**Project:** Project 0, BIOS 6623

**Report:** Descriptive Statistics & Analysis

**Investigator:** Rachel Johnson

**Date:** 13 September, 2017

**Introduction**

The goal of this project is to determine whether a new gel aimed at treating gum disease improves 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 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 attachments 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 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, 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; the 27 individuals with missing outcomes were excluded as they were excluded from the analysis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Placebo** | **Control** | **Low Dose** | **Medium Dose** | **High Dose** |
| **Sex (n (%))** |  |  |  |  |  |
| 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 (%))** |  |  |  |  |  |
| 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))** | 50.15 (9.95) | 50.15 (9.95) | 50.15 (9.95) | 50.15 (9.95) | 50.15 (9.95) |
| **Smoker (n (%))** |  |  |  |  |  |
| 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) |

**Conclusion**

* Introduction: Briefly describe the project (including data received from the investigator) and the scientific hypothesis of interest. Rephrase the scientific hypothesis of interest into testable statistical hypotheses (~.5 page).  Note this is NOT a scientific introduction to a paper and should give information pertinent to the data analysis, not more general biology background.
* Methods: Describe the methods used to clean and analyze the data. Justify and explain your data analysis approach (~ 2 pages).  Should be written in past tense and should not include results.  Do NOT include equations.
* Results: Present results for analyses described in the methods (~1-1.5 pages). Use Tables and Figures as appropriate, including in the text the full interpretation of statistical results for the main findings (i.e. point estimates, confidence bounds, p-values, interpretation of results of test).
* Conclusions: Interpret your results (~.5 to 1 page) in context of scientific question(s). Also discuss any limitations to your analysis that may affect interpretation or that require additional consideration by the investigator.

**Reproducible Research:** Good biostatistical practice requires documentation of data sources and code such that another biostatistician could reproduce your analysis if given the appropriate data and code you used to produce your report.  The last pages of your report (not included in the 6-page limit) should include a directory and file name listing of the location of your data and statistical code you used to create the report (your github directory link).  The top of your statistical code should include a note about the directory location of the data being read into your statistical program.  For instance, using the data import function in SAS without copying the code used to import the data does not constitute reproducible research.  The submitted code should be limited to that used for justifying and obtaining the results you present in the report and should be clearly commented so that another biostatistician can easily follow your steps. Exploratory work done to justify the final analysis may also be included. But it should be commented out (i.e., will not run if the code is run) and include comments on the rationale of the work.