575 Final Project: Data Analysis

Lizzie, Suhaib, Emma

2021-11-04

Contents

1 Load packages

```
library(tidyverse)
library(haven)
library(ggplot2)
library(here)
library(jtools)
library(lme4)
library(lmerTest)
library(glmmTMB)
library(modelsummary)
library(knitr)
library(kableExtra)
theme_set(jtools::theme_apa())
```

2 Load Data

2.1 Import Data

```
# Read in the data
df.hatch <- read_sav(here("HATCH 09.29.21.sav"))
```

2.2 Tidy Data

```
# Subset Variables
df.mlm <- df.hatch %>%
  select(c(CoupID,                                    # Couple ID

           # Couple Level                             ## COUPLE ##
           DelMod, ModeofDeliverySpecific,            # Delivery Method
           GesAgeWk,                                  # Gestational age
           Bb.sex,                                    # Baby sex

           # Person Level                             ## PERSON ##
           contains("pnAge"),                         # Parent age
           contains("Ethn"),                          # Parent ethnicity
           contains("Educ"),                          # Parent level of education
           contains("peritot"),                       # BEQ (birth stress)

           # Time Level                               ## TIME ##
           bage3pp.1, bage6pp, bage12pp.1,            # Baby age
           contains("PSI_t"),                         # Parenting Stress
              contains("PSIt"),
              contains("PSI.t")
           ))

# Rename Variables
colnames(df.mlm) <- c("CoupID",

                      # Couple Level
                      "DelMeth", "DelMeth.Specific",
                      "GestationAge",
                      "BabySex",

                      # Person Level
                      "age.mom", "age.dad",
                      "ethnicity.mom", "ethnicity.dad",
                      "education.mom", "education.dad",
                      "beq.mom", "beq.dad",

                      # Time Level
                      "BabyAge.3", "BabyAge.6", "BabyAge.12",
                      "PSI.3.mom", "PSI.12.mom", "PSI.12.dad",
                          "PSI.3.dad", "PSI.6.dad",
                          "PSI.6.mom"
```

```
                      )

# Reorder PSI Variables
df.mlm <- df.mlm %>%
  relocate(c("PSI.3.mom", "PSI.3.dad",
             "PSI.6.mom", "PSI.6.dad",
             "PSI.12.mom", "PSI.12.dad"),
           .after = BabyAge.12)
```

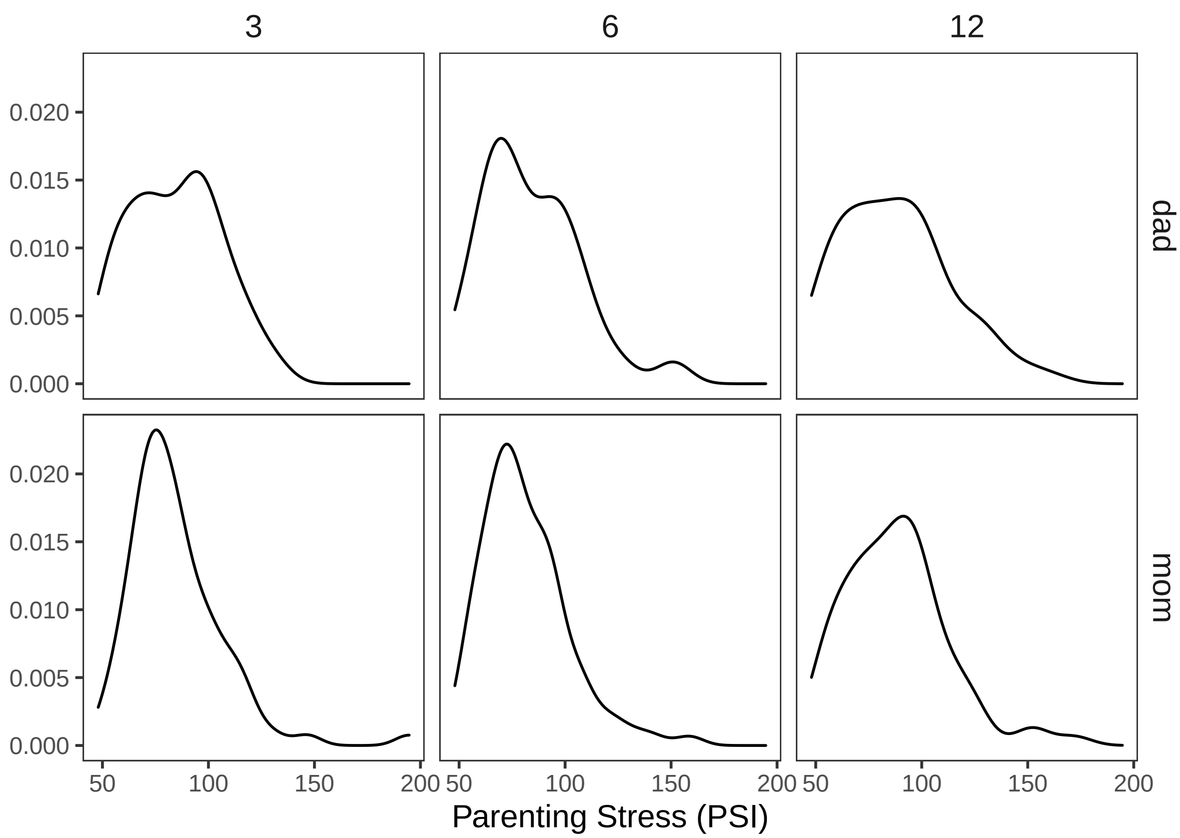
2.3 Wide to Long

```
# Person-Level and Time-Level Variables
df.long1 <- df.mlm %>%
  # Parent-level variables
  pivot_longer(
    cols = age.mom:beq.dad,
    names_to = c(".value", "parent"),
    names_pattern = "(age|ethnicity|education|beq).(mom|dad)",
    names_transform = list(parent = as.factor)
  ) %>%
  # Assign participant IDs
  mutate(PersonID = seq(1:200)) %>%
  # Time-level variables
  pivot_longer(
    cols = BabyAge.3:BabyAge.12,
    names_to = c(".value", "time"),
    names_pattern = "(BabyAge).(3|6|12)",
    names_transform = list(time = as.integer)
  ) %>%
  # Remove PSI variables
  select(-starts_with("PSI"))

# PSI Variable
df.long2 <- df.mlm %>%
  pivot_longer(
    cols = PSI.3.mom:PSI.12.dad,
    names_to = c(".value", "time", "parent"),
    names_pattern = "(PSI).(3|6|12).(mom|dad)",
    names_transform = list(time = as.integer, parent = as.factor)
  ) %>%
  # Remove variables in `df.long1` that aren't identifying variables
  select(c(CoupID, time, parent, PSI))


# Combine Data Frames
df.long <- left_join(df.long1, df.long2, by = c("CoupID", "parent", "time")) %>%
  # Move identifying variable
  relocate(c(PersonID, parent, time), .after = CoupID) %>%
  # Remove funky formatting
  mutate(beq = na_if(beq, -97.450))
```

3 Density Plots

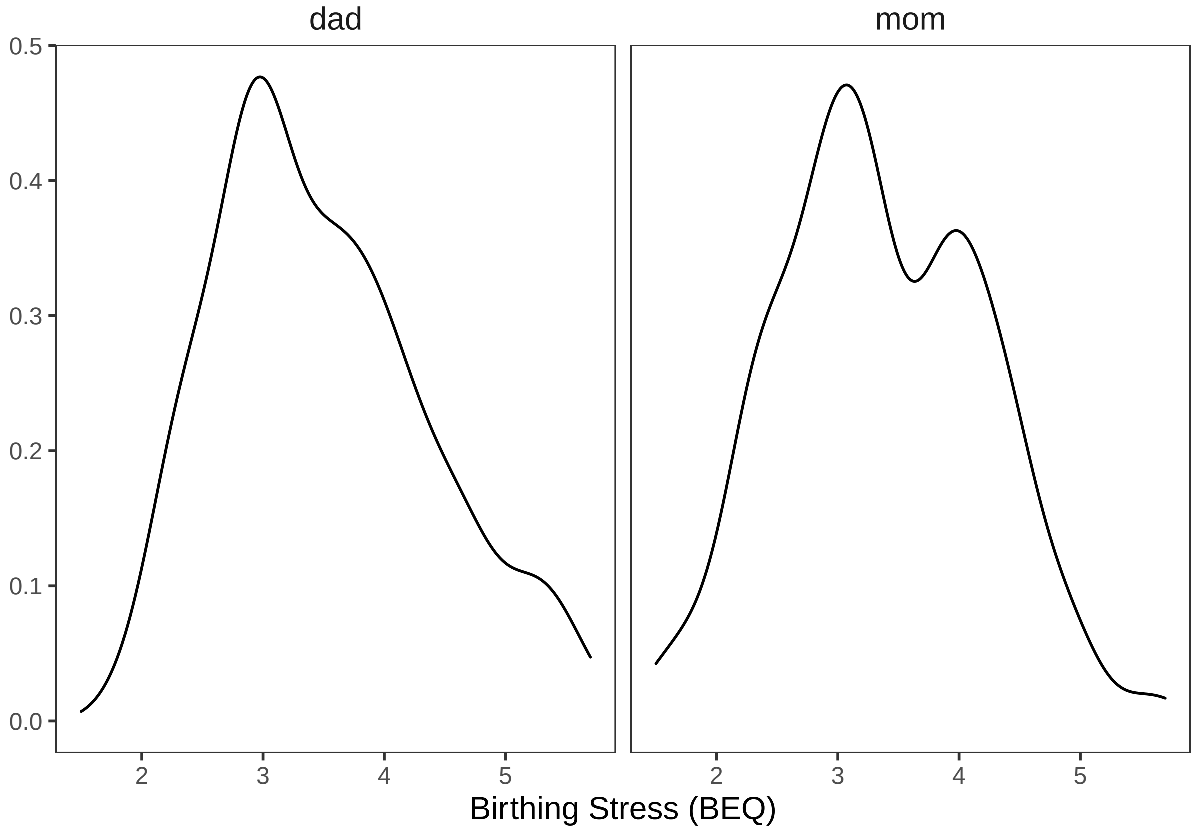
```
ggplot(data = df.long, aes(x = PSI)) +
  geom_density() +
  facet_grid(parent~time) +
  labs(x = "Parenting Stress (PSI)", y = "")
```

```
## Warning: Removed 98 rows containing non-finite values (stat_density).
```

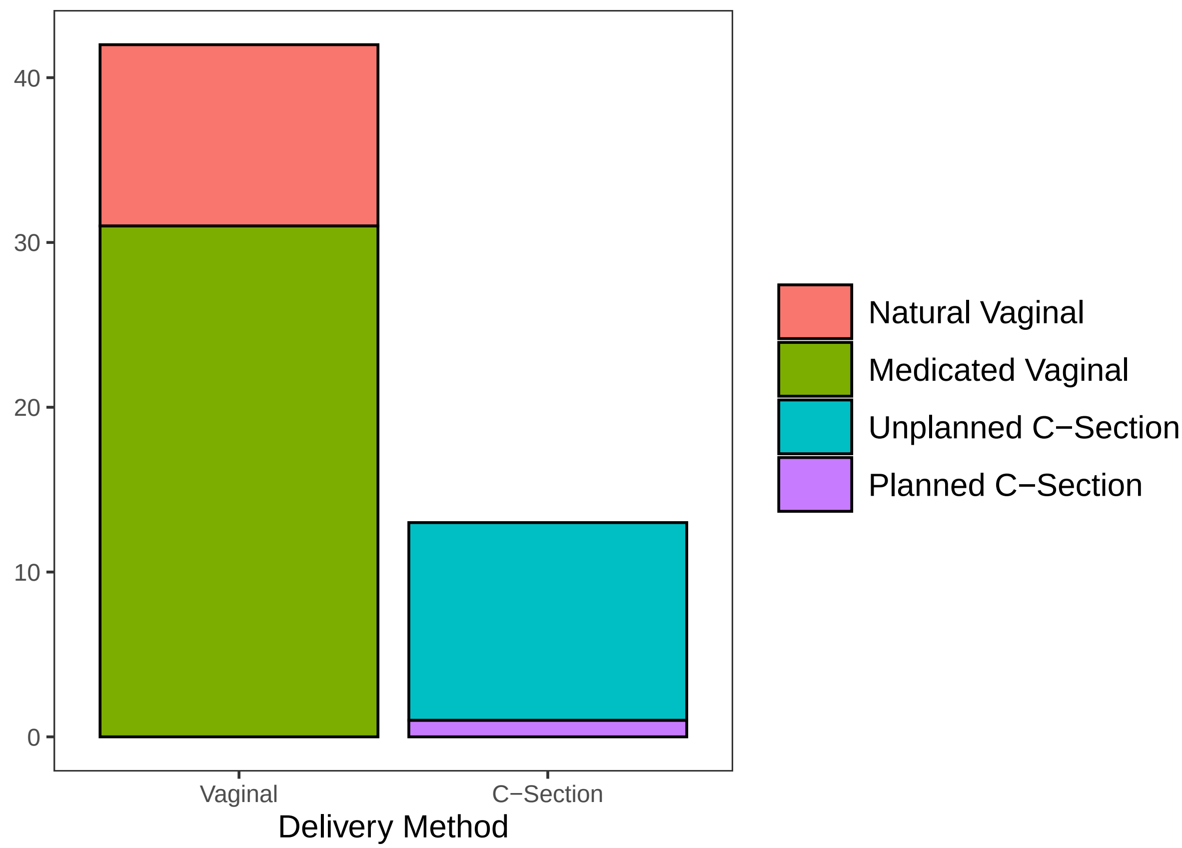



```
ggplot(data = df.long, aes(x = beq)) +
  geom_density() +
  facet_wrap(~parent) +
  labs(x = "Birthing Stress (BEQ)", y = "")
```

```
## Warning: Removed 45 rows containing non-finite values (stat_density).
```




```
ggplot(data = na.omit(distinct(df.long, CoupID, .keep_all = T)),
       aes(x = as.factor(DelMeth),
           fill = as.factor(DelMeth.Specific))) +
  geom_bar(position = "stack", color = "black") +
  scale_x_discrete(labels = c("Vaginal", "C-Section")) +
  scale_fill_discrete(labels = c("Natural Vaginal", "Medicated Vaginal", "Unplanned C-Section", "Planned
  labs(x = "Delivery Method", y = "")
```



3.1 Data Processing

```
#scaling and decomposing of variables
df_scaled <- df.long %>% mutate(log_psi = scale(log(PSI))) %>%
  group_by(CoupID) %>% mutate(beq_cm = mean(beq)) %>% ungroup() %>% mutate(beq_cmc = beq - beq_cm)
```

3.2 Models

3.2.1 Model Equation:

Nomenclature: j: Couples i: persons t: time intervals

$$psi_{tij} = \gamma_{000} + \gamma_{100} \cdot TIME_{tij} + \gamma_{010} \cdot beq_{ij}^{cmc} + u_{0i0} + \gamma_{001} \cdot beq_j^{cm} + \gamma_{002} \cdot DelMeth_j + \gamma_{003} \cdot DelMeth \cdot beq_j^{cm} + u_{00j} + e_{tij} \tag{1}$$

3.2.2 Modelling

```
m0 <- lmer(PSI ~ 1 + (time|CoupID), data=df_scaled)
ranova(m0) #not significant!
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## PSI ~ (time | CoupID)
##                         npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>                     5 -2189.3 4388.6
## time in (time | CoupID)    3 -2191.2 4388.5 3.8782  2     0.1438
```

```
#random slope insignificant for Couples and does not converge for Persons
```

The random slope on the couple level is not significant, so it will not be included in the model. Note, that the model did not converge for person-level random slopes, which is why they have been excluded for now, however, they are planned to be included in future versions.

```
m1 <- glmmTMB(PSI ~ time + beq_cmc  + DelMeth*beq_cm + (1|PersonID) + (1|CoupID), data = df_scaled)
summary(m1) #robust se estimation (faster than brms)
```

```
##  Family: gaussian  ( identity )
## Formula:
## PSI ~ time + beq_cmc + DelMeth * beq_cm + (1 | PersonID) + (1 |      CoupID)
## Data: df_scaled
##
##      AIC      BIC   logLik deviance df.resid
##   3300.1   3335.6  -1641.1   3282.1      372
##
## Random effects:
##
## Conditional model:
##  Groups   Name        Variance Std.Dev.
##  PersonID (Intercept) 118.4    10.88
##  CoupID   (Intercept) 169.7    13.03
##  Residual             182.3    13.50
## Number of obs: 381, groups:  PersonID, 147; CoupID, 74
##
## Dispersion estimate for gaussian family (sigma^2):  182
##
## Conditional model:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    74.35352   11.74448   6.331 2.44e-10 ***
## time            0.24808    0.19832   1.251    0.211
## beq_cmc        -0.04171    2.49317  -0.017    0.987
## DelMeth        28.49545   25.16737   1.132    0.258
## beq_cm          1.77115    3.53338   0.501    0.616
## DelMeth:beq_cm -4.25052    6.49776  -0.654    0.513
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# m2 <- glmmTMB(PSI ~ time + beq_cmc  + DelMeth*beq_cm + age + parent + (1|PersonID) + (time|CoupID), d
# summary(m2) #robust se estimation (faster than brms)
```

```
modelsummary(m1, group = group + term ~ model, title = "Model 1 Coefficient Summary") %>% kable_styling
```

Table 1: Model 1 Coefficient Summary

		Model 1
	(Intercept)	74.354
		(11.744)
	time	0.248
		(0.198)
	beq_cmc	−0.042
		(2.493)
	DelMeth	28.495
		(25.167)
	beq_cm	1.771
		(3.533)
	DelMeth × beq_cm	−4.251
		(6.498)
PersonID	SD (Intercept)	10.881
CoupID		13.026
Residual	SD (Observations)	13.500
	Num.Obs.	381
	R2 Marg.	0.063
	R2 Cond.	0.637
	AIC	3300.1
	BIC	3335.6
	ICC	0.6
	RMSE	11.42

Table 2.1: Table shows the intercept, level-1 fixed effects (time), level-2 fixed effects (beq_cmc), level-3 fixed effects (beq_cm, delivery method & interaction), random intercept (variance of intercept) and measures of fit (e.g., RMSE).

3.3 Plots

```
effect_plot(m1, pred = beq_cmc,interval = TRUE, plot.points = TRUE)
```

```
# ggplot(data = df_scaled, aes(x = beq, y=PSI, group=CoupID))+
#   geom_point()+
#   geom_smooth(method = "lm", se = F, aes(colour = CoupID))+
#   xlab("Birth experience")+ylab("Parental Stress")+
#   theme(legend.position = "none")
```

3.4 Interpretation

We investigated the model described in equation 1. The model output in table 1 shows the fixed effects as follows: For time (level 1) as 0.25 (SE=0.198), for birth experience (cmc, level 2) as -0.04 (SE=2.5), for birth experience (cm, level 3) as 1.8 (SE=3.5), for delivery method (level 3) as 28.5 (SE=25.2) and the interaction of delivery method and birth experience as -4.3 (SE=6.5). As such, none of the main effects were significant. The random intercept for the person-level was 10.9 and on the couple level 13. Figure @1 emphasizes this

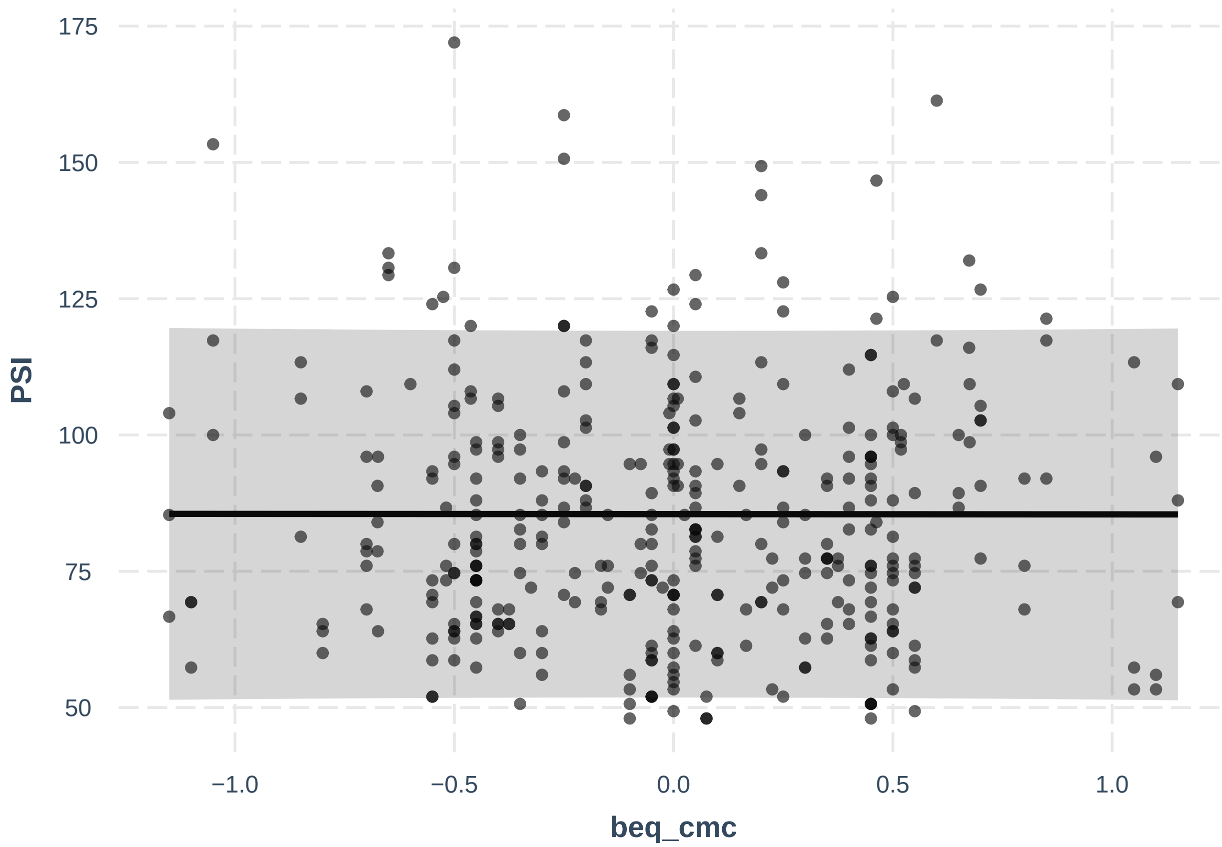


Figure 1: Scatter plot with regression line and 95% confidence interval of parental stress over birth experience (cluster mean centered). Plot shows no significant relation between both variables.

by showing the connection between birth experience (cmc) and parental stress. The plot shows that there is no significant relation.

These preliminary findings/analysis do not support our hypothesis that birth experience and delivery method relate to post partum parental stress. We will look further into our modelling to be sure that the findings are not caused by our model choices. It could be the case that previous research found relations between post partum depression and birth experience or delivery methods due to unsuitable models (e.g. OLS isntead of MLMs).