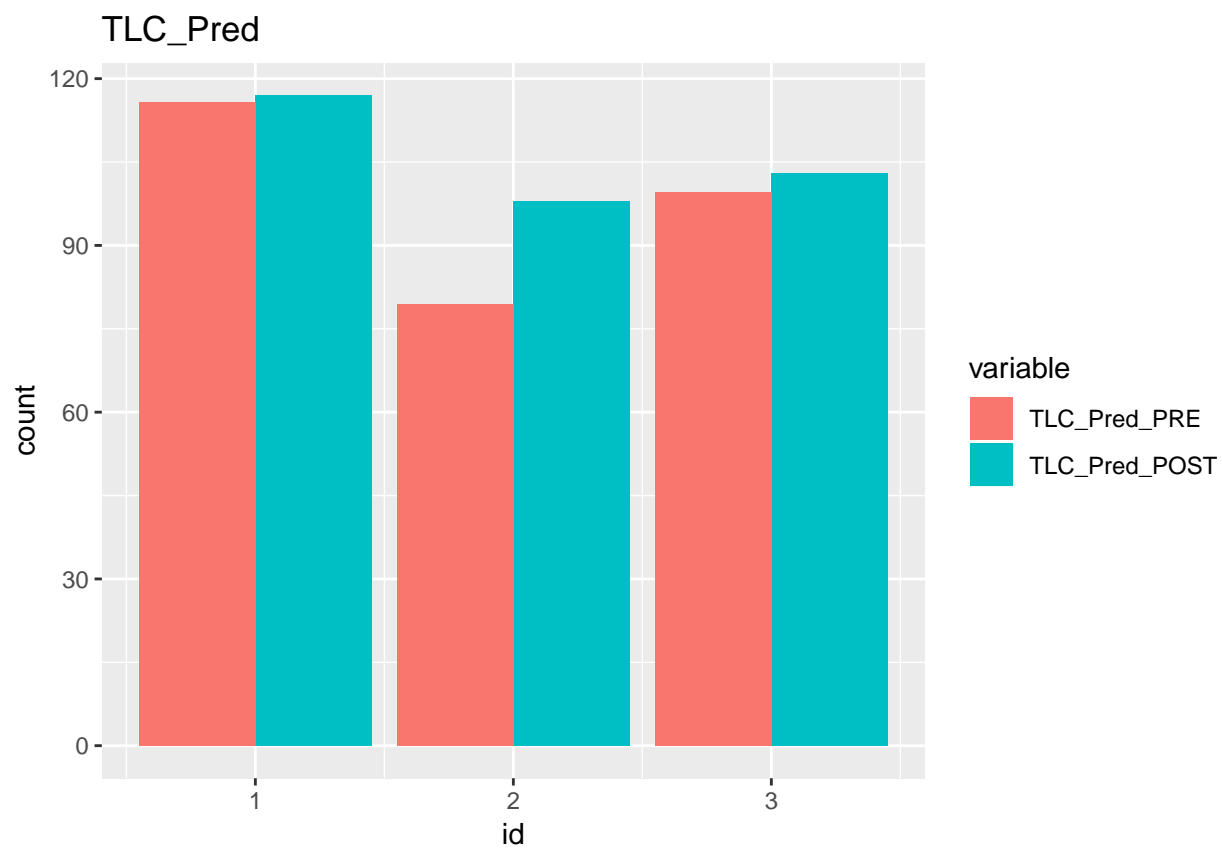


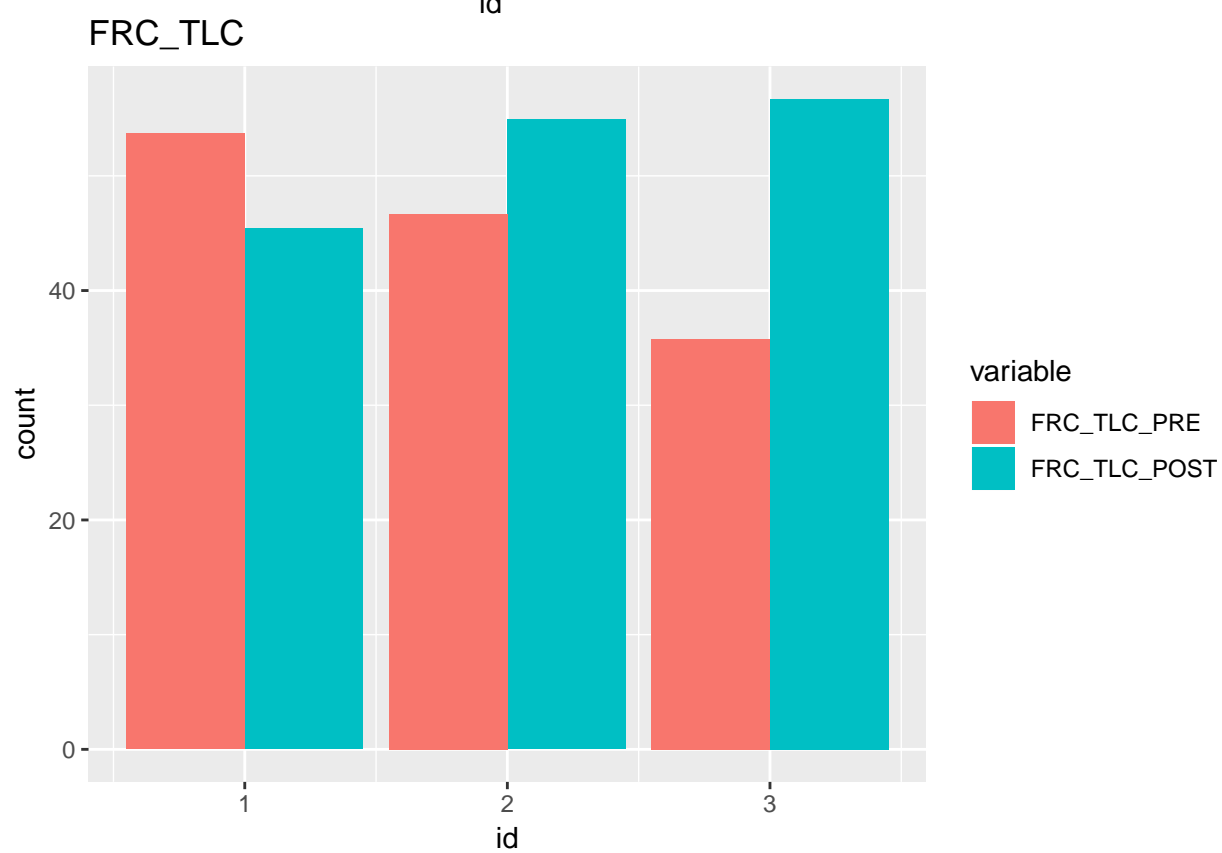
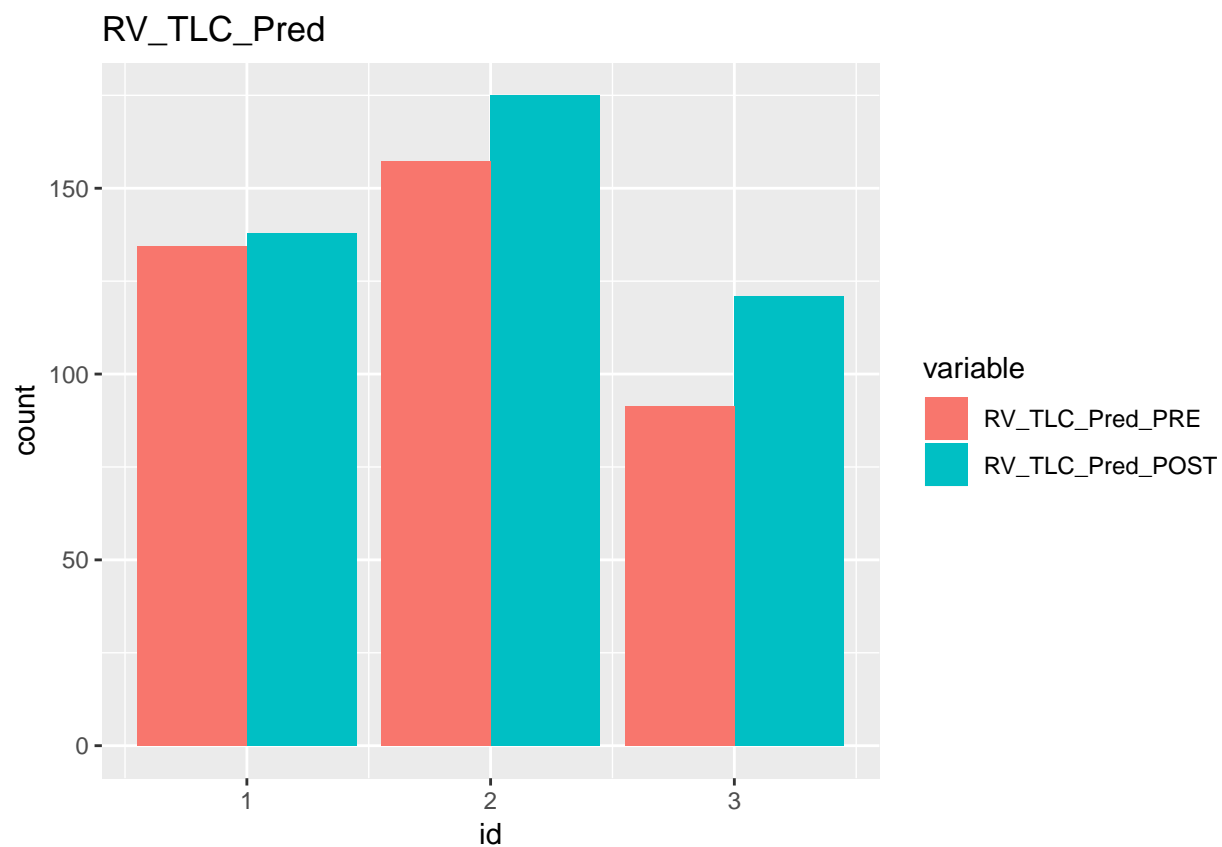


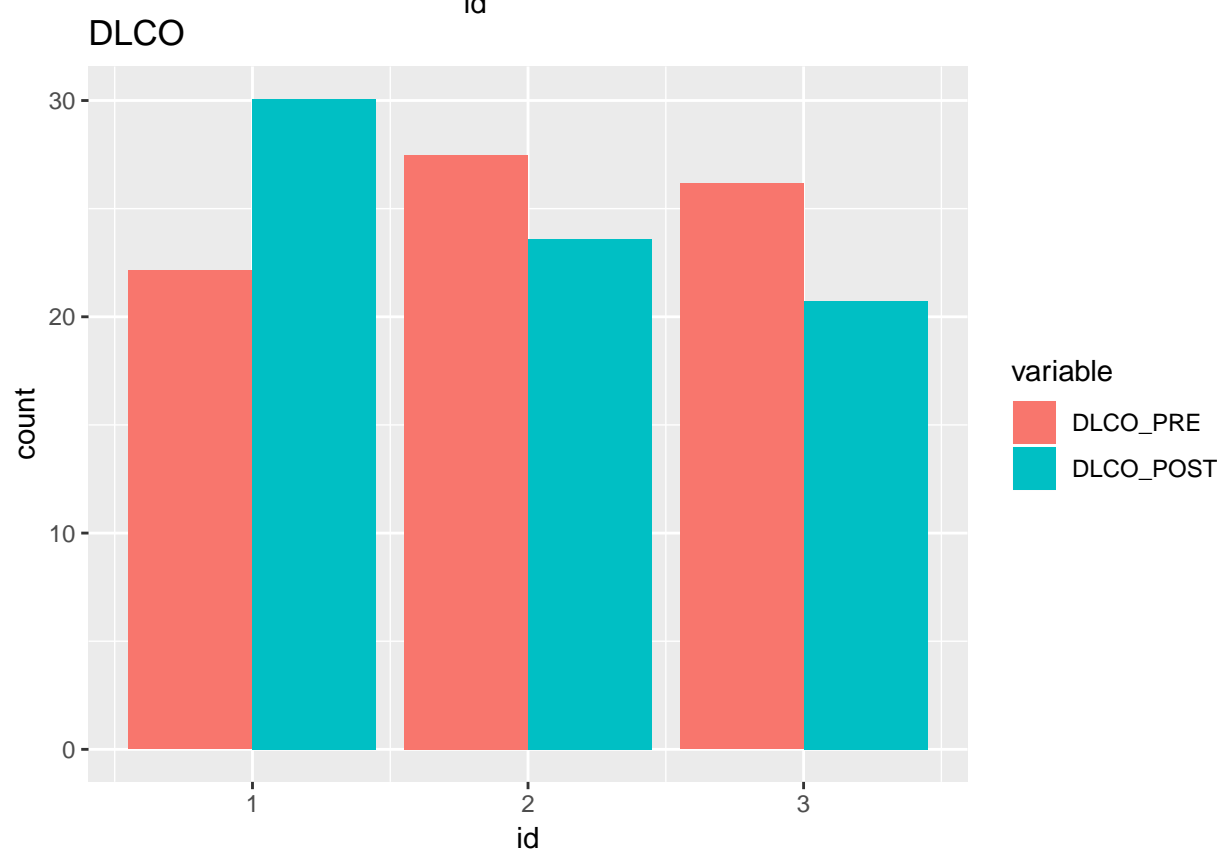
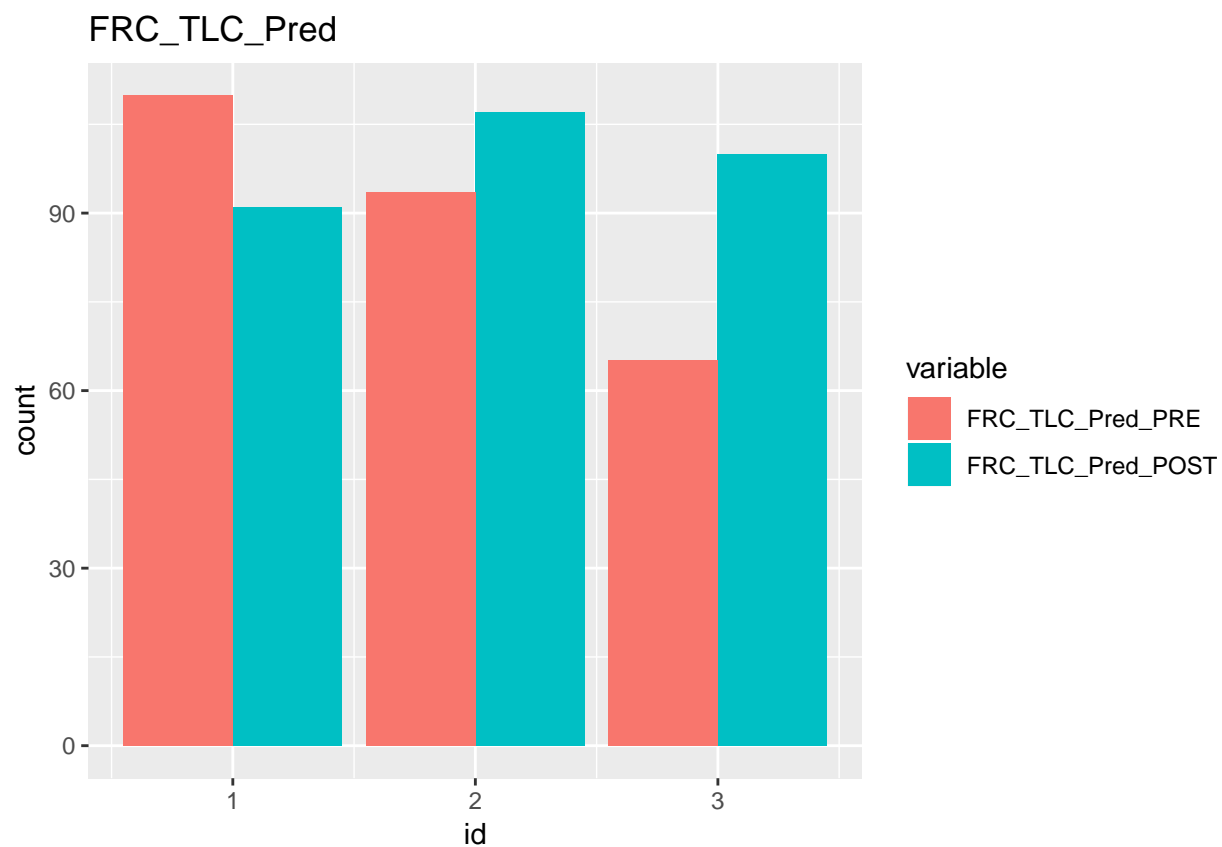


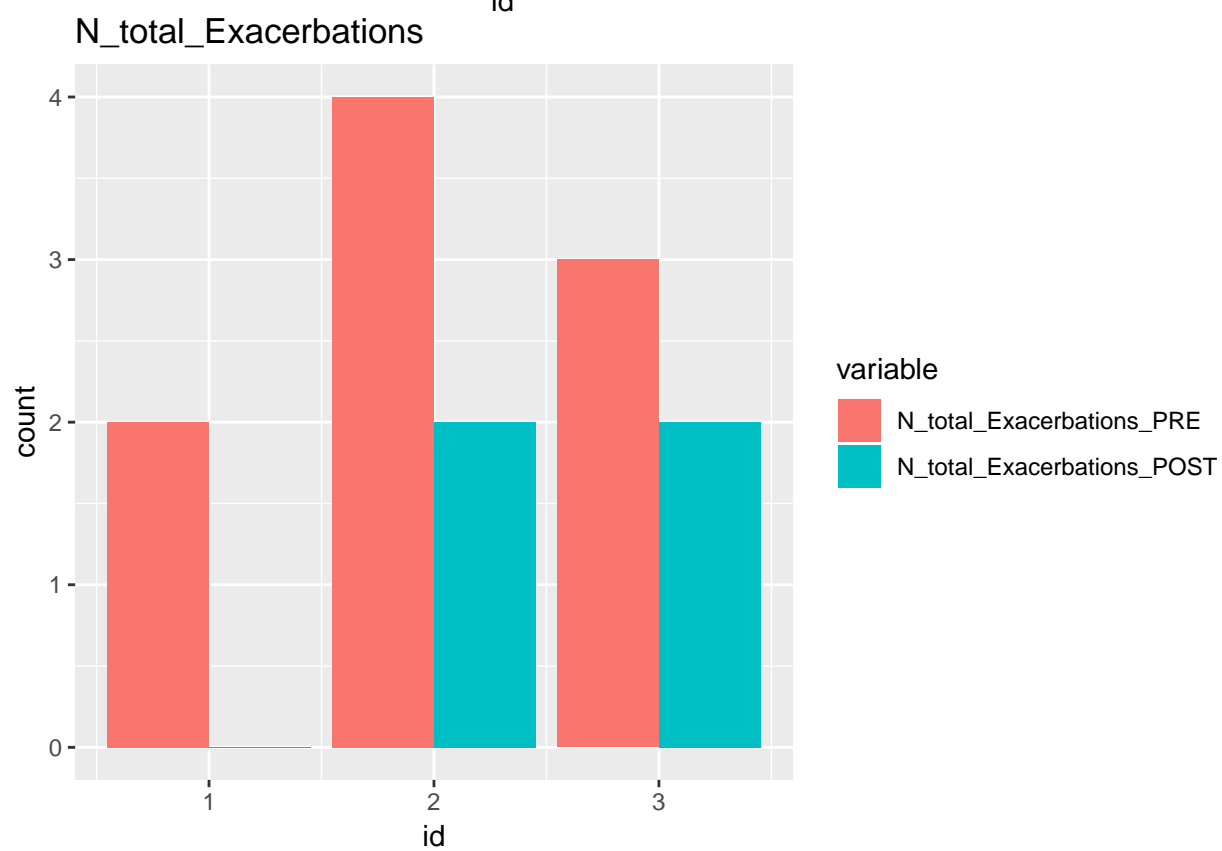
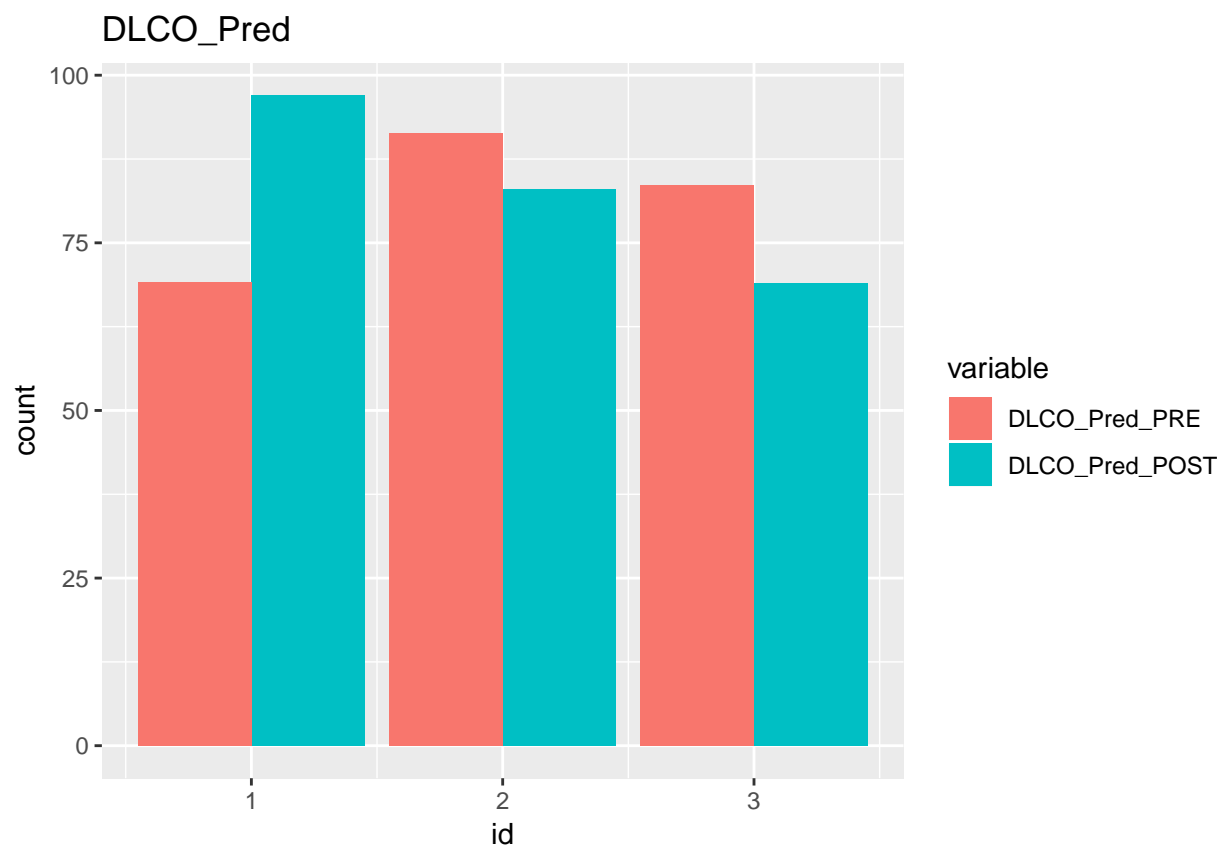


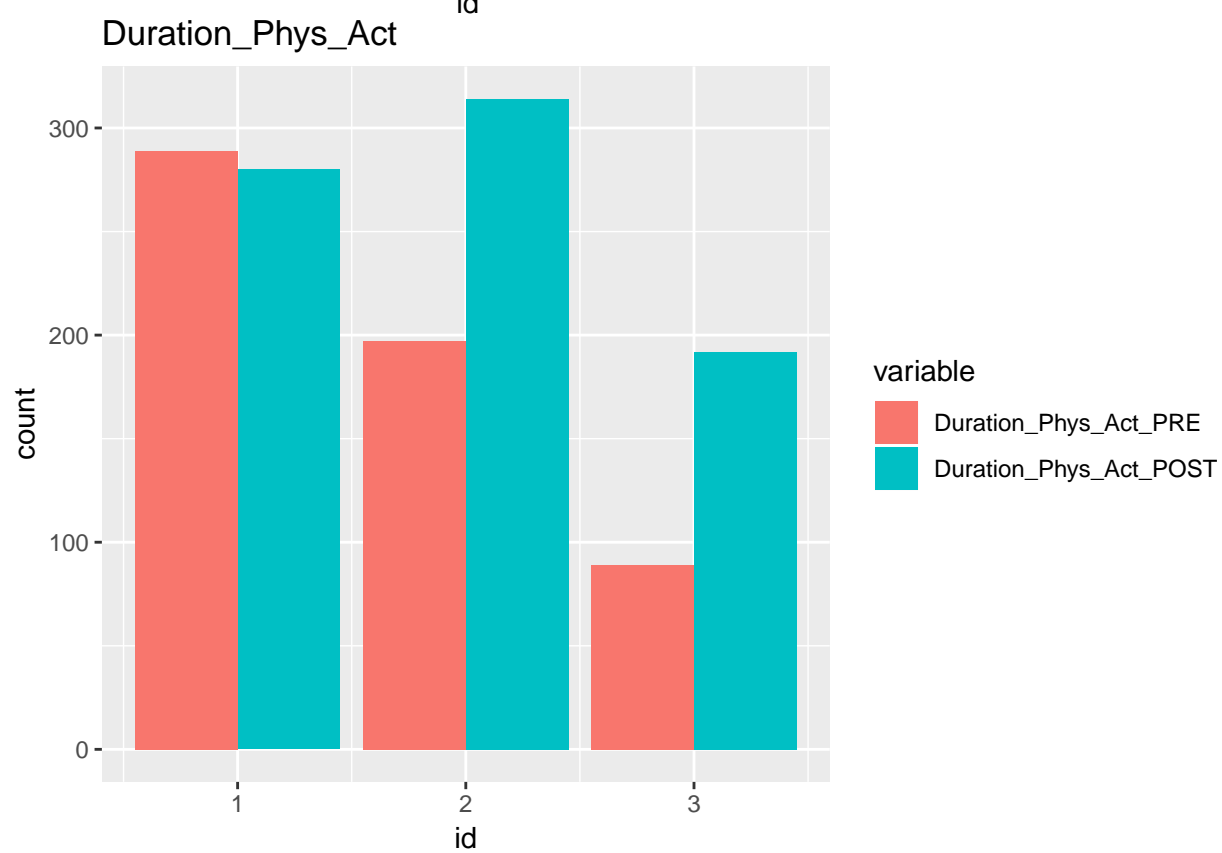
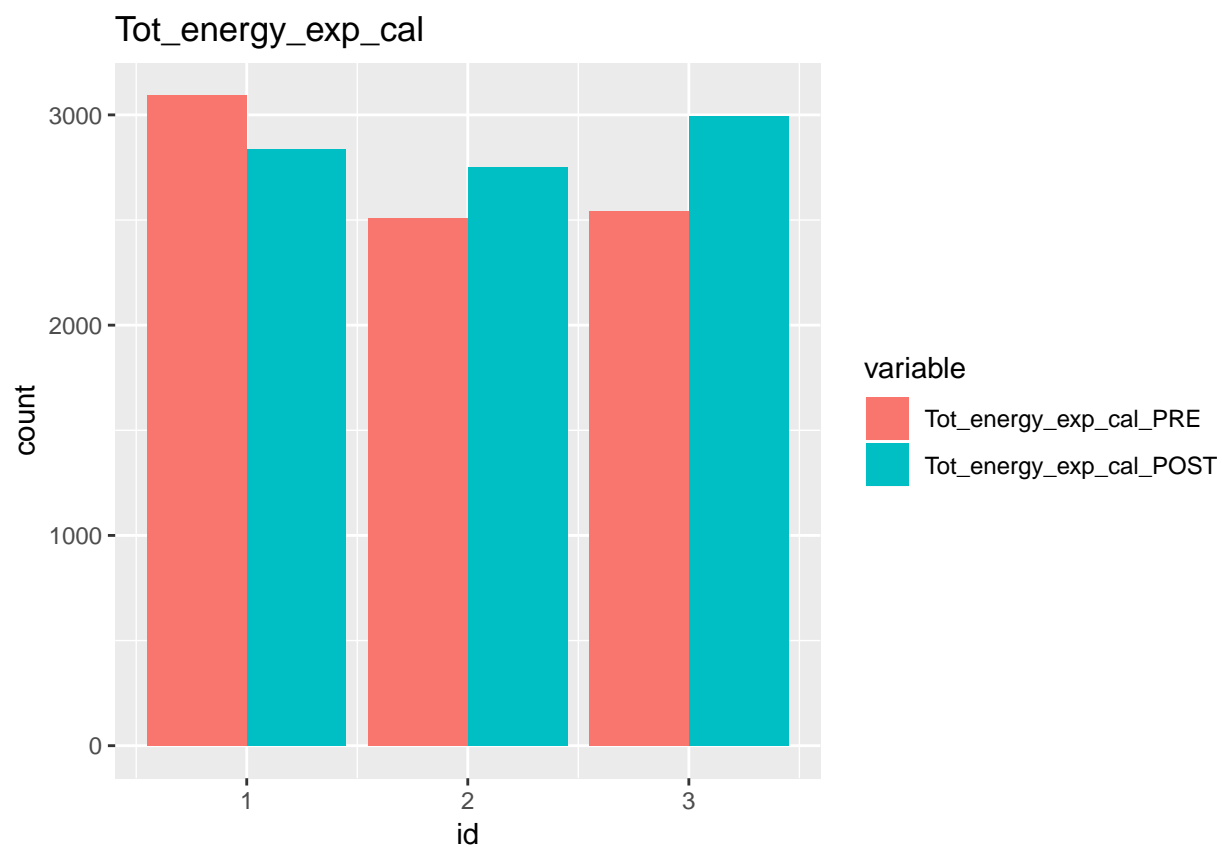


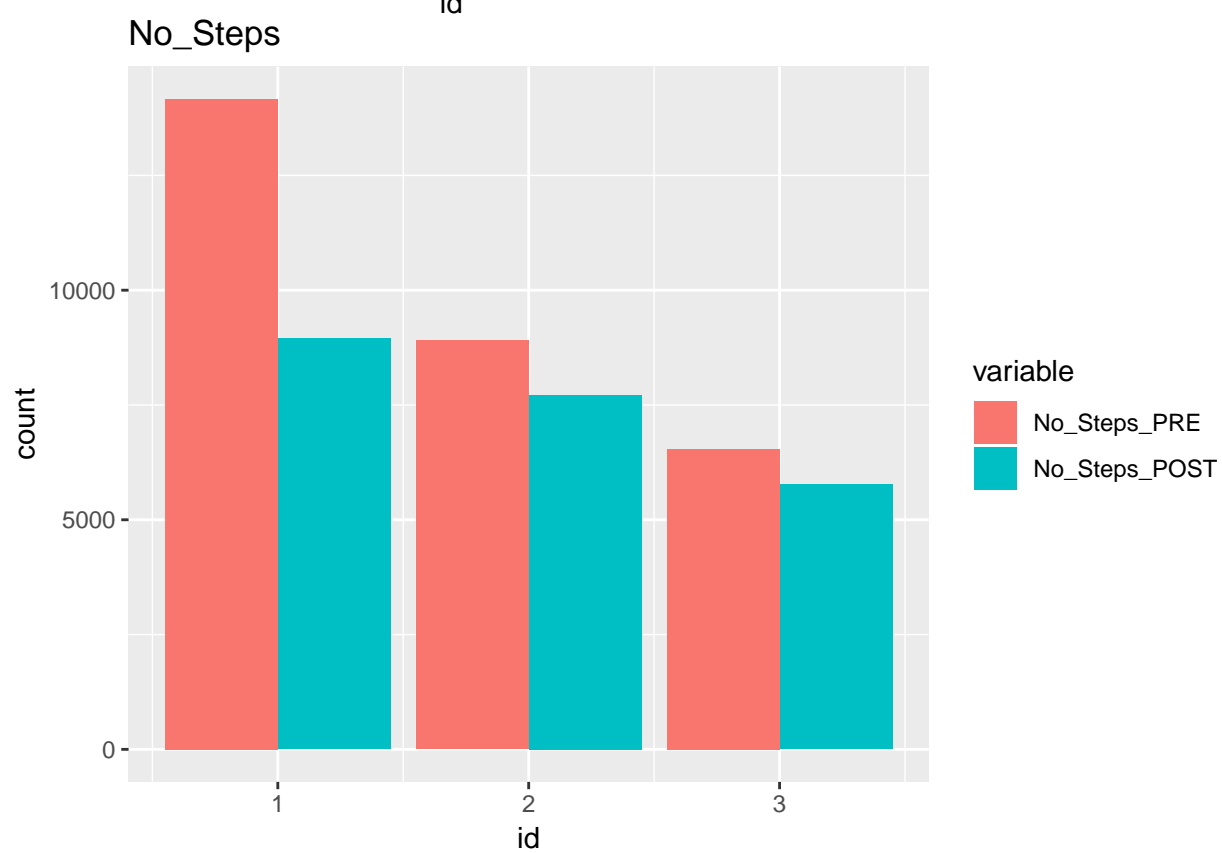
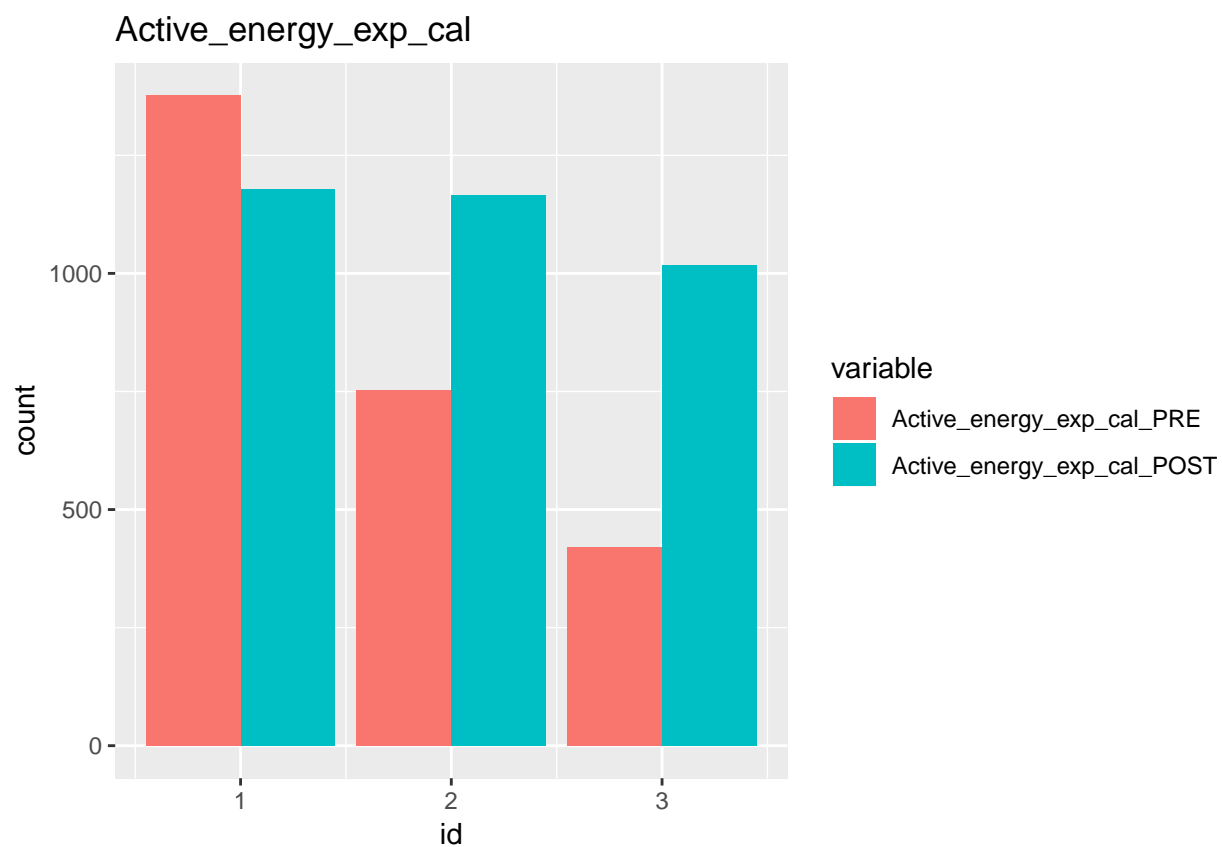


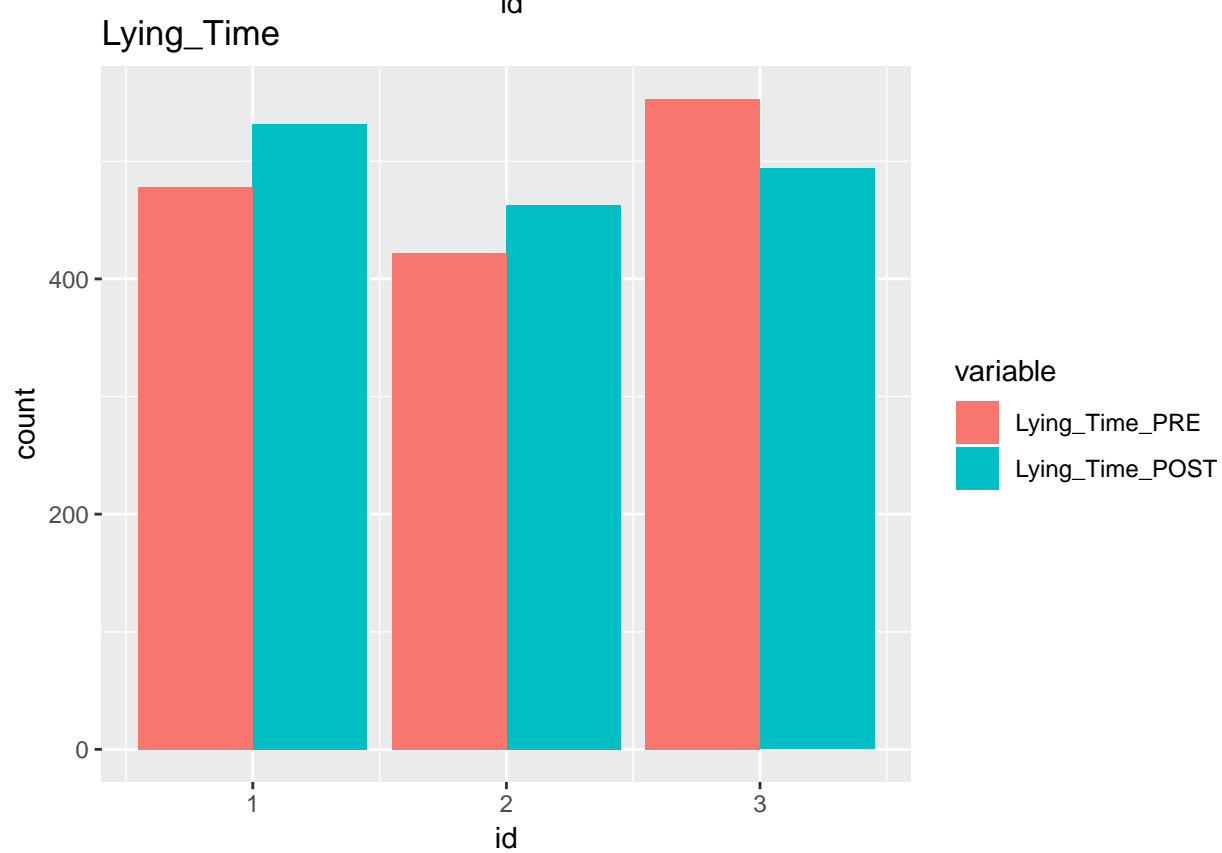
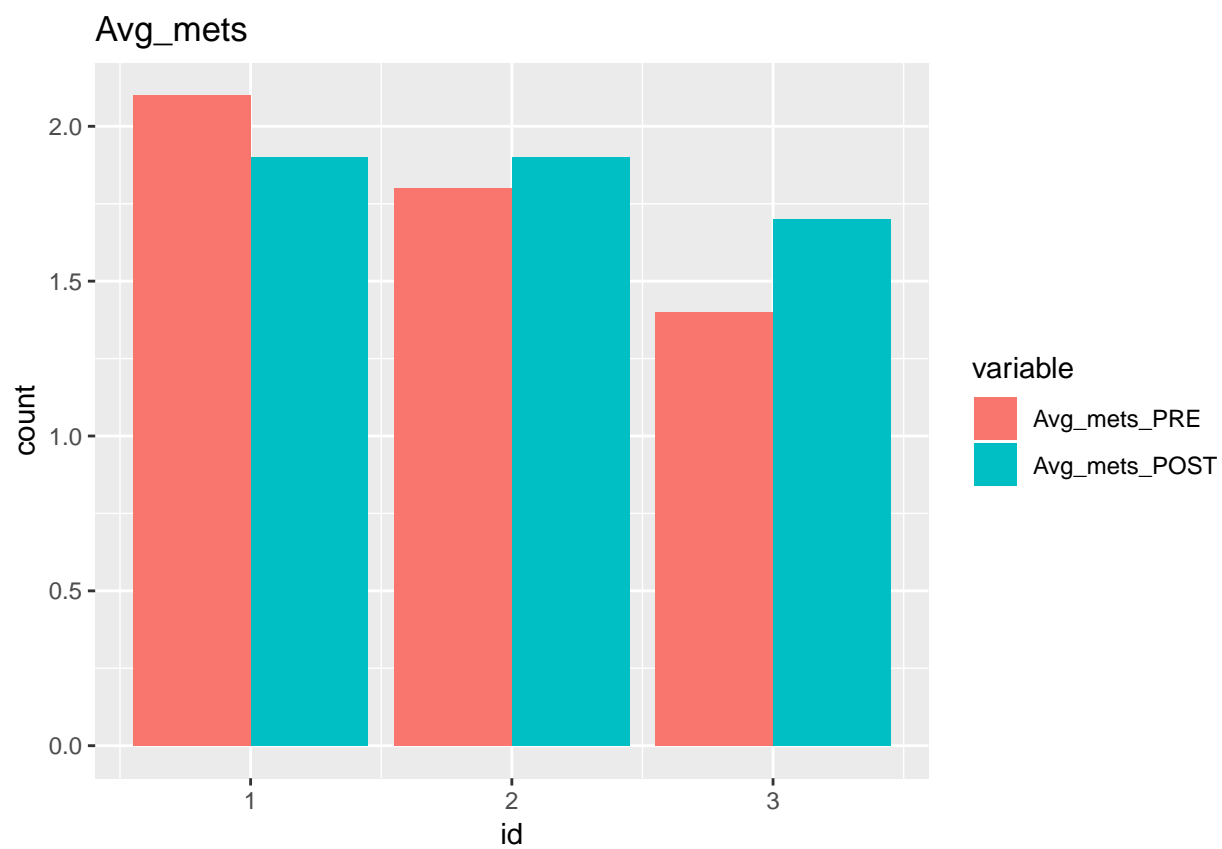


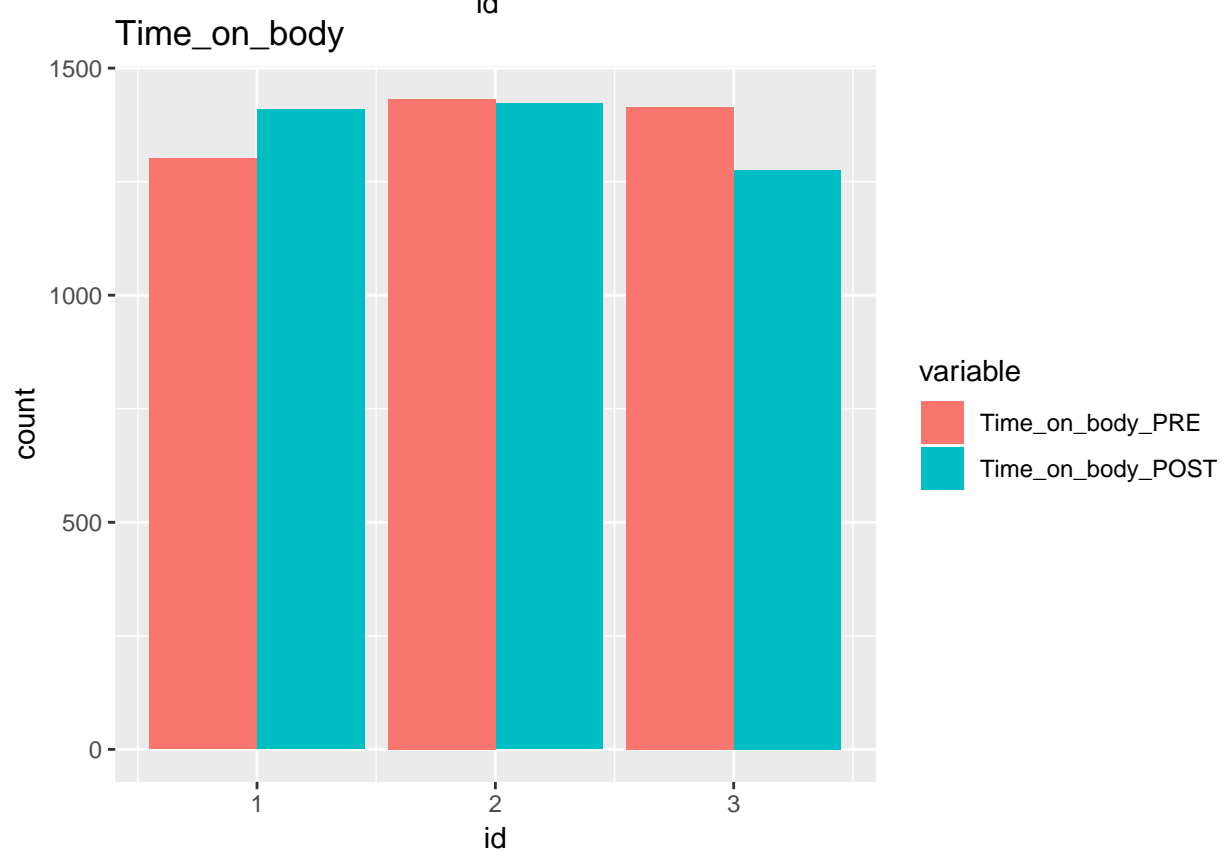
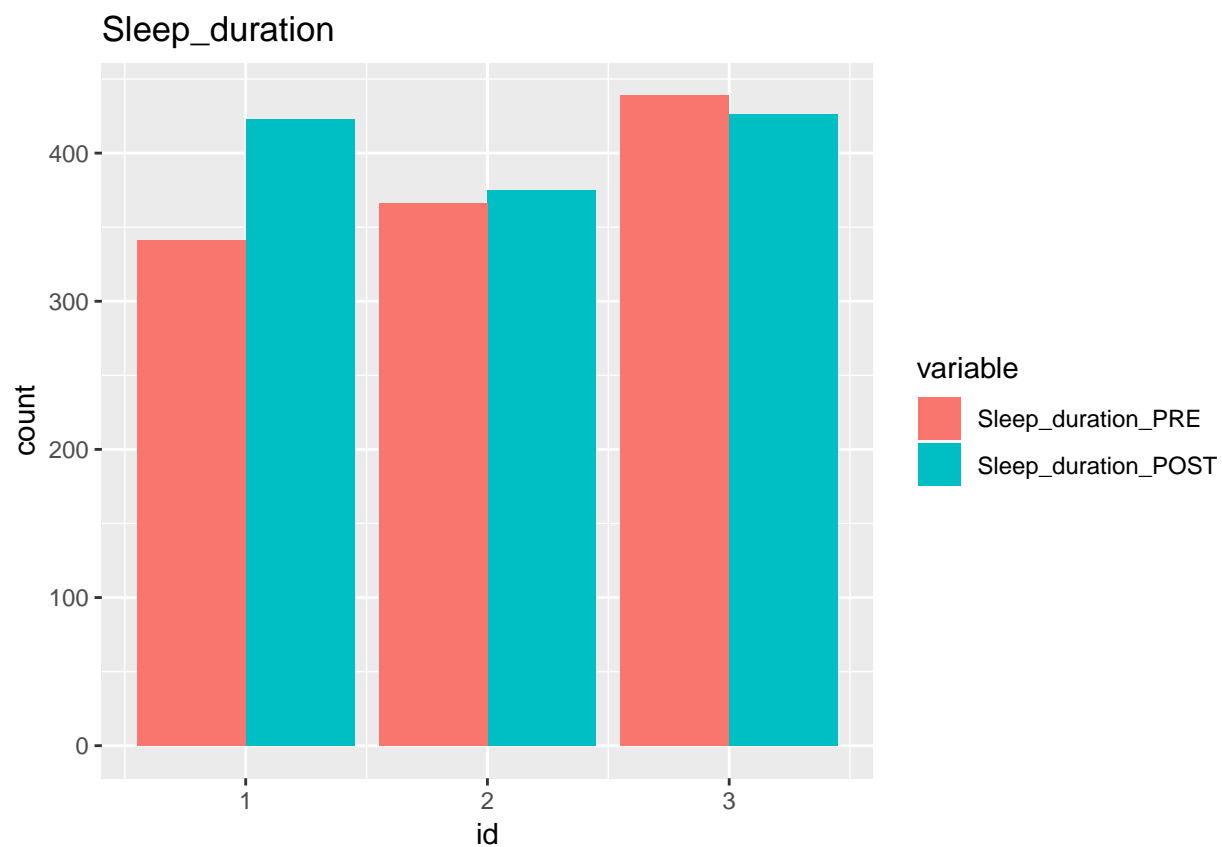


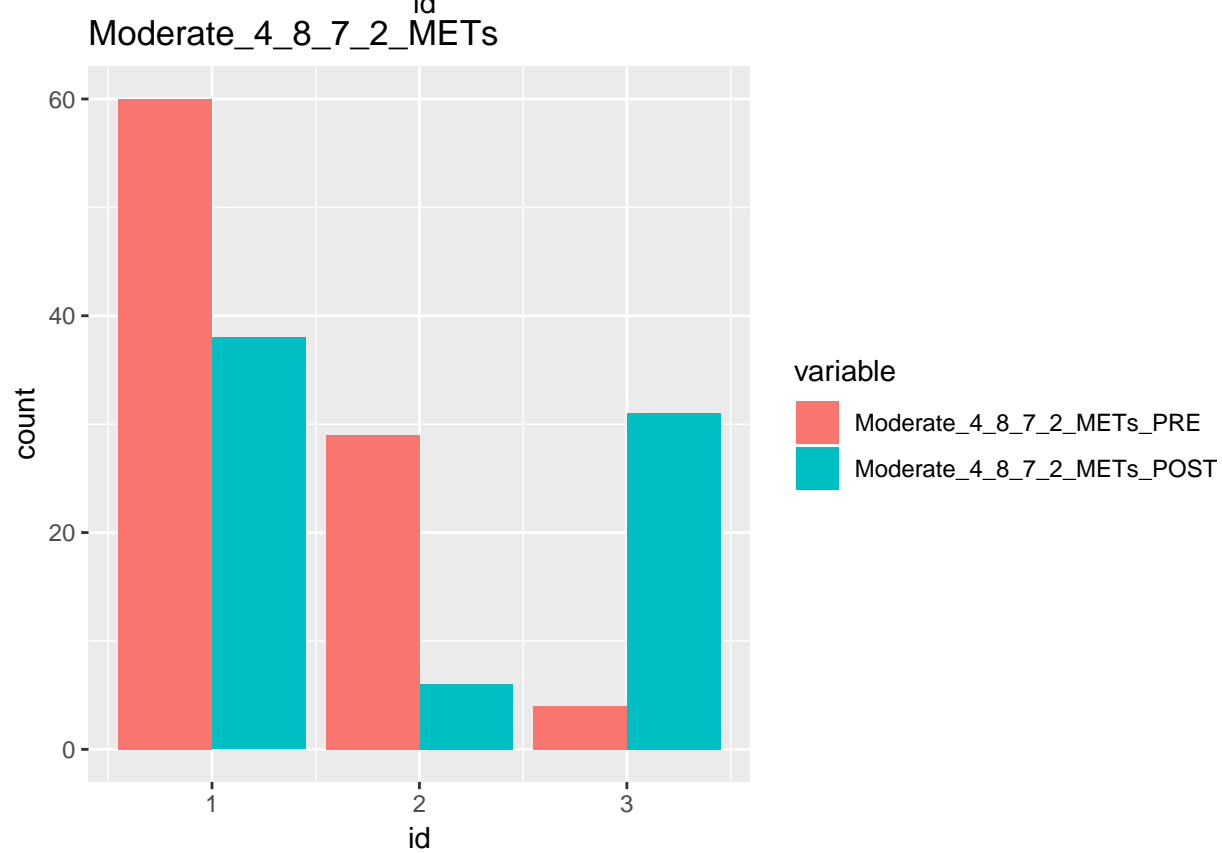
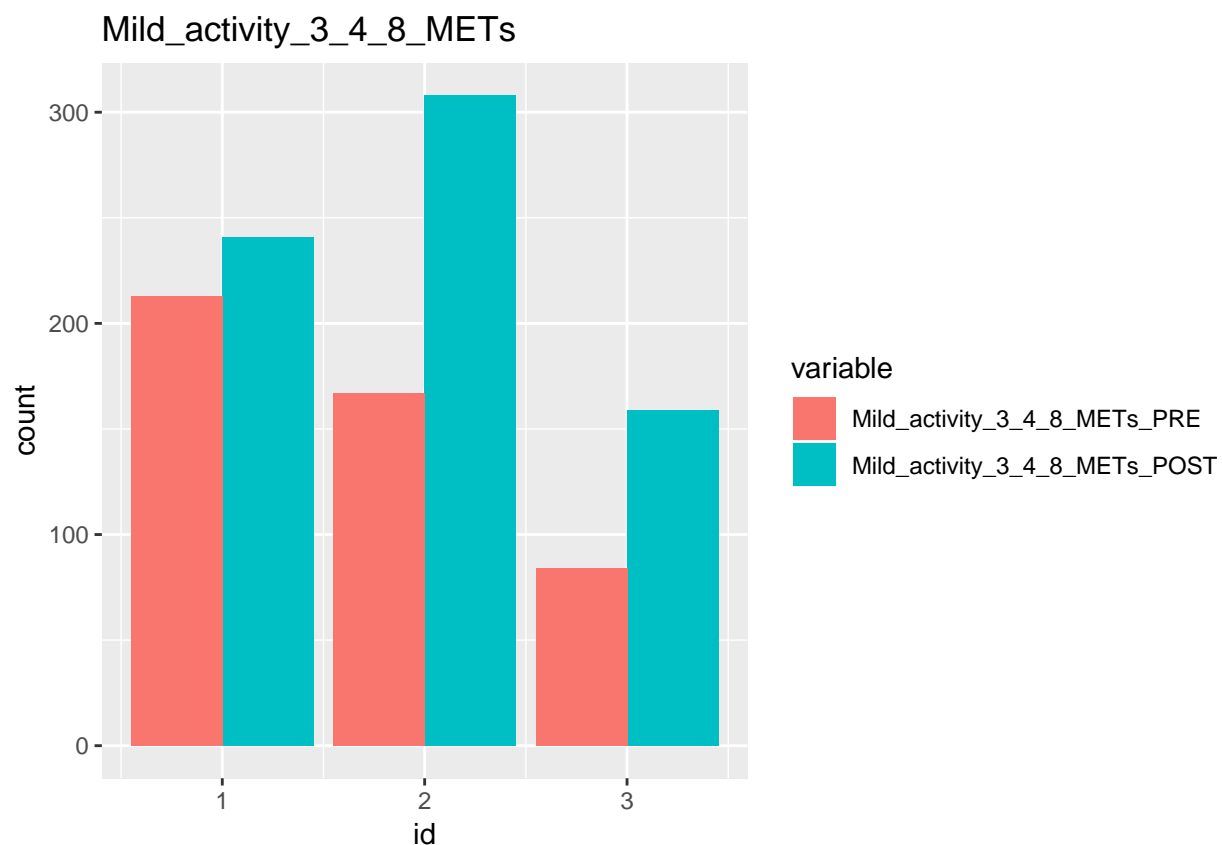


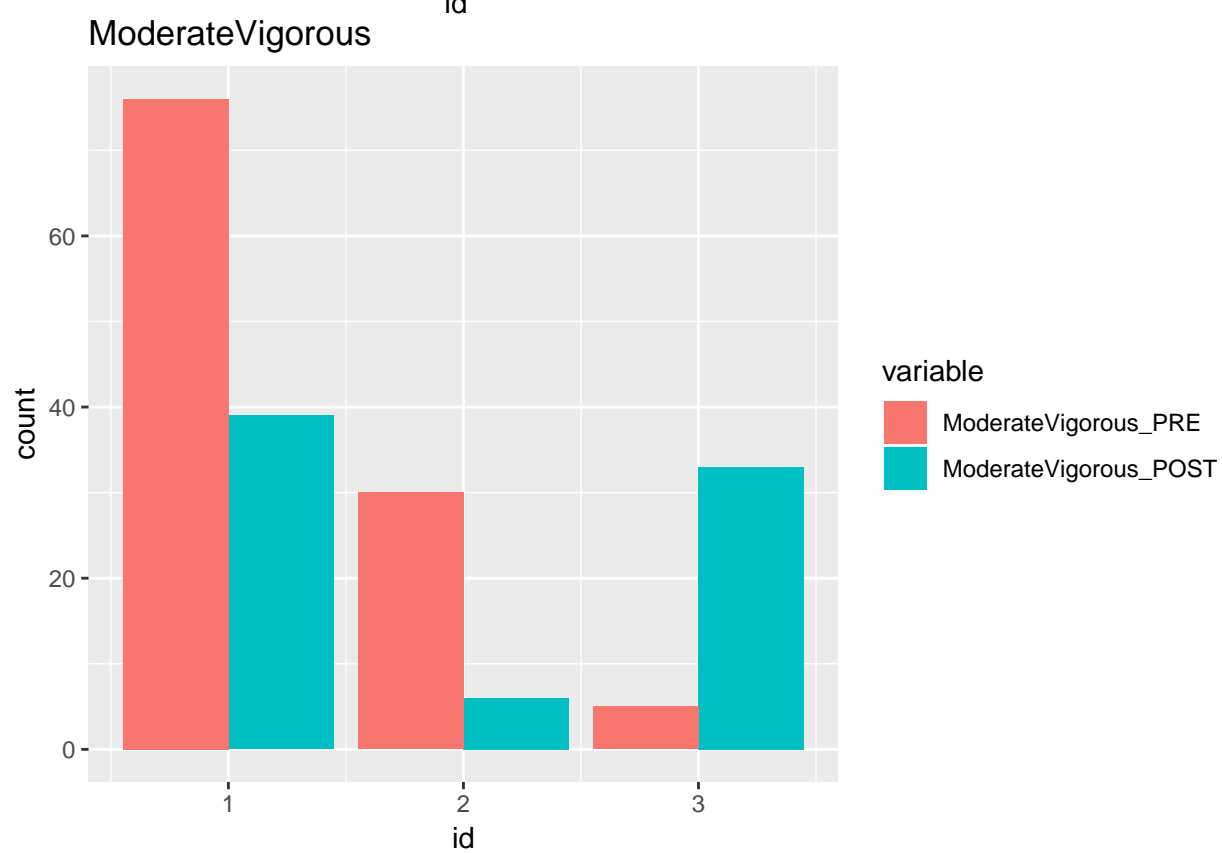
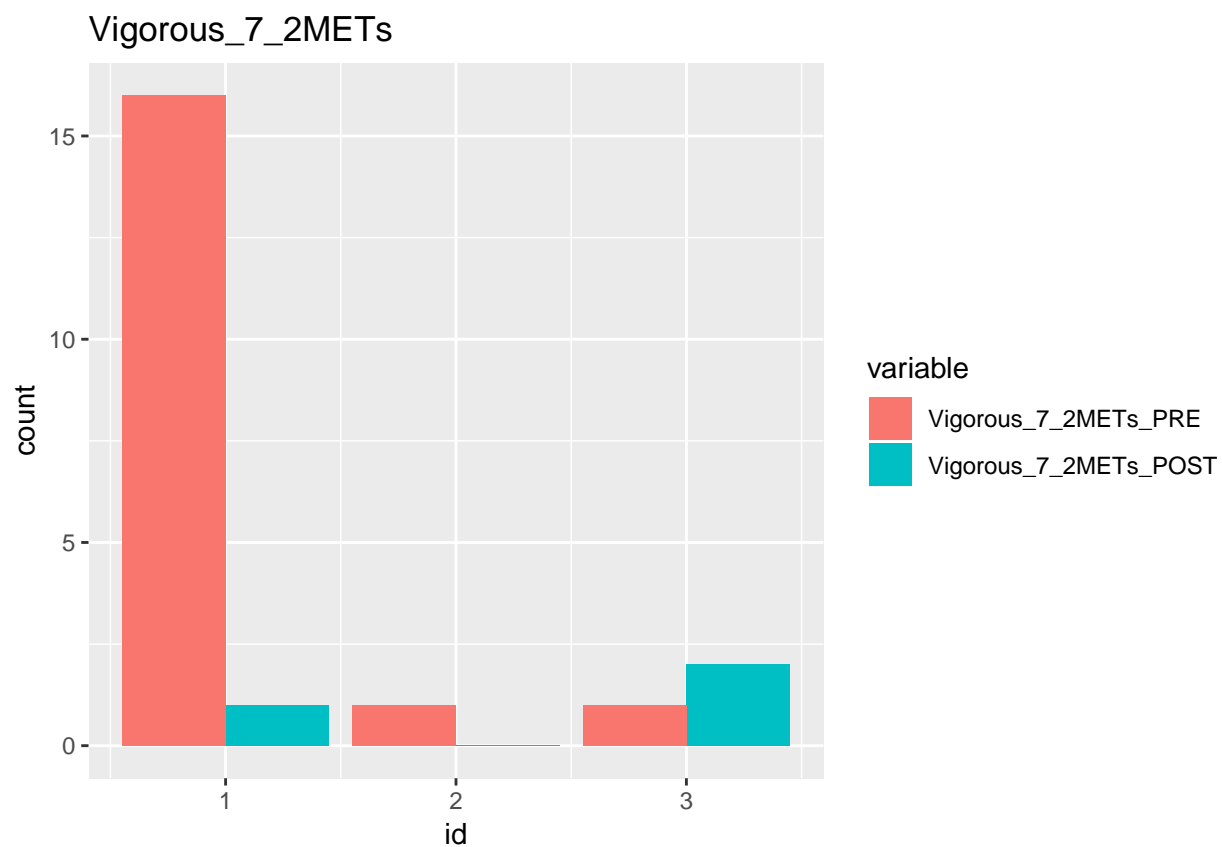










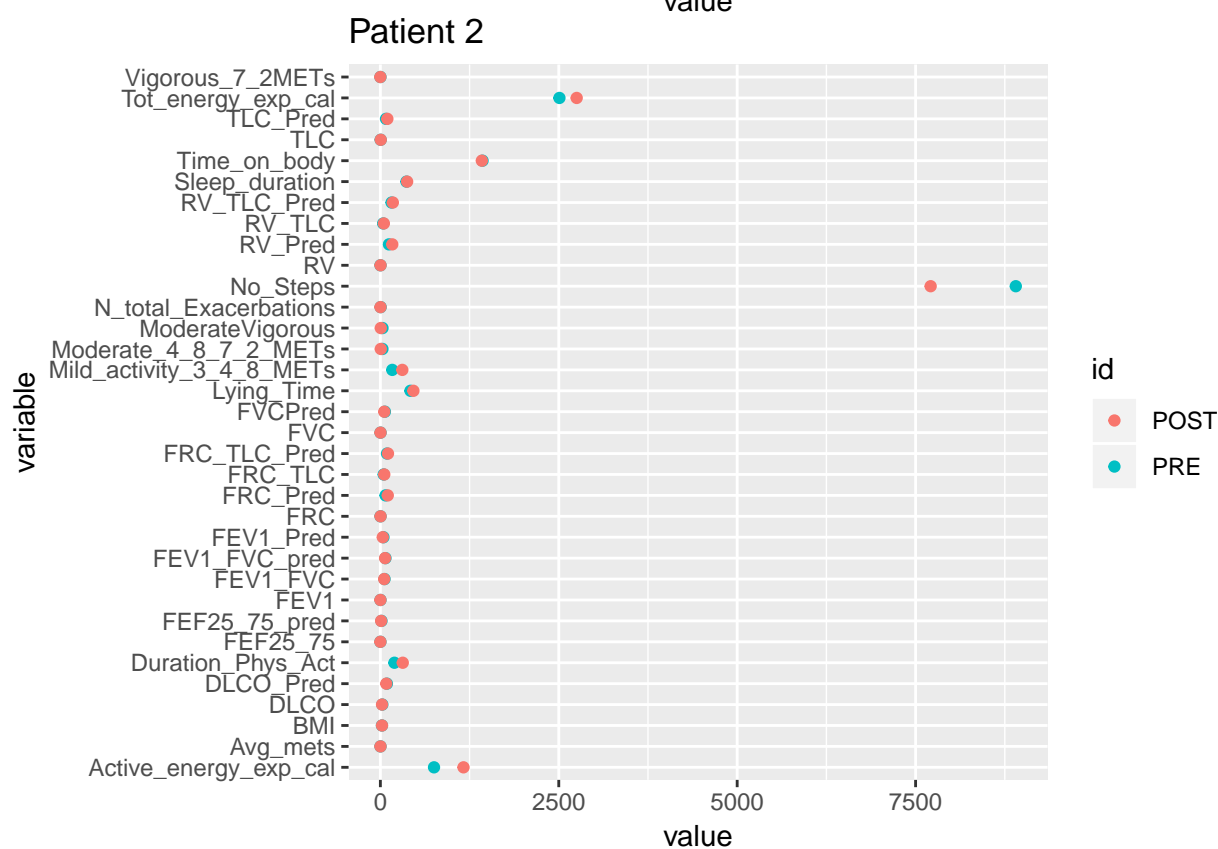
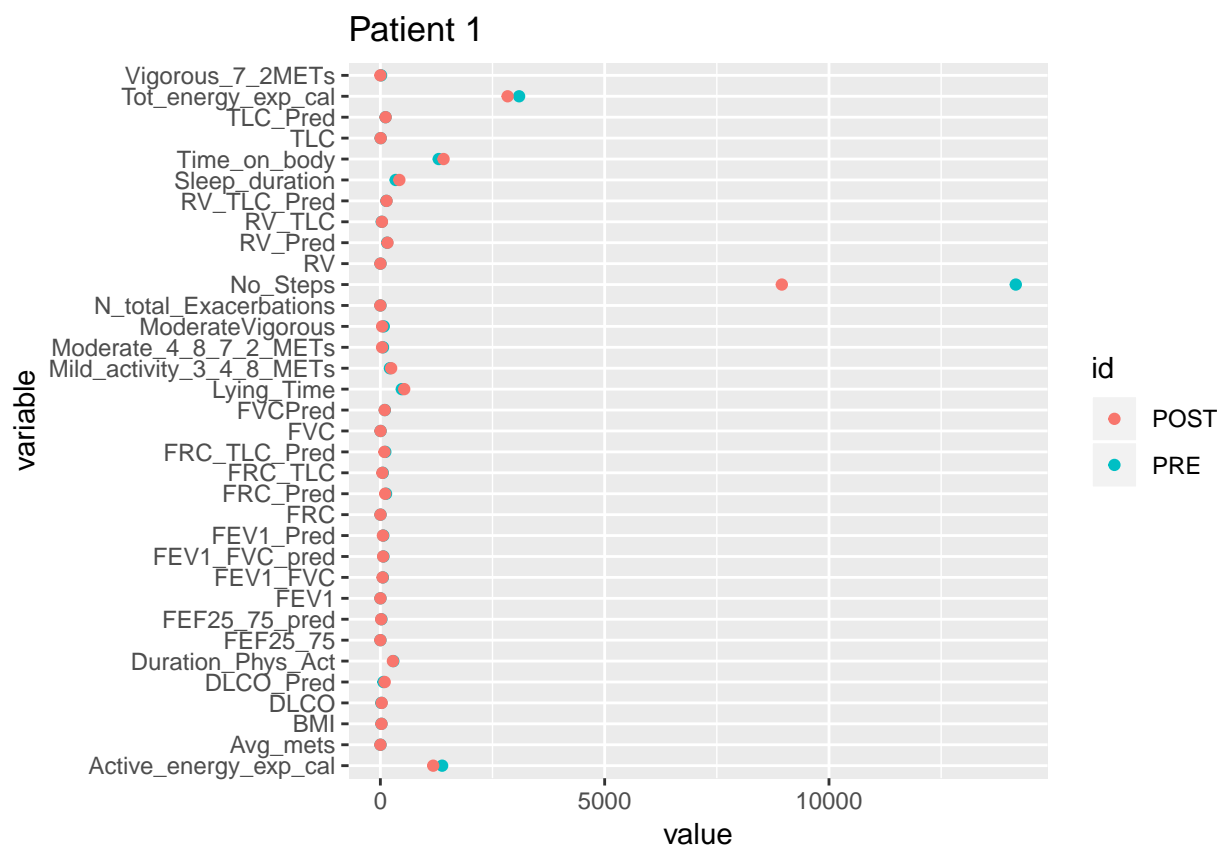


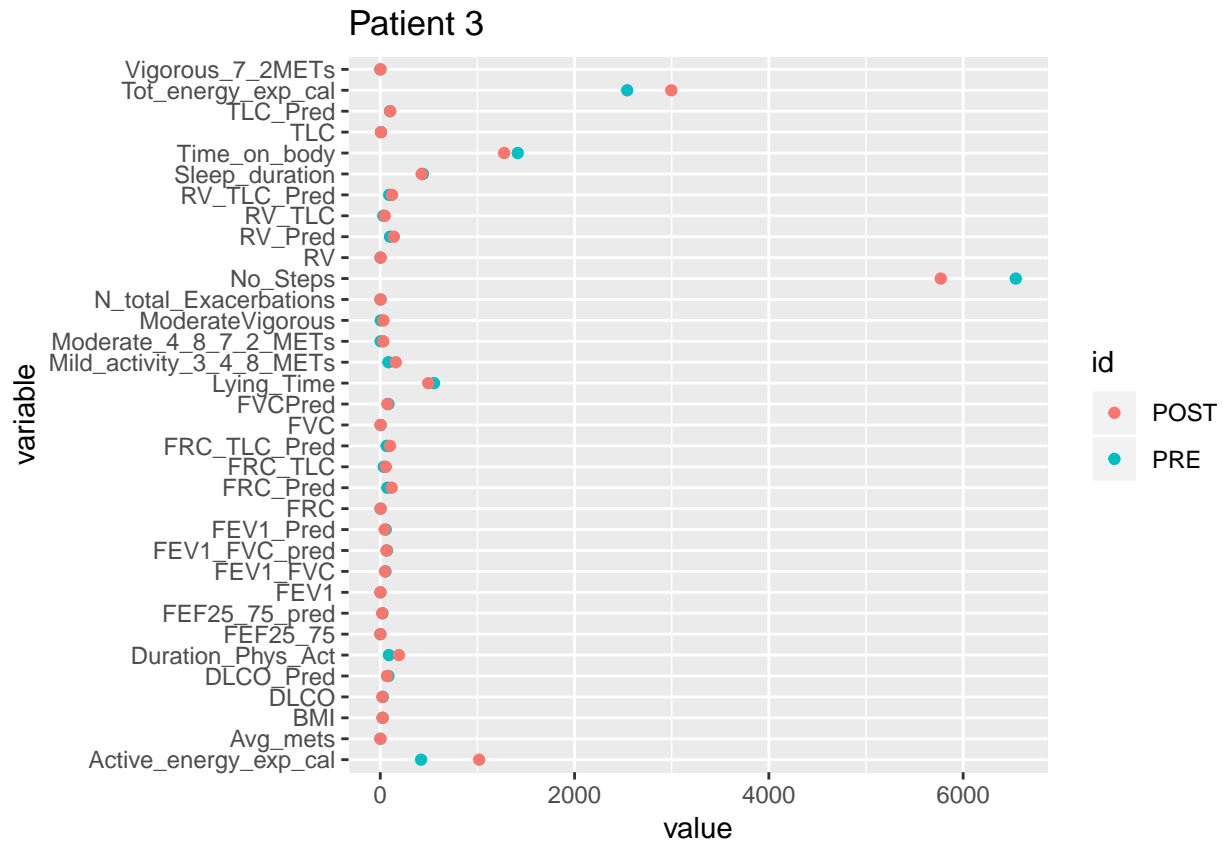
patient visualization

Here we plot for each patient his level of all variables together in order to have an overview of the treatment.

```
dataset.pre <- dataset[,idx.pre]
dataset.post <- dataset[,idx.pre+1]

for (i in 1:dim(dataset)[1])
{
  pre <- melt(dataset.pre[i,])
  pre$id <- "PRE"
  post <- melt(dataset.post[i,])
  post$id <- "POST"
  pp <- rbind(pre,post)
  pp$variable <- gsub(pattern="_PRE","",pp$variable)
  pp$variable <- gsub(pattern="_POST","",pp$variable)
  g <- ggplot(pp) +
    geom_point(mapping=aes(x=variable, y=value, color=id)) +
    ggtitle(label=paste0("Patient ", i)) + coord_flip()
  print(g)
}
```





my tiny conclusions

From the barplots we can notice, in some cases, a difference between the variables in each patient. Also from the scatterplots we can notice in some cases for each patient a difference in some variables.

But even if we are able to see a difference in the variables between the PRE and POST conditions, we are not able to catch it with a t-test, maybe because of the low number of patients.