Smoking cessation for CNCP Pharmaceutical Opioid Users

Phillip Hungerford

27 May 2019

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

CNCP Opioid Users Who Smoke

```
# EDA
smokers <- df %>%
  select(participant_id, wave, cig_12m)

for (i in 0:5){
  tmp <- subset(smokers, wave== i)
  summation <- sum(tmp$cig_12m == "Yes", na.rm=T)
  proportion <- summation / length(tmp$wave)
  cat("\nWave: ", i, "\nTotal\t\t", summation, "\nProportion:\t", round(proportion,
2)*100, "%\n")
}
```

```
##
## Wave:  0
## Total        556
## Proportion:   37 %
##
## Wave:  1
## Total        412
## Proportion:   33 %
##
## Wave:  2
## Total        432
## Proportion:   34 %
##
## Wave:  3
## Total        400
## Proportion:   33 %
##
## Wave:  4
## Total        398
## Proportion:   33 %
##
## Wave:  5
## Total        351
## Proportion:   31 %
```

```
num <- data.frame("Wave" = c(0,1,2,3,4,5),
                  "num" = c(556, 412, 432, 400, 398, 351))
prop <- data.frame("Wave" = c(0,1,2,3,4,5),
                   "prop" = c(37, 33, 34, 33, 33, 31))

ggplot() +
  geom_line(data=num, aes(Wave, num), color='red')+
  theme_classic() +
  ylab("Total")+
  ggtitle("Number of Smokers")
```

Number of Smokers
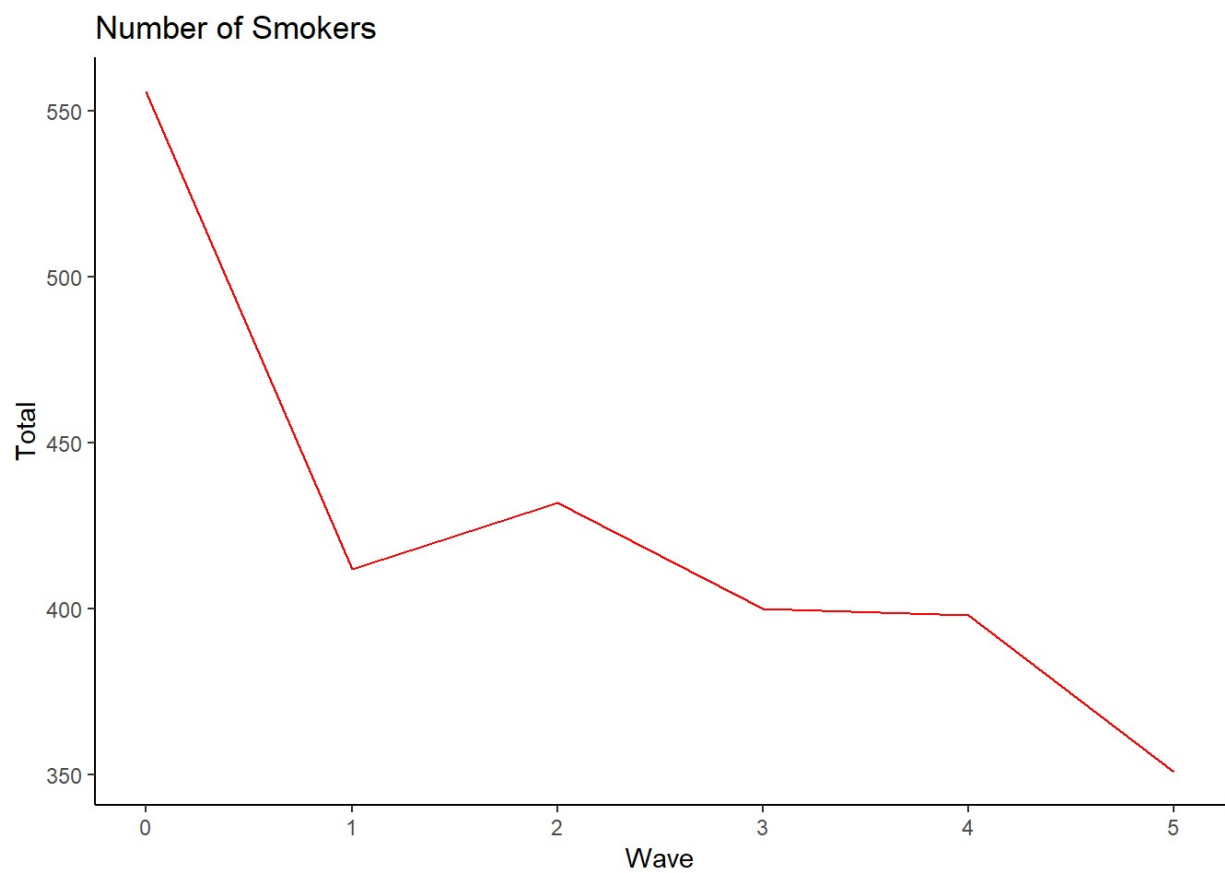

```
ggplot() +
  geom_line(data=prop, aes(x=Wave, y=prop), color="blue") +
  theme_classic() +
  ylab("Proportion")+
  ggtitle("Proportion of Smokers")
```

Proportion of Smokers
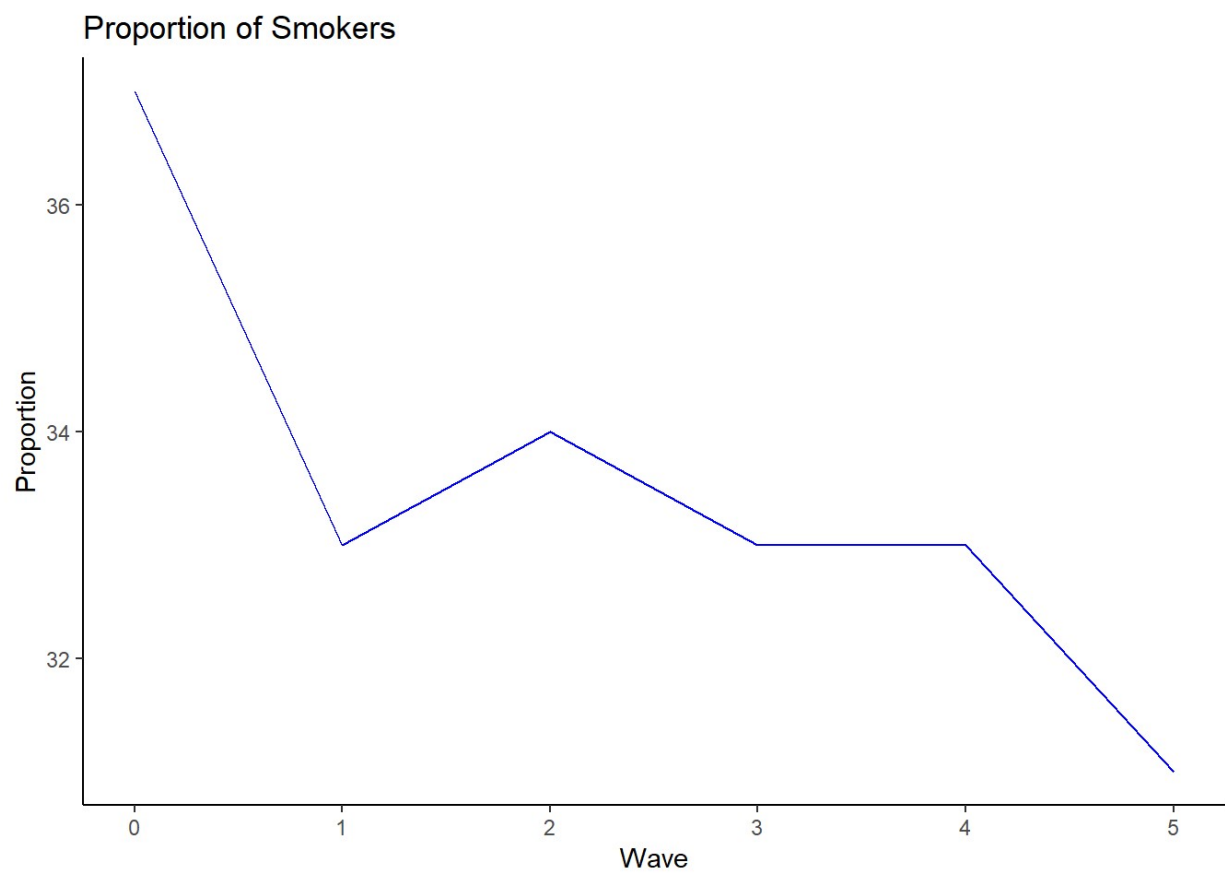
Calculate Baseline Smokers

We will now find those who were smoking at baseline.

```
###############################################################################
# 2. Create smoker flag
smokers <- df %>%
  select(participant_id, wave, cig_12m)

# find people who smoke in baseline
tmp <- subset(smokers, wave==0)
tmp <- subset(tmp, cig_12m == "Yes")
bs_smokers <- tmp$participant_id

# Create data frame of people who smoked from baseline onwards
smokers <- subset(smokers, participant_id %in% bs_smokers)

# filter through to have participants who have data for all years
keep <- NULL
for (i in 1:length(tmp$participant_id)){
  tmp2 <- subset(smokers, participant_id == tmp$participant_id[i])
  if (sum(tmp2[,1], na.rm=T) / 6 == tmp$participant_id[i]){
    keep[i] <- tmp$participant_id[i]
    }
}

keep <- unique(keep) # 323 participants who stay full term 5 years
ft_smokers <- subset(df, participant_id %in% keep)
```

Of the 556 participants who stated that they smoked at baseline, 332 have remained in the study across all waves (have data points from Baseline -> Year 5).

Participants Who Quit Smoking

```r
################################################################################
# 3. Find smokers who quit at each interval

# Create function that can detect who quit at what year
quit_detector <- function(ft_smokers, year){
  # Create placeholder to flag
  ft_smokers$quit <- 0
  # Get list of participants who smoke into a list
  participants <- unique(ft_smokers$participant_id)
  # Iterate through each participant

  for (participant in participants){
    tmp <- subset(ft_smokers, participant_id == participant)

    smoking <- tmp[["cig_12m"]]
    status1 <- unique(smoking[1:year])
    status2 <- unique(smoking[(year+1):6])

    if (length(status1) ==1){
      if (length(status2) ==1){
        if (status1 == "Yes" & status2 == "No"){
          ft_smokers$quit[ft_smokers$participant_id == participant] <- "Yes"}
    }}}

  ft_smokers$quit[ft_smokers$quit == 0] <- "No"
  ft_smokers$quit <- as.factor(ft_smokers$quit)
return(ft_smokers)
}

# People who quit at Year1
print("Participants who quit at:")
```

```
## [1] "Participants who quit at:"
```

```r
for (i in 1:4){
  quitters <- quit_detector(ft_smokers, year=i)
  candidates <- subset(quitters, quit == "Yes")
  num_cand <- unique(candidates$participant_id)
  cat("\nWave", i, ":", length(num_cand))
  candidates$wave <- as.numeric(candidates$wave)
}
```

```
##
## Wave 1 : 17
## Wave 2 : 8
## Wave 3 : 6
## Wave 4 : 8
```

Sample Size calculator

```
# population size
p <- 1514
# proportion with target
p2 <- 0.01
# Critical Z value
z <- 1.96
# Margin of error
MOE <- 0.05

top <- (p2 * (1 - p2) * z**2) / MOE**2
bot <- 1 + (top-1)/p # adjusts for finite population
n <- top/bot
n <- ceiling(n)
cat("Minimum number of positive cases for analysis:", n)
```

```
## Minimum number of positive cases for analysis: 16
```

Based on our sample size calculator we would need a minimum of 16 positive classes which we just made by 1.

Baseline -> Year 5

```
###############################################################################
# 4. EDA
# Baseline -> Year 5
quitters <- quit_detector(ft_smokers, year=1)
candidates <- subset(quitters, quit == "Yes")
candidates$wave <- as.numeric(candidates$wave)

ggplot(candidates, aes(x=wave, y=bpi_pscore)) +
 geom_point() + geom_smooth(method ='lm')+
  ggtitle("Bpi Score") +
  theme_classic()
```
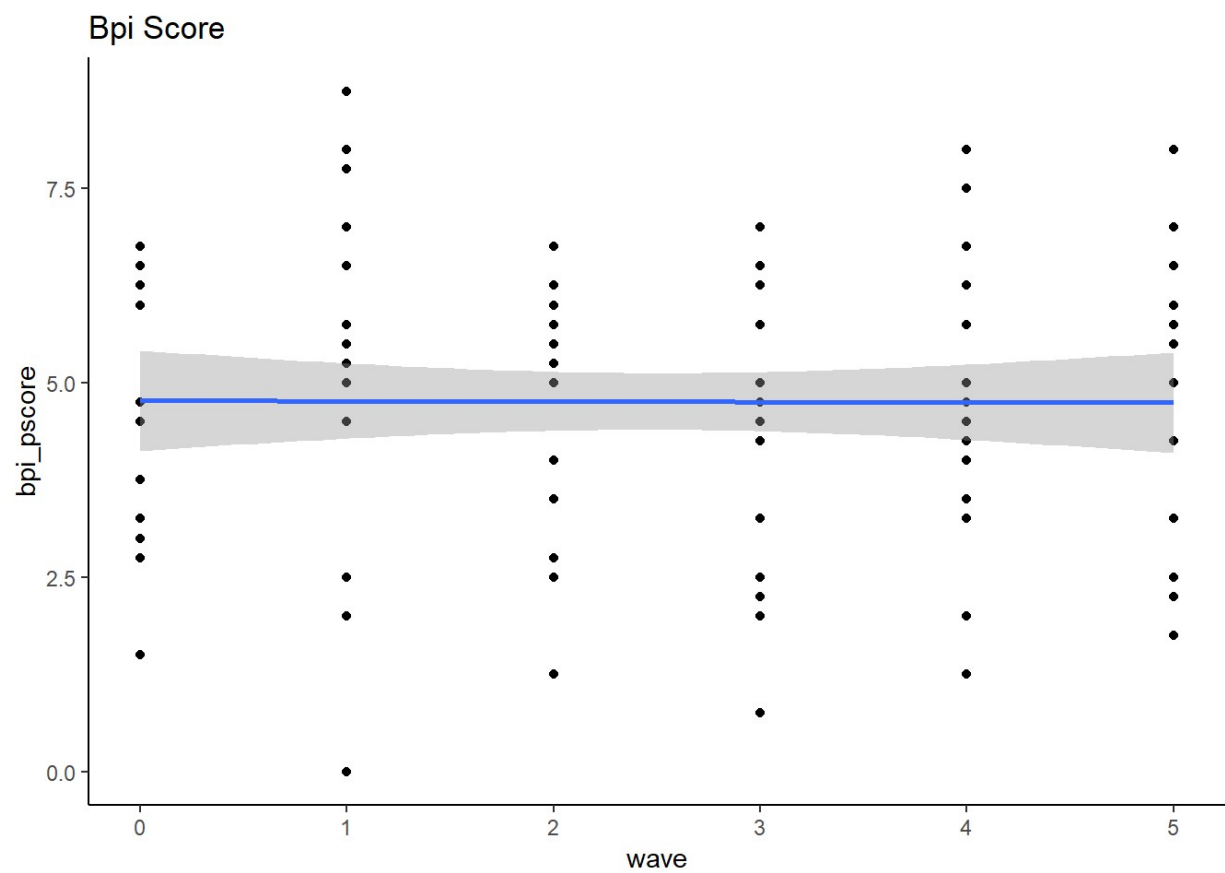
Bpi Score


```
ggplot(candidates, aes(x=wave, y=bpi_interference)) +
  geom_point() + geom_smooth(method="lm")+
  ggtitle("Bpi Interference") +
  theme_classic()
```
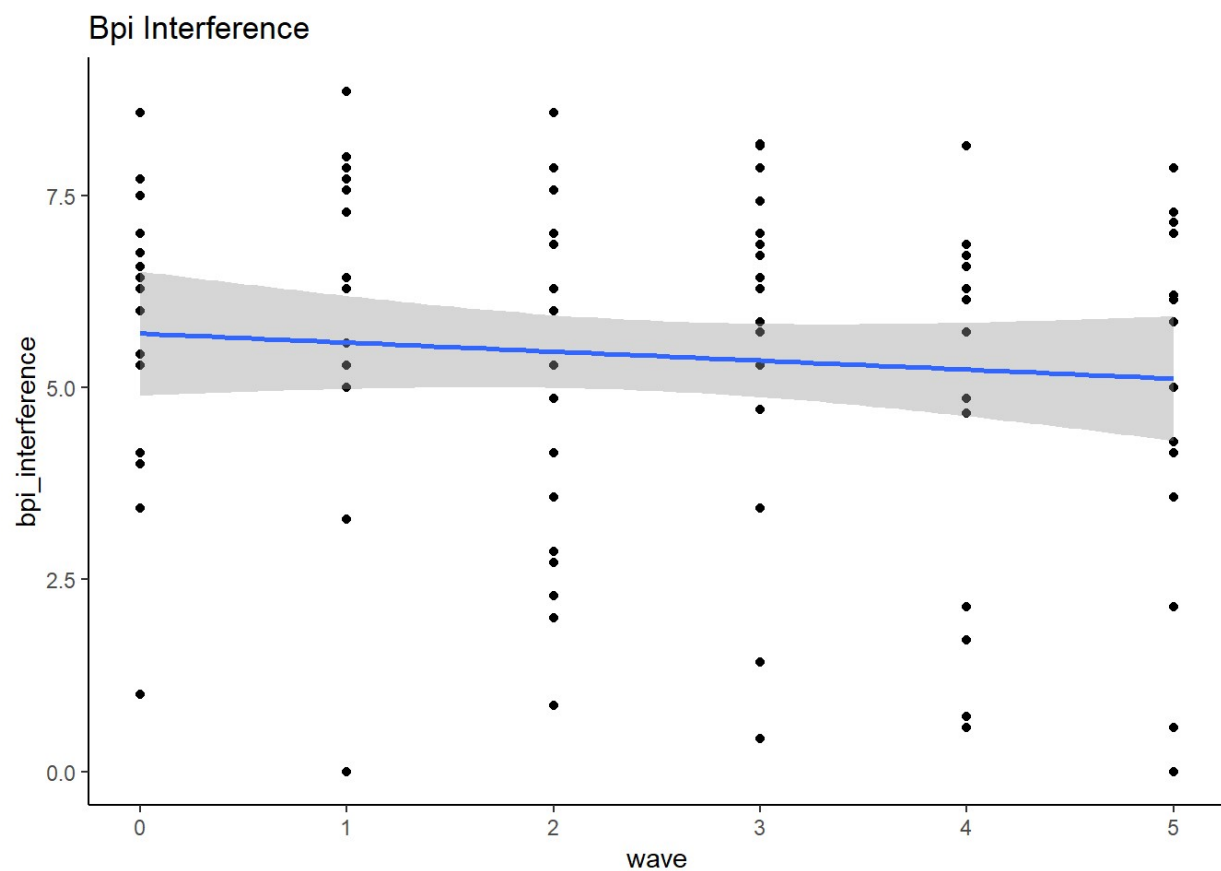
Bpi Interference

Mean BPI score appears to remain constant across all waves. Interference appears to show a slight decrease across the 5 years.

```
ggplot(candidates, aes(x=wave, y=sf12_mcs)) +
  geom_point() + geom_smooth(method="lm")+
  ggtitle("SF12 MCS") +
  theme_classic()
```
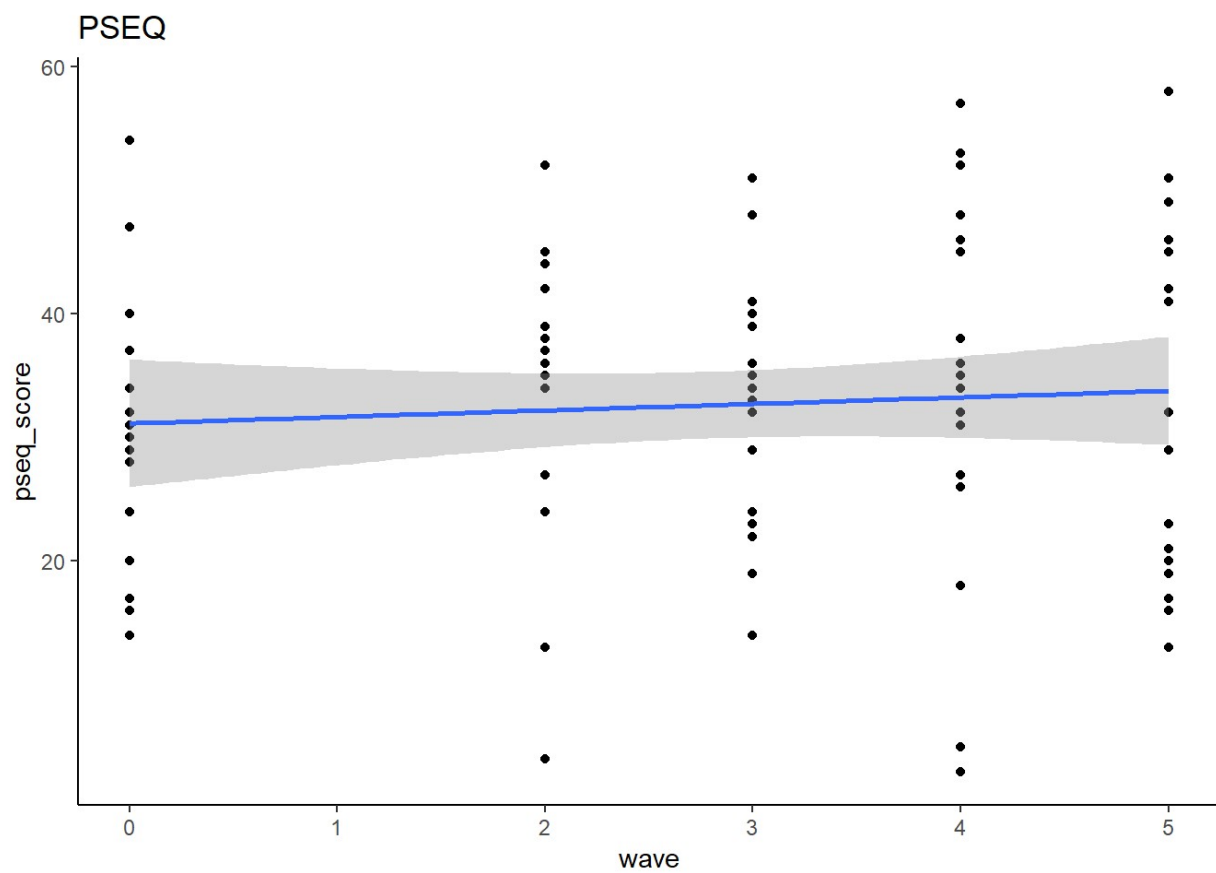
SF12 MCS
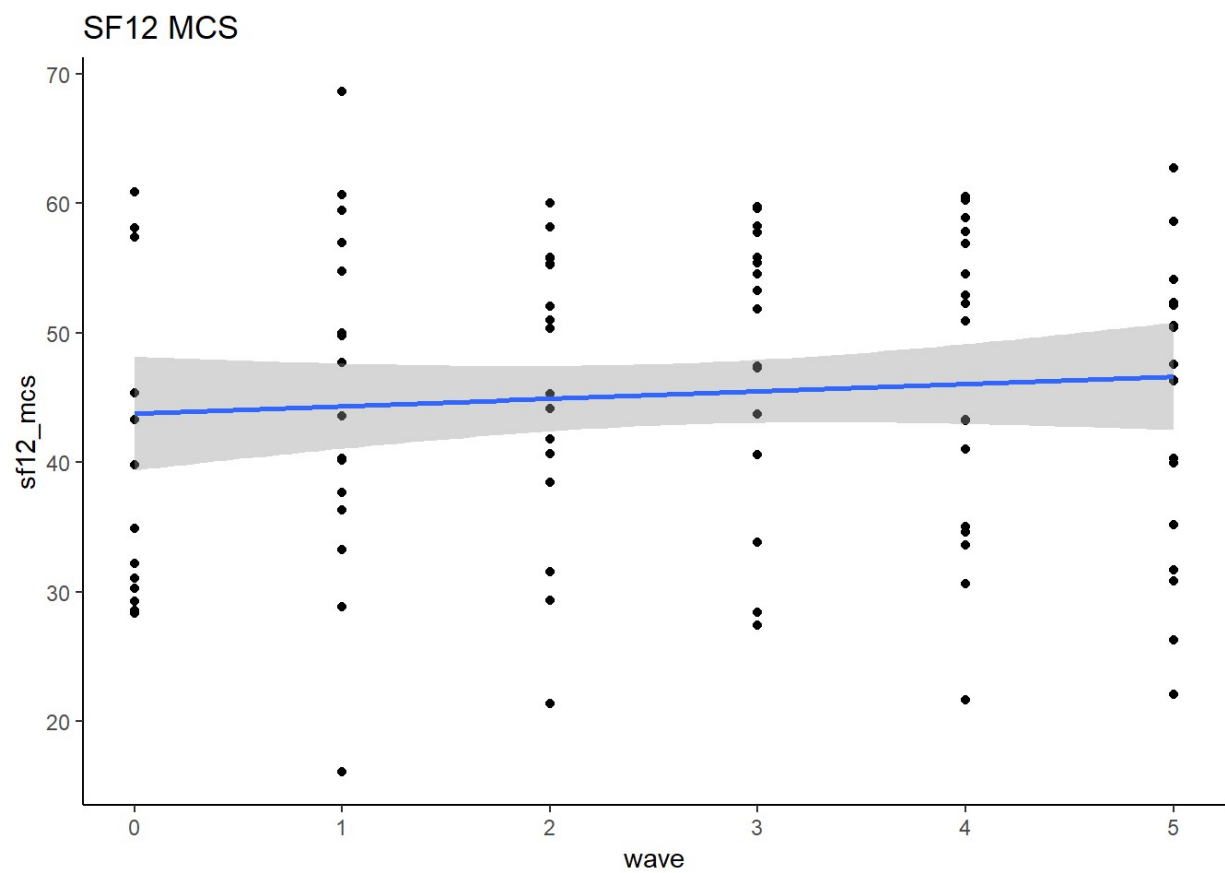
SF12 scores appear to slightly increase.

```
ggplot(candidates, aes(x=wave, y=sf12_pcs)) +
  geom_point() + geom_smooth(method="lm")+
  ggtitle("SF12 PCS") +
  theme_classic()
```
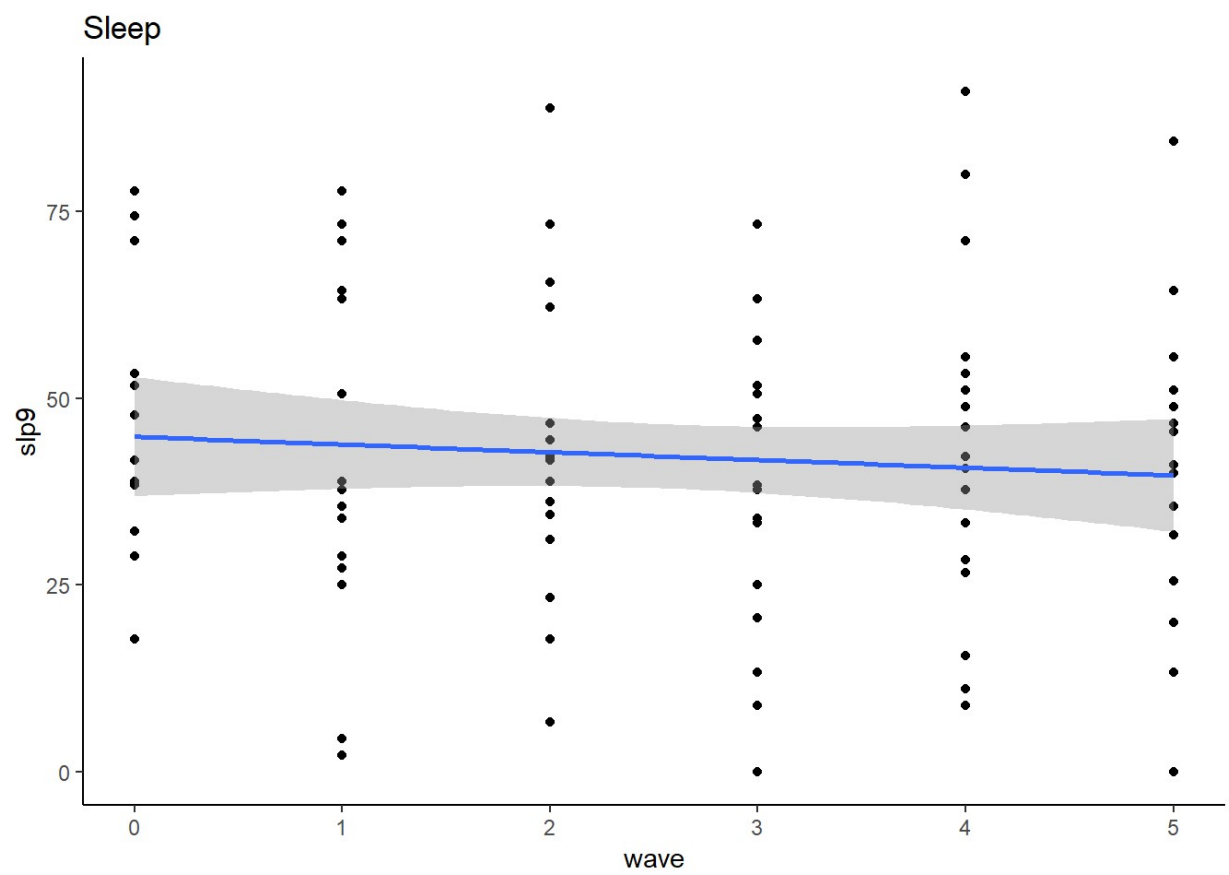
SF12 PCS
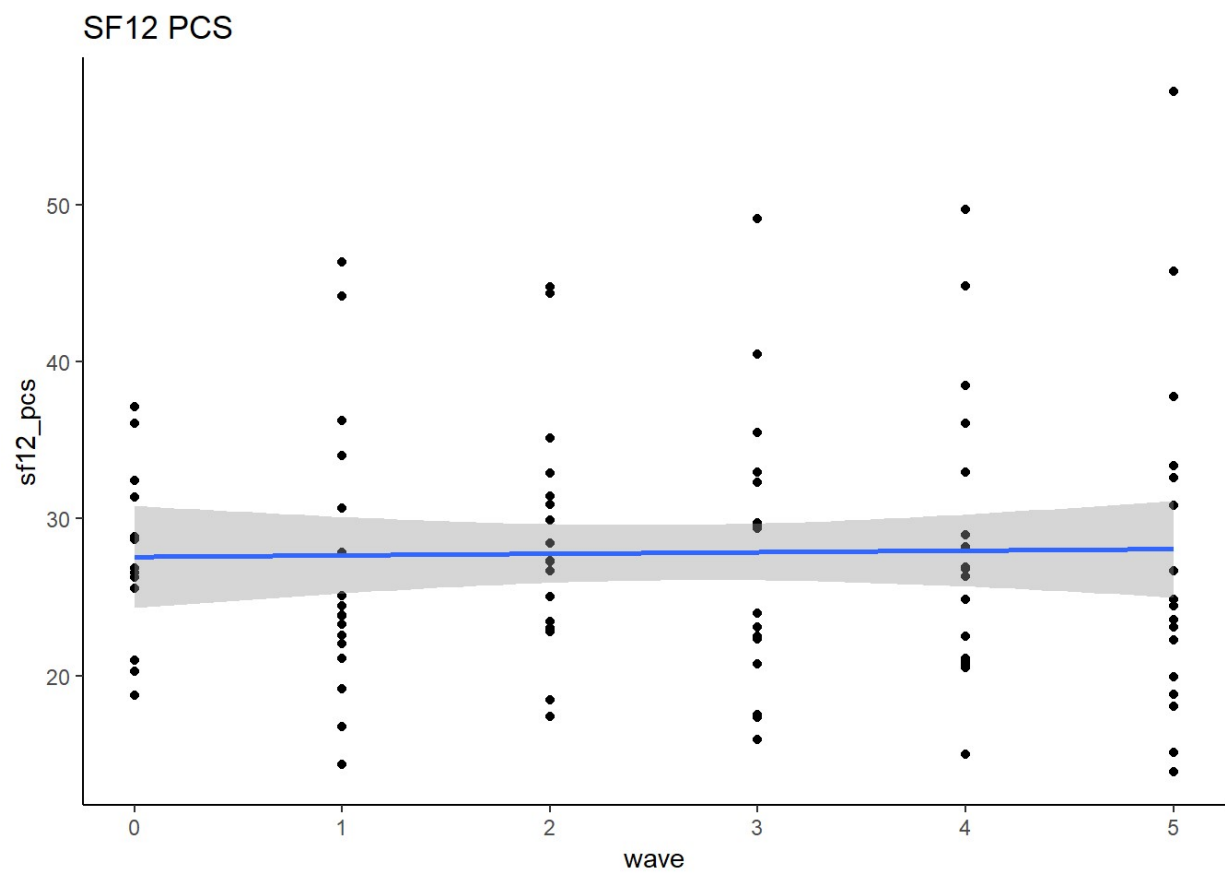
SF12 shows a slight increase across time.

```
ggplot(candidates, aes(x=wave, y=pseq_score)) +
  geom_point() + geom_smooth(method="lm")+
  ggtitle("PSEQ") +
  theme_classic()
```

PSEQ

PSEQ shows a slight increase across time.

```
ggplot(candidates, aes(x=wave, y=slp9)) +
  geom_point() + geom_smooth(method="lm")+
  ggtitle("Sleep") +
  theme_classic()
```

Sleep quality shows a slight decrease across time.