Project 1: Exploratory Data Analysis

Hannah Eglinton

2023-09-25

Data Collection

Around 15 years ago, approximately 800 pregnant people were recruited in a study that investigated whether a tailored video intervention reduced smoking and environmental tobacco exposure during and after pregnancy [1]. The study did not find that this intervention had an effect on smoking during or after pregnancy.

The current study's sample is a subset of the mothers recruited in the previous study and their children, who are now 12 to 16 years old. The study aims to examine the association between prenatal and postnatal smoke exposure with the teenage children's substance use, self regulation, and externalizing behaviors. The study collected data during baseline, 6-month, and 12-month laboratory sessions, though this report uses only the baseline data. The children's internalizing and externalizing behaviors were measured using computerized assessments completed by both the mothers and the children themselves.

This report uses data about the mothers' smoking status collected during the original study, along with parent and adolescent covariates, parent's recalled smoke exposure when their child was 0-5 years, and adolescent outcome measures collected during the current study.

Data Preprocessing

The raw data included 1,282 variables on 49 mother/adolescent pairs. The data were processed to include only 78 variables by dropping variables that were irrelevant or redundant for this portion of the project or by combining multiple variables into summary variables. Variables that were dropped included the Youth Self-Report survey responses, adolescent survey responses about non-marijuana drug use or drug norms, questions about physical development, disregulation, adolescent and adult life stress, adolescent and adult temperament, diet questions, Connor's ADHD test responses, and Chaos test responses. Sets of variables that were summarized included cigarette usage, marijuana usage, alcohol usage, Brief Problem Monitor scores, ERQ scores, parental monitoring scores, and SWAN scores.

A few variables were not reported in a standardized way, which caused some additional cleaning to be required. One `income` value was reported as "250, 000", which caused this variable to be interpreted as characters rather than numeric. This value was changed to "250000" and the variable was converted to numeric. The `mom_numcig` variable included text ("2 black and miles a day", "None"), a range of values ("20-25"), and an implausible value ("44989"). These values were changed to 2, 0, 23, and NA, respectively, and the variable was converted to numeric. The smoking during pregnancy (SDP) variables were recorded as either "2=No" or "1=Yes". These were converted to binary values 0 and 1, respectively.

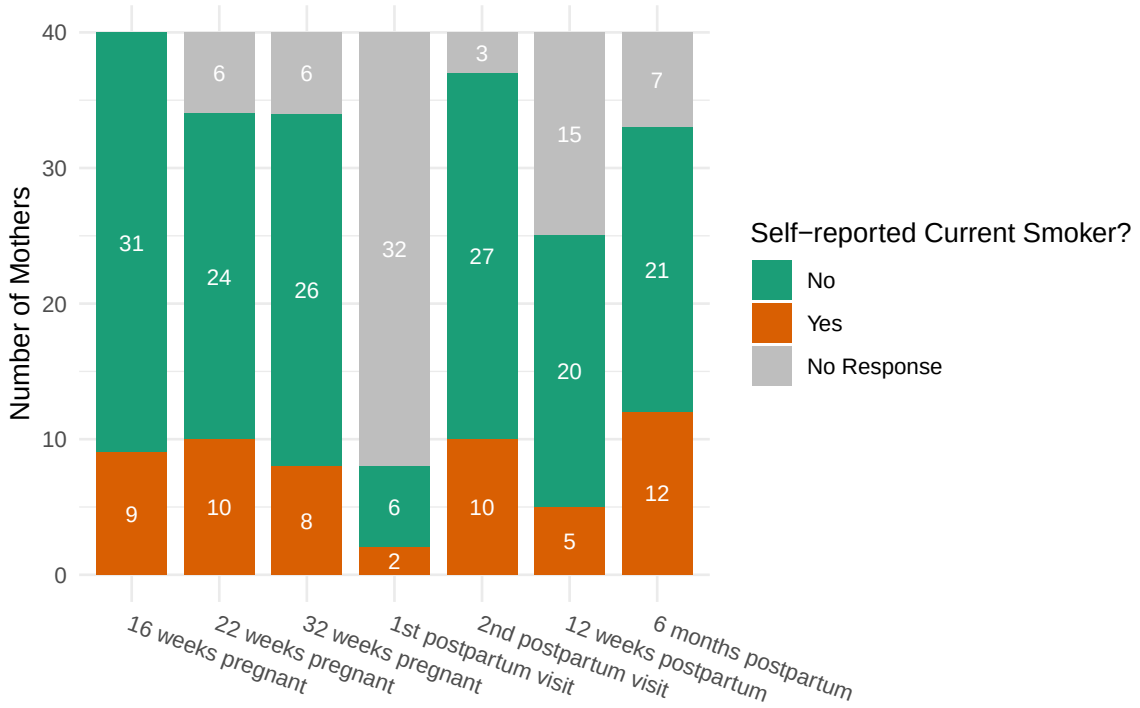
Eight parent/adolescent pairs were missing data for all variables associated with the current study – they only had reported responses for the variables collected during the original study (mother's prenatal and postpartum smoking status and mean urine cotinine). Since these subjects were missing the data that is necessary to answer the questions in this study (i.e. postpartum smoke exposure, covariates, and outcomes), they were removed from this analysis. An additional parent/adolescent pair was missing all variables except for the covariates age, sex, race, and income. As this pair had no exposure or outcome variables that could be used in this analysis, it was also dropped. Although some missing data can be handled through approaches like multiple imputation, these nine pairs do not have enough non-missing data to estimate the values of the

missing variables. The dataset contained 40 parent/adolescent pairs after removing pairs that had missing data for almost all of the variables.

Exposure Variables (Original Study)

In the original study, pregnant women were asked to self-report whether they were a current smoker at six timepoints: at 16 weeks pregnant, at 22 weeks pregnant, at 23 weeks pregnant, at first postpartum visit, at second postpartum visit, at 12 weeks postpartum, and at 6 months postpartum. Figure 1 visualizes the responses to these questions. At each timepoint, between 20% and 36% of responding mothers reported that they were current smokers. We see that the there was no missing data when the mothers were 16 weeks pregnant, while the first postpartum visit had a large amount of missing data with 80% of mothers reporting no response. It is possible that many of these mothers only attended one postpartum check-up during the 12 weeks after birth. If we wanted to impute these missing values for further analysis, we would need to confirm that the mothers that missed postpartum check-ups weren't characteristically different than the mothers that attended two check-ups between 0 and 12 weeks.

Figure 1. Self−Reported Smoking During Pregnancy and Postpartum



The original study also measured the mothers' cotinine levels at 34 weeks gestation and 6 months postpartum, which measured the amount of nicotine in their urine. Although test results can vary depending on a number of factors, nonsmokers typically have cotinine levels less than 10 ng/mL, light smokers or those exposed to secondhand smoke are usually between 11 and 30 ng/mL, and heavy smokers have levels greater than 30 ng/mL (though levels can be greater than 500 ng/mL) [2].

Figure 2a visualizes the distribution in urine cotinine levels at 34 weeks gestation. The cuttoffs for light and heavy smokers are shown with dotted orange lines (at 10 and 30 ng/mL, respectively). We can see that most mothers had cotinine levels that indicated that they were nonsmokers. The heavy smokers were very widely distributed with one mother reporting cotinine levels of 329.19 ng/mL (which still remains within the range of reasonable values for heavy smokers).

Figure 2b visualizes the distribution in cotinine at 6 months postpartum. There is a slight shift to the right in comparison to the pregnancy levels, with fewer mothers in the nonsmoking range and higher levels among heavy smokers (up to 465.56 ng/mL).

Figure 2a. Distribution in Pregnancy Cotinine Levels

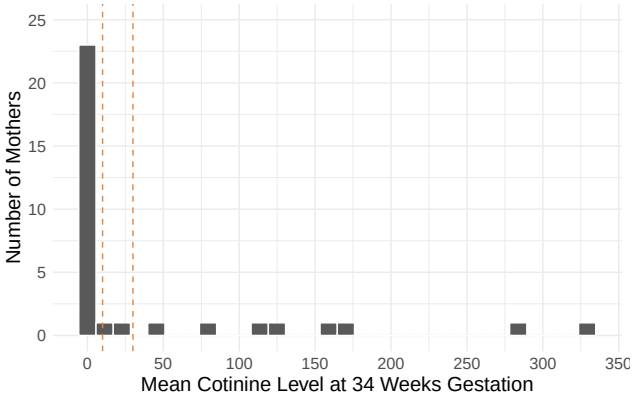


Figure 2b. Distribution in Postpartum Cotinine Levels

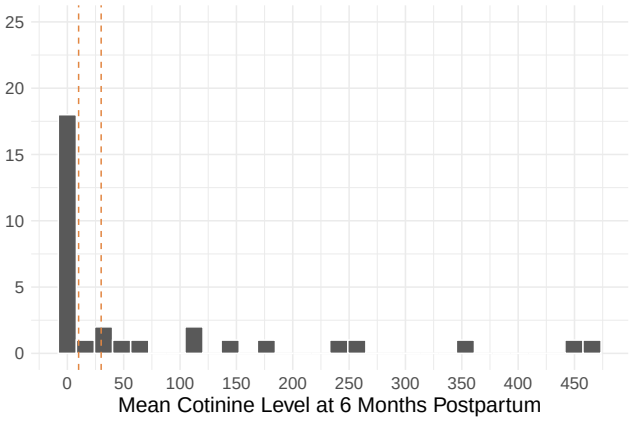


Table 1 compares the self-reported smoking status during pregnancy to the smoking level that was determined by the mother's 34 week cotinine measurement (nonsmoker, light, heavy, missing). Most mothers that were classified as a "nonsmoker" from their pregnancy cotinine level (N = 23) reported that they were not smokers during pregnancy. All mothers classified as "heavy" smokers (N = 8) reported smoking at every pregnancy time point. This may indicate that the mothers were honest in their reporting, as their cotinine levels match their responses. The self-reported smoking status at 16 weeks pregnant seems to be a good proxy for overall SDP, as there is no missing data, most responses do not change over time, and they mostly align with the measured cotinine levels.

Table 1: Comparing 34 Week Gestation Cotinine Levels to Self-reported Smoking Status

Characteristic	**Nonsmoker**, N = 23	**Light**, N = 2	**Heavy**, N = 8	**Missing**, N = 7
16 Weeks Pregnant				
0	22 (96%)	2 (100%)	0 (0%)	7 (100%)
1	1 (4.3%)	0 (0%)	8 (100%)	0 (0%)
22 Weeks Pregnant				
0	19 (95%)	1 (50%)	0 (0%)	4 (100%)
1	1 (5.0%)	1 (50%)	8 (100%)	0 (0%)
Missing	3	0	0	3
32 Weeks Pregnant				
0	23 (100%)	2 (100%)	0 (0%)	1 (100%)
1	0 (0%)	0 (0%)	8 (100%)	0 (0%)
Missing	0	0	0	6

[1] n (%)

Table 2 uses the six-month postpartum cotinine measurements to stratify mothers by smoking level (nonsmoker, light, heavy, missing) and compares these categories to the self-reported postpartum smoking statuses. We can see that there were more light smokers and fewer nonsmokers at 6 months postpartum compared to during pregnancy (Table 1). This suggests that some women may return to smoking after withholding during pregnancy, showing that it is important to evaluate the effects of postpartum smoke exposure separately than gestational smoke exposure, since these values may differ. Again, it appears that women were honest about their smoking status, as their self-reported statuses tend to align with their cotinine categories.

Table 2: Comparing Six Month Postpartum Cotinine Levels to Self-reported Smoking Status

Characteristic	**Nonsmoker**, N = 18	**Light**, N = 7	**Heavy**, N = 7	**Missing**, N = 3
First postpartum visit				
0	3 (100%)	3 (100%)	0 (0%)	0 (NA%)

Table 2: Comparing Six Month Postpartum Cotinine Levels *(continued)*

Characteristic	**Nonsmoker**, N = 18	**Light**, N = 7	**Heavy**, N = 7	**Missing**, N = 3
1	0 (0%)	0 (0%)	2 (100%)	0 (NA%)
Unknown	15	4	5	3
Second postpartum visit				
0	12 (100%)	4 (100%)	2 (40%)	0 (NA%)
1	0 (0%)	0 (0%)	3 (60%)	0 (NA%)
Unknown	6	3	2	3
12 weeks postpartum				
0	17 (94%)	6 (86%)	1 (14%)	0 (NA%)
1	1 (5.6%)	1 (14%)	6 (86%)	0 (NA%)
Unknown	0	0	0	3
6 months postpartum				
0	16 (89%)	5 (71%)	0 (0%)	0 (NA%)
1	2 (11%)	2 (29%)	7 (100%)	0 (NA%)
Unknown	0	0	0	3

[1] n (%)

Unfortunately, many of the subjects that were missing cotinine measurements were also missing self-reported smoking statuses, especially during postpartum where none of the mothers missing cotinine levels reported their postpartum smoking. This would make it difficult to impute missing cotinine values based on self-reported scores, or vice versa.

The current study measured a variety of covariates that may be associated with smoking during pregnancy. Table 3 reports a summary of each covariate stratified by self-reported smoking status at 32 weeks. It is important to emphasize that the covariate data was collected around 15 years after pregnancy, so may not reflect each woman's situation at the time that they reported their smoking status.

There were no statistically significant differences between women who reported smoking at 32 weeks pregnant and those who reported not smoking at 32 weeks pregnant. One notable observation is that the women who reported smoking during pregnancy were less likely to have a college degree and had a lower average income.

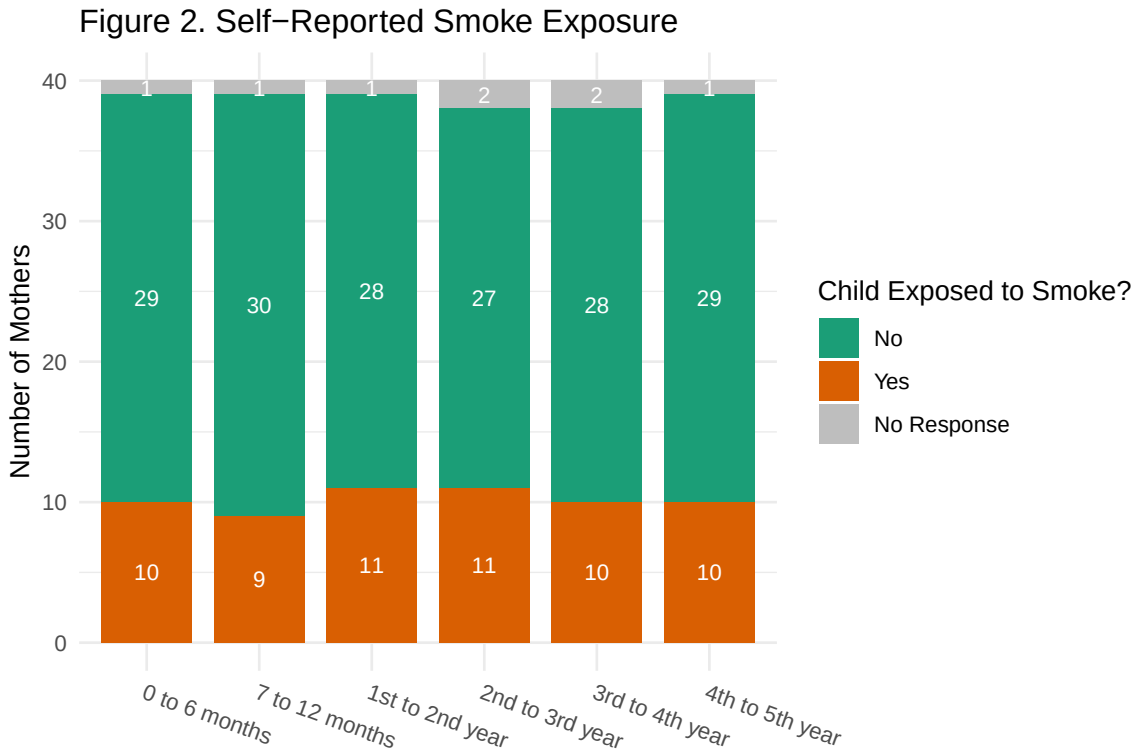
Table 3: Parent Covariates by Self-Reported 32-Week Smoking Status

Covariate	Nonsmoker	Smoker
Speaks another language at home	12 (39%)	3 (33%)
Missing	0	0
Identifies as white	19 (61%)	6 (67%)
Missing	0	0
Is employed (part-time or full-time)	22 (71%)	7 (78%)
Missing	0	0
Has a college degree (2-year or 4-year)	14 (45%)	1 (11%)
Missing	0	0
Has used alcohol in the past 6 months (4 or more drinks per day)	15 (50%)	3 (33%)
Missing	1	0
Has used prescription drugs in the past 6 months	3 (10%)	0 (0%)
Missing	1	1
Has used illegal drugs in the past 6 months	1 (3.3%)	1 (11%)
Missing	1	0
Family's estimated annual household income ($)	73,560 (66,723)	37,222 (16,045)
Missing	4	0

[1] n (%); Mean (SD)

Exposure Variables (Current Study)

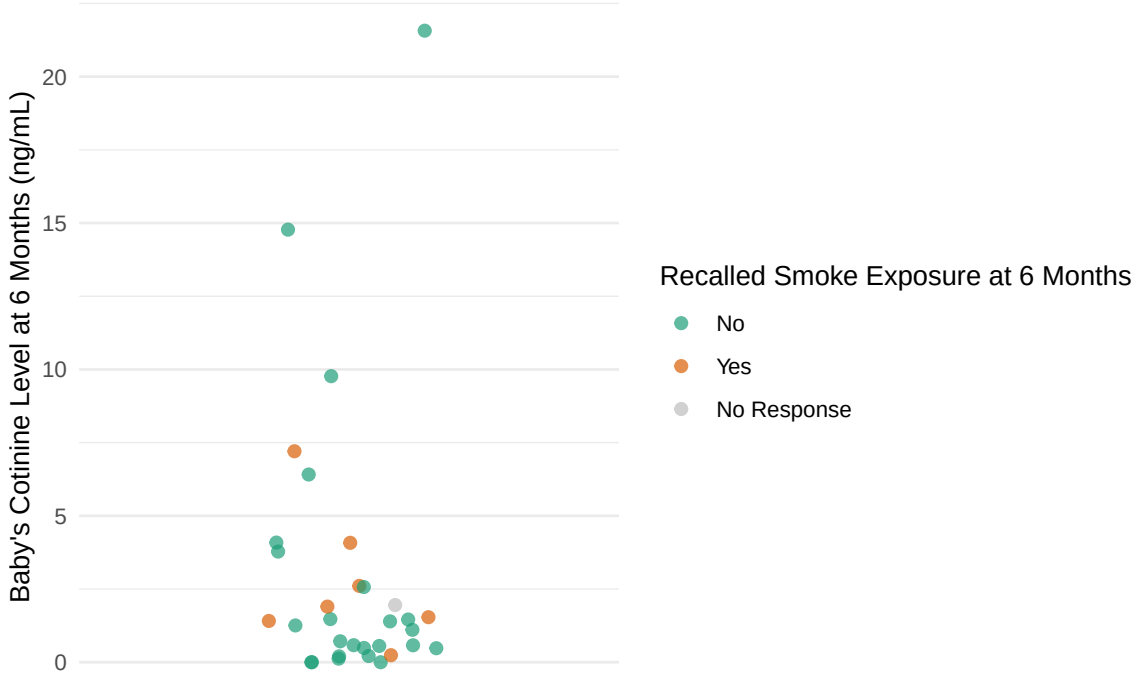
In the current study, mothers recalled their child's smoke exposure (from either the mother or their partner) from birth to age 5 (Figure 2). The proportions of each answer (yes, no, missing response) stayed relatively constant throughout each recalled time period.



The recalled smoke exposure may be more relevant than the self-reported postpartum smoking status from the original study because a baby's cotinine levels would be influenced by any secondary smoke exposure, not just from the mother. However, this variable is less likely to be accurate since it requires the parent's memory to be correct.

One way to validate the accuracy of the mothers' recall is to compare the recalled smoke exposure from 0 to 6 months to the baby's 6-month continine levels (that were measured in the previous study). Figure 3 plots each baby's cotinine level at 6 months, colored by the mother's response to the 6-month smoke exposure question from the current study. We would expect that the babies whose mothers recalled being exposed to smoke would have higher levels of cotinine. However, the babies with the highest cotinine levels supposedly had no smoke exposure, according to the parents around 15 years later. The mean cotinine levels among babies with no recalled smoke exposure was 3.06 (SD 5.29) ng/mL while the levels among babies with recalled smoke exposure was 2.71 (SD 2.31) ng/mL.

Figure 3. Reported Smoke Exposure vs Baby's Cotinine Levels



There are a few potential reasons for this pattern. First, the mother could have inaccurately remembered her or her partner's smoking status. Or, the mother could have been correct that neither she or her partner smoked, but the baby was exposed to secondary smoke due to other people smoking nearby. Regardless, these data may suggest that the recalled smoke exposure responses collected in the current study may not accurately reflect the child's exposure to secondary smoke.

Outcome Variables

The outcomes in this study are the adolescents' externalizing behaviors and self-regulation. Externalizing behaviors include substance use, attention-deficit/hyperactivity disorder, and conduct disorder. Substance use was measured by surveying the children on their use of cigarettes, e-cigarettes, marijuana, and alcohol. Attention-deficit, hyperactivity, and conduct disorders were evaluated through both the child's and parent's responses to the Brief Problem Monitor and through the parent's responses to the SWAN rating scale.

Self-regulation refers to an individual's ability to understand and manage their own behavior and reactions, and can involve executive function, emotion regulation, effortful control, and vagal tone. Self-regulation was evaluated using the Brief Problem Monitor and the Emotion Regulation Questionnaire.

Substance Use

The children of this study were asked if they had *ever* tried cigarettes, e-cigarettes, marijuana, or alcohol. Four subjects (out of 40) did not provide a response to these questions, and one subject did not provide a response to this question in regard to alcohol but responded for cigarettes, e-cigarettes, and marijuana (and answered 'no' to each).

Only one subject said that they had ever tried cigarettes and reported having 0 cigarettes in the last 30 days. Three subjects said that they had ever tried e-cigarettes. One of these subjects had 2 e-cigarettes in the past 30 days, one had zero, and the third did not respond. Three subjects said that they had ever tried marijuana and reported the number of days using marijuana in the past 30 days to be 3, 12, and 18. Five subjects said that they had ever tried alcohol. Two of these subjects drank alcohol zero days in the past 30 days, one subject drank alcohol one day in the past 30 days, one subject drank alcohol 10 days in the past 30 days, and

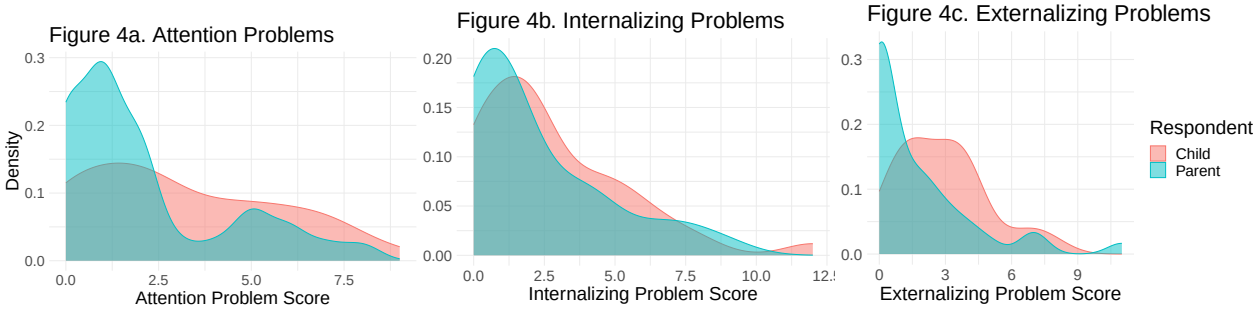
one subject did not provide the number of days drinking alcohol in the past 30 days. Among the 6 subjects that reported ever trying substances, 1 had tried only marijuana, 3 had tried only alcohol, 1 had tried both marijuana and alcohol, and 1 had tried marijuana, alcohol, and cigarettes.

Four subjects report using at least one substance in the past 30 days. Due to this small sample size, it may be difficult to determine the effect of smoke exposure on regular substance use.

Brief Problem Monitor

Adolescents and parents reported how true each statement in the Brief Problem Monitor (BPM) was about themselves or their child, respectively. Questions were asked on a 0-1-2 rating (0 = Not True, 1 = Somewhat True, 2 = Very True) and responses were summed over each category to yield a BPM score. The categories included attention problems, externalizing problems, and internalizing problems.

Figure 4 compares the child's responses about themselves and the parent's responses about their child. The parents' responses were more heavily right-skewed; in other words, parents were more likely to report that their child had little to no problems than the child would report about themselves. This may indicate that the children were more aware of their own difficulties than their parents were. This pattern was most visible in the attention and externalizing problem categories, while the distributions of internalizing problem scores were fairly similar between the children and the parents.



Emotion Regulation Questionnaire

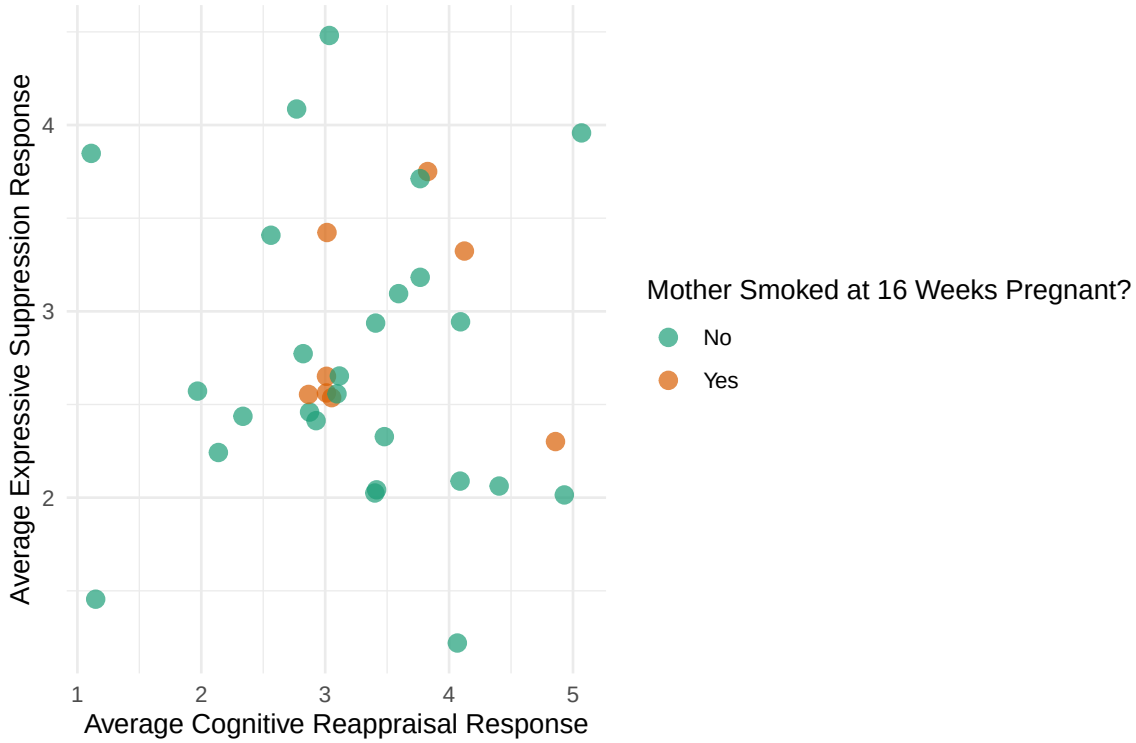
The Emotion Regulation Questionnaire (ERQ) is a 10-item scale designed to measure respondents' tendency to regulate their emotions through cognitive reappraisal or expressive suppression. The children in this study answered each item on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). Cognitive reappraisal is considered a positive emotional response, where one reinterprets an emotional situation to change its meaning and impact. Expressive suppression is considered a negative emotional response, where one attempts to hide or inhibit emotional expression [3].

One child was missing responses for the cognitive reappraisal questions, one child was missing responses for the expressive suppression questions and four children were missing responses for both sections.

Figure 5 visualizes the relationship between each adolescent's average cognitive reappraisal and their average expression suppression. We would expect adolescents with emotion regulation difficulties to have higher expressive suppression and lower cognitive reappraisal (i.e. in the top left of the figure). We don't see a distinct group with this pattern, though it is difficult to quantify since the ERQ does not define cutoffs for high or low average responses.

We may have expected there to be a strong negative correlation between these two variables, with each individual considered either "healthy" or "unhealthy" in both types of emotional response. Instead, we observe a very weak, slightly negative correlation (r = -0.009), indicating that these two types of emotional regulation are not necessarily related. In other words, having high cognitive reappraisal does not mean that one has low expressive suppression, and vice versa. We also do not see a clear pattern of these scores' relationship with smoking during pregnancy. Adolescents whose mothers smoked at 16-weeks pregnant have comparable scores to those whose mothers did not report smoking.

Figure 5. Cognitive Reappraisal vs. Expressive Suppression



One potential confounder with this outcome could be the parents' emotional regulation. Kids are may be likely to mimic how their parents' are responding to emotional situations, so this study also measured the parents' ERQ scores. However, there was a weak positive correlation (r = 0.09) between parent and child cognitive reappraisal scores and a weak positive correlation (r = 0.26) between the parent and child expressive suppression scores, suggesting that this association is not as strong as one might think.

SWAN Scale

The SWAN rating scale, completed by parents about their child, is an 18-item questionnaire that assesses Attention-Deficit Hyperactivity Disorder (ADHD). Nine questions are used to assess the hyperactive/impulsive type of ADHD. Another nine questions are used to assess the inattentive type of ADHD. Each response is on a 0-1-2-3 scale (0 = Not at all, 1 = Just a little, 2 = Quite a bit, 3 = Very much). When calculating the score, responses of 0 and 1 are coded as 0 and responses of 2 and 3 are coded as 1. These coded values are summed to create a score for each subset. A score of 6 or greater in either subset indicates that the child likely has ADHD of the respective type. A score of 6 or greater in both subsets indicates that the child likely has combined-type ADHD [4].

Figure 6 visualizes the relationship between each adolescent's hyperactive score and their inattentive score. The cutoff value of 6 is shown with the dashed orange lines. Individuals in the bottom left quadrant (N = 26) likely do not have ADHD. Individuals in the top left quadrant (N = 4) are likely ADHD-Inattentive type. Individuals in the bottom right quadrant (N = 2) are likely ADHD-Hyperactive/Impulsive type. Individuals in the top right quadrant (N = 6) are likely ADHD-Combined type. The correlation between hyperactive and inattentive scores is 0.78, indicating a positive association. Two individuals were missing all SWAN responses.

Figure 6. Hyperactive ADHD vs. Inattentive ADHD

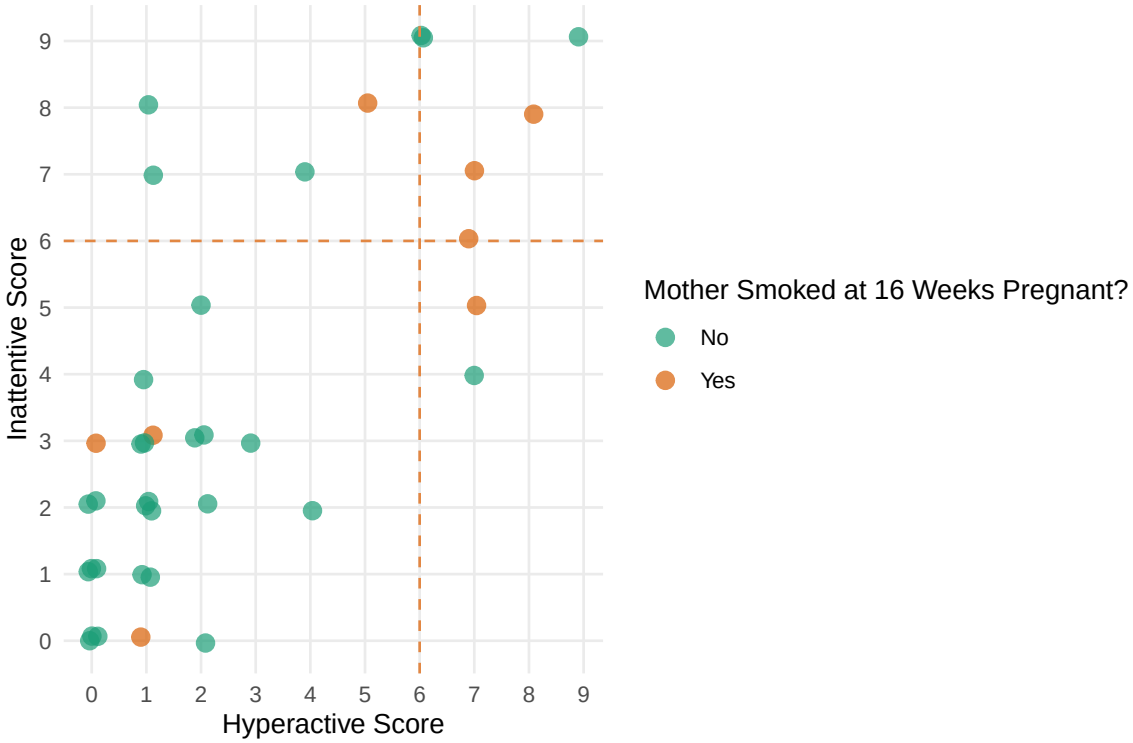


Figure 6 also shows whether each adolescent had a mother who reported smoking when she was 16 weeks pregnant. 12% of mothers of adolescents without ADHD reported smoking during pregnancy, while 42% of mothers of adolescents with ADHD reported smoking during pregnancy. This may suggest that smoking during pregnancy is associated with a higher risk of ADHD, but there are likely confounders that need to be adjusted for.

Table 4 reports the differences in child covariates by ADHD status (defined by the SWAN scale). There was a statistically significant difference in assigned sex at birth between those that likely have ADHD (9.1% female) and those that likely don't have ADHD (50% female). Adolescents that like have ADHD were also less likely to speak another language at home, more likely to be white, more likely to have Autism Spectrum Disorder, and had a lower mean family income, though none of these differences were statistically significant.

Table 4: Child Covariates by ADHD Status

Covariate	Likely Doesn't Have ADHD	Likely Has ADHD	**p-value**
Age	13 (1)	13 (1)	>0.9
Unknown	3	1	
Assigned female at birth	11 (50%)	1 (9.1%)	0.027
Unknown	4	1	
Speaks another language at home	9 (39%)	1 (9.1%)	0.11
Unknown	3	1	
White	11 (42%)	7 (58%)	0.4
Diagnosed or suspected of having Austism Spectrum Disorder	1 (7.7%)	1 (17%)	>0.9
Unknown	13	6	
Family's estimated annual household income	75,056 (66,537)	47,276 (41,671)	0.2
Unknown	2	2	

[1] Mean (SD); n (%)

[2] Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

Summary of Internalizing Behaviors

Table 5 reports a summary of the measured internalizing behaviors in adolescents, stratified by whether the mother reported smoking at 16 weeks pregnant. Although not a statistically significant difference, adolescents whose mothers smoked at 16 weeks pregnant had lower median cognitive reappraisal scores and higher median expressive suppression scores, indicating "unhealthier" emotional regulation. However, more rigorous tests that adjust for confounders will be necessary to make any conclusions.

Table 5: Internalizing Behaviors by Mother's 16-week Smoking Status

Characteristic	Nonsmoker	Smoker	**p-value**
BPM Internalizing Score	2 (1, 5)	2 (0, 3)	0.8
Unknown	6	0	
ERQ Cognitive Reappraisal	3.17 (2.83, 3.83)	3.00 (3.00, 3.92)	0.6
Unknown	4	1	
ERQ Expressive Suppression	2.50 (2.06, 3.19)	2.75 (2.50, 3.50)	0.4
Unknown	5	0	

[1] Median (IQR)

[2] Wilcoxon rank sum test

Summary of Externalizing Behaviors

Table 6 reports a summary of the measured externalizing behaviors in adolescents, stratified by whether the mother reported smoking at 16 weeks pregnant. Although not statistically significant, the adolescents whose mothers smoked during pregnancy had a higher prevalence of trying substances, higher BPM externalizing and attention scores, and higher SWAN scores (ADHD). The most pronounced difference was observed in the SWAN scores, suggesting that ADHD is the externalizing response that is most closely associated with smoking during pregnancy.

Table 6: Externalizing Behaviors by Mother's 16-week Smoking Status

Characteristic	Nonsmoker	Smoker	**p-value**
Ever tried cigarettes, e-cigarettes, alcohol, or marijuana	5 (19%)	2 (25%)	0.6
Unknown	4	1	
BPM Externalizing Score	2 (1, 4)	3 (1, 4)	0.6
Unknown	4	0	
BPM Attention Score	2 (1, 5)	5 (2, 7)	0.14
Unknown	4	0	
SWAN Hyperactive Score	1 (0, 2)	6 (1, 7)	0.067
Unknown	1	1	
SWAN Inattentive Score	2 (1, 4)	6 (3, 7)	0.11
Unknown	1	1	

[1] n (%); Median (IQR)

[2] Fisher's exact test; Wilcoxon rank sum test

Discussion

The current study's hypothesis is that prenatal and postpartum smoke exposure causes self-regulation deficits that can then cause early substance use initiation and externalizing behaviors. This report's exploratory data analysis revealed that there wasn't an association between smoking during pregnancy and the measured self-regulation behaviors, while there was a stronger association between smoking during pregnancy and the measured externalizing behaviors, particularly the incidence of ADHD. If this pattern holds with more statistical tests and after adjustment for confounders, this would make the study's hypothesis unlikely to be true, since smoking during pregnancy appears to affect externalizing behaviors more than internalizing behaviors.

One strength of this study is its access to self-reported smoking status and cotinine levels of the adolescents' mothers while pregnant. However, this real-time data was only collected up to 6 months postpartum, while smoke exposure is still relevant to development far after 6 months. The current study collected information about postpartum smoke exposure up to 5 years postpartum, but relied on the mothers' memories at each timepoint. By comparing the mothers' recalled smoke exposure at 6 months to the baby's measured cotinine levels at 6 months, we learned that these recalled responses may not be reliable, which is a major limitation of this study since it is interested in postpartum smoke exposure.

Another limitation is this study's small sample size and the prevalence of missing data. Nine out of 49 parent/adolescent pairs were missing all exposure and outcome data, leaving 40 parent/adolescent pairs with sufficient data to answer the study's questions. Still, almost all variables contained at least some missing data. Further analysis will need to be conducted to determine whether this missingness is Missing At Random and whether imputation techniques can be used to address this missing data.

Finally, the current study is limited by the lack of covariates measured in the original study. Potential confounders like the parents' substance use, employment, and income were collected in the current study and not measured at the actual time of pregnancy. Also, since only cigarette use was evaluated at pregnancy, we do not have information about other behaviors that can affect development, such as alcohol use during pregnancy. Alcohol use during pregnancy is known to be associated with hyperactivity, mood disorders, difficulty with attention, poor reasoning, and ADHD [5], likely making this a very influential unmeasured confounder.

In conclusion, although this exploratory data analysis revealed some potentially interesting associations between smoking during pregnancy and externalizing behaviors, it also revealed study limitations that will need to be addressed in order to reach meaningful or generalizable conclusions about smoke exposure during development.

References

[1] Risica, P. M., Gavarkovs, A., Parker, D. R., Jennings, E., & Phipps, M. (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. https://doi.org/10.1016/j.cct.2016.10.010

[2] Nicotine Cotinine (Urine) - Health Encyclopedia - University of Rochester Medical Center. https://www.urmc.rochester.edu/encyclopedia/content.aspx?contentid=nicotine_cotinine&contenttypeid=167

[3] Cutuli, D. (2014). Cognitive reappraisal and expressive suppression strategies role in the Emotion Regulation: An overview on their modulatory effects and neural correlates. Frontiers in Systems Neuroscience, 8. https://doi.org/10.3389/fnsys.2014.00175

[4] The SWAN rating scale for ADHD - AmeriHealth. https://www.amerihealth.com/pdfs/providers/resources/worksheets/prevhealth_swan.pdf

[5] Peadon, E., & Elliott, E. J. (2010). Distinguishing between attention-deficit hyperactivity and fetal alcohol spectrum disorders in children: clinical guidelines. Neuropsychiatric disease and treatment, 6, 509–515. https://doi.org/10.2147/ndt.s7256

Code Appendix:

```
knitr::opts_chunk$set(echo = FALSE)
library(tidyverse)
library(mice)
library(gtsummary)
library(RColorBrewer)

# read in data
```

```r
data <- read.csv("project1.csv")

### DATA CLEANING

# clean income variable and convert to numeric
data$income[which(data$income == "250, 000")] <- "250000"
data$income <- as.numeric(data$income)

# clean mom_numcig variable and convert to numeric
data$mom_numcig[which(data$mom_numcig == "2 black and miles a day")] <- "2"
data$mom_numcig[which(data$mom_numcig == "None")] <- "0"
data$mom_numcig[which(data$mom_numcig == "20-25")] <- "23"
data$mom_numcig[which(data$mom_numcig == "44989")] <- ""

data$mom_numcig <- as.numeric(data$mom_numcig)

# convert mom smoking variables to binary
data <- data %>%
  mutate_at(
  vars(mom_smoke_16wk:mom_smoke_pp6mo),
  funs(case_when(
  . == "2=No" ~ 0,
  . == "1=Yes" ~ 1,
  . == "" ~ NA)))

### MISSING DATA
# remove subjects with no data from current study
data <- data %>%
  filter(!parent_id %in% c(50502, 51202, 51602, 52302, 53502, 54402, 54602,
                           54702, 53902))

### SDP VARIABLES

# Mom smoking variables
mom_smoking <- data %>%
  select(mom_smoke_16wk, mom_smoke_22wk, mom_smoke_32wk,
         mom_smoke_pp1, mom_smoke_pp2, mom_smoke_pp12wk, mom_smoke_pp6mo,
         cotimean_34wk, cotimean_pp6mo, cotimean_pp6mo_baby,
         smoke_exposure_6mo, smoke_exposure_12mo, smoke_exposure_2yr,
         smoke_exposure_3yr, smoke_exposure_4yr, smoke_exposure_5yr)


mom_smoking_long <- data %>%
  select(mom_smoke_16wk:mom_smoke_pp6mo) %>%
  pivot_longer(mom_smoke_16wk:mom_smoke_pp6mo, names_to = "time",
               values_to = "smoking")

# Figure 1
ggplot(mom_smoking_long, aes(x = time, fill = factor(smoking,
                                                     levels = c(NA, 0, 1),
                                                     exclude = NULL))) +
  geom_bar(width = 0.8) +
  scale_fill_brewer(palette = "Dark2", na.value = "grey",
```

```
                    labels = c("No", "Yes", "No Response")) +
  geom_text(aes(label = after_stat(count)), stat = "count",
            position = position_stack(vjust = 0.5), color = "white", size = 3) +
  labs(x = "", y = "Number of Mothers", fill = "Self-reported Current Smoker?",
       title = "Figure 1. Self-Reported Smoking During Pregnancy and Postpartum") +
  scale_x_discrete(labels = c("16 weeks pregnant", "22 weeks pregnant",
                              "32 weeks pregnant", "1st postpartum visit",
                              "2nd postpartum visit", "12 weeks postpartum",
                              "6 months postpartum")) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = -20, hjust = 0))

# Figure 2a
ggplot(data) +
  geom_histogram(aes(x = cotimean_34wk), color = "white", bins = 30) +
  lims(x = c(0, 475), y = c(0, 25)) +
  labs(x = "Mean Cotinine Level at 34 Weeks Gestation", y = "Number of Mothers",
       title = "Figure 2a. Distribution in Pregnancy Cotinine Levels") +
  geom_vline(xintercept = c(10, 30), linetype = "dashed", color = "#e28743") +
  scale_x_continuous(n.breaks = 10) +
  theme_minimal() +
  theme(text = element_text(size=15))

# Figure 2b
ggplot(data) +
  geom_histogram(aes(x = cotimean_pp6mo), color = "white", bins = 30) +
  lims(x = c(0, 475), y = c(0, 25)) +
  labs(x = "Mean Cotinine Level at 6 Months Postpartum", y = "",
       title = "Figure 2b. Distribution in Postpartum Cotinine Levels") +
  geom_vline(xintercept = c(10, 30), linetype = "dashed", color = "#e28743") +
  scale_x_continuous(n.breaks = 10) +
  theme_minimal() +
  theme(text = element_text(size=15))
# Reformat/Factor SDP and cotinine data
data <- data %>%
  mutate(across(where(~all(. %in% c(0, 1, NA))), factor))

data <- data %>%
  mutate(smoking_level_34wk = case_when(cotimean_34wk < 10 ~ "Nonsmoker",
                                        (cotimean_34wk > 10 &
                                           cotimean_34wk < 30) ~ "Light",
                                        cotimean_34wk > 30 ~ "Heavy",
                                        is.na(cotimean_34wk) ~ "Missing"),
         smoking_level_6mo = case_when(cotimean_pp6mo < 10 ~ "Nonsmoker",
                                        (cotimean_pp6mo > 10 &
                                           cotimean_34wk < 30) ~ "Light",
                                        cotimean_pp6mo > 30 ~ "Heavy",
                                        is.na(cotimean_34wk) ~ "Missing"))

data$smoking_level_34wk <- factor(data$smoking_level_34wk,
                                        levels = c("Nonsmoker", "Light",
                                                   "Heavy", "Missing"))
```

```r
# Table 1
tbl_summary(data, include = c(mom_smoke_16wk, mom_smoke_22wk,
                              mom_smoke_32wk),
            by = smoking_level_34wk,
            label = list(mom_smoke_16wk ~ "16 Weeks Pregnant",
                         mom_smoke_22wk ~ "22 Weeks Pregnant",
                         mom_smoke_32wk ~ "32 Weeks Pregnant"),
            missing_text = "Missing") %>%
   as_kable_extra(booktabs = TRUE,
                  caption = "Comparing 34 Week Gestation Cotinine Levels to
                  Self-reported Smoking Status",
                 longtable = TRUE) %>%
  kableExtra::kable_styling(font_size = 8,
                            latex_options = c("repeat_header", "HOLD_position"))
# Table 2
data$smoking_level_6mo <- factor(data$smoking_level_6mo,
                                 levels = c("Nonsmoker", "Light",
                                            "Heavy", "Missing"))

tbl_summary(data, include = c(mom_smoke_pp1, mom_smoke_pp2,
                              mom_smoke_pp12wk, mom_smoke_pp6mo),
            by = smoking_level_6mo,
            label = list(mom_smoke_pp1 ~ "First postpartum visit",
                         mom_smoke_pp2 ~ "Second postpartum visit",
                         mom_smoke_pp12wk ~ "12 weeks postpartum",
                         mom_smoke_pp6mo ~ "6 months postpartum")) %>%
   as_kable_extra(booktabs = TRUE,
                  caption = "Comparing Six Month Postpartum Cotinine Levels to
                  Self-reported Smoking Status",
                 longtable = TRUE) %>%
  kableExtra::kable_styling(font_size = 8,
                            latex_options = c("repeat_header", "HOLD_position"))
# dichotomize covariates
data <- data %>%
  mutate(college = factor(ifelse(pedu >= 4, 1, 0)),
         employed = factor(ifelse(employ == 0, 0, 1)),
         alc_bin = factor(ifelse(nidaalc > 0, 1, 0)),
         pres_bin = factor(ifelse(nidapres > 0, 1, 0)),
         drugs_bin = factor(ifelse(nidaill > 0, 1, 0)))

# Table 3
tbl_summary(data,
            include = c(plang, pwhite, employed, college, alc_bin, pres_bin,
                        drugs_bin, income),
            by = mom_smoke_16wk,
            type = list(plang ~ "dichotomous", pwhite ~ "dichotomous",
                        employed ~ "dichotomous", college ~ "dichotomous",
                        alc_bin ~ "dichotomous", pres_bin ~ "dichotomous",
                        drugs_bin ~ "dichotomous"),
            statistic = list(all_continuous() ~  c("{mean} ({sd})")),
            label = c(plang ~ "Speaks another language at home",
                      pwhite ~ "Identifies as white",
                      employed ~ "Is employed (part-time or full-time)",
```

```
                      college ~ "Has a college degree (2-year or 4-year)",
                      income ~ "Family's estimated annual household income ($)",
                      alc_bin ~ "Has used alcohol in the past 6 months (4 or more drinks per day)",
                      pres_bin ~ "Has used prescription drugs in the past 6 months",
                      drugs_bin ~ "Has used illegal drugs in the past 6 months"),
            missing = "always", missing_text = "Missing") %>%
  modify_header(list(label ~ "Covariate", stat_1 ~ "Nonsmoker", stat_2 ~ "Smoker")) %>%

   as_kable_extra(booktabs = TRUE,
                  caption = "Parent Covariates by Self-Reported 32-Week Smoking Status",
                 longtable = TRUE) %>%
  kableExtra::kable_styling(font_size = 8,
                            latex_options = c("repeat_header", "HOLD_position"))
### SMOKE EXPOSURE VARIABLES

# Reformat smoke exposure variables
smoke_exposure_long <- data %>%
  select(smoke_exposure_6mo:smoke_exposure_5yr) %>%
  pivot_longer(smoke_exposure_6mo:smoke_exposure_5yr, names_to = "time",
               values_to = "smoking") %>%
  mutate_at("time", factor, levels = c("smoke_exposure_6mo", "smoke_exposure_12mo",
                                       "smoke_exposure_2yr", "smoke_exposure_3yr",
                                       "smoke_exposure_4yr", "smoke_exposure_5yr"))

# Figure 2
ggplot(smoke_exposure_long, aes(x = time, fill = factor(smoking,
                                                        levels = c(NA, 0, 1),
                                                        exclude = NULL))) +
  geom_bar(width = 0.8) +
  scale_fill_brewer(palette = "Dark2", na.value = "grey",
                    labels = c("No", "Yes", "No Response")) +
  geom_text(aes(label = after_stat(count)), stat = "count",
            position = position_stack(vjust = 0.5), color = "white", size = 3) +
  labs(x = "", y = "Number of Mothers", fill = "Child Exposed to Smoke?",
       title = "Figure 2. Self-Reported Smoke Exposure") +
  scale_x_discrete(labels = c("0 to 6 months", "7 to 12 months",
                              "1st to 2nd year","2nd to 3rd year",
                              "3rd to 4th year",  "4th to 5th year")) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = -20, hjust = 0))

# Figure 3
ggplot(data) +
  geom_point(aes(x = "x", y = cotimean_pp6mo_baby, color = smoke_exposure_6mo),
             position = position_jitter(w = 0.2), alpha = 0.7, size = 2)  +
  scale_color_brewer(palette = "Dark2", na.value = "grey",
                     labels = c("No", "Yes", "No Response")) +
  theme_minimal() +
  theme(panel.grid.major.x = element_blank(), axis.ticks.x = element_blank(),
        axis.text.x = element_blank()) +
  labs(x = "", y = "Baby's Cotinine Level at 6 Months (ng/mL)",
       color = "Recalled Smoke Exposure at 6 Months",
       title = "Figure 3. Reported Smoke Exposure vs Baby's Cotinine Levels")
```

```r
# Mean cotinine levels
data %>%
  group_by(smoke_exposure_6mo) %>%
  summarize(mean = mean(cotimean_pp6mo_baby, na.rm = TRUE),
            sd = sd(cotimean_pp6mo_baby, na.rm = TRUE))

### OUTCOME VARIABLES
outcomes <- data %>%
  select(parent_id, cig_ever, num_cigs_30, num_e_cigs_30, mj_ever, num_mj_30,
         alc_ever, num_alc_30, bpm_att, bpm_ext, bpm_int, erq_cog, erq_exp,
         pmq_parental_knowledge, pmq_child_disclosure, pmq_parental_solicitation,
         pmq_parental_control, erq_cog_a, erq_exp_a, bpm_att_p, bpm_ext_p,
         bpm_int_p, swan_hyperactive, swan_inattentive)

# Format BPM responses
bpm_child <- data %>%
  select(parent_id, bpm_att, bpm_ext, bpm_int) %>%
  mutate(type = "Child")

bpm_parent <- data %>%
  select(parent_id, bpm_att_p, bpm_ext_p, bpm_int_p) %>%
  mutate(type = "Parent") %>%
  rename(bpm_att = bpm_att_p, bpm_ext = bpm_ext_p, bpm_int = bpm_int_p)

bpm <- bind_rows(bpm_child, bpm_parent)

# Figure 4a
ggplot(bpm) +
  geom_density(aes(x = bpm_att, color = type, fill = type), alpha = 0.5) +
  labs(x = "Attention Problem Score", y = "Density",
       title = "Figure 4a. Attention Problems") +
  theme_minimal() +
  theme(legend.position = "none", text = element_text(size=20))

# Figure 4b
ggplot(bpm) +
  geom_density(aes(x = bpm_int, color = type, fill = type), alpha = 0.5)  +
  labs(x = "Internalizing Problem Score", y = "",
       title = "Figure 4b. Internalizing Problems") +
  theme_minimal() +
  theme(legend.position = "none", text = element_text(size=20))

# Figure 4c
ggplot(bpm) +
  geom_density(aes(x = bpm_ext, color = type, fill = type), alpha = 0.5)  +
  labs(x = "Externalizing Problem Score", y = "",
       title = "Figure 4c. Externalizing Problems", color = "Respondent",
       fill = "Respondent") +
  theme_minimal() +
  theme(text = element_text(size=20))


# Figure 5
```

```r
ggplot(data) +
  geom_point(aes(x = erq_cog, y = erq_exp,
                 color = factor(mom_smoke_16wk, levels = c(NA, 0, 1),
                                exclude = NULL)),
             size = 3, alpha = 0.7,
             position = position_jitter(width = 0.1, height = 0.1, seed = 2)) +
  scale_color_brewer(palette = "Dark2", na.value = "grey",
                     labels = c("No", "Yes", "No Response")) +
  labs(x = "Average Cognitive Reappraisal Response",
       y = "Average Expressive Suppression Response",
       title = "Figure 5. Cognitive Reappraisal vs. Expressive Suppression",
       color = "Mother Smoked at 16 Weeks Pregnant?") +
  theme_minimal()

# get correlation between cognitive reappraisal and expressive suppression
cor(data$erq_cog, data$erq_exp, use = "complete.obs")

# get correlation between child score and the parent's score about themselves
cor(data$erq_cog, data$erq_cog_a, use = "complete.obs")
cor(data$erq_exp, data$erq_exp_a, use = "complete.obs")

# Figure 6
ggplot(data, aes(x = swan_hyperactive, y = swan_inattentive,
                 color = factor(mom_smoke_16wk, levels = c(NA, 0, 1),
                                                        exclude = NULL))) +
  geom_point(size = 3, alpha = 0.7,
             position = position_jitter(width = 0.13, height = 0.1, seed = 15)) +
  labs(x = "Hyperactive Score", y = "Inattentive Score",
       title = "Figure 6. Hyperactive ADHD vs. Inattentive ADHD",
       color = "Mother Smoked at 16 Weeks Pregnant?") +
  scale_color_brewer(palette = "Dark2", na.value = "grey",
                     labels = c("No", "Yes", "No Response")) +
  geom_vline(xintercept = 6, linetype = "dashed", color = "#e28743") +
  geom_hline(yintercept = 6, linetype = "dashed", color = "#e28743") +
  scale_x_continuous(breaks = seq(0,10), minor_breaks = NULL) +
  scale_y_continuous(breaks = seq(0,10), minor_breaks = NULL) +
  theme_minimal()

# get correlation between hyperacctive and inattentive scores
cor(data$swan_hyperactive, data$swan_inattentive, use = "complete.obs")

# create ADHD variable
data <- data %>%
  mutate(adhd = case_when(swan_hyperactive >= 6 | swan_inattentive >= 6 ~ 1,
                          swan_hyperactive < 6 & swan_inattentive < 6 ~ 0))

# get proportion of SDP by ADHD status
data %>%
  group_by(adhd) %>%
  summarize(n = n(), mean = mean(mom_smoke_16wk))

# dichotomize childasd variable
data <- data %>%
```

```r
  mutate(asd = factor(ifelse(childasd >= 1, 1, 0)))

# Table 4
tbl_summary(data,
            include = c(tage, tsex, language, twhite, asd, income),
            by = adhd,
            type = list(tage ~ "continuous",
                        tsex ~ "dichotomous",
                        language ~ "dichotomous",
                        twhite ~ "dichotomous",
                        asd ~ "dichotomous",
                        income ~ "continuous"),
            statistic = list(all_continuous() ~  c("{mean} ({sd})")),
            label = c(language ~ "Speaks another language at home",
                      tage ~ "Age",
                      tsex ~ "Assigned female at birth",
                      twhite ~ "White",
                      asd ~ "Diagnosed or suspected of having Austism Spectrum Disorder",
                      income ~ "Family's estimated annual household income",
            missing = "always", missing_text = "Missing")) %>%
  modify_header(list(label ~ "Covariate", stat_1 ~ "Likely Doesn't Have ADHD",
                     stat_2 ~ "Likely Has ADHD")) %>%
  add_p() %>%

   as_kable_extra(booktabs = TRUE,
                  caption = "Child Covariates by ADHD Status",
                 longtable = TRUE) %>%
  kableExtra::kable_styling(font_size = 8,
                            latex_options = c("repeat_header", "HOLD_position"))

# Table 5
tbl_summary(data,
            include = c(bpm_int, erq_cog, erq_exp),
            by = mom_smoke_16wk,
            label = list(bpm_int ~ "BPM Internalizing Score",
                         erq_cog ~ "ERQ Cognitive Reappraisal",
                         erq_exp ~ "ERQ Expressive Suppression")) %>%
  add_p() %>%
  modify_header(list(stat_1 ~ "Nonsmoker", stat_2 ~ "Smoker")) %>%
   as_kable_extra(booktabs = TRUE,
                  caption = "Internalizing Behaviors by Mother's 16-week Smoking Status",
                 longtable = TRUE) %>%
  kableExtra::kable_styling(font_size = 8,
                            latex_options = c("repeat_header", "HOLD_position"))

# create substance use variable
data <- data %>%
  mutate(su_ever = case_when(cig_ever == 1 | e_cig_ever == 1 | mj_ever == 1 |
                               alc_ever == 1 ~ 1,
                             cig_ever == 0 & e_cig_ever == 0 & mj_ever == 0 &
                               alc_ever == 0 ~ 0))

# Table 6
```

```r
tbl_summary(data,
            include = c(su_ever, bpm_ext, bpm_att, swan_hyperactive,
                        swan_inattentive),
            by = mom_smoke_16wk,
            type = list(bpm_ext ~ "continuous",
                        bpm_att ~ "continuous"),
            label = list(su_ever ~ "Ever tried cigarettes, e-cigarettes, alcohol, or marijuana",
                         bpm_ext ~ "BPM Externalizing Score",
                         bpm_att ~ "BPM Attention Score",
                         swan_hyperactive ~ "SWAN Hyperactive Score",
                         swan_inattentive ~ "SWAN Inattentive Score")) %>%
  add_p() %>%
  modify_header(list(stat_1 ~ "Nonsmoker", stat_2 ~ "Smoker")) %>%
   as_kable_extra(booktabs = TRUE,
                  caption = "Externalizing Behaviors by Mother's 16-week Smoking Status",
                 longtable = TRUE) %>%
  kableExtra::kable_styling(font_size = 8,
                            latex_options = c("repeat_header", "HOLD_position"))
```