Interim Project Update

Key Research Question of Interest

- Aim 1: Evaluate the relationship between time to obesity and exposure to gestational diabetes.
  - Hypothesis: Children who have been exposed to gestational diabetes during pregnancy will have a higher hazard of becoming obese.
- Aim 2: Evaluate the relationship between time to becoming overweight and exposure to gestational diabetes.
  - Hypothesis: Children who have been exposed to gestational diabetes during pregnancy will have a higher hazard of becoming overweight

Description of the Dataset

The dataset that I will utilize is from the LEAD center. It is from the Exploring Perinatal Outcomes among Children (EPOCH) study, which is a retrospective cohort study. Participants in the study were enrolled from the Kaiser Permanente of Colorado Health Plan system and were aged 6-13 years of age upon enrollment. All participants were invited to a total of two research office visits and medical record abstraction was performed to obtain height and weight data from birth to the last visit (which occurred at the end of puberty). The first office visit occurred at the mean age of 10.4 (SD=1.5) years and the second at 16.7 (SD=1.2) years. At both visits demographic, anthropometric, and adiposity measures were collected.

Proposed Analysis Method Kaplan Meier curves will also be generated to visualize the survival curves of the exposed and unexposed populations as a descriptive. A cox proportional hazards model will be used to investigate both aims. The main predictor to be assessed will be exposure to gestational diabetes during pregnancy. A Wald test will be conducted to assess a difference in the hazards at an $\alpha = 0.05$ level. Sociodemographic features will be selected for inclusion into the model. I will also vestigate education level at birth and at the second EPOCH visit as a time varying covariate.

I will also investigate stratifying the cox proportional hazards model by gender. I will first see if this makes sense by examining the survival curves by gender. If there is a large difference between the two shapes, which could also indicate violation of the proportional hazards assumption, I will stratify by gender which will modify the baseline hazards for each group to make them more flexible.

Exploratory Plots

Here is a table 1 of the participants included for analysis:

```
# First, let's show what the data looks like
obesity %>%
  head()
```

```
## # A tibble: 6 x 8
##     PID   sex ageyearsnew   gdm  FH28  FH33 t_event status
##   <dbl> <dbl>       <dbl> <dbl> <dbl> <dbl>   <dbl>  <dbl>
## 1 20015     1       1.99      0    13     5   1.99        1
## 2 20018     2      11.2       0     8     4  11.2         1
## 3 20024     2       9.33      0    12     3   9.33        1
```

Table 1: Participant Characteristics

Characteristic	**No Gestational Diabetes** N = 505	**Exposure to Gestational Diabetes** N = 99
Sex		
Female	257 (51%)	46 (46%)
Male	248 (49%)	53 (54%)
Race		
Hispanic	202 (40%)	30 (30%)
Non-Hispanic Black	43 (8.5%)	5 (5.1%)
Non-Hispanic Other	31 (6.1%)	4 (4.0%)
Non-Hispanic White	229 (45%)	60 (61%)
Household Income		
$0-$24,999	49 (9.7%)	6 (6.1%)
$25,000-$49,999	189 (37%)	33 (33%)
$50,000-$74,999	156 (31%)	33 (33%)
$75,000+	110 (22%)	27 (27%)
Don't know or Refused	1 (0.2%)	0 (0%)
Mother's Education at Child's Birth		
Associate Degree	30 (6.0%)	4 (4.0%)
Bachelor's Degree	165 (33%)	31 (31%)
Business/Technical School	18 (3.6%)	4 (4.0%)
High School Diploma	93 (19%)	14 (14%)
Less than High School Diploma	13 (2.6%)	5 (5.1%)
Masters Degree	54 (11%)	9 (9.1%)
Professional or Doctorate Degree	11 (2.2%)	2 (2.0%)
Some College	117 (23%)	30 (30%)
Unknown	4	0

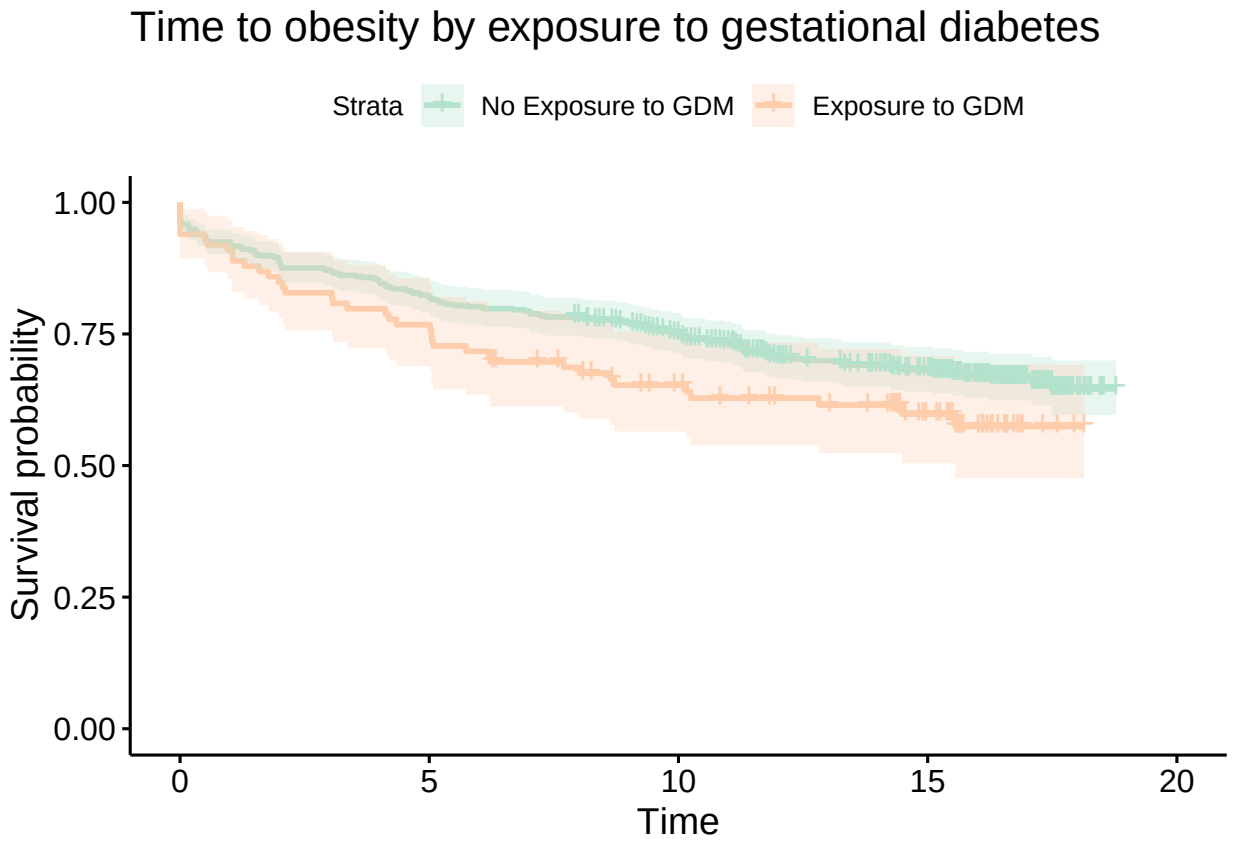
[1] n (%)

```
## 4 20029     1      11.3       0    10     4  11.3        1
## 5 20030     2       0.361     0    12     3   0.361      1
## 6 20033     2       2.11      1    12     4   2.11       1
```

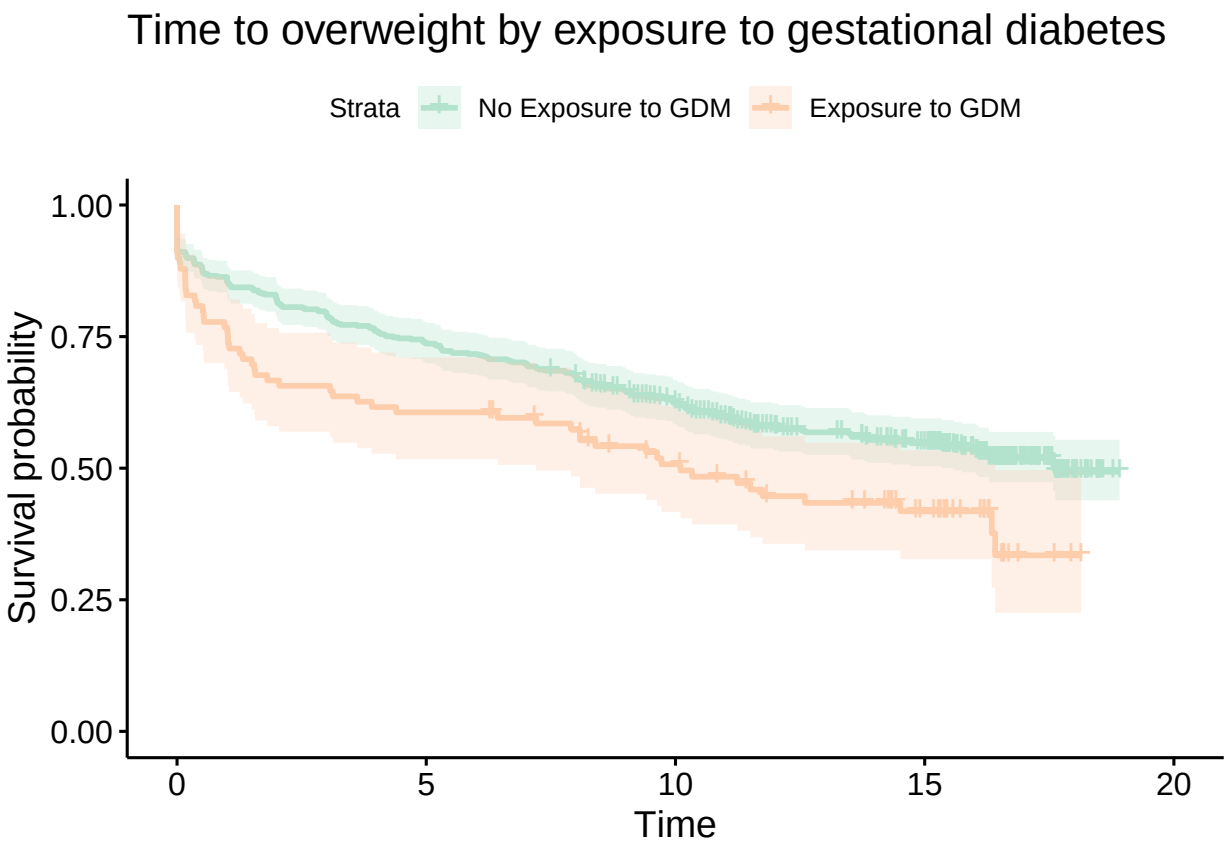
```
overweight %>%
  head()
```

```
## # A tibble: 6 x 8
##     PID   sex ageyearsnew   gdm  FH28  FH33 t_event status
##   <dbl> <dbl>       <dbl> <dbl> <dbl> <dbl>   <dbl>  <dbl>
## 1 20009     1       0         0    13     4   0          1
## 2 20013     1       0         0     8     3   0          1
## 3 20014     1       0.348     0    13     2   0.348      1
## 4 20025     1      11.0       0    10     3  11.0        1
## 5 20028     2       0         0    13     5   0          1
## 6 20030     2       0.498     0    12     3   0.498      1
```

```
##############################################
# Kaplan-Meier Curve for the Obesity Outcome
##############################################
obese_outcomeGDM <- survfit(Surv(t_event, status) ~ gdm,
                            data= obesity)
ggsurvplot(obese_outcomeGDM,
           palette = "Pastel2",
           conf.int = TRUE,
           title = "Time to obesity by exposure to gestational diabetes",
           legend.labs = c("No Exposure to GDM", "Exposure to GDM"))
```



```
##############################################
# Kaplan-Meier Curve for the Overweight Outcome
##############################################
overweight_outcomeGDM <- survfit(Surv(t_event, status) ~ gdm, data= overweight)
ggsurvplot(overweight_outcomeGDM,
           palette = "Pastel2",
           conf.int = TRUE,
           title = "Time to overweight by exposure to gestational diabetes",
           legend.labs = c("No Exposure to GDM", "Exposure to GDM"))
```




```
# Conduct a log-rank test to determine if the curves differ by the exposure
logrank_obesity <- survdiff(Surv(t_event, status) ~ gdm, data= obesity)
logrank_overweight <- survdiff(Surv(t_event, status) ~ gdm, data= overweight)
logrank_obesity
```

```
## Call:
## survdiff(formula = Surv(t_event, status) ~ gdm, data = obesity)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## gdm=0 505      158    167.2     0.503      3.35
## gdm=1  99       39     29.8     2.820      3.35
##
##  Chisq= 3.3  on 1 degrees of freedom, p= 0.07
```

```
logrank_overweight
```

```
## Call:
## survdiff(formula = Surv(t_event, status) ~ gdm, data = overweight)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## gdm=0 505      228    244.1      1.06      7.53
## gdm=1  99       57     40.9      6.32      7.53
##
##  Chisq= 7.5  on 1 degrees of freedom, p= 0.006
```

Above, I've presented the Kaplan-Meier curves for the two outcomes that I have outlined in the paper above. From examining these graphs, we can see that there is likely no evidence to reject that the proportionality of hazards is upheld when we simply consider the exposure. I have also conducted log-rank tests to see if the two survival curves differ by exposure to gestational diabetes. We've found that there is a significant difference in the overweight survival curves by exposure to gestational diabetes. Those who have been exposed to gestational diabetes have a lower median survival survival time (10.1 years of age) than those who were not exposed to gestational diabetes (17.6 years of age).

Supplementary

(Please don't dock points because I went over, I had to submit a proposal for the LEAD center and I ended up writing a lot, I want to use it here too!!!)

Outcome

The outcome is time to obesity for the first aim, and time to becoming overweight for the secondary hypothesis. I will create obesity and overweight covariates using the following methods:

- Both outcomes will be generated using the CDC growth charts. The methods are as follows:
  - Obesity will be classified as a BMI of >= 95th percentile
  - Overweight will be classified as a BMI of >= 85th percentile and < 95th percentile
  - Reference: https://www.cdc.gov/bmi/child-teen-calculator/bmi-categories.html