Project 3: Periodontal Disease and Preterm Birth

Jin Cao, Henry Wong, Jennifer Zhang

13 March, 2024

Table of Contents

Investigator

Ronald G. Thomas (rgthomas@health.ucsd.edu)

Primary Analysts

- Jin Cao (j9cao@ucsd.edu)
- Henry Wong (hjwong@ucsd.edu)
- Jennifer Zhang (jxz009@ucsd.edu)

Introduction

Periodontitis, also called gum disease, is a serious bacterial infection that destroys connective tissue and bone around teeth. Recent studies have indicated that periodontitis during pregnancy is associated with an increased risk of preterm birth, low birth weight, and preeclampsia, a serious hypertension condition that could develop during pregnancy.

Investigators speculate that the transport of bacteria from sites of periodontal infection into the placenta, fetal membranes, and amniotic cavity leads to these adverse outcomes (Srinivas 2012). The Obstetrics and Periodontal Therapy study, which evaluated the effect of nonsurgical periodontal treatment on preterm birth, failed to find a significant association between maternal periodontal treatment and preterm birth via a survival analysis model. (Michalowicz et al., 2006)

In this study, our group plans to re-analyze the original data in the Obstetrics and Periodontal Therapy. We hypothesize that nonsurgical periodontal treatment is associated with preterm birth when the baseline periodontal health is adjusted for.

Variables

Overall dataset contains 823 observations and 171 variables.

Primary Variable	Variable Label	Description
PID	Participant ID	
Clinic	Enrollment center	Four centers
Group	Randomized treatment assignment	T=Intervention;C=Control
Birth.outcome	Birth outcome	Elective abortion; Live birth; Lost to follow-up; Non-live birth
GA.at.outcome	Gestational age at end of pregnancy	days

Dataset also contains variables for baseline risk factors, periodontal therapy and health summary, dental care, obstetrical outcomes, risk factors for pregnancies, microbiological information, and immunological information. Since we are investigating the association between periodontal disease and preterm birth, we may also need to account for initial periodontal health. Therefore, we are also interested in the following variables measured during baseline in the dataset.

Baseline Periodontal Summary	Variable Label
N.qualifying.teeth	Number of teeth meeting OPT
BL.GE	Whole mouth average gingival index
BL..BOP	Fraction of sites bleeding on probing

Baseline Periodontal Summary	Variable Label
BL.PD.avg	Whole mouth average pocket depth at baseline
BL.CAL.avg	Whole mouth average clinical attachment level
BL.Calc.I	Whole mouth average calculus index
BL.Pl.I	Whole mouth average plaque index

In the dataset, there were 103 preterm births out of 814 (9 lost to followup and were censored at 37 weeks). Below are two table for Normal/Preterm birth stratified by treatment groups. Unfortunately, there does not appear to a treatment effect when compared between the groups.

Table 1: Contingency table for extremely preterm birth

	NOT Extremely Preterm	Extremely Preterm
Control	393(48%)	17(2%)
Treatment	400(49%)	13(2%)

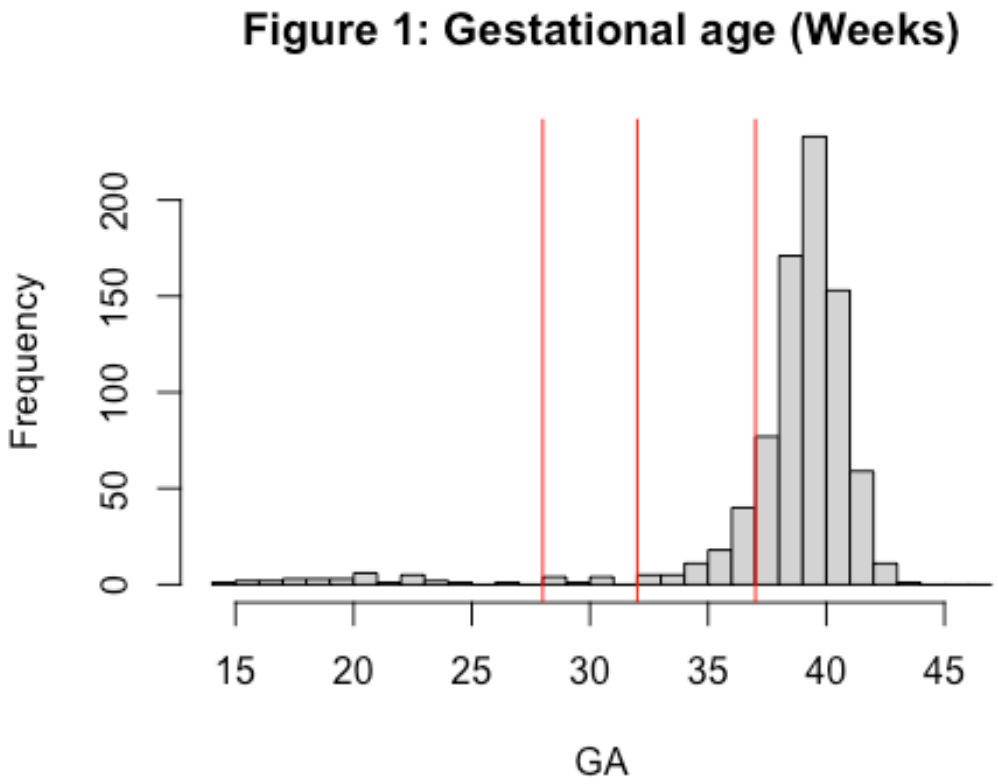
Table 2: Contingency table for very preterm birth

	NOT Very Preterm	Very Preterm
Control	387(47%)	23(3%)
Treatment	397(48%)	16(2%)

Table 3: Contingency table for late preterm birth

	Normal	Late Preterm
Control	353(43%)	57(7%)
Treatment	358(43%)	55(7%)

The figure below shows the distribution of gestational age in weeks. The red vertical lines indicates the 28th, 32nd, and 37th week so any births left of the red lines which indicate extreme preterm, very preterm, and moderate to late preterm respectively.



Lastly, gestation age and baseline periodontal summaries are summarized below. There does not appear to be any significance differences between control and treatment groups.

Table 4: Descriptive statistics table

	C / Mean	T / Mean	C / SD	T / SD	p-value
Gestational Age (Weeks)	38.26	38.45	4.25	3.81	0.498
Number of Teeth meeting OPT	14.36	15.16	6.69	6.79	0.494
Avg Gingival Index	1.42	1.45	0.40	0.44	0.499
% Bleeding on Probing	69.10	69.76	17.07	17.43	0.497
Avg Pocket Depth	2.84	2.90	0.53	0.59	0.498
Avg Clinical Attachment Level	1.37	1.45	0.66	0.71	0.498
Avg Calculus Index	1.13	1.15	0.62	0.62	0.500
Avg Plaque Index	1.23	1.24	0.48	0.50	0.500

Methods

(PART 1) For our proposed analyses, we first plan to replicate the results from the original study using a Cox Proportional hazards model while controlling for the four enrollment

centers. We do not expect the results to be significant, rather, we would like to verify the lack of findings in the original study. Moreover, we are interested in the following regression model formula:

Late Preterm Birth ~ Group + Clinic

(PART 2) We are primarily interested in performing further exploratory analyses. We hypothesize that it may be important to adjust for baseline periodontal health and possible Clinic site. Thus, we plan to perform additional exploratory analyses using Firth's penalized logistic regression to handle rare events while adjusting for periodontal health at baseline. Furthermore, we will be estimating the treatment effect using doubly robust estimation. We may also be interested in adjusting for clinic sites in the logistic regression if we find a statistical significance. In addition, we would like to explore different categories of preterm birth (extremely preterm at 28 weeks, very preterm at 32 weeks, and late preterm at 37 weeks). The p-values for treatment effect will also be adjusted for multiple comparisons.

(1) Extremely Preterm Birth ~ Group + (Clinic + Baseline Periodontal Summaries)

(2) Very Preterm Birth ~ Group + (Clinic + Baseline Periodontal Summaries)

(3) Late Preterm Birth ~ Group + (Clinic + Baseline Periodontal Summaries)

PART 1: Survival Analysis

To validate the results of the original paper by Michalowicz et al., we propose employing a Cox proportional hazard model. In the original study, log-rank tests were conducted on the Kaplan-Meier curves between groups, stratified by clinic sites. A log-partial-likelihood test was utilized to assess model performance, revealing Clinic as a statistically significant variable to adjust for ($p=0.048$). However, adjusting for baseline periodontal summaries was determined to not significantly enhance the model fit ($p=0.242$). The model we will interpret is:

Late Preterm Birth ~ Group + Clinic.

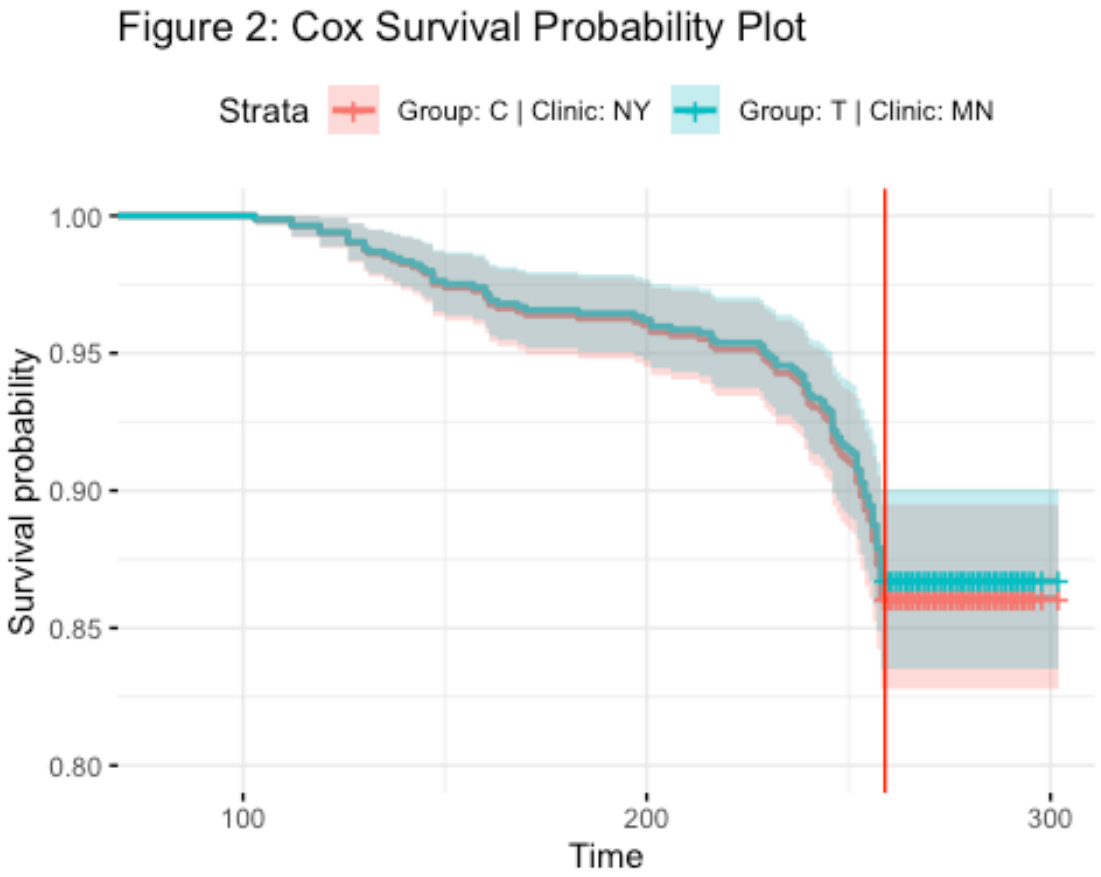
We also tested the proportional hazards assumption using the cox.zph() function and found that the assumptions hold.

Upon examining the model summary, it is evident that treating for periodontal disease reduced the estimated hazard by a factor of 0.954 (4.6% decrease) compared to no treatment. Furthermore, the estimated hazard for individuals treated at the Hennepin County Center (ClinicMN) decreased by a factor of 0.873 (12.7% decrease) compared to those treated at the University of Kentucky (ClinicKY). Interestingly, the estimated hazard increased by factors of 1.557 (55.7% increase) and 1.590 (59.0% increase) for those treated at the University of Mississippi Medical Center (ClinicMS) and Harlem Hospital (ClinicNY), respectively. Unfortunately, none of the variables reached statistical significance, and the results are presented in Table 5.

Table 5: Cox PH model summary for late preterm births with the control group and KY as the references

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
GroupT	-0.047	0.954	0.189	-0.251	0.802
ClinicMN	-0.136	0.873	0.286	-0.476	0.634
ClinicMS	0.443	1.557	0.268	1.651	0.099
ClinicNY	0.464	1.590	0.274	1.694	0.090

Finally, using the Cox proportional hazard model, we were able to fit the survival probabilities for individuals most at risk for late preterm birth and those least at risk. The model indicates that individuals in the control group treated at Harlem Hospital are at the highest risk (as shown by the red curve), while those in the treatment group treated at Hennepin County Center are at the lowest risk (as indicated by the blue curve). The fitted survival probabilities are illustrated in Figure 2 below. It appears that, to a small extent, individuals least at risk (blue curve) have a slightly higher survival probability than those most at risk (red curve); however, there is a significant overlap in the 95% confidence intervals.



PART 2: Firth Logistic Regression

After removing subjects who had an elective abortion or were lost to follow-up, we obtain 12 subjects in the control group and 7 in the treatment group who had extremely preterm birth. There does not appear to be a significant treatment effect in those who had extremely preterm birth.

From fitting the Firth logistic regression model on the extremely preterm birth group, we see that enrollment center contributes significantly to risk of extremely preterm birth. Compared to the University of Kentucky (KY), one's risk of extremely preterm birth at the Hennepin County Medical Center (MN), University of Mississippi Medical Center (MS), and Harlem Hospital (NY) increase by an estimated multiplicative factor of 5.666 (95% CI: 1.144, 27.989), 7.589 (95% CI: 2.329, 28.900), and 33.718 (95% CI: 8.891, 143.811), respectively.

Another covariate that stood out in the model summary was whole mouth average plaque index at baseline (BL.Pl.I). For every 1 unit increase in score on the Silness-Löe Plaque Index, one's risk of extremely preterm birth increases by an estimated multiplicative factor of 8.798, with a 95% confidence interval of (3.219, 24.763).

Whole mouth average gingival index at baseline (BL.GE) appeared to be associated with a decreased risk of extremely preterm birth. For every 1 unit increase in score on the Silness-Löe Gingival Index, one's risk of extremely preterm birth decreases by an estimated 97.8%, with a 95% confidence interval of (-99.5%, -91.2%).

The results for the extremely preterm birth group is summarized in Table 7.

Table 6: Contingency table for extremely preterm birth after removing elective abortion and loss to follow-up

	NOT Extremely Preterm	Extremely Preterm
Control	393(48%)	12(1%)
Treatment	400(49%)	7(1%)

Table 7: Firth logisitic regression of extremely preterm including 95% confidence intervals and pvalue computed using profile penalized log-likelihood

	Coefficients	lowerCI	upperCI	pvalue
(Intercept)	0.419	0.006	35.284	0.694
GroupT	0.559	0.277	1.093	0.089
ClinicMN	5.666	1.144	27.989	0.034
ClinicMS	7.589	2.329	28.900	0.001
ClinicNY	33.718	8.981	143.811	0.000
N.qualifying.teeth	1.133	1.015	1.273	0.026
BL.GE	0.022	0.005	0.088	0.000

	Coefficients	lowerCI	upperCI	pvalue
BL..BOP	1.048	1.020	1.077	0.001
BL.PD.avg	0.113	0.013	0.839	0.032
BL.Calc.I	0.385	0.146	0.943	0.036
BL.Pl.I	8.798	3.219	24.763	0.000

After removing subjects who had an elective abortion or were lost to follow-up, we obtain 18 subjects in the control group and 10 in the treatment group who had very preterm birth. There is a significant treatment effect in those who had very preterm birth. However, when the result is no longer significant after adjusting for multiple comparisons.

From fitting the Firth logistic regression model on the very preterm birth group, we see that enrollment center contributes significantly to risk of very preterm birth. Compared to the University of Kentucky (KY), one's risk of very preterm birth at the Hennepin County Medical Center (MN), University of Mississippi Medical Center (MS), and Harlem Hospital (NY) increase by an estimated multiplicative factor of 4.957 (95% CI: 1.560, 16.043), 5.150 (95% CI: 1.988, 14.163), and 12.099 (95% CI: 4.043, 37.481), respectively.

Another covariate that stood out in the model summary was the whole mouth average plaque index at baseline (BL.Pl.I). For every 1 unit increase in score on the Silness-Löe Plaque Index, one's risk of very preterm birth increases by an estimated multiplicative factor of 3.672, with a 95% confidence interval of (1.585, 8.552).

Whole mouth average gingival index at baseline (BL.GE) appeared to be associated with a decreased risk of very preterm birth. For every 1 unit increase in score on the Silness-Löe Gingival Index, one's risk of very preterm birth decreases by an estimated 93.2%, with a 95% confidence interval of (-97.9%, -77.5%).

The results for the very preterm birth group is summarized in Table 9.

Table 8: Contingency table for very preterm birth after removing elective abortion and loss to follow-up

	NOT Very Preterm	Very Preterm
Control	387(48%)	18(2%)
Treatment	397(49%)	10(1%)

Table 9: Firth logisitic regression of very preterm including 95% confidence intervals and pvalue computed using profile penalized log-likelihood

	Coefficients	lowerCI	upperCI	pvalue
(Intercept)	0.195	0.010	4.737	0.309
GroupT	0.539	0.303	0.933	0.027
ClinicMN	4.957	1.560	16.043	0.007
ClinicMS	5.150	1.988	14.163	0.001

	Coefficients	lowerCI	upperCI	pvalue
ClinicNY	12.099	4.043	37.481	0.000
N.qualifying.teeth	1.094	1.005	1.195	0.037
BL.GE	0.068	0.021	0.225	0.000
BL..BOP	1.038	1.014	1.062	0.001
BL.PD.avg	0.270	0.055	1.133	0.076
BL.Calc.I	0.400	0.179	0.857	0.018
BL.Pl.I	3.672	1.585	8.552	0.002

After removing subjects who had an elective abortion or were lost to follow-up, we obtain 52 subjects in the control group and 49 in the treatment group who had late preterm birth. There does not appear to be a significant treatment effect in those who had late preterm birth.

From fitting the Firth logistic regression model on the late preterm birth group, we see that those enrolled at University of Mississippi Medical Center (MS) and Harlem Hospital (NY) had a significantly increased risk of late preterm birth. Compared to the University of Kentucky (KY), one's risk of late preterm birth at MS and NY increase by an estimated multiplicative factor of 3.093 (95% CI: 1.830, 5.264) and 3.504 (95% CI: 1.857, 6.618), respectively. The estimated risk of late preterm birth at Hennepin County Center (MN) compared to KY was 1.422 (95% CI: 0.785, 2.580); this estimate was not statistically significant, unlike the extremely and very preterm birth groups.

Another covariate that stood out in the model summary was whole mouth average plaque index at baseline (BL.Pl.I). For every 1 unit increase in score on the Silness-Löe Plaque Index, one's risk of late preterm birth increases by an estimated multiplicative factor of 1.994, with a 95% confidence interval of (1.253, 3.185).

Whole mouth average gingival index at baseline (BL.GE) appeared to be associated with a decreased risk of late preterm birth. For every 1 unit increase in score on the Silness-Löe Gingival Index, one's risk of late preterm birth decreases by an estimated 73.4%, with a 95% confidence interval of (-86.4%, -47.9%).

The results for the late preterm birth group is summarized in Table 11.

Table 10: Contingency table for late preterm birth after removing elective abortion and loss to follow-up

	NOT Preterm	Late Preterm
Control	353(43%)	52(6%)
Treatment	358(44%)	49(6%)

Table 11: Firth logisitic regression of late preterm including 95% confidence intervals and pvalue computed using profile penalized log-likelihood

	Coefficients	lowerCI	upperCI	pvalue
(Intercept)	0.058	0.021	0.171	0.000
GroupT	0.953	0.707	1.284	0.753
ClinicMN	1.422	0.785	2.580	0.245
ClinicMS	3.093	1.830	5.264	0.000
ClinicNY	3.504	1.857	6.618	0.000
N.qualifying.teeth	0.992	0.958	1.029	0.658
BL.GE	0.266	0.136	0.521	0.000
BL..BOP	1.010	0.998	1.023	0.105
BL.PD.avg	1.267	0.767	1.977	0.337
BL.Calc.I	0.994	0.651	1.509	0.978
BL.Pl.I	1.994	1.253	3.185	0.004

Compared to KY, we see that MN, MS, and NY had the least to greatest increase in risk of preterm birth for all 3 categories (extremely preterm, very preterm, late preterm). All coefficient estimates were highly significant, with the exception of ClinicMN for the late preterm group. We can also see that the coefficient estimates decreased in extremity as we progressed from extremely to late preterm birth.

Coefficient estimates for whole mouth average plaque index at baseline (BL.Pl.I) and whole mouth average gingival index at baseline (BL.GE) were also highly significant for all 3 categories. The estimates followed a similar pattern of decreasing in extremity as we progressed from extremely to late preterm birth.

The coefficient estimates for BL.GE suggest that an increase on the score on the Silness-Löe Gingival Index is associated with a decrease in risk of preterm birth, which seems counterintuitive. Since the Firth's logistic regression model already reduces bias due to collinearity, we are unsure of how this may have occurred.

Overall, we observed that treating for periodontal diseases decreases the risk of extremely, very, and late preterm birth by a factor of 0.559 (44.1% decrease), 0.539 (46.1% decrease), and 0.953 (4.7% decrease), respectively. Unfortunately, we do not observed any statistically significant results after adjusting the p-values for multiple comparisons.

	ExtremelyPreterm	VeryPreterm	LatePreterm
(Intercept)	0.419	0.195	0.058
GroupT	0.559	0.539	0.953
ClinicMN	5.666	4.957	1.422
ClinicMS	7.589	5.150	3.093
ClinicNY	33.718	12.099	3.504

	ExtremelyPreterm	VeryPreterm	LatePreterm
N.qualifying.teeth	1.133	1.094	0.992
BL.GE	0.022	0.068	0.266
BL..BOP	1.048	1.038	1.010
BL.PD.avg	0.113	0.270	1.267
BL.Calc.I	0.385	0.400	0.994
BL.Pl.I	8.798	3.672	1.994

Conclusion and Scientific Interpretation

In the survival analysis setting, we confirmed that there is not a statistically significant treatment effect when using the Cox proportional hazards model. Moreover, we investigated the effects of periodontal health and late preterm birth while adjusting for clinic sites.

For our primary analysis, we wanted to further investigate the effects of periodontal health on the different categories of preterm birth. Using Firth's logistic regression with doubly robust estimation, we found a significant treatment effect for the very preterm group, but not for the extremely preterm and late preterm groups. However, significance is lost after adjusting the p-value for multiple comparisons, but we do notice a trend that the treatment may have an impact on the more severe preterm births. Further research on these subgroups with more participants is recommended.

In the context of Firth's logistic regression, we observed that the clinic, baseline whole mouth average plaque index (BL.Pl.I), and baseline whole mouth average gingival index (BL.GE) emerged as the most significant contributors to the risk of preterm birth. The notably high odds ratios associated with different clinics suggest the necessity of delving deeper into how treatment received at specific clinics may be linked to preterm birth. Across all variables, we observe a gradual decrease in odds ratios from extremely to late preterm birth, converging towards 1. This implies that factors such as group, clinic choice, and baseline periodontal summaries may carry greater significance earlier in pregnancy when the risk of extremely preterm birth is at its peak. Future research endeavors might explore the impacts of periodontal health across various categories of preterm birth.

Reference

Michalowicz BS, Hodges JS, DiAngelis AJ, Lupo VR, Novak MJ, Ferguson JE, Buchanan W, Bofill J, Papapanou PN, Mitchell DA, Matseoane S, Tschida PA; OPT Study. Treatment of periodontal disease and the risk of preterm birth. N Engl J Med. 2006 Nov 2;355(18):1885-94. doi: 10.1056/NEJMoa062249. PMID: 17079762.

Srinivas SK, Parry S. (2012). Periodontal disease and pregnancy outcomes: time to move on? Journal of Women's Health, 21(2), 121-5.

Quinn JA, Munoz FM, Gonik B, Frau L, Cutland C, Mallett-Moore T, Kissou A, Wittke F, Das M, Nunes T, Pye S, Watson W, Ramos AA, Cordero JF, Huang WT, Kochhar S, Buttery J; Brighton Collaboration Preterm Birth Working Group. Preterm birth: Case definition & guidelines for data collection, analysis, and presentation of immunisation safety data. Vaccine. 2016 Dec 1;34(49):6047-6056. doi: 10.1016/j.vaccine.2016.03.045. Epub 2016 Oct 13. PMID: 27743648; PMCID: PMC5139808.

Firth D (1993). Bias reduction of maximum likelihood estimates. Biometrika 80, 27-38. Heinze G, Schemper M (2002). A solution to the problem of separation in logistic regression. Statistics in Medicine 21: 2409-2419.

Puhr, R., Heinze, G., Nold, M., Lusa, L., and Geroldinger, A. (2017) Firth's logistic regression with rare events: accurate effect estimates and predictions?. Statist. Med., 36: 2302–2317. doi: 10.1002/sim.7273.

Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. Am J Epidemiol. 2011 Apr 1;173(7):761-7. doi: 10.1093/aje/kwq439. Epub 2011 Mar 8. PMID: 21385832; PMCID: PMC3070495.

Appendix R Code

Setup

```
rm(list = ls())

library(tidyverse)
library(alr4)
library(modelsummary)
library(survival)
library(survminer)
library(logistf)

# Load data for preprocessing
df <- medicaldata::opt %>%
  mutate_if(is.factor,as.character)

# Create new preterm variables
df2 <- df %>%
  mutate(ExtremelyPreterm = ifelse(GA.at.outcome/7 < 28, 1,0),
         VeryPreterm = ifelse(GA.at.outcome/7 < 32, 1,0),
         LatePreterm = ifelse(GA.at.outcome/7 < 37, 1,0))

# Select variables of interest
df3 <- df2 %>%
  select(PID,Clinic,Group,Birth.outcome,GA.at.outcome,
         ExtremelyPreterm,VeryPreterm,LatePreterm,

N.qualifying.teeth,BL.GE,BL..BOP,BL.PD.avg,BL.CAL.avg,BL.Calc.I,BL.Pl.I)
```

Variable Description

```
tab = with(df3,table(Group,ExtremelyPreterm))
ptab = round(prop.table(tab),2)

ctab = matrix(paste0(tab,"(",100*ptab,"%)"),ncol = 2, nrow = 2)
colnames(ctab) <- c("NOT Extremely Preterm","Extremely Preterm")
rownames(ctab) <- c("Control","Treatment")

knitr::kable(ctab, caption = "Table 1: Contingency table for extremely
preterm birth")

tab = with(df3,table(Group,VeryPreterm))
ptab = round(prop.table(tab),2)

ctab = matrix(paste0(tab,"(",100*ptab,"%)"),ncol = 2, nrow = 2)
colnames(ctab) <- c("NOT Very Preterm","Very Preterm")
rownames(ctab) <- c("Control","Treatment")

knitr::kable(ctab, caption = "Table 2: Contingency table for very preterm
```

```
birth")

tab = with(df3,table(Group,LatePreterm))
ptab = round(prop.table(tab),2)

ctab = matrix(paste0(tab,"(",100*ptab,"%)"),ncol = 2, nrow = 2)
colnames(ctab) <- c("Normal","Late Preterm")
rownames(ctab) <- c("Control","Treatment")

knitr::kable(ctab, caption = "Table 3: Contingency table for late preterm
birth")

hist(df3$GA.at.outcome/7,breaks = 14:47,
     main = "Figure 1: Gestational age (Weeks)",
     xlab = "GA")
abline(v=28, col = "red")
abline(v=32, col = "red")
abline(v=37, col = "red")

dfoutput <- datasummary(
  GA.at.outcome/7 + N.qualifying.teeth + BL.GE + BL..BOP + BL.PD.avg +
BL.CAL.avg + BL.Calc.I + BL.Pl.I ~ Group*Mean + Group*SD + Group*N,
  data = df3, output = "data.frame")

rownames(dfoutput) <- c("Gestational Age (Weeks)","Number of Teeth meeting
OPT","Avg Gingival Index","% Bleeding on Probing","Avg Pocket Depth","Avg
Clinical Attachment Level","Avg Calculus Index","Avg Plaque Index")

dfoutput <- dfoutput %>%
  mutate(`C / Mean` = as.numeric(`C / Mean`),
         `T / Mean` = as.numeric(`T / Mean`),
         `C / SD` = as.numeric(`C / SD`),
         `T / SD` = as.numeric(`T / SD`),
         `C / N` = as.numeric(`C / N`),
         `T / N` = as.numeric(`T / N`))

dfoutput$`p-value` <- round(with(dfoutput,pnorm( (`C / Mean` - `T /
Mean`)/sqrt((`C / SD`+`T / SD`)/2 * (1/`C / N`+`T / N` )) )),3)

knitr::kable(dfoutput[,c(2:5,8)], caption = "Table 4: Descriptive statistics
table")
```

PART 1: Survival Analysis

```
cox_fit3 <- coxph(Surv(time = GA.at.outcome, event = LatePreterm) ~ Group,
data = df3)

cox_fit3.1 <- coxph(Surv(time = GA.at.outcome, event = LatePreterm) ~ Group +
Clinic, data = df3)
```

```r
cox_fit3.2 <- coxph(Surv(time = GA.at.outcome, event = LatePreterm) ~ Group +
                      Clinic + N.qualifying.teeth + BL.GE + BL..BOP +
                      BL.PD.avg + BL.Calc.I + BL.Pl.I, data = df3)

anova(cox_fit3,cox_fit3.1)

anova(cox_fit3.1,cox_fit3.2)

cox.zph(cox_fit3.1)

scox_fit3.1 <- summary(cox_fit3.1)

knitr::kable(round(scox_fit3.1$coefficients,3), caption = "Table 5: Cox PH
model summary for late preterm births with the control group and KY as the
references")

newdf <- data.frame(Group = c("C","T"), Clinic = c("NY","MN"))

coxplot <- ggsurvplot(survfit(cox_fit3, newdata = newdf), data = df3,
                      ggtheme = theme_minimal(),
                      legend.labs = c("Group: C | Clinic: NY","Group: T |
Clinic: MN"),
                      title = "Figure 2: Cox Survival Probability Plot",
                      xlim = c(80,300),
                      ylim = c(.8,1))

coxplot$plot + geom_vline(xintercept = 37*7, col = "red")
```

PART 2: Firth Logistic Regression

```r
df3 <- df3 %>%
  filter(Birth.outcome == "Live birth          " |
           Birth.outcome == "Non-live birth      ") %>%
  mutate(GroupBi = ifelse(Group == "T",1,0))

propensity_fit <- glm(GroupBi ~ Clinic + N.qualifying.teeth + BL.GE + BL..BOP
+ BL.PD.avg + BL.Calc.I + BL.Pl.I, data = df3, family = binomial)
propensity <- predict(propensity_fit, type = "response")
ws <- ifelse(df3$Group == "T", 1/(propensity), 1/(1-propensity))

tab = with(df3,table(Group,ExtremelyPreterm))
ptab = round(prop.table(tab),2)

ctab = matrix(paste0(tab,"(",100*ptab,"%)"),ncol = 2, nrow = 2)
colnames(ctab) <- c("NOT Extremely Preterm","Extremely Preterm")
rownames(ctab) <- c("Control","Treatment")

knitr::kable(ctab, caption = "Table 6: Contingency table for extremely
preterm birth after removing elective abortion and loss to follow-up")
```

```r
#firth_fit1 <- glm(ExtremelyPreterm ~ Group + N.qualifying.teeth + BL.GE + BL..BOP + BL.PD.avg + BL.Calc.I + BL.Pl.I, data = df3, weights = ws, family = binomial)

firth_fit1.1 <- logistf(ExtremelyPreterm ~ Group + Clinic + N.qualifying.teeth + BL.GE +  BL..BOP + BL.PD.avg + BL.Calc.I + BL.Pl.I, data = df3, weights = ws)

#summary(firth_fit1)
#summary(firth_fit1.1)

output_df_EPreterm <- data.frame(Coefficients = round(exp(firth_fit1.1$coefficients),3),
                                 lowerCI = round(exp(firth_fit1.1$ci.lower),3),
                                 upperCI = round(exp(firth_fit1.1$ci.upper),3),
                                 pvalue = round(firth_fit1.1$prob,3))

knitr::kable(output_df_EPreterm, caption = "Table 7: Firth logisitic regression of extremely preterm including 95% confidence intervals and pvalue computed using profile penalized log-likelihood")

tab = with(df3,table(Group,VeryPreterm))
ptab = round(prop.table(tab),2)

ctab = matrix(paste0(tab,"(",100*ptab,"%)"),ncol = 2, nrow = 2)
colnames(ctab) <- c("NOT Very Preterm","Very Preterm")
rownames(ctab) <- c("Control","Treatment")

knitr::kable(ctab, caption = "Table 8: Contingency table for very preterm birth after removing elective abortion and loss to follow-up")

#firth_fit2 <- glm(VeryPreterm ~ Group + N.qualifying.teeth + BL.GE + BL..BOP + BL.PD.avg + BL.Calc.I + BL.Pl.I, data = df3, weights = ws, family = binomial)

firth_fit2.1 <- logistf(VeryPreterm ~ Group + Clinic + N.qualifying.teeth + BL.GE +  BL..BOP + BL.PD.avg + BL.Calc.I + BL.Pl.I, data = df3, weights = ws)

#summary(firth_fit2)
#summary(firth_fit2.1)

output_df_VPreterm <- data.frame(Coefficients = round(exp(firth_fit2.1$coefficients),3),
                                 lowerCI = round(exp(firth_fit2.1$ci.lower),3),
                                 upperCI = round(exp(firth_fit2.1$ci.upper),3),
```

```
                                  pvalue = round(firth_fit2.1$prob,3))

knitr::kable(output_df_VPreterm, caption = "Table 9: Firth logisitic
regression of very preterm including 95% confidence intervals and pvalue
computed using profile penalized log-likelihood")

tab = with(df3,table(Group,LatePreterm))
ptab = round(prop.table(tab),2)

ctab = matrix(paste0(tab,"(",100*ptab,"%)"),ncol = 2, nrow = 2)
colnames(ctab) <- c("NOT Preterm","Late Preterm")
rownames(ctab) <- c("Control","Treatment")

knitr::kable(ctab, caption = "Table 10: Contingency table for late preterm
birth after removing elective abortion and loss to follow-up")

#firth_fit3 <- glm(ExtremelyPreterm ~ Group + N.qualifying.teeth + BL.GE +
BL..BOP + BL.PD.avg + BL.Calc.I + BL.Pl.I, data = df3, weights = ws, family =
binomial)

firth_fit3.1 <- logistf(LatePreterm ~ Group + Clinic + N.qualifying.teeth +
BL.GE +  BL..BOP + BL.PD.avg + BL.Calc.I + BL.Pl.I, data = df3, weights = ws)

#summary(firth_fit3)
#summary(firth_fit3.1)

output_df_LPreterm <- data.frame(Coefficients =
round(exp(firth_fit3.1$coefficients),3),
                                 lowerCI =
round(exp(firth_fit3.1$ci.lower),3),
                                 upperCI =
round(exp(firth_fit3.1$ci.upper),3),
                                 pvalue = round(firth_fit3.1$prob,3))

knitr::kable(output_df_LPreterm, caption = "Table 11: Firth logisitic
regression of late preterm including 95% confidence intervals and pvalue
computed using profile penalized log-likelihood")

output_df_Preterm <- data.frame(ExtremelyPreterm =
round(exp(firth_fit1.1$coefficients),3),
                                VeryPreterm =
round(exp(firth_fit2.1$coefficients),3),
                                LatePreterm =
round(exp(firth_fit3.1$coefficients),3))

knitr::kable(output_df_Preterm, label = "Table 12: Coefficients of preterm
categories")
```