

Project One: Exploratory Data Analysis

Tara Amruthur

2023-10-08

Abstract

Smoking during pregnancy is associated with adverse birth outcomes, yet its prevalence persists. A study involving 800 pregnant individuals examined the effectiveness of a tailored video intervention to reduce smoking and environmental tobacco exposure. This analysis focuses on a subset of 100 mothers and children from the study, investigating demographics, substance use, ADHD scores, and self-regulation assessments. The dataset reveals missing data and a predominantly white sample.

Race-stratified smoking prevalence is explored, showing disparities from national trends. Self-reported smoking decreases postpartum, but the data's reliability is acknowledged. Urine cotinine values indicate a majority as passive or non-smokers. Substance use among children is higher for passive smokers. ADHD scores, particularly hyperactive, peak with trimester smoking.

Self-regulation, measured by the Brief Problem Monitor and Emotion Regulation Questionnaire, presents nuanced results. Internalizing problems are higher per parent reports, but child self-reports show trends with attention and externalizing problems, suggesting potential behavioral effects of smoking.

Postpartum exposure analysis highlights increased ERQ scores with more visits, implying an association with child behavior. The study emphasizes the need for further research on smoking's impact, especially postpartum.

Introduction

Pregnant individuals are advised to refrain from smoking throughout their pregnancy. This is due to the fact that smoking during pregnancy can be linked to several adverse birth outcomes, such as low birth weight, restricted head growth, placental problems, and other issues (NIH 2021).

Unfortunately, smoking during pregnancy is a fairly common phenomenon. As of 2014, it has been reported that 8.4% of women smoked at any time during pregnancy (NIH 2021). Additionally, 7-15% of U.S. infants born per year are exposed to smoking during pregnancy (SDP). Not only does this have a negative effect on the children and mothers themselves, but it has an economical cost as well. SDP imposes a \$4 billion annual burden on the U.S. economy due to health-care costs (Micalizzi 2023). As such, it is important to conduct research on this phenomenon.

A study published in January 2017 involved 800 pregnant people who were smoke exposed (either current smokers, smokers who quit on their own, or exposed to smoke of others), pregnant with only one baby, had access to a telephone and video player, and were randomized to either experimental or control conditions. This study aimed to compare the effectiveness of a tailored video intervention at reducing smoking and environmental tobacco exposure during and after pregnancy (Risica et al. 2017).

A subset of 100 mothers and children involved in this study make up the sample for the current study, conducted by Dr. Micalizzi. This study consists of three laboratory sessions: baseline, 6 month, and 12 month sessions. As part of this study, the parents and children complete self-regulatory assessments and provide self-reported information on substance use.

Using the information collected from these mothers at the baseline laboratory session, we want to see if we can identify a link between smoke exposure during and after pregnancy and adverse behavioral outcomes, such as issues with self-regulation, conduct disorder, substance use, and ADHD.

Data

The data provided consists of data collected from 49 unique parents and their children at the baseline laboratory session. Demographic and financial information has been collected from each mother, along with other information regarding their cigarette use both during and after pregnancy. The demographic information of the mother is provided in Table 1. Additionally, results from questionnaires such as the Brief Problem Monitor Questionnaire, Emotion Regulation Questionnaire, and Parental Knowledge Questionnaire were collected. The sum of responses on the SWAN rating scale for both types of ADHD (inattentive and hyperactive) are also present in the data.

For the child, demographic information as well as information related to their substance use is collected. Results from the Brief Problem Monitor Questionnaire, Emotion Regulation Questionnaire, and Parental Knowledge Questionnaire are in the data as well.

From Table 1, we are able to see that the data is overwhelmingly white, which could be indicative that our results might not be generalizable to the target population of mothers who smoked during pregnancy and their children. We also see that the majority of mothers in this dataset have at least attended some college and have either full or part-time employment.

Of all the variables in Table 1, with the exception of annual income, we see that there are 8 unknown variables. Upon further investigation, we are able to conclude that if an individual has a missing value in one of these columns, they have a missing value in the other seven columns. This could be due to any number of confounders, but they are not accounted for in our data, so it is fair to conclude that the data is missing at random.

Table 1: Data Demographics

Characteristic	N = 49 [†]
Age	37.0 (35.0, 39.0)
Unknown	8
Sex	
F	40 (98%)
M	1 (2.4%)
Unknown	8
Race	
White	25 (61%)
Native Hawaiian/Pacific Islander	6 (15%)
Other	6 (15%)
Biracial	3 (7.3%)
American Indian/Alaska Native	1 (2.4%)
Unknown	8
Hispanic/Latino	13 (32%)
Unknown	8
Employed	
Full-Time	22 (54%)
No	12 (29%)
Part-Time	7 (17%)
Unknown	8
Highest Level of Education	
Some College	15 (37%)
4-Year Degree	10 (24%)

GED	5 (12%)
2-Year Degree	3 (7.3%)
High School	3 (7.3%)
Some High School	3 (7.3%)
Post-Graduate Degree	2 (4.9%)
Unknown	8
Annual Income	46,848 (20,000, 70,000)
Unknown	12

¹Median (IQR); n (%)

Exploratory Data Analysis

Smoking Differences by Race

In the original study, data on the smoking habits of mothers was collected at three different time points during the pregnancy: 16 weeks, 22 weeks, and 32 weeks (the trimesters). Additionally, this information was collected at four different time points postpartum: the first two postpartum visits, 12 weeks postpartum, and 6 months postpartum.

In a study conducted by the CDC in 2016, they found that Non-Hispanic American Indian or Alaska Native women had the highest prevalence of smoking during pregnancy, indicating potential for race-based confounding (Drake, Driscoll, and Mathews 2018). As such, I wanted to look at the prevalence of smoking at each of the different time points, stratified by race, which can be seen in Figure 1. Due to the limited amount of data, race was grouped into three categories: white, non-white, and not provided. Our results don't seem to align with the CDC, as in this data, white women are the predominant smokers at each of the time points. However, as our sample size is very small and there are several individuals who did not report their race, it is hard to say if this is representative of the general population.

Figure 1 shows that the total number of self-reported smokers at each trimester hovers around 10, peaking during the second trimester. However, at the first postpartum visit, the numbers plummet, with only 3 individuals identifying as smokers during this time period. As time goes on, though, we see the numbers start to increase, with 16 total smokers by the 6th month postpartum visit. However, this is a self-reported measure of smoking, so it may be inaccurate; many mothers may not want to admit their smoking behaviors during or immediately following the birth of their child. Additionally, race is a self-reported measure in this data, where individuals check off their race. As they can check off multiple races, there may be a few cases of double-counting.

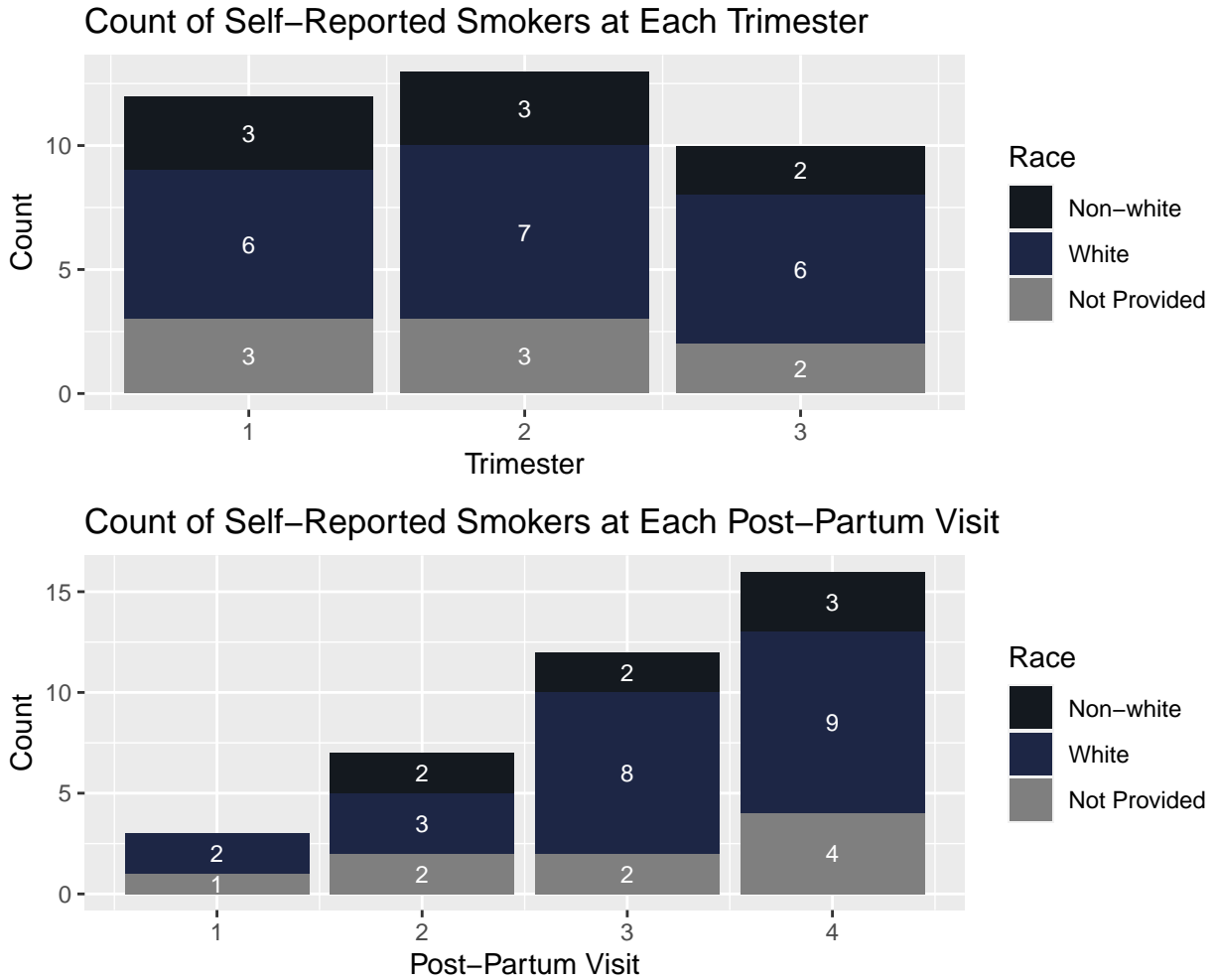


Figure 1. Count of Self-Reported Smokers During and After Pregnancy by Race

The original study only provides information on smoke exposure up to 6 months postpartum. The current study aims to expand upon this work by collecting information on smoke exposure up to five years post-pregnancy. Therefore, we wanted to look at differences in reported smoke exposure post-pregnancy, again stratified by race. The results are in Figure 2.

Similar to the results from the original study, at each time point, the smokers are predominantly white mothers. Additionally, the number of smokers across each of these time points seems to stay relatively constant, indicating continuous smoking behavior by these individuals.

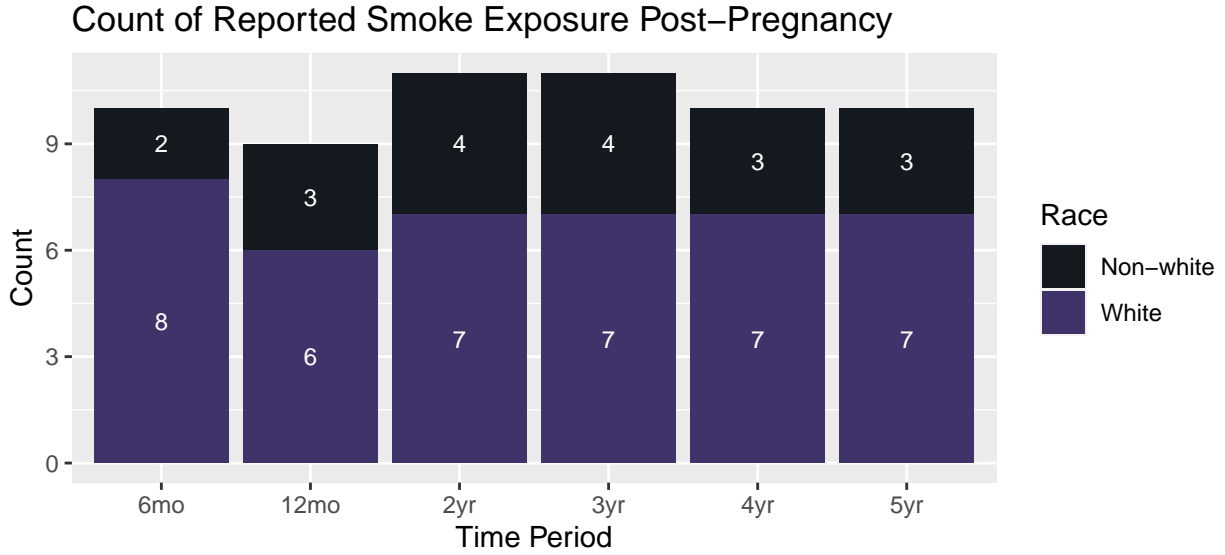


Figure 2. Count of Reported Smoke Exposure Post-Pregnancy by Race

Classification of Smokers

These self-reported measures provide information on whether or not an individual was smoking, but they do not tell us how heavy of a smoker these mothers are. There are other self-reported measures that provide this information for the past 30 days, but not during pregnancy. However, the urine cotinine values collected at 34 weeks into pregnancy and 6 months postpartum can be used for this purpose.

Cut-off values of urine cotinine for different levels of smokers have been subject to lots of research, but for this analysis, the following criteria will be used: nonsmokers will have levels less than 100 ng/mL, passive smokers would have levels between 100 to 500 ng/mL, and active smokers would have levels greater than 500 ng/mL (Hellemons et al. 2015). Table 2 shows the proportion of mothers in each of these categories at both time points. Surprisingly, there are no active smokers at the 34 week gestation period; in fact, the majority of mothers at each of these time points are either passive smokers or non-smokers. These numbers differ slightly with the values in Figure 1 and Figure 2, but as the cut-off numbers are only estimates, that could be the reason for this difference.

Missing data does appear to be a problem for this data, with 11 missing values at both time points. That is a substantial percentage of the data ($\approx 20\%$), indicating that the real proportion could be noticeably different than what is presented in Table 2.

Table 2: Smoking Level by Urine Cotinine Pre- and Post- Pregnancy

Characteristic	N = 49 ¹
Smoking Level (34 Week Gestation)	
nonsmoker	30 (79%)
passive	8 (21%)
Unknown	11
Smoking Level (6 Months Postpartum)	
active smoker	1 (2.6%)
nonsmoker	26 (68%)
passive	11 (29%)
Unknown	11

¹n (%)

Effects of Smoking on Child

Some consequences of SDP exposure include externalizing behavior such as attention-deficit/hyperactivity disorder (ADHD), conduct disorder, or substance use. We wanted to see if these patterns are prevalent in the data. To start, Table 3 was constructed to look at the percent of substance use among children based on their mothers' smoking status at the two different time periods.

Substance Use Table 3 shows that children of passive smokers are more likely to engage in substance use compared to children of non-smokers. Surprisingly, we see no substance use from the child of the active smoker. As there is only one active smoker in the dataset, this could just be an anomaly and not indicative of the real relationship between smoking status and child substance use, when the mother was an active smoker.

Table 3: Proportion of Child Substance Use by Smoking Level at Different Periods

Smoking Level	Time Period	Proportion of Child Substance Use
passive	34 weeks	0.25
passive	6 months postpartum	0.36
nonsmoker	34 weeks	0.13
nonsmoker	6 months postpartum	0.12
active smoker	6 months postpartum	0.00

Furthermore, we wanted to see which substances made up the majority of total substance use. Do children of smokers tend to prefer certain substances? These values can be seen in Table 4. Of children engaged in substance use, we see that alcohol is the most popular, with 71% of substance use representing alcohol. Surprisingly, the substance used least is cigarettes, with only 14% of substance use attributed to cigarettes.

As these are self-reported measures, there is no guarantee that it is completely accurate. There is also a possibility of double-counting as children could check off multiple mediums of substance use.

Table 4: Substance Use by Medium

Proportion of Substance Use	
E-Cigarette	0.43
Marijuana	0.43
Alcohol	0.71
Cigarette	0.14

ADHD As mentioned previously, there is a link between smoking during pregnancy and ADHD. In order to see if this trend persists in our data, a variable for the number of trimesters a mother smoked was created. All mothers either did not smoke during the pregnancy, smoked during the first trimesters, or smoked for all three trimesters.

A child's potential for being ADHD can be determined using the SWAN scores. There are two different SWAN scores provided: one for ADHD-Inattentive and one for ADHD-Hyperactive. It is worth noting that the cut-off for SWAN scores is 6, with a value above 6 indicating that the child is likely ADHD.

In Figure 3, the distribution of SWAN scores was plotted by the number of trimesters a mother smoked. The cut-off value of 6 is denoted by a dotted black line. For the SWAN Inattentive scores, we see that all three distributions peak after the cut-off. However, the scores peak furthest to the right for children whose mothers smoked during all three trimesters of their pregnancy.

We can also see that for the SWAN Hyperactive scores, the distribution of SWAN scores for mothers who did not smoke during pregnancy peaks before the cut-off, while the distributions for mothers who smoked

during pregnancy peak after the cut-off. The difference between the distributions of mothers who smoked during all three semester and only one semester appears to be less pronounced with the Hyperactive scores as opposed to the Inattentive scores. However a clearer difference between the mothers who did not smoke during pregnancy and those who did can be seen with Hyperactive scores as opposed to Inattentive scores.

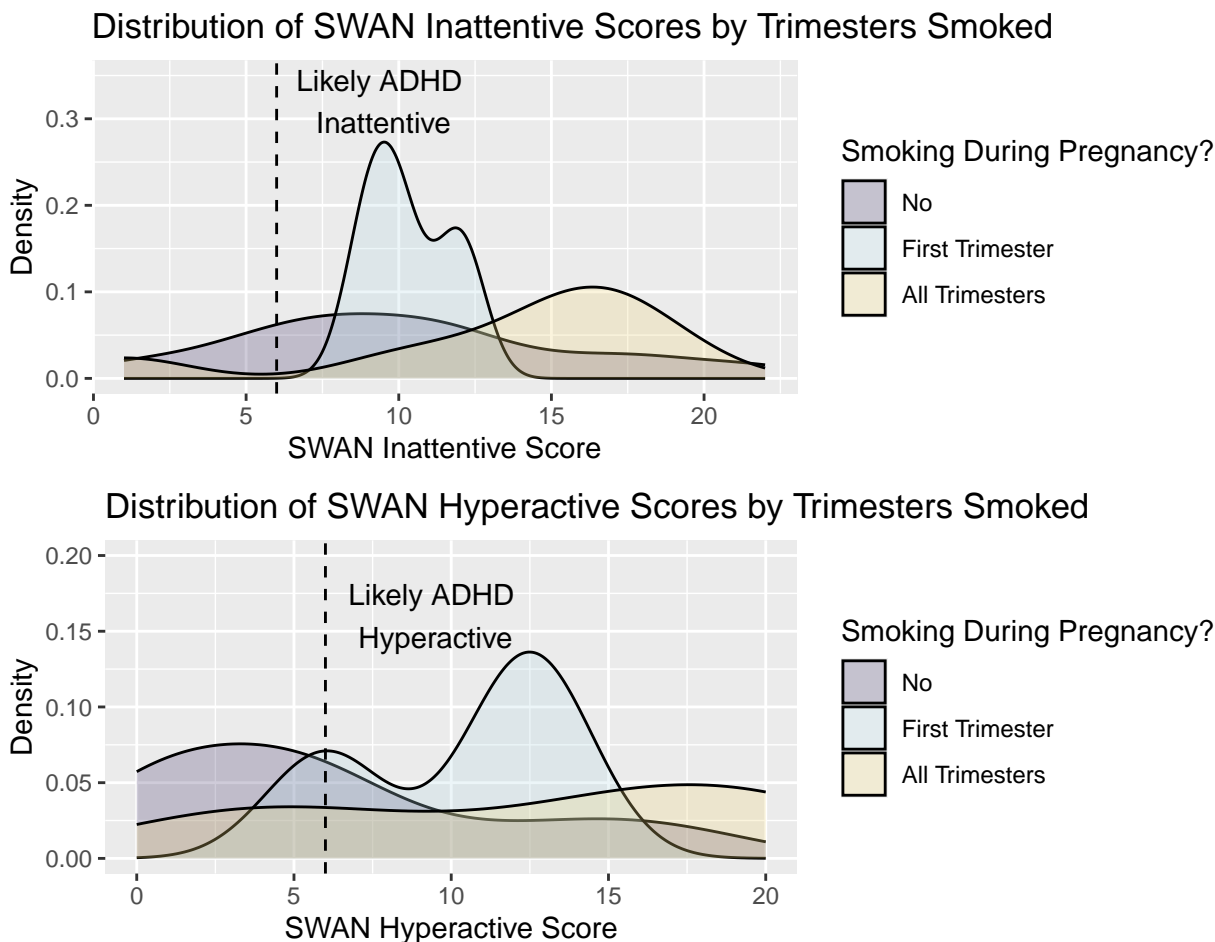


Figure 3. SWAN Scores by SDP Status

Self-Regulatory Problems Problems with self-regulation as a result of smoking during pregnancy are also of interest. Contrary to ADHD, conduct disorder, and substance use, self-regulation is an internalizing behavior. This was measured in two different ways: the Brief Problem Monitor (BPM) and Emotion Regulation Questionnaire (ERQ).

The Brief Problem Monitor is used for normed multi-informant assessment of children’s functioning & responses to interventions. In this study, it has been completed both by parents and children. There are separate scales for attention problems, internalizing problems, and externalizing problems (ASEBA 2022).

Figure 4 shows the average BPM scores across all three different scales by the number of trimesters the mother smoked. The plot on the top uses the scores from the child’s assessment of themselves and the plot on the bottom uses the scores from the mother’s assessment of their child.

For mothers who never smoked and those who smoked all three semesters, the difference in BPM scores from the child and the mother are not very pronounced. The largest differences can be seen among mothers who smoked for one trimester; specifically for internalizing problems. In fact, regardless of the mother’s smoking behavior, the score for internalizing problems from the parent was significantly higher than the score from the child. We also see something similar with attention problems, though the differences are not as large.

In the plot of scores collected from the child, we see that BPM scores related to attention problems and externalizing problems on self appear to increase somewhat as the number of trimesters the mother smoked increases. This trend does not appear in the plot of scores collected from the mother, but we do see that the average scores from all three sections are higher for children whose mother smoked during all three trimesters compared to no trimesters. This could be indicative of an effect on smoking during pregnancy on the child's behavior problems.

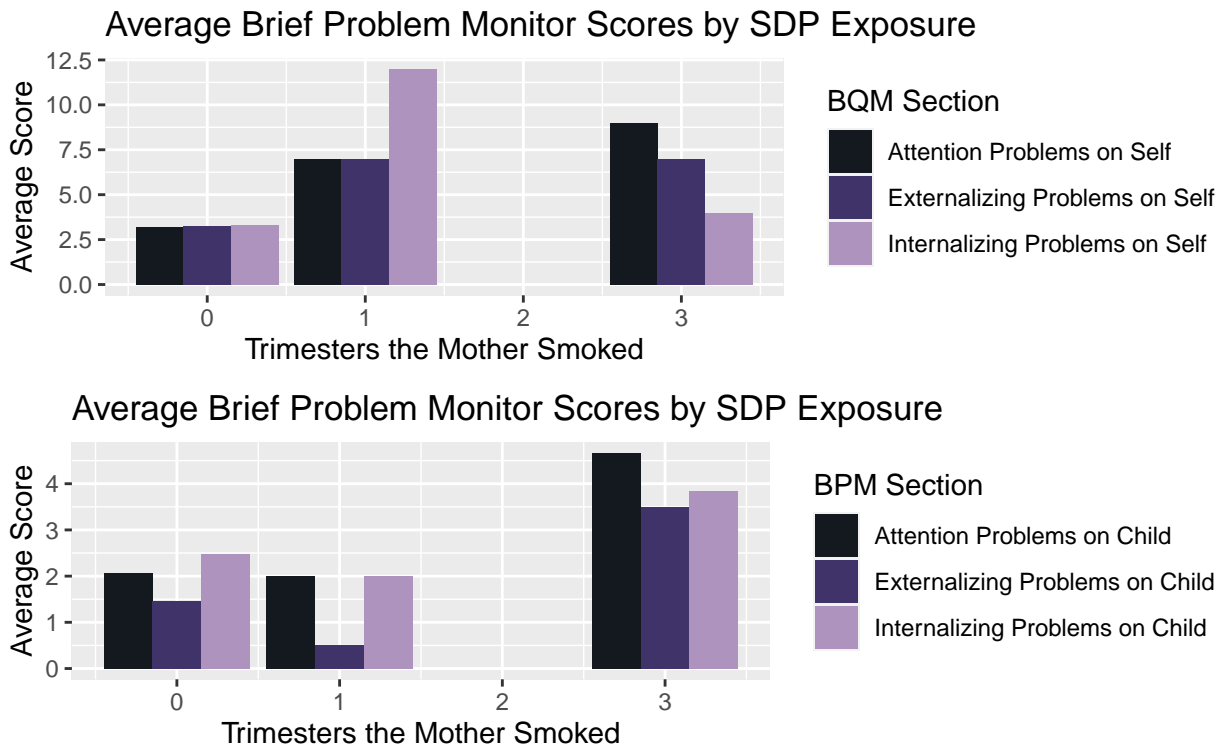


Figure 4. Average BPM Scores by SDP Reported by Child (Top) and Parent (Bottom)

We have also been provided scores from the Emotion Regulation Questionnaire. The ERQ consists of 10 questions that involve two distinct aspects of an individual's emotional life: their emotional experience (i.e. what they feel inside) and their emotional expression (i.e. how they show their emotions in the way they talk, gesture, or behave). Each of the 10 questions are answered using a scale ranging from 1-7, with 1 indicating they strongly disagree with a statement and 7 indicating they strongly agree. The higher the score, the greater the use of cognitive reappraisal or emotion suppression (Gross and John 2003). Similarly, this has been completed by both the parent and the child.

Figure 5 shows the Average ERQ scores by the number of trimesters the mother smoked. The plot on the top uses ERQ scores from the child, while the plot on the bottom uses ERQ scores from the mother. Unlike the BPM scores, we see that there does not appear to be a significant difference between the scores collected from the child and the scores collected from the mother.

There also does not appear to be an effect of smoking during pregnancy on ERQ scores. As the number of trimesters the mother smoked during increases, we don't see a clear trend. However, as there is no information on mothers who smoked during just two semesters, this could just be due to the lack of data available.

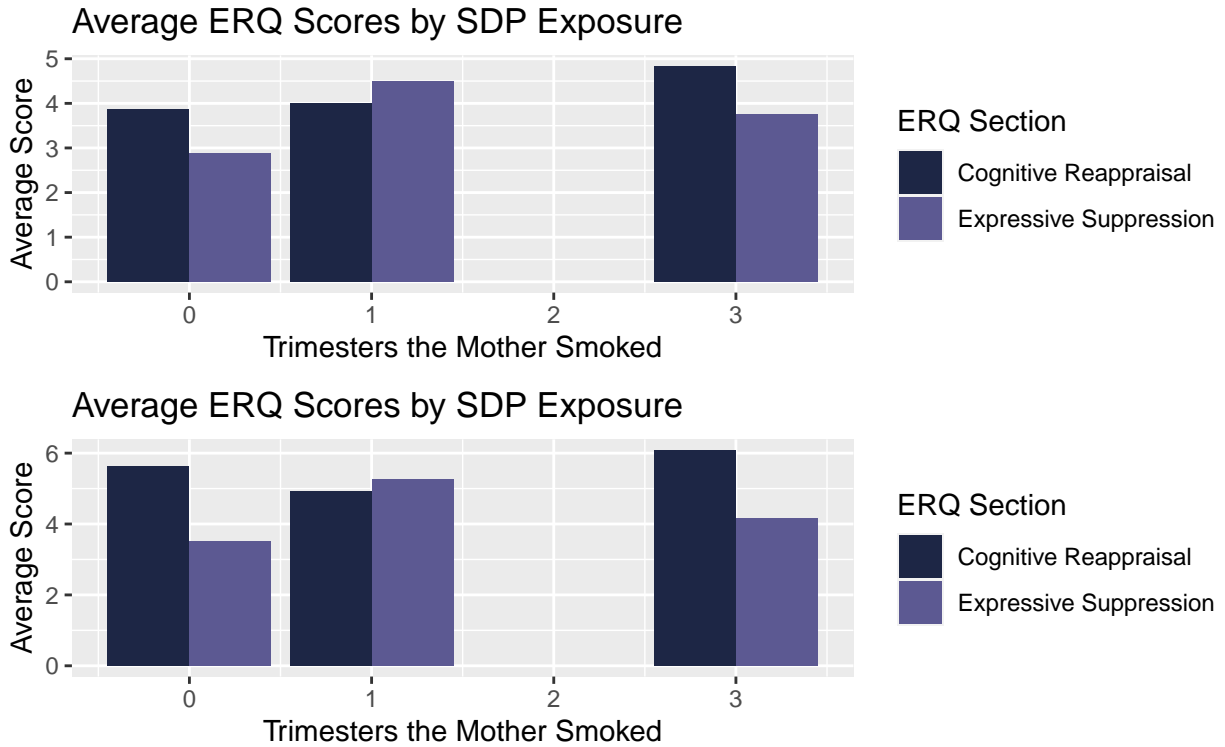


Figure 5. Average ERQ Scores by SDP Reported by Child (Top) and Parent (Bottom)

Postpartum Smoking The effect of smoking during pregnancy has been investigated, but what about the effect of smoking after pregnancy? Self-reported smoke exposure, by either the mother or partner, was collected at six different time points after the pregnancy, ranging from 6 months postpartum to 5 years postpartum. The values from each of these time points were combined to quantify at exactly how many time points postpartum the child was exposed to smoke. Based on these values, we wanted to look at how this impacted the child’s behavior.

Table 5 shows the average score from the BPM, ERQ, and SWAN Inattentive/Hyperactive assessments by the number of postpartum visits with self-reported smoke exposure. The clearest pattern is seen with the average ERQ score, where we see a clear increase in the score as the number of postpartum visits increases. We see a similar pattern for the SWAN Inattentive score, but for BPM and SWAN Hyperactive, we see that the values oscillate between increases and decreasing. This could be due to the data, as the majority of individuals reported no smoke exposure at any of the postpartum visits, so the average values are heavily influenced by one or two data points that may not be representative of the target population. Still, based on the data we have, there does appear to be a link between postpartum smoke exposure and a child’s behavior.

Table 5: Effects of Postpartum Smoking on Internalizing and Externalizing Behaviors

Count of Postpartum Smoke Exposure	Average BPM Score	Average ERQ Score	Average SWAN Inattentive Score	Average SWAN Hyperactive Score
0	7.11	5.80	9.83	6.71
1	13.50	5.92	12.50	8.50
3	14.00	5.75	15.00	20.00
4	10.67	6.36	16.50	15.75
5	8.50	6.54	10.50	8.50
6	11.60	7.19	9.20	4.20

Discussion

This analysis has provided several points worth further research in analyzing the effects of pre- and post-natal smoke exposure to children. Not only that, but we have investigated the rates of smoking among the women in the study both during and after pregnancy. We have analyzed race-based differences, as it could be a potential confounder, while also determining how heavy of a smoker these individuals are.

The analysis confirmed the initial belief that smoking during pregnancy can be associated with different externalizing behaviors, such as ADHD. We saw that smoking during pregnancy can lead to higher scores on the SWAN scores, indicating either Inattentive or Hyperactive ADHD. However, these differences seem more pronounced with Hyperactive ADHD.

Additionally, there appear to be more problems with attention and externalizing behaviors among children exposed to SDP compared to those not exposed, though we do not see a similar pattern when it comes to internalizing behaviors. This could be an area for further research, as the sample size in this study is too small to make any generalizations.

Conclusion

This analysis unveils complex relationships between maternal smoking, child behavior, and self-regulation. The study raises awareness of the persistent issue of smoking during pregnancy, its consequences, and the need for targeted interventions. Despite the limitations of a small sample size and missing data, the findings underscore the importance of continuous research in understanding the long-term effects of prenatal and postnatal smoke exposure on children.

Code Appendix

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE,
  message = FALSE)
suppressPackageStartupMessages(library(tidyverse, warn.conflicts = FALSE))
library(ggplot2)
library(gtsummary)
suppressPackageStartupMessages(library(kableExtra))
library(ghibli)
suppressPackageStartupMessages(library(gridExtra))

data <- read.csv("/Users/taram/Documents/Grad School/Year 2/Practical Data Analysis/Projects/project_1.

# Data Pre-Processing (1) Turn race columns from
# dummy-coded into single race column. Note: in
# cases where there are two races checked, store
# them as biracial. (2) Convert sex, ethnicity,
# employment, and education columns from numbers
# into the actual values (taken from the data
# codebook). (3) For income value 250,000 remove
# the extra space and column. Convert income
# column to numeric. (4) Convert smoking column
# values to 1 or 0 for smoking or no smoking.
# (5) Create new column for smoking level at 34
# weeks gestation period and six months
# postpartum. The smoking level is calculated
# based on the urine cotinine values. The cut-off
# values for each level have been taken from The
# Transplantation Journal. See citations for more
# details.
data <- data %>%
  mutate(parent_race = case_when(paian == 1 & pwhite ==
    1 | paian == 1 & pnhipi == 1 ~ "Biracial", paian ==
    1 ~ "American Indian/Alaska Native", pasian ==
    1 ~ "Asian", pnhipi == 1 ~ "Native Hawaiian/Pacific Islander",
    pblack == 1 ~ "Black", pwhite == 1 ~ "White",
    prace_other == 1 ~ "Other")) %>%
  mutate(parent_sex = case_when(psex == 1 ~ "F",
    psex == 0 ~ "M", TRUE ~ NA)) %>%
  mutate(ethnic = case_when(pethnic == 1 ~ "Yes",
    pethnic == 0 ~ "No", TRUE ~ NA)) %>%
  mutate(employed = case_when(employ == 0 ~ "No",
    employ == 1 ~ "Part-Time", employ == 2 ~ "Full-Time",
    TRUE ~ NA)) %>%
  mutate(education = case_when(pedu == 0 ~ "Some High School",
    pedu == 1 ~ "High School", pedu == 2 ~ "GED",
    pedu == 3 ~ "Some College", pedu == 4 ~ "2-Year Degree",
    pedu == 5 ~ "4-Year Degree", pedu == 6 ~ "Post-Graduate Degree")) %>%
  mutate(income = case_when(income == "250, 000" ~
    "250000", TRUE ~ income)) %>%
  mutate(income = as.numeric(income)) %>%
  mutate(mom_smoke_16wk = case_when(mom_smoke_16wk ==
    "1=Yes" ~ 1, mom_smoke_16wk == "2=No" ~ 0),
    mom_smoke_22wk = case_when(mom_smoke_22wk ==
```

```

    "1=Yes" ~ 1, mom_smoke_22wk == "2=No" ~
    0), mom_smoke_32wk = case_when(mom_smoke_32wk ==
    "1=Yes" ~ 1, mom_smoke_32wk == "2=No" ~
    0), mom_smoke_pp1 = case_when(mom_smoke_pp1 ==
    "1=Yes" ~ 1, mom_smoke_pp1 == "2=No" ~
    0), mom_smoke_pp2 = case_when(mom_smoke_pp2 ==
    "1=Yes" ~ 1, mom_smoke_pp2 == "2=No" ~
    0), mom_smoke_pp12wk = case_when(mom_smoke_pp12wk ==
    "1=Yes" ~ 1, mom_smoke_pp12wk == "2=No" ~
    0), mom_smoke_pp6mo = case_when(mom_smoke_pp6mo ==
    "1=Yes" ~ 1, mom_smoke_pp6mo == "2=No" ~
    0)) %>%
mutate(smoking_level_34wk = case_when(is.na(cotimean_34wk) ~
    NA, cotimean_34wk < 100 ~ "nonsmoker", cotimean_34wk <
    500 ~ "passive", TRUE ~ "active")) %>%
mutate(smoking_level_pp6mo = case_when(is.na(cotimean_pp6mo) ~
    NA, cotimean_pp6mo < 100 ~ "nonsmoker", cotimean_pp6mo <
    500 ~ "passive", TRUE ~ "active smoker"))

# Create Demographic Table
data %>%
  mutate(page = as.numeric(page)) %>%
  dplyr::select(page, parent_sex, parent_race, ethnic,
    employed, education, income) %>%
  tbl_summary(type = c(income ~ "continuous", page ~
    "continuous"), label = list(page ~ "Age", parent_sex ~
    "Sex", parent_race ~ "Race", ethnic ~ "Hispanic/Latino",
    employed ~ "Employed", education ~ "Highest Level of Education",
    income ~ "Annual Income"), sort = list(everything() ~
    "frequency")) %>%
  as_gt() %>%
  gt::tab_header(title = "Table 1: Data Demographics")

# Creating a race lookup table for the parent
race_lookup <- data %>%
  gather(key = "race", value = "value", c(paian,
    pasian, pblack, pwhite, pnhipi, prace_other)) %>%
  filter(!is.na(value)) %>%
  filter(value == 1) %>%
  dplyr::select(parent_id, race) %>%
  mutate(race = case_when(race == "pblack" ~ "black",
    race == "pwhite" ~ "white", !(race %in% c("pblack",
    "pwhite")) ~ "other", is.na(race) ~ "not provided"))

# Pivot the data to get whether or not the mother
# smoked during each trimester. 16 weeks
# represents trimester 1, 22 weeks represents
# trimester 2, and 32 weeks represents trimester
# 3. This is used to create the trimester column.
smoking_trimester <- data %>%
  gather(key = "time", value = "smoking", c(mom_smoke_16wk,
    mom_smoke_22wk, mom_smoke_32wk)) %>%
  mutate(trimester = case_when(time == "mom_smoke_16wk" ~
    1, time == "mom_smoke_22wk" ~ 2, time == "mom_smoke_32wk" ~

```

```

3))

# Get the number of smokers at each trimester
# stratified by race.
smoking_by_race <- smoking_trimester %>%
  dplyr::select(parent_id, trimester, smoking) %>%
  left_join(race_lookup, by = "parent_id", relationship = "many-to-many") %>%
  group_by(trimester, race) %>%
  summarize(count = sum(smoking, na.rm = TRUE), .groups = "drop_last") %>%
  filter(count > 0)

# Pivot the data to get whether or not the mother
# smoked during each postpartum visit.
post_partum_smoking <- data %>%
  gather(key = "time", value = "smoking", c(mom_smoke_pp1,
    mom_smoke_pp2, mom_smoke_pp12wk, mom_smoke_pp6mo)) %>%
  mutate(pp_visit = case_when(time == "mom_smoke_pp1" ~
    1, time == "mom_smoke_pp2" ~ 2, time == "mom_smoke_pp12wk" ~
    3, time == "mom_smoke_pp6mo" ~ 4))

# Get the number of smokers at each postpartum
# visit stratified by race.
pp_smoking_by_race <- post_partum_smoking %>%
  dplyr::select(parent_id, pp_visit, smoking) %>%
  left_join(race_lookup, by = "parent_id", relationship = "many-to-many") %>%
  group_by(pp_visit, race) %>%
  summarize(count = sum(smoking, na.rm = TRUE), .groups = "drop_last") %>%
  filter(count > 0)

# Make two plots: one for smoking during
# pregnancy & one for postpartum smoking.
smoking_plot_1 <- ggplot(smoking_by_race, aes(x = trimester,
  y = count, fill = race, label = count)) + geom_bar(stat = "identity") +
  geom_text(size = 3, position = position_stack(vjust = 0.5),
    color = "white") + scale_fill_manual(name = "Race",
  labels = c("Non-white", "White", "Not Provided"),
  values = ghibli_palette("LaputaMedium")) + labs(title = "Count of Self-Reported Smokers at Each Trimester",
  x = "Trimester", y = "Count")

smoking_plot_2 <- ggplot(pp_smoking_by_race, aes(x = pp_visit,
  y = count, fill = race, label = count)) + geom_bar(stat = "identity") +
  geom_text(size = 3, position = position_stack(vjust = 0.5),
    color = "white") + scale_fill_manual(name = "Race",
  labels = c("Non-white", "White", "Not Provided"),
  values = ghibli_palette("LaputaMedium")) + labs(title = "Count of Self-Reported Smokers at Each Postpartum Visit",
  x = "Post-Partum Visit", y = "Count")
grid.arrange(smoking_plot_1, smoking_plot_2, nrow = 2)

# Pivot the data to get values for smoke exposure
# at each time point postpartum. This data comes
# from the newer study and includes time periods
# at 6 months, 12 months, 2 years, 3 years, 4
# years, and 5 years.
smoke_exp <- data %>%

```

```

gather(key = "time", value = "smoke_exposure",
        c(smoke_exposure_6mo, smoke_exposure_12mo,
          smoke_exposure_2yr, smoke_exposure_3yr,
          smoke_exposure_4yr, smoke_exposure_5yr)) %>%
dplyr::select(parent_id, time, smoke_exposure) %>%
mutate(time = substring(time, 16))

# Get the number of self-reported smoke exposures
# at each time point stratified by race.
smoke_exp_over_time <- smoke_exp %>%
  left_join(race_lookup, by = "parent_id", relationship = "many-to-many") %>%
  filter(!is.na(smoke_exposure)) %>%
  group_by(race, time) %>%
  summarize(count = sum(smoke_exposure, na.rm = TRUE),
            .groups = "drop_last") %>%
  filter(count > 0)

# Order the x-axis
level_order <- c("6mo", "12mo", "2yr", "3yr", "4yr",
                 "5yr")

# Plot the smoke exposure postpartum over time by
# race.
smoke_exp_plot <- ggplot(smoke_exp_over_time, aes(x = factor(time,
  level = level_order), y = count, fill = race, label = count)) +
  geom_bar(stat = "identity") + geom_text(size = 3,
    position = position_stack(vjust = 0.5), color = "white") +
  scale_fill_manual(name = "Race", labels = c("Non-white",
    "White"), values = ghibli_palette("LaputaMedium")[c(1,
    3)]) + labs(title = "Count of Reported Smoke Exposure Post-Pregnancy",
    x = "Time Period", y = "Count")
grid.arrange(smoke_exp_plot, nrow = 1)

# Show the number of individuals that fall into
# each of the smoking levels at both time periods
# that urine cotinine levels are provided for.
data %>%
  dplyr::select(smoking_level_34wk, smoking_level_pp6mo) %>%
  tbl_summary(label = list("smoking_level_34wk" ~
    "Smoking Level (34 Week Gestation)", "smoking_level_pp6mo" ~
    "Smoking Level (6 Months Postpartum)")) %>%
  as_gt() %>%
  gt::tab_header(title = "Table 2: Smoking Level by Urine Cotinine Pre- and Post- Pregnancy")

# Get the percentage of child substance use each
# substance contributes to based on the smoking
# level of the parent at 34 weeks gestation.
substance_use_34wk <- data %>%
  mutate(smoking_level = smoking_level_34wk) %>%
  group_by(smoking_level) %>%
  filter(!is.na(smoking_level)) %>%
  mutate(substance_use = ifelse(e_cig_ever == 1 |
    cig_ever == 1 | alc_ever == 1 | mj_ever ==
    1, 1, 0)) %>%
  summarize(sub_use_prop = round(sum(substance_use,
    na.rm = TRUE)/n(), 2)) %>%

```

```

mutate(time = "34 weeks") %>%
  dplyr::select(smoking_level, time, sub_use_prop)

# Get the percentage of child substance use each
# substance contributes to based on the smoking
# level of the parent at 6 months postpartum.
# Join it with the results from 34 weeks
# gestation.
substance_use_total <- data %>%
  mutate(smoking_level = smoking_level_pp6mo) %>%
  group_by(smoking_level) %>%
  filter(!is.na(smoking_level)) %>%
  mutate(substance_use = ifelse(e_cig_ever == 1 |
    cig_ever == 1 | alc_ever == 1 | mj_ever ==
    1, 1, 0)) %>%
  summarize(sub_use_prop = round(sum(substance_use,
    na.rm = TRUE)/n(), 2)) %>%
  mutate(time = "6 months postpartum") %>%
  dplyr::select(smoking_level, time, sub_use_prop) %>%
  rbind(substance_use_34wk) %>%
  arrange(desc(smoking_level), time)

colnames(substance_use_total) <- c("Smoking Level",
  "Time Period", "Proportion of Child Substance Use")

kable(substance_use_total, booktabs = T, escape = F,
  caption = "Proportion of Child Substance Use by Smoking Level at Different Periods") %>%
  kable_styling(latex_options = "HOLD_position")
# Get the rate of total substance use by medium.
# Filter to include only individuals who have
# self-reported substance use and get the rate of
# use among each medium.
substance_use <- data %>%
  filter(e_cig_ever == 1 | cig_ever == 1 | alc_ever ==
    1 | mj_ever == 1) %>%
  dplyr::select(e_cig_ever, mj_ever, alc_ever, cig_ever) %>%
  summarize(e_cig = round(sum(e_cig_ever, na.rm = TRUE)/n(),
    2), mj = round(sum(mj_ever, na.rm = TRUE)/n(),
    2), alc = round(sum(alc_ever, na.rm = TRUE)/n(),
    2), cig = round(sum(cig_ever, na.rm = TRUE)/n(),
    2)) %>%
  t()

colnames(substance_use) <- c("Proportion of Substance Use")
rownames(substance_use) <- c("E-Cigarette", "Marijuana",
  "Alcohol", "Cigarette")

kbl(substance_use, booktabs = T, escape = F, caption = "Substance Use by Medium") %>%
  kable_styling(latex_options = "HOLD_position")
# The SWAN scores for these individuals should be
# NA. In the data pre-processing, they became 0.
# Changing them back to NA.
parent_ids <- c(50502, 51202, 51602, 52302, 53002,

```

```

53502, 53902, 54402, 54602, 54702)

for (id in parent_ids) {
  data[data$parent_id == id, ]["swan_inattentive"] = NA
  data[data$parent_id == id, ]["swan_hyperactive"] = NA
}

# Smoking during pregnancy If they smoked at any
# point during pregnancy, set to SDP. If not,
# set to No SDP.
sdp <- smoking_trimester %>%
  filter(!is.na(smoking)) %>%
  group_by(parent_id) %>%
  summarise(sdp = sum(smoking, na.rm = TRUE))

data_sdp <- data %>%
  left_join(sdp, by = "parent_id")

# Make two plots: distribution of both types of
# SWAN scores by SDP status
sdp_plot_1 <- ggplot(data_sdp, aes(x = swan_inattentive,
  fill = as.factor(sdp))) + geom_density(alpha = 0.25) +
  ylim(0, 0.35) + geom_vline(xintercept = 6, linetype = "dashed",
  color = "black", linewidth = 0.5) + annotate(geom = "text",
  x = 9.5, y = 0.32, label = "Likely ADHD \nInattentive") +
  scale_fill_manual(name = "Smoking During Pregnancy?",
    labels = c("No", "First Trimester", "All Trimesters"),
    values = ghibli_palette("LaputaMedium")[c(3,
      6, 7)]) + labs(title = "Distribution of SWAN Inattentive Scores by Trimesters Smoked",
    x = "SWAN Inattentive Score", y = "Density")

sdp_plot_2 <- ggplot(data_sdp, aes(x = swan_hyperactive,
  fill = as.factor(sdp))) + geom_density(alpha = 0.25) +
  ylim(0, 0.2) + geom_vline(xintercept = 6, linetype = "dashed",
  color = "black", linewidth = 0.5) + annotate(geom = "text",
  x = 9.5, y = 0.16, label = "Likely ADHD \nHyperactive") +
  scale_fill_manual(name = "Smoking During Pregnancy?",
    labels = c("No", "First Trimester", "All Trimesters"),
    values = ghibli_palette("LaputaMedium")[c(3,
      6, 7)]) + labs(title = "Distribution of SWAN Hyperactive Scores by Trimesters Smoked",
    x = "SWAN Hyperactive Score", y = "Density")

grid.arrange(sdp_plot_1, sdp_plot_2, nrow = 2)

# Create a separate column to store the races of
# the children along with the SDP effects. There
# are two categories of biracial children
# (AIAN/Black and Black/ White) so I just defined
# these as biracial.
sdp_race <- data_sdp %>%
  mutate(child_race = case_when(taian == 1 & tblack ==
    "white" | twhite == 1 & tblack == 1 ~ "biracial",
    taian == 1 ~ "aian", tasian == 1 ~ "asian",
    tnhpi == 1 ~ "nhpi", tblack == 1 ~ "black",

```



```

    twhite == 1 ~ "white", trace_other == 1 ~ "other"))

# For each category of SDP and Race, calculate
# the average ERQ scores for expressive
# suppression and cognitive reappraisal.
erq_child <- sdp_race %>%
  filter(!is.na(child_race)) %>%
  group_by(sdp, child_race) %>%
  summarize(cog = mean(erq_cog, na.rm = TRUE), exp = mean(erq_exp,
    na.rm = TRUE), .groups = "drop_last") %>%
  gather(key = "erq", value = "value", c(cog, exp))

erq_parent <- sdp_race %>%
  left_join(race_lookup, by = "parent_id") %>%
  filter(!is.na(race)) %>%
  group_by(sdp, race) %>%
  summarize(cog = mean(erq_cog_a, na.rm = TRUE),
    exp = mean(erq_exp_a, na.rm = TRUE), .groups = "drop_last") %>%
  gather(key = "erq", value = "value", c(cog, exp))

# For each category of SDP and Race, calculate
# the average BPM scores for attention,
# externalizing, and internalizing.
bpm_child <- sdp_race %>%
  filter(!is.na(child_race)) %>%
  group_by(sdp, child_race) %>%
  summarize(att = mean(bpm_att, na.rm = TRUE), ext = mean(bpm_ext,
    na.rm = TRUE), int = mean(bpm_int, na.rm = TRUE),
    .groups = "drop_last") %>%
  gather(key = "bpm", value = "value", c(att, ext,
    int))

bpm_parent <- sdp_race %>%
  left_join(race_lookup, by = "parent_id") %>%
  filter(!is.na(race)) %>%
  group_by(sdp, race) %>%
  summarize(att = mean(bpm_att_p, na.rm = TRUE),
    ext = mean(bpm_ext_p, na.rm = TRUE), int = mean(bpm_int_p,
    na.rm = TRUE), .groups = "drop_last") %>%
  gather(key = "bpm", value = "value", c(att, ext,
    int))

# Plot average BPM values by SDP status
p1 <- ggplot(bpm_child, aes(x = sdp, y = value, fill = bpm)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  scale_fill_manual(name = "BQM Section", labels = c("Attention Problems on Self",
    "Externalizing Problems on Self", "Internalizing Problems on Self"),
    values = ghibli_palette("LaputaMedium")[c(1,
      3, 5)]) + labs(title = "Average Brief Problem Monitor Scores by SDP Exposure",
    x = "Trimesters the Mother Smoked", y = "Average Score")

p2 <- ggplot(bpm_parent, aes(x = sdp, y = value, fill = bpm)) +
  geom_bar(stat = "identity", position = position_dodge()) +

```

```

    scale_fill_manual(name = "BPM Section", labels = c("Attention Problems on Child",
      "Externalizing Problems on Child", "Internalizing Problems on Child"),
      values = ghibli_palette("LaputaMedium")[c(1,
        3, 5)]) + labs(title = "Average Brief Problem Monitor Scores by SDP Exposure",
      x = "Trimesters the Mother Smoked", y = "Average Score")
grid.arrange(p1, p2, nrow = 2, ncol = 1)
# Plot average ERQ values by SDP status
erq_plot <- ggplot(erq_child, aes(x = sdp, y = value,
  fill = erq)) + geom_bar(stat = "identity", position = position_dodge()) +
  scale_fill_manual(name = "ERQ Section", labels = c("Cognitive Reappraisal",
    "Expressive Suppression"), values = ghibli_palette("LaputaMedium")[c(2,
      4)]) + labs(title = "Average ERQ Scores by SDP Exposure",
    x = "Trimesters the Mother Smoked", y = "Average Score")

erq_plot_parent <- ggplot(erq_parent, aes(x = sdp,
  y = value, fill = erq)) + geom_bar(stat = "identity",
  position = position_dodge()) + scale_fill_manual(name = "ERQ Section",
  labels = c("Cognitive Reappraisal", "Expressive Suppression"),
  values = ghibli_palette("LaputaMedium")[c(2, 4)]) +
  labs(title = "Average ERQ Scores by SDP Exposure",
    x = "Trimesters the Mother Smoked", y = "Average Score")

grid.arrange(erq_plot, erq_plot_parent, nrow = 2, ncol = 1)
smoke_pp <- data %>%
  mutate(smoke_exp_pp = smoke_exposure_6mo + smoke_exposure_12mo +
    smoke_exposure_2yr + smoke_exposure_3yr + smoke_exposure_4yr +
    smoke_exposure_5yr) %>%
  dplyr::select(parent_id, smoke_exp_pp, bpm_att,
    bpm_ext, bpm_int, erq_cog, erq_exp, swan_inattentive,
    swan_hyperactive)

behavior_by_smoke_pp <- smoke_pp %>%
  filter(!is.na(smoke_exp_pp)) %>%
  group_by(smoke_exp_pp) %>%
  summarise(avg_bpm = round(mean(bpm_att + bpm_ext +
    bpm_int, na.rm = TRUE), 2), avg_erq = round(mean(erq_cog +
    erq_exp, na.rm = TRUE), 2), avg_swan_inattentive = round(mean(swan_inattentive,
    na.rm = TRUE), 2), avg_swan_hyperactive = round(mean(swan_hyperactive,
    na.rm = TRUE), 2))

colnames(behavior_by_smoke_pp) <- c("Count of Postpartum Smoke Exposure",
  "Average BPM Score", "Average ERQ Score", "Average SWAN Inattentive Score",
  "Average SWAN Hyperactive Score")

kbl(behavior_by_smoke_pp, booktabs = T, escape = F,
  caption = "Effects of Postpartum Smoking on Internalizing and Externalizing Behaviors") %>%
  kable_styling(latex_options = "HOLD_position") %>%
  column_spec(1:5, width = "3cm")

```

Bibliography

- ASEBA. 2022. *ASEBA*. [https://aseba.org/school-age-bpm/#:~:text=The%20Brief%20Problem%20Monitor%20\(BPM,youths%20\(BPM%2DY\).%20Problem%20Monitor%20\(BPM,youths%20\(BPM%2DY\).](https://aseba.org/school-age-bpm/#:~:text=The%20Brief%20Problem%20Monitor%20(BPM,youths%20(BPM%2DY).%20Problem%20Monitor%20(BPM,youths%20(BPM%2DY).)
- Drake, Patrick, Anne K Driscoll, and TJ Mathews. 2018. “Cigarette Smoking During Pregnancy: United States, 2016.”
- Gross, James J, and Oliver P John. 2003. “Individual Differences in Two Emotion Regulation Processes: Implications for Affect, Relationships, and Well-Being.” *Journal of Personality and Social Psychology* 85 (2): 348.
- Hellemons, Merel E, Jan-Stephan F Sanders, Marc AJ Seelen, Rijk OB Gans, Anneke C Muller Kobold, Willem J van Son, Douwe Postmus, Gerjan J Navis, and Stephan JL Bakker. 2015. “Assessment of Cotinine Reveals a Dose-Dependent Effect of Smoking Exposure on Long-Term Outcomes After Renal Transplantation.” *Transplantation* 99 (9): 1926–32.
- Micalizzi, Lauren. 2023. “Micalizzi PHP2550.pdf.”
- NIH. 2021. *National Institutes of Health*. U.S. Department of Health; Human Services. <https://nida.nih.gov/publications/research-reports/tobacco-nicotine-e-cigarettes/what-are-risks-smoking-during-pregnancy>.
- Risica, Patricia Markham, Adam Gavarkovs, Donna R Parker, Ernestine Jennings, and Maureen Phipps. 2017. “A Tailored Video Intervention to Reduce Smoking and Environmental Tobacco Exposure During and After Pregnancy: Rationale, Design and Methods of Baby’s Breath.” *Contemporary Clinical Trials* 52: 1–9.