**Project:** Project 0, BIOS 6623

**Report:** Descriptive Statistics & Analysis

**Investigator:** Rachel Johnson

**Date:** 13 September 2017

**Introduction**

The goal of this project was to determine whether a new gel aimed at treating gum disease improved dental measurements after 1 year of use. Subjects were randomized to one of 5 groups, which included both a control group and placebo gel treatment group. The other three groups had differing levels of the active ingredient in the gel treatment: low, medium, or high. 26 subjects were originally randomized to each of these groups, resulting in a total of 130 subjects measured at baseline.

At baseline and at the 1 year follow-up visit, each subject’s gum attachment and gum pocket depth were measured at a number of sites, and these measurements were averaged to create an average attachment and average pocket depth score for each subject at each time point. The primary research question was whether the treatment results in lower average pocket depths and attachment losses after one year.

**Methods**

Originally, there were 130 subjects included in the subject who were randomized to 5 treatment groups. Since 27 subjects were missing both pocket depth and attachment measurements at 1 year, they were excluded from the analysis, including the descriptive analyses.

Variables were created to quantify the differences in pocket depth and attachment measurements between the baseline and 1 year follow-up time points, and these were used as the primary outcomes for the analyses.

For each of the two clinical outcomes, the difference between the baseline and 1 year follow-up measurements were modeled as a function of the baseline values for the respective measurement and the treatment group to best answer the primary research question. Additional covariates were not included in the models, since the trial was randomized and the univariate analyses were not significant at the level = 0.05, but the summaries of these demographic variables are included in Table 1.

All analyses were performed in R version 3.4.0.

**Results**

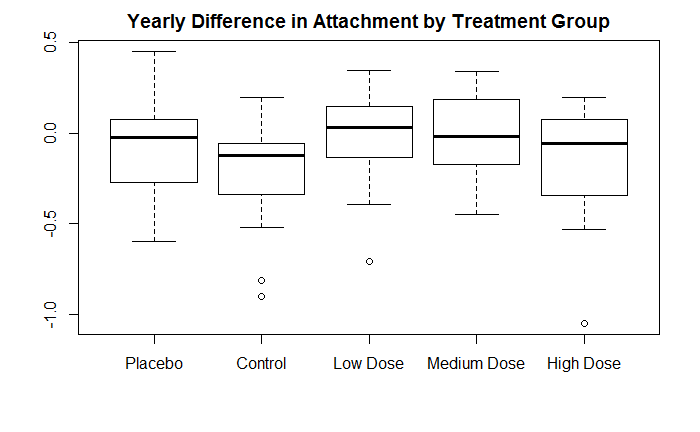
**Table 1.** This table describes the demographics of the 103 individuals who remained in the study for the entire year. Categorical variables are shown with counts and proportions for each level, and continuous variables are displayed with mean and standard deviations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Placebo** | **Control** | **Low Dose** | **Medium Dose** | **High Dose** |
| **Sex (n (prop))** |  |  |  |  |  |
| Male | 10 (0.43) | 7 (0.3) | 9 (0.43) | 7 (0.35) | 3 (0.19) |
| Female | 13 (0.57) | 16 (0.7) | 12 (0.57) | 13 (0.65) | 13 (0.81) |
| **Race (n (prop))** |  |  |  |  |  |
| Native American | 0 (0) | 1 (0.04) | 1 (0.05) | 0 (0) | 1 (0.06) |
| African American | 1 (0.04) | 1 (0.04) | 3 (0.14) | 0 (0) | 1 (0.06) |
| Asian | 1 (0.04) | 1 (0.04) | 0 (0) | 1 (0.05) | 0 (0) |
| White | 21 (0.91) | 20 (0.87) | 17 (0.81) | 19 (0.95) | 14 (0.88) |
| **Age (mean (sd))** | 47.18 (8.86) | 51.42 (10.16) | 51.47 (10.26) | 48.63 (9.87) | 52.61 (10.71) |
| **Smoker (n (prop))** |  |  |  |  |  |
| No | 13 (0.57) | 16 (0.7) | 15 (0.71) | 12 (0.6) | 11 (0.69) |
| Yes | 10 (0.43) | 7 (0.3) | 6 (0.29) | 9 (0.45) | 5 (0.31) |
| **Sites measured (mean (sd))** | 159.65 (10.66) | 153.91 (11.24) | 161.71 (8.16) | 153.5 (17.12) | 158.25 (8.73) |
| **Attachment at baseline (mean (sd))** | 1.83 (0.66) | 2.55 (0.65) | 2.1 (1.09) | 2.24 (0.67) | 2.31 (1.01) |
| **Attachment at 1 year (mean (sd))** | 1.74 (0.54) | 2.33 (0.55) | 2.08 (1.06) | 2.24 (0.65) | 2.15 (0.92) |
| **Pocket depth at baseline (mean (sd))** | 3.1 (0.39) | 3.29 (0.49) | 3.22 (0.64) | 3.05 (0.42) | 3.18 (0.28) |
| **Pocket depth at 1 year (mean (sd))** | 2.75 (0.48) | 2.95 (0.46) | 3.02 (0.58) | 2.84 (0.47) | 2.8 (0.42) |

**Outcome 1: Gum attachment**

First, the yearly difference in gum attachment—the baseline attachment subtracted from the attachment at 1 year—was compared among the five treatment groups.

**Figure 1.** This graph displays the yearly difference in attachment among the control, placebo, and three levels of treatment groups. The mean difference in attachment loss was -0.099.



**Table 2.** This table displays the beta estimates, 95% confidence intervals, and p-values for the treatment groups and baseline attachment measurement covariates in this multiple linear regression model.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Estimate** | **95% Confidence Interval** | **p-value** |
| **Intercept (Control)** | 0.107 | (-0.075, 0.289) | 0.253 |
| **Placebo** | 0.042 | (-0.107, 0.191) | 0.582 |
| **Low Dose** | 0.146 | (-0.003, 0.295) | 0.057 |
| **Medium Dose** | 0.176 | (0.027, 0.325) | 0.023 |
| **High Dose** | 0.027 | (-0.132, 0.186) | 0.742 |
| **Baseline Attachment** | -0.129 | (-0.188, -0.07) | <0.001 |

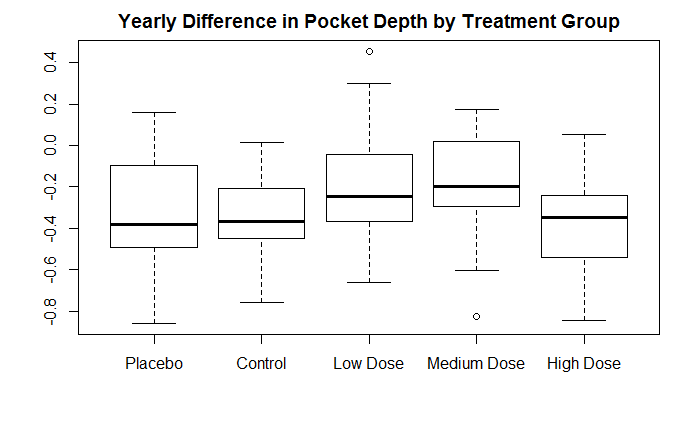
There were only two significant predictors of the difference in gum attachment: medium concentration of the active ingredient compared to the control group (p = 0.023) and baseline attachment (p < 0.001). However, the medium concentration in the gel did not reduce gum attachment measurements; compared to the control group, those in the medium concentration group had differences in gum attachment that were 0.176 points higher (95% CI: 0.027, 0.325), when controlling for baseline values.

Therefore, although there was an overall decrease in whole-mouth average gum attachment of approximately -0.010, this difference was only significantly explained by baseline gum attachment values. For every one point increase in baseline gum attachment, there was a decrease of 0.129 points (95% CI: -0.188, -0.07) in the yearly difference in gum attachment.

**Outcome 2: Pocket depth**

Second, the yearly difference in pocket depth—the baseline pocket depth subtracted from the pocket depth at 1 year—was compared among the five treatment groups.

**Figure 2.** This graph displays the yearly difference in pocket depth among the control, placebo, and three levels of treatment groups. The mean difference in pocket depth was -0.294.



**Table 3.** This table displays the beta estimates, 95% confidence intervals, and p-values for the treatment groups and baseline pocket depth covariates in this multiple linear regression model.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Estimate** | **95% Confidence Interval** | **p-value** |
| **Intercept (Control)** | 0.033 | (-0.343, 0.409) | 0.862 |
| **Placebo** | -0.033 | (-0.184, 0.118) | 0.67 |
| **Low Dose** | 0.125 | (-0.028, 0.278) | 0.113 |
| **Medium Dose** | 0.108 | (-0.049, 0.265) | 0.179 |
| **High Dose** | -0.056 | (-0.221, 0.109) | 0.506 |
| **Baseline Pocket Depth** | -0.113 | (-0.223, -0.003) | 0.046 |

There was an overall decrease in pocket depth of approximately -0.30, which was only significantly explained by baseline pocket depth (p = 0.046). For every 1 point increase in baseline pocket depth, there was a -0.113 decrease in yearly difference in pocket depth (95% CI: -0.223, -0.003).

**Conclusion**

Based on the results of the two models, treatment does not have a significant effect on producing lower average pocket depth and attachment loss at one year. The overall differences in attachment and pocket depth were negative, as hypothesized, but their magnitude was not explained by treatment group.

This study is potentially limited by the 27 individuals who had missing outcomes. Based on a brief analysis of these 27 individuals, there was a significant difference in proportions of men and women in the missing and non-missing data (p = 0.0058). The greatest number of individuals were missing from the high concentration of active ingredient group and the majority of these were male, so this should be noted as a limitation.

Furthermore, it should be noted that the results of this study are not generalizable to all populations: the majority of the population is white and middle-aged, and the proportion of smokers in this study is higher than that of the general population.

**Code**

#Load packages, import data, create difference variables

library(knitr)

setwd("~/School/AdvancedData")

gums <- read.csv("Project0\_dental\_data.csv", header = T)

gums$age <- as.numeric(gums$age)

gums$diffattach <- gums$attach1year - gums$attachbase

gums$diffpd <- gums$pd1year - gums$pdbase

#Create dataset with appropriate labels (from README)

labels <- gums

names(labels)

labels$trtgroup <- factor(labels$trtgroup, levels = c("1", "2", "3", "4", "5"),

labels = c("Placebo", "Control", "Low Dose",

"Medium Dose", "High Dose"))

labels$sex <- factor(labels$sex, levels = c("1", "2"), labels =

c("Male", "Female"))

labels$race <- factor(labels$race, levels = c("1", "2", "4", "5"),

labels = c("Native American", "African American",

"Asian", "White"))

labels$smoker <- factor(labels$smoker, levels = c("0", "1"), labels = c("No", "Yes"))

#Remove missing outcomes rows

labels <- labels[is.na(labels$pd1year) == FALSE, ]

gums <- gums[is.na(gums$pd1year) == FALSE, ]

missing <- gums[is.na(gums$pd1year) == TRUE, ]

missing$incl <- "No"

gums$incl <- "Yes"

tests <- rbind.data.frame(missing, gums)

table(tests$sex, tests$incl)

chisq.test(tests$sex, tests$incl) #p = 0.005827

#Divide datasets by tx

group1 <- labels[labels$trtgroup == "Placebo", ]

group2 <- labels[labels$trtgroup == "Control", ]

group3 <- labels[labels$trtgroup == "Low Dose", ]

group4 <- labels[labels$trtgroup == "Medium Dose", ]

group5 <- labels[labels$trtgroup == "High Dose", ]

#Create matrix and fill in table 1 (demographics)

num <- 3 + 5 + 1 + 3 + 1 + 1 + 1 + 1 + 1

tab <- matrix(data = NA, ncol = 6, nrow = 17)

colnames(tab) <- c("", levels(labels$trtgroup))

tab[, 1] <- c("Sex (n (%))", levels(labels$sex), "Race (n (%))", levels(labels$race),

"Age (mean (sd))", "Smoker (n (%))", levels(labels$smoker),

"Sites measured (mean (sd))", "Attachment at baseline (mean (sd))",

"Attachment at 1 year (mean (sd))", "Pocket depth at baseline (mean (sd))",

"Pocket depth at 1 year (mean (sd))")

for(i in 1:5){

tab[2, i+1] <- paste(nrow(gums[gums$trtgroup == i & gums$sex == 1, ]), "(",

round(nrow(gums[gums$trtgroup == i & gums$sex == 1, ])/nrow(gums[gums$trtgroup == i, ]), 2), ")")

tab[3, i+1] <- paste(nrow(gums[gums$trtgroup == i & gums$sex == 2, ]), "(",

round(nrow(gums[gums$trtgroup == i & gums$sex == 2, ])/nrow(gums[gums$trtgroup == i, ]), 2), ")")

tab[5, i+1] <- paste(nrow(gums[gums$trtgroup == i & gums$race == 1, ]), "(",

round(nrow(gums[gums$trtgroup == i & gums$race == 1, ])/nrow(gums[gums$trtgroup == i, ]), 2), ")")

tab[6, i+1] <- paste(nrow(gums[gums$trtgroup == i & gums$race == 2, ]), "(",

round(nrow(gums[gums$trtgroup == i & gums$race == 2, ])/nrow(gums[gums$trtgroup == i, ]), 2), ")")

tab[7, i+1] <- paste(nrow(gums[gums$trtgroup == i & gums$race == 4, ]), "(",

round(nrow(gums[gums$trtgroup == i & gums$race == 4, ])/nrow(gums[gums$trtgroup == i, ]), 2), ")")

tab[8, i+1] <- paste(nrow(gums[gums$trtgroup == i & gums$race == 5, ]), "(",

round(nrow(gums[gums$trtgroup == i & gums$race == 5, ])/nrow(gums[gums$trtgroup == i, ]), 2), ")")

tab[9, i+1] <- paste(round(mean(gums[gums$trtgroup == i,]$age, na.rm = T), 2), "(",

round(sd(gums[gums$trtgroup == i, ]$age, na.rm = T), 2), ")")

tab[11, i+1] <- paste(nrow(gums[gums$trtgroup == i & gums$smoker == 0, ]), "(",

round(nrow(gums[gums$trtgroup == i & gums$smoker == 0, ])/nrow(gums[gums$trtgroup == i, ]), 2), ")")

tab[12, i+1] <- paste(nrow(gums[gums$trtgroup == i & gums$smoker == 1, ]), "(",

round(nrow(gums[gums$trtgroup == i & gums$smoker == 1, ])/nrow(gums[gums$trtgroup == i, ]), 2), ")")

tab[13, i+1] <- paste(round(mean(gums[gums$trtgroup == i, ]$sites, na.rm = T), 2), "(",

round(sd(gums[gums$trtgroup == i, ]$sites, na.rm = T), 2), ")")

tab[14, i+1] <- paste(round(mean(gums[gums$trtgroup == i, ]$attachbase, na.rm = T), 2), "(",

round(sd(gums[gums$trtgroup == i, ]$attachbase, na.rm = T), 2), ")")

tab[15, i+1] <- paste(round(mean(gums[gums$trtgroup == i, ]$attach1year, na.rm = T), 2), "(",

round(sd(gums[gums$trtgroup == i, ]$attach1year, na.rm = T), 2), ")")

tab[16, i+1] <- paste(round(mean(gums[gums$trtgroup == i, ]$pdbase, na.rm = T), 2), "(",

round(sd(gums[gums$trtgroup == i, ]$pdbase, na.rm = T), 2), ")")

tab[17, i+1] <- paste(round(mean(gums[gums$trtgroup == i, ]$pd1year, na.rm = T), 2), "(",

round(sd(gums[gums$trtgroup == i, ]$pd1year, na.rm = T), 2), ")")

}

#Run Model 1 (attachment)

model1\_data <- gums[, c("id", "trtgroup", "attachbase", "attach1year", "diffattach")]

model1\_data$trtgroup <- factor(model1\_data$trtgroup, levels = c("2", "1", "3", "4", "5"))

model1 <- lm(diffattach ~ trtgroup + attachbase, data = model1\_data)

summary(model1)

#Run Model 2 (pocket depth)

model2\_data <- gums[, c("id", "trtgroup", "pdbase", "pd1year", "diffpd")]

model2\_data$trtgroup <- factor(model2\_data$trtgroup, levels = c("2", "1", "3", "4", "5"))

model2 <- lm(diffpd ~ trtgroup + pdbase, data = model2\_data)

summary(model2)

#Output table 1

kable(tab)

#Create boxplot for outcome 1 (attachment)

boxplot(gums$diffattach ~ labels$trtgroup,

main = "Yearly Difference in Attachment by Treatment Group")

#Model 1 analysis table

model1res <- summary(model1)$coef[, -3]

model1res <- round(model1res, 3)

rownames(model1res) <- c("Intercept (Control)", "Placebo", "Low Dose",

"Medium Dose", "High Dose", "Baseline Attachment")

model1res <- as.data.frame(model1res)

model1res$CI <- round(model1res$`Std. Error` \* 1.96, 3)

model1res$CIlow <- model1res$Estimate - model1res$CI

model1res$CIhigh <- model1res$Estimate + model1res$CI

model1res$CIfull <- paste("(", paste(model1res$CIlow, ",", model1res$CIhigh, sep = " "), ")", sep = "")

model1tab <- model1res[, c(1, 7, 3)]

colnames(model1tab) <- c("Estimate", "95% Confidence Interval", "p-value")

model1tab[6, 3] <- "<0.001" #because was rounding to 0

kable(model1tab)

#Create boxplot for outcome 2 (pocket depth)

boxplot(gums$diffpd ~ labels$trtgroup,

main = "Yearly Difference in Pocket Depth by Treatment Group")

#Create analysis table for model 2

model2res <- summary(model2)$coef[, -3]

model2res <- round(model2res, 3)

rownames(model2res) <- c("Intercept (Control)", "Placebo", "Low Dose",

"Medium Dose", "High Dose", "Baseline Pocket Depth")

model2res <- as.data.frame(model2res)

model2res$CI <- round(model2res$`Std. Error` \* 1.96, 3)

model2res$CIlow <- model2res$Estimate - model2res$CI

model2res$CIhigh <- model2res$Estimate + model2res$CI

model2res$CIfull <- paste("(", paste(model2res$CIlow, ",", model2res$CIhigh, sep = " "), ")", sep ="")

model2tab <- model2res[, c(1, 7, 3)]

colnames(model2tab) <- c("Estimate", "95% Confidence Interval", "p-value")

kable(model2tab)