PHP2550_Project1

Jialin Liu

2023-10-02

Abstract

This work presents an exploratory data analysis of smoking during pregnancy and postpartum periods and examines effects of smoking prenatally and postnatally on adolescent's externalizing behaviors, substance use and self-regulation.The relationship between responses on SWAN Rating Scale, including ADHA- Hyperactive/Impulsive type and ADHD- Inattentive type, and smoking prenatally and postnatally variables show significantly differences between two groups and indicates adolescents are more likely to have ADHD-Hyperactive/Impulsive externalizing problems if their mothers smoked during pregnancy. Also, postnatal smoking exposure is more likely to adversely impact adolescent's externalizing behaviors especially for ADHD. Limitations are also examined to give insights into potential drawbacks of conclusions.

Introduction

Maternal smoking during pregnancy (MSDP) as a major public health concern contributes to a variety of adverse outcomes for both mothers and unborn children present at birth, such as low birth weights and breathing difficulties, as well as long lasting behavioral and neuro-related (cognitive and brain development) impairment throughout life. If mothers smoke while pregnant their newborns are at increased risk of a wide range of externalizing and self-regulatory problems, including Attention-Deficit/Hyperactivity Disorder (ADHD), substance use, poor executive function, less emotional control and so on. To examine associations between self-regulatory mechanisms and externalizing behaviors, some findings suggest that self-regulatory problems, especially for behavioral inhibitory control and emotion regulation, are particularly important in distinguishing between children who show externalizing behaviors across early childhood after exposure from MSDP. Therefore, deficits in the early acquisitions of self-regulatory skills impact externally behavioral problems, substance use and subsequent progression.

The original study developed a randomized controlled trail and aimed to test the efficacy of tailored video as an intervention versus usual care approaches to reduce smoking and environmental tobacco smoke (ETS) exposure during and after pregnancy among low-income women (N = 738) who were pregnant with only one baby. In control conditions, participants received newsletters with content of smoking cessation and behavioral theory-based survey questions. Participants in treatment group receive both newsletters and videos on healthy pregnancy topics. One measurement of outcome was salivary cotinine of both mother and baby (32 weeks gestation and 6 months postpartum) to evaluate if this low cost intervention could reduce environmental tobacco exposure, as well as self-reported behaviors as secondary outcomes, and to describe the utility of this innovative intervention technology with potential benefits to this type of intervention. Based on the strength of the previous study, our focus of the current study is aiming to examine a random subset of 100 mothers and their children who are youth aging between 12 and 16 years old now with respect to self-regulation, substance use and externalizing behavioral problems from parent- and child-report.

Exploring Univariates and Bivariate Relationships and Checking Missing Values

We randomly recruited 100 mothers and their children from the original study on smoke avoidance intervention to reduce smoking and low-income women's ETS exposure during pregnancy and their children's ETS exposure in the immediate postpartum period. The current version of data comes from two processed data sets for parent and child-reports by performing the inner join with respect to parent's identification number on two data frames in which some variables were dropped or filtered observations by baseline to simplify analysis so that this is a useful data set to illustrate exploratory data analysis (EDA) we advocate. The data consists of 49 participants who were random samples of the original study and records two dimensional assessments: adolescents and their mothers. The rest of this report is devoted to describing briefly the statistics and exploring relationships between variables as well as missing values and finally answering our research goal about adverse effects of prenatal exposure to smoking in youth in terms of three assessments: self-regulation, substance use and externalizing behaviors.

The data set gives information regarding individual characteristics such as age, sex, language, ethnicity, race for both adolescents and their mothers. It also contains information about parents regarding employment status, income level, and educational background. Information on past substance use including cigarettes, e-cigarettes (child-report only), marijuana (child-report only), alcohol, prescription and illegal drugs for non-medical reasons (parent-report only) has been recorded from parent- and child-report based on current study's survey questions. In addition, information about self-reported smoking status while pregnant and postpartum at several time points (for example, 16, 22, and 32 weeks pregnant; 12 weeks, 6 months postpartum) is provided, and urine cotinine levels in gestation and 6 months postpartum from mom and baby are also recorded from the previous study. Other than pre-collected data, our current study also contains tracing back questions which ask participants to recall their smoking status at 6 months, 12 months, 2 years to 5 years in the postnatal period and names these variables starting with `smoke_exposure`.

The three main outcome variables of interest in our data set are substance use, externalizing behaviors, and self-regulation. In terms of self-regulation, responses are measured on the Emotional Regulation Questionnaire (ERQ) related to cognitive reappraisal and expressive suppression (`erq_cog` and `erq_exp` measured from adolescents; `erq_cog_a` and `erq_exp_a` from parental side). In the context of the ERQ, a higher score on the cognitive reappraisal questions represents that an individual has a strong ability of emotion adjustment strategy since cognitive reappraisal is considered as a healthier and more positive strategy to alter someone's negative thoughts and reduce adverse impacts of emotional instability. Conversely, a higher score on the expressive suppression questions potentially indicates that an individual is more likely to inhibit negative expression and mask emotional behaviors, which is often associated with worse emotionally regulatory behaviors. Thus, it is important to better understand those underlying meanings of scores in the context of the ERQ and then examine effects of early exposure to smoking in youth. Based on behavioral results by child- and parent-report, substance use in youth could be identified four general categories: smoking cigarettes, e-cigarettes, marijuana, and alcohol, similarly in adults, smoking cigarettes falls into one type of substance use of our interests. Then, information on externalizing behaviors is captured from multiple sides including child's self-reported scores on self, parent's reported scores on child, and parent's reported scores on self. The externalizing issues are measured by the Brief Problem Monitor (BPM) on items related to attention (`bpm_att`) and externalizing (`bpm_ext`) problems, respectively. Since ADHD belongs to externalizing behaviors, responses on SWAN Rating Scale with scores of 6 or greater are also recorded as main outcomes of child's attention problems (`swan_hyperactive` and `swan_inattentive`).

Before looking for descriptive statistics, we process categorical variables by transforming them to statistical form, for example, answer of smoking status is "1 = Yes" converted to a factorized level "1", and number of cigarettes with "None" are converted to "0", similarly, answer with "20-25" is selected in middle "23". Other than transforming informal answers to statistically meaningful values, we are aware of missing values in follow-up questions about substance use. Regarding the question about ever used cigarettes in the past 30 days or not, the follow-up question about exact number of cigarettes or days of substance use should be consist with previous question. For example, the number of cigarettes or days are provided only if the answer about ever used cigarettes is yes. Following this logic, we will fill missing values in those follow-up

questions to be 0 if children and parents ever used substance in the past 30 days.

We firstly analyze descriptive statistics with respect to variables for parents (starting with **p**) and adolescents (starting with **t**): age, sex (0 = Male, 1 = Female), ethnicity (0 = Hispanic or Latino, 1 = No), race(`aian` = American Indian/Alaska Native, `nhpi` = Native Hawaiian or Pacific Islander), employment (0=No, 1 = Part-Time, 2 = Full-Time), and income. These variables are traditionally giving us insights into demo-graphical information such as diversity or aging of samples. Note, however, at least 8 missing values for those variables without complete cases for all participants are observed. Surprisingly, if we take a detailed look at the data set, it is too obvious to find a missing pattern, which is mainly formed by eight participants who did not answer all other personal characteristics related variables and left them as 0, besides that one participant answered child's age and race. As such, other missing values inspection of smoking-related variables with respect to those eight participants is necessary.

Table 1: Participants Characteristics

Characteristic	**N**	**N = 49**[1]
page	41 / 49	
Mean (Minimum,Maximum, SD)		37.5 (32.0,45.0, 3.6)
N missing (% missing)		8 (16%)
psex	41 / 49	
0		1 / 41 (2.4%)
1		40 / 41 (98%)
pethnic	41 / 49	
0		28 / 41 (68%)
1		13 / 41 (32%)
paian	49 / 49	
0		45 / 49 (92%)
1		4 / 49 (8.2%)
pnhpi	49 / 49	
0		41 / 49 (84%)
1		8 / 49 (16%)
pblack	49 / 49	
0		49 / 49 (100%)
pwhite	49 / 49	
0		23 / 49 (47%)
1		26 / 49 (53%)
prace_other	49 / 49	
0		43 / 49 (88%)
1		6 / 49 (12%)
employ	41 / 49	

[1]n / N (%)

Table 1: Participants Characteristics

Characteristic	**N**	**N = 49**[1]
0		12 / 41 (29%)
1		7 / 41 (17%)
2		22 / 41 (54%)
income	36 / 49	
Mean (Minimum,Maximum, SD)		57,947 (760,265,000, 51,607)
N missing (% missing)		13 (27%)
tage	37 / 49	
12		8 / 37 (22%)
13		10 / 37 (27%)
14		9 / 37 (24%)
15		8 / 37 (22%)
16		2 / 37 (5.4%)
tsex	36 / 49	
0		23 / 36 (64%)
1		13 / 36 (36%)
tethnic	37 / 49	
0		21 / 37 (57%)
1		15 / 37 (41%)
2		1 / 37 (2.7%)
taian	49 / 49	
0		44 / 49 (90%)
1		5 / 49 (10%)
tnhpi	49 / 49	
0		49 / 49 (100%)
tblack	49 / 49	
0		34 / 49 (69%)
1		15 / 49 (31%)
twhite	49 / 49	
0		30 / 49 (61%)
1		19 / 49 (39%)
trace_other	49 / 49	
0		44 / 49 (90%)
1		5 / 49 (10%)

[1]n / N (%)

Table 1: Participants Characteristics

Characteristic	**N**	**N = 49**[1]

[1]n / N (%)

We extract information about number of cigarettes per day ever used in the past 30 days by mother `mom_numcig` and number of days adolescents ever sued cigarette, e-cigarette, marijuana, and alcohol, splitting adolescents by SDP (0 = No smoking during pregnancy, 1 = smoking during pregnancy). Table 2 reports descriptive statistics of the number of cigarettes smoked by mother per day and indicates an abnormal outlier (44,989 cigarettes) which is an extremely unreliable data and highly inflate the mean. Without that unreasonable outlier, mother smoked 2.5 cigarettes on average per day in the past 30 days. Among 10 missing values of this variable, 8 of them are the same group of participants who did not answer individual characteristics related questions as well, and more 2 parents' records are lost.

Table 2: How many cigarettes per day do mothers smoke in the past 30 days?

Characteristic	**N**	**N = 49**
mom_numcig	39 / 49	
Mean (Range, SD)		1,156.0 (0.0, 44,989.0, 7,203.6)
N missing (% missing)		10 (20%)

Table 3 depicts summary statistics of adolescents' substance use given by mother's smoking status while pregnant. We can find that whatever mothers were smoking while pregnant substance use among those adolescents are not really severe since most of them did not ever smoke, vape, use marijuana, and drink alcohol in the past 30 days. In other words, even though their mothers smoked during pregnancy, only 10% or less of them have substance use behaviors, which may help us explore the effects of SDP in youth with respect to substance use.

Table 3: How many of the past 30 days do adolescents smoke/vape/use marijuana/drink alcohol? (Stratified by SDP)

Characteristic	**N**	**0**, N = 24[1]	**1**, N = 14[1]
num_cigs_30	29 / 38	0 / 18 (0%)	0 / 11 (0%)
num_e_cigs_30	28 / 38		
0		18 / 18 (100%)	9 / 10 (90%)
2		0 / 18 (0%)	1 / 10 (10%)
num_mj_30	29 / 38		
0		17 / 18 (94%)	9 / 11 (82%)
3		0 / 18 (0%)	1 / 11 (9.1%)
12		0 / 18 (0%)	1 / 11 (9.1%)
18		1 / 18 (5.6%)	0 / 11 (0%)

[1]n / N (%)

Table 3: How many of the past 30 days do adolescents smoke/vape/use marijuana/drink alcohol? (Stratified by SDP)

Characteristic	**N**	**0**, N = 24[1]	**1**, N = 14[1]
num_alc_30	27 / 38		
0		17 / 17 (100%)	8 / 10 (80%)
1		0 / 17 (0%)	1 / 10 (10%)
10		0 / 17 (0%)	1 / 10 (10%)

[1]n / N (%)

The below Table 4 represents the relationship between cognitive appraisal subscale and expressive suppression scores on the ERQ and SDP variables. Two assessments with respect to youth's self-regulation do not show significantly differences between SDP group and non-SDP group since we notice the p-values are all larger than 0.05 and conclude that two groups have the same distribution with the same median in terms of cognitive appraisal and expressive suppression scores. However, we can also notice that the mean cognitive reappraisal score in SDP group is higher than the corresponding mean value in non-SDP group, which is a counter-intuitive observation since higher scores in cognitive appraisal indicates individuals are more likely to healthier adjust emotional instability. As for expressive suppression, the average score (3.14) in SDP group is higher than the average score (2.54) in non SDP group, which indicates that adolescents have a greater tendency to inhibit outward expression if their mother smoked during pregnancy.

Table 4: Adolescent's Average response on the Emotion Regulation Questionairre related to Cognitive Reappraisal (Stratified by SDP)

Characteristic	**0**, N = 24	**1**, N = 14	**p-value**[1]
erq_cog			0.4
Mean (Minimum,Maximum,Median)	2.97 (1.00,5.00,3.00)	3.45 (2.83,4.83,3.00)	
N missing (% missing)	6 (25%)	4 (29%)	
erq_exp			0.084
Mean (Minimum,Maximum,Median)	2.54 (1.25,4.00,2.50)	3.14 (2.25,4.50,3.25)	
N missing (% missing)	7 (29%)	3 (21%)	

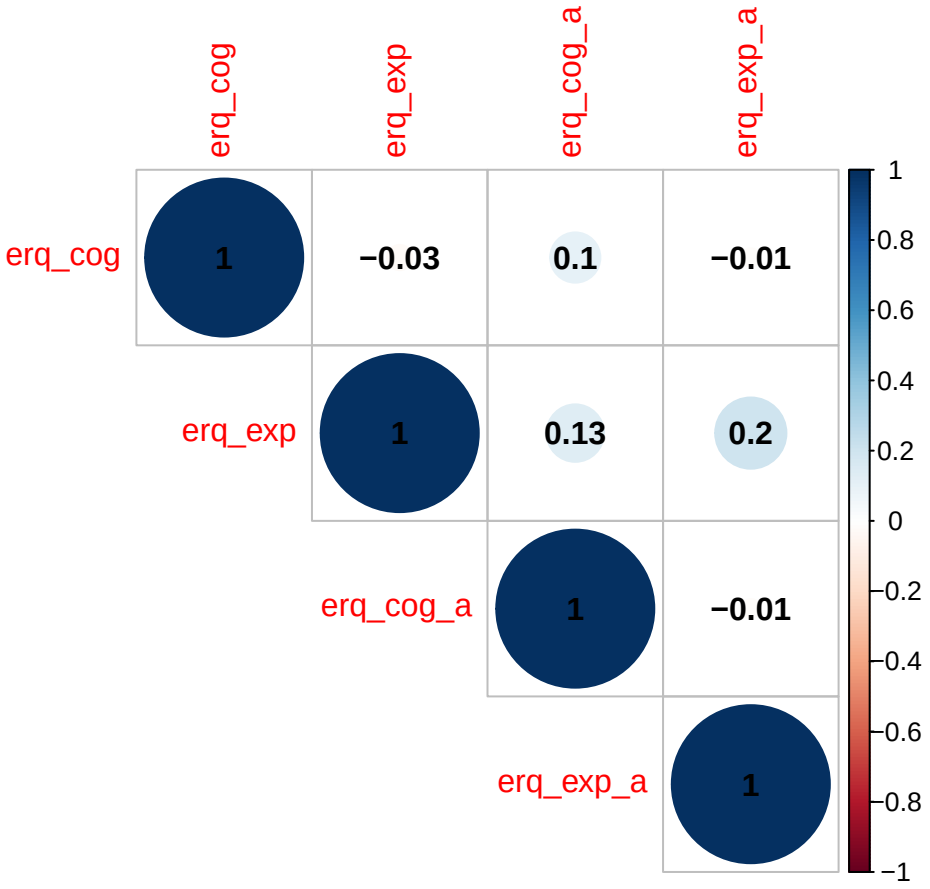
[1]Wilcoxon rank sum test

Table 5 shows the relationship between cognitive appraisal subscale and expressive suppression scores on the ERQ and postpartum smoking exposure variable. ETS consists with two parts of exposure: prenatal (SDP) and postnatal exposure. Here, two scores related to youth's self-regulation do not show significantly differences between postnatal exposure group and non-postnatal exposure group since we can see that two p-values are greater than 0.05 and conclude that two groups have the same distribution with the same median in terms of cognitive appraisal and expressive suppression scores. Similar to observation in Table 4, we can also notice that the mean cognitive reappraisal score in postnatal exposure group is slightly higher than the corresponding mean value in non-postnatal exposure group. The average score of expressive suppression (2.91) in postnatal exposure group is higher than the average score (2.50) in non-postnatal exposure group, which indicates that adolescents are more likely to mask their outward expression under ETS exposure.

Table 5: Adolescent's Average response on the Emotion Regulation Questionairre related to Cognitive Reappraisal (Stratified by Postpartum Smoking Exposure)

Characteristic	**0**, N = 12	**1**, N = 30	**p-value**[1]
erq_cog			0.8
Mean (Minimum,Maximum,Median)	3.06 (1.17,4.17,3.25)	3.26 (1.00,5.00,3.00)	
N missing (% missing)	0 (0%)	7 (23%)	
erq_exp			0.2
Mean (Minimum,Maximum,Median)	2.50 (1.25,3.75,2.50)	2.91 (1.25,4.50,2.75)	
N missing (% missing)	0 (0%)	7 (23%)	

[1]Wilcoxon rank sum test



To explore **interrelations among self-regulation variables**, we use heat maps style to visualize correlations. The above figure showing correlations between adolescents' and parents' average responses on the ERQ related to cognitive reappraisal and expressive suppression. Positive correlations are represented in blue, negative correlations in red. Since the correlation coefficient between `erq_cog` and `erq_cog_a` is only

0.1 and 0.13 as for expressive suppression scale, we can realize that self-regulation variables are not found to be significantly correlated between parent- and child-levels.

The below Table 6 represents the relationship between responses on SWAN Rating Scale, including ADHA-Hyperactive/Impulsive type and ADHD- Inattentive type, and SDP variables. We notice `swan_hyperactive` shows a significantly difference between SDP group and non-SDP group as the p-value (0.025) is greatly less than 0.05, which give us insights into hyperactive/impulsive type of poor attention behaviors for adolescents whose mothers were smoking while pregnant. We can also see that the mean scale with respect to hyperactive/impulsive type in SDP group (11.8) is almost as twice as the corresponding mean value in non-SDP group (6.3), which is a meaningful observation saying adolescents are more likely to have ADHD-Hyperactive/Impulsive externalizing problems if their mothers smoked during pregnancy compared to those youths in non-SDP group. As for ADHD- Inattentive type, it does not show significantly different between two groups about SDP so that we can conclude that two groups have the same distribution with the same median in terms of inattentive externalizing behaviors.

Table 6: Responses on SWAN Rating Scale (Stratified by SDP)

Characteristic	**0**, N = 24	**1**, N = 14	**p-value**[1]
swan_hyperactive			0.025
Mean (Minimum,Maximum,SD)	6.3 (0.0,19.0,5.8)	11.8 (2.0,20.0,6.5)	
N missing (% missing)	5 (21%)	3 (21%)	
swan_inattentive			0.4
Mean (Minimum,Maximum,SD)	11.5 (4.0,22.0,4.8)	12.6 (1.0,19.0,5.1)	
N missing (% missing)	5 (21%)	3 (21%)	

[1]Wilcoxon rank sum test

Table 7 evaluates the relationship between responses on SWAN Rating Scale, including ADHA- Hyperactive/Impulsive type and ADHD- Inattentive type, and postpartum smoking exposure variable. Here, two assessments related to ADHD scales show significantly differences between postnatal exposure group and non-postnatal exposure group because two p-values are smaller than 0.05 and conclude that two groups have different distribution with various medians in terms of hyperactive (p-value = 0.044) and inattentive types (p-value = 0.004). Based on the mean aspect, the average hyperactive/impulsive subscale in postnatal exposure group (9.6) is much higher than the corresponding mean value in non-postnatal exposure group (5.1) as well as the average inattentive scale (12.8) in postnatal exposure group versus mean scale in non-postnatal exposure group (8.2), which indicates that postnatal smoking exposure is more likely to adversely impact adolescent's externalizing behaviors especially for ADHD.

Table 7: Responses on SWAN Rating Scale (Stratified by Postpartum Smoking Exposure)

Characteristic	**0**, N = 12	**1**, N = 30	**p-value**[1]
swan_hyperactive			0.044
Mean (Minimum,Maximum,SD)	5.1 (0.0,19.0,5.6)	9.6 (0.0,20.0,6.5)	
N missing (% missing)	0 (0%)	6 (20%)	

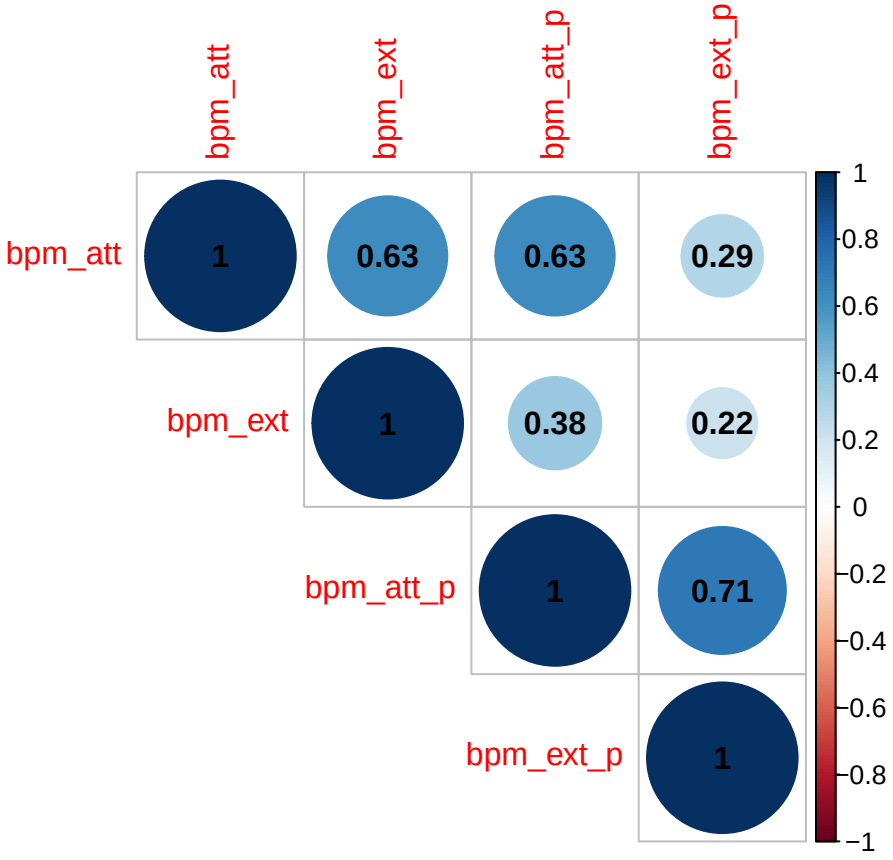
[1]Wilcoxon rank sum test

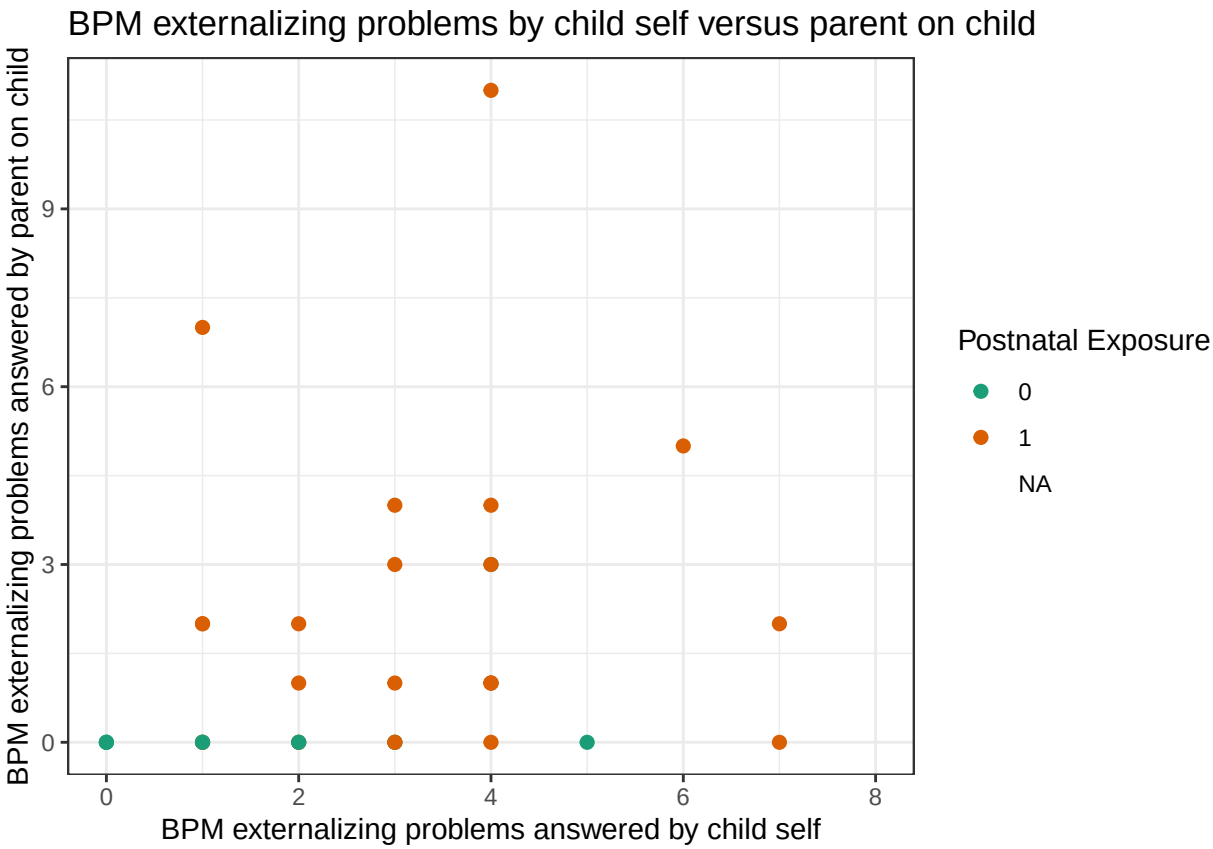
Table 7: Responses on SWAN Rating Scale (Stratified by Postpartum Smoking Exposure)

Characteristic	**0**, N = 12	**1**, N = 30	**p-value**[1]
swan_inattentive			0.004
Mean (Minimum,Maximum,SD)	8.2 (1.0,22.0,5.1)	12.8 (1.0,21.0,4.9)	
N missing (% missing)	0 (0%)	6 (20%)	

[1]Wilcoxon rank sum test

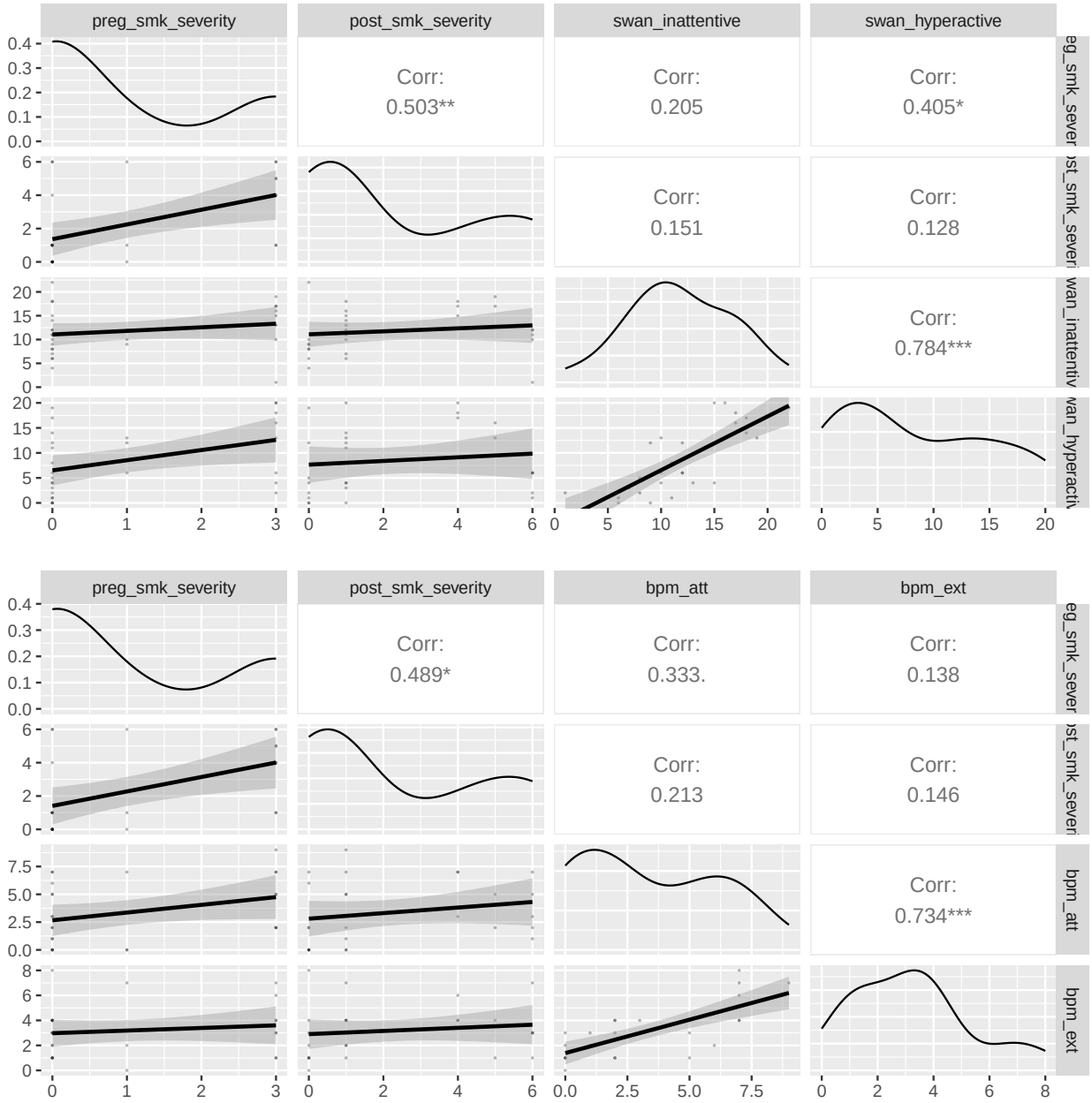
Another assessments on externalizing behaviors are measured by the Brief Problem Monitor (BPM) related to attention and externalizing problems reported by child self, parent self, and parent's child respectively. Before analyzing any relationships between this type of assessment and SDP/postnatal smoking exposure, we tend to examine interrelations among BPM variables, particularly for parents' responses on child and children's responses on self. The below figure about correlation indicates a strong correlation between `bpm_att` and `bpm_att_p` ($\rho = 0.63$), which verifies a consistency about attention problems evaluated by child self and parents observation to some extent.





Particularly, we're interested in exploring the trend of BPM externalizing responses by child self versus parent on child under postnatal smoking exposure groups. Interestingly, in postnatal exposure group, adolescents have a greater tendency to self-report externalizing problems with higher scores, however, their mothers seem not to realize their babies attention problems with such low scores. Based on this observation, we tend to create table summary with respect to child's self-reported externalizing behaviors under SDP and postnatal smoking exposure groups separately. By using fisher exact test, p-values are not significantly smaller than 0.05, therefore, we do not have sufficient evidence to reject the null hypothesis and conclude that there is a strong association between BPM externalizing and attention problems reported by children with SDP and postnatal smoking exposure, respectively.

To examine **interrelatedness of prenatal and postnatal exposure**, we constructed a new variable named `post_smk_severity` which cumulative adds smoking status up at several postpartum time points (6 months, 12 months, 2 year to 5 year). Similarly, another new variable named `preg_smk_severity` is calculated by cumulative sum of smoking status at three pregnant time periods (16, 22 and 32 weeks). This prenatal exposure severity variable is informed by prospective maternal report of SDP and ETS throughout the prenatal period, and the postnatal exposure severity variable is informed by infant and mother cotinine and maternal smoking report of across postpartum assessments. After calculating the correlation between `post_smk_severity` and `preg_smk_severity`, $\rho = 0.5$ suggests a weak correlation between prenatal and postnatal exposure variables. The below figure represents the bivariate relationships between prenatal and postnatal severity exposure, and BPM externalizing and attention problems answered by child self. We find that higher levels of severity during pregnancy are associated with higher attention problems, and similarly, higher levels of severity durting postnatal periods are often related to higher values of exteralizing problems on BPM scale.



Limitations

There are several limitations of this study. Firstly, the sample size is quite small so that we cannot advocate more advanced EDA tools due to limitation of observed cases and records. Secondly, there are a lot of unobserved confounders which may play an important role in explaining effects of SDP/EST in youth and impacting our current conclusions. Thirdly, missing values presented in this dataset are easily to track since we've found particular eight participants did not complete those survey questions rather than leaving most of questions as either 0 or NA, which also reduce the amount of useful observational outcomes and have an adverse impact on our analysis. Finally, a lot of inconsistency issues have been realized, such as some participants answered no cigarette used in the past 30 days, however, at the follow-up question, they answered 5 or 3 cigarettes per day in past.

Code Appendix:

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)
library(tidyverse)
library(dplyr)
library(ggplot2)
library(hrbrthemes)
library(kableExtra)
library(gridExtra)
library(gtsummary)
library(mice)
library(flextable)
library(corrplot)
library(psych)
library(GGally)
#=== Read the data file
TobExp_df <- read.csv("project1.csv", na.strings=c("","NA"))
#=== Preprocessing variables
##=== Converting descriptive answers to factor levels
TobExp_df$mom_smoke_16wk[TobExp_df$mom_smoke_16wk=="1=Yes"] <- "1"
TobExp_df$mom_smoke_22wk[TobExp_df$mom_smoke_22wk=="1=Yes"] <- "1"
TobExp_df$mom_smoke_32wk[TobExp_df$mom_smoke_32wk=="1=Yes"] <- "1"
TobExp_df$mom_smoke_pp1[TobExp_df$mom_smoke_pp1=="1=Yes"] <- "1"
TobExp_df$mom_smoke_pp2[TobExp_df$mom_smoke_pp2=="1=Yes"] <- "1"
TobExp_df$mom_smoke_pp12wk[TobExp_df$mom_smoke_pp12wk=="1=Yes"] <- "1"
TobExp_df$mom_smoke_pp6mo[TobExp_df$mom_smoke_pp6mo=="1=Yes"] <- "1"

TobExp_df$mom_smoke_16wk[TobExp_df$mom_smoke_16wk=="2=No"] <- "0"
TobExp_df$mom_smoke_22wk[TobExp_df$mom_smoke_22wk=="2=No"] <- "0"
TobExp_df$mom_smoke_32wk[TobExp_df$mom_smoke_32wk=="2=No"] <- "0"
TobExp_df$mom_smoke_pp1[TobExp_df$mom_smoke_pp1=="2=No"] <- "0"
TobExp_df$mom_smoke_pp2[TobExp_df$mom_smoke_pp2=="2=No"] <- "0"
TobExp_df$mom_smoke_pp12wk[TobExp_df$mom_smoke_pp12wk=="2=No"] <- "0"
TobExp_df$mom_smoke_pp6mo[TobExp_df$mom_smoke_pp6mo=="2=No"] <- "0"

##=== Create a new column about smoking status during pregnancy
TobExp_df$mom_smoke_preg <- ifelse(TobExp_df$mom_smoke_16wk=="1" | TobExp_df$mom_smoke_22wk == "1" |Tob
TobExp_df$mom_smoke_preg <- as.factor(TobExp_df$mom_smoke_preg)

##=== Preprocessing mom_numcig column with error cells
TobExp_df$mom_numcig[TobExp_df$mom_numcig == 'None'] <- '0'
TobExp_df$mom_numcig[TobExp_df$mom_numcig == '20-25'] <- '23'
TobExp_df$mom_numcig[TobExp_df$mom_numcig == '2 black and miles a day'] <- '2'

##=== Swan_inattentive and swan_hyperactive values should be NA instead of 0
TobExp_df$swan_inattentive[TobExp_df$parent_id %in% c("50502", "51202", "51602", "52302", "53002", "535
TobExp_df$swan_hyperactive[TobExp_df$parent_id %in% c("50502", "51202", "51602", "52302", "53002", "535
##=== Create a new column about pregnant smoking severity
subset_preg_expo <- subset(TobExp_df[, c("mom_smoke_16wk", "mom_smoke_22wk", "mom_smoke_32wk", "mom_smo
subset_preg_expo <- subset_preg_expo %>% mutate_if(is.character, as.numeric)
subset_preg_expo$preg_smk_severity <- subset_preg_expo$mom_smoke_16wk + subset_preg_expo$mom_smoke_22wk
##=== Create a new column about postpartum smoking severity if any of original study or recalled study
```

```r
subset_post_expo <- subset(TobExp_df[, c("mom_smoke_pp1", "mom_smoke_pp2", "mom_smoke_pp12wk", "mom_smo

## Overlapping column at 6 month
# subset_post_expo$mom_smoke_expo_6mo <- ifelse(subset_post_expo$mom_smoke_pp6mo=="1" | subset_post_exp
## Column about exposure before 6 months
# subset_post_expo$mom_smoke_expo_0to6 <- ifelse(subset_post_expo$mom_smoke_pp1=="1" | subset_post_expo
# TobExp_df[, c("mom_smoke_pp6mo", "smoke_exposure_6mo", "mom_smoke_pp1", "mom_smoke_pp2", "mom_smoke_p
subset_post_expo$smoke_expo_6mo_new <- ifelse(subset_post_expo$mom_smoke_pp1=="1" |
                                              subset_post_expo$mom_smoke_pp2=="1" |
                                              subset_post_expo$mom_smoke_pp12wk=="1" |
                                              subset_post_expo$mom_smoke_pp6mo=="1" |
                                              subset_post_expo$cotimean_pp6mo_baby > 10 |
                                              subset_post_expo$cotimean_pp6mo > 10, "1", "0")

subset_post_expo <- subset_post_expo %>% mutate(across(smoke_expo_6mo_new, ~ replace(., is.na(.) & subs
                                              subset_post_expo$smoke_exposure_6mo=="0" &
                                              subset_post_expo$cotimean_pp6mo < 10 &
                                              subset_post_expo$cotimean_pp6mo_baby < 10, 0)))

subset_post_expo <- subset_post_expo %>% mutate(across(smoke_expo_6mo_new, ~ replace(., is.na(.) & subs

# subset_post_expo[, c("mom_smoke_pp6mo", "smoke_exposure_6mo", "mom_smoke_pp1", "mom_smoke_pp2", "mom_
subset_post_expo <- subset_post_expo %>% dplyr::select(-c("mom_smoke_pp1", "mom_smoke_pp2", "mom_smoke_p

subset_post_expo$post_smk_severity <- subset_post_expo$smoke_expo_6mo_new +
subset_post_expo$smoke_exposure_12mo +
 subset_post_expo$smoke_exposure_2yr +
 subset_post_expo$smoke_exposure_3yr +
 subset_post_expo$smoke_exposure_4yr +
subset_post_expo$smoke_exposure_5yr

TobExp_df$smoke_expo_6mo_new <- subset_post_expo$smoke_expo_6mo_new
TobExp_df$post_smk_severity <- subset_post_expo$post_smk_severity
TobExp_df$preg_smk_severity <- subset_preg_expo$preg_smk_severity
##=== Factorizing all categorical variables
factorCols <- c("psex","plang","pethnic","paian","pasian","pnhpi","pblack","pwhite","prace_other","empl
TobExp_df[,factorCols] <- lapply(TobExp_df[,factorCols], as.factor)

##=== Converting to numeric values for remaining variables
TobExp_df[,!(names(TobExp_df) %in% factorCols)] <- lapply(TobExp_df[,!(names(TobExp_df) %in% factorCols

##=== Create a new column about smoking status postpartum
TobExp_df$mom_smoke_post <- ifelse(TobExp_df$smoke_expo_6mo_new =="1" |
                                   TobExp_df$smoke_exposure_12mo == "1" |
                                   TobExp_df$smoke_exposure_2yr=="1" |
                                   TobExp_df$smoke_exposure_3yr == "1" |
                                   TobExp_df$smoke_exposure_4yr == "1" |
                                   TobExp_df$smoke_exposure_5yr == "1", "1", "0")
TobExp_df$mom_smoke_post <- as.factor(TobExp_df$mom_smoke_post)
##=== Filling missing values in follow-up questions
##=== If answer is no, the number of substance use should be 0 instead of A
childCol <- c("tage", "tsex", "language", "tethnic", "taian", "tasian", "tnhpi", "tblack", "twhite", "t
```

```r
TobExp_df$num_cigs_30[TobExp_df$cig_ever=="0"] <- "0"
TobExp_df$num_e_cigs_30[TobExp_df$e_cig_ever=="0"] <- "0"
TobExp_df$num_mj_30[TobExp_df$mj_ever=="0"] <- "0"
TobExp_df$num_alc_30[TobExp_df$alc_ever=="0"] <- "0"
TobExp_df$mom_numcig[TobExp_df$momcig=="0"] <- "0"

TobExp_df <- TobExp_df %>% mutate_if(is.character, as.numeric)
TobExp_df %>%
  dplyr::select(c('page', 'psex','pethnic', 'paian', 'pnhpi', 'pblack', 'pwhite', 'prace_other', 'emplo
  tbl_summary(
    missing = "no",
    statistic = list(
    all_continuous() ~ c("{mean} ({min},{max}, {sd})",
                         "{N_miss} ({p_miss}%)"),
    all_categorical() ~ "{n} / {N} ({p}%)"
    ),
    type = all_continuous() ~ "continuous2"
    ) %>%
  add_n(statistic = "{n} / {N}") %>%
  as_flex_table() %>%
  flextable::set_caption(caption="Participants Characteristics")
  # as_gt() %>%
  # gt:::as.tags.gt_tbl()
TobExp_df %>%
  dplyr::select(c('mom_numcig')) %>%
  tbl_summary(
    missing = "no",
    statistic = list(
    all_continuous() ~ c("{mean} ({min}, {max}, {sd})",
                         "{N_miss} ({p_miss}%)"),
    all_categorical() ~ "{n} / {N} ({p}%)"
    ),
    type = all_continuous() ~ "continuous2"
    ) %>%
  add_n(statistic = "{n} / {N}") %>%
  as_flex_table() %>%
  flextable::set_caption(caption="How many cigarettes per day do mothers smoke in the past 30 days?")
TobExp_df %>%
  dplyr::select(c('num_cigs_30', 'num_e_cigs_30', 'num_mj_30', 'num_alc_30', 'mom_smoke_preg')) %>%
  tbl_summary(
    by = mom_smoke_preg,
    missing = "no",
    statistic = list(
    all_continuous() ~ c("{mean} ({min},{max}, {sd})",
                         "{N_miss} ({p_miss}%)"),
    all_categorical() ~ "{n} / {N} ({p}%)"
    ),
    type = all_continuous() ~ "continuous2"
    ) %>%
  add_n(statistic = "{n} / {N}") %>%
  as_flex_table() %>%
  flextable::set_caption(caption="How many of the past 30 days do adolescents smoke/vape/use marijuana/
TobExp_df %>%
```

```
  dplyr::select(c('erq_cog', 'erq_exp',  'mom_smoke_preg')) %>%
  tbl_summary(
    by = mom_smoke_preg,
    missing = "no",
    statistic = list(
    all_continuous() ~ c("{mean} ({min},{max},{median})",
                         "{N_miss} ({p_miss}%)"),
    all_categorical() ~ "{n} / {N} ({p}%)"
    ),
    type = all_continuous() ~ "continuous2"
    ) %>%
  add_p() %>%
  as_flex_table() %>%
  flextable::set_caption(caption="Adolescent's Average response on the Emotion Regulation Questionairre
TobExp_df %>%
  dplyr::select(c('erq_cog', 'erq_exp',  'mom_smoke_post')) %>%
  tbl_summary(
    by = mom_smoke_post,
    missing = "no",
    statistic = list(
    all_continuous() ~ c("{mean} ({min},{max},{median})",
                         "{N_miss} ({p_miss}%)"),
    all_categorical() ~ "{n} / {N} ({p}%)"
    ),
    type = all_continuous() ~ "continuous2"
    ) %>%
  add_p() %>%
  as_flex_table() %>%
  flextable::set_caption(caption="Adolescent's Average response on the Emotion Regulation Questionairre
##=== Display a correlation plot using the corrplot package
# First calculate correlation coefficients to be visualised
cor_matrix = cor(TobExp_df[ ,c("erq_cog", "erq_exp", "erq_cog_a", "erq_exp_a")], method='pearson',use='

# Now produce the plot
corrplot(cor_matrix, method='circle', type='upper', addCoef.col = "black")

TobExp_df %>%
  dplyr::select(c('swan_hyperactive', 'swan_inattentive',  'mom_smoke_preg')) %>%
  tbl_summary(
    by = mom_smoke_preg,
    missing = "no",
    statistic = list(
    all_continuous() ~ c("{mean} ({min},{max},{sd})",
                         "{N_miss} ({p_miss}%)"),
    all_categorical() ~ "{n} / {N} ({p}%)"
    ),
    type = all_continuous() ~ "continuous2"
    ) %>%
  add_p() %>%
  as_flex_table() %>%
  flextable::set_caption(caption="Responses on SWAN Rating Scale (Stratified by SDP)")
TobExp_df %>%
  dplyr::select(c('swan_hyperactive', 'swan_inattentive',  'mom_smoke_post')) %>%
```

```
  tbl_summary(
    by = mom_smoke_post,
    missing = "no",
    statistic = list(
    all_continuous() ~ c("{mean} ({min},{max},{sd})",
                         "{N_miss} ({p_miss}%)"),
    all_categorical() ~ "{n} / {N} ({p}%)"
    ),
    type = all_continuous() ~ "continuous2"
    ) %>%
  add_p() %>%
  as_flex_table() %>%
  flextable::set_caption(caption="Responses on SWAN Rating Scale (Stratified by Postpartum Smoking Expos
cor_matrix_2 = cor(TobExp_df[ ,c("bpm_att", "bpm_ext", "bpm_att_p", "bpm_ext_p")], method='pearson',use=

# Now produce the plot
corrplot(cor_matrix_2, method='circle', type='upper', addCoef.col = "black")
ggplot(data=TobExp_df,
       aes(x = bpm_att,
           y=bpm_att_p,
           colour=mom_smoke_preg)) +
  geom_point(na.rm=TRUE, size=2) +
  scale_colour_brewer(name = 'SDP',
                      palette = 'Dark2') +        # Use colorbrewer palette Dark2
  labs(x='BPM related to attention problems answered by child self',
       y='BPM related to attention problems answered by parent on child') +
  theme_bw()
ggplot(data=TobExp_df,
       aes(x = bpm_ext,
           y=bpm_ext_p,
           colour=mom_smoke_preg)) +
  geom_point(na.rm=TRUE, size=2) +
  scale_colour_brewer(name = 'SDP',
                      palette = 'Dark2') +
  labs(x='BPM related to attention problems answered by child self',
       y='BPM related to attention problems answered by parent on child') +
  theme_bw()
ggplot(data=TobExp_df,
       aes(x = bpm_att,
           y=bpm_att_p,
           colour=mom_smoke_post)) +
  geom_point(na.rm=TRUE, size=2) +
  scale_colour_brewer(name = 'SDP',
                      palette = 'Dark2') +
  labs(x='BPM related to attention problems answered by child self',
       y='BPM related to attention problems answered by parent on child') +
  theme_bw()
ggplot(data=TobExp_df,
       aes(x = bpm_ext,
           y=bpm_ext_p,
           colour=mom_smoke_post)) +
  geom_point(na.rm=TRUE, size=2) +
  scale_colour_brewer(name = 'Postnatal Exposure',
```

```
                      palette = 'Dark2') +
  labs(x='BPM externalizing problems answered by child self',
       y='BPM externalizing problems answered by parent on child',
       title = 'BPM externalizing problems by child self versus parent on child') +
  theme_bw()
TobExp_df %>%
  dplyr::select(c('bpm_att', 'bpm_ext',  'mom_smoke_preg')) %>%
  tbl_summary(
    by = mom_smoke_preg,
    missing = "no",
    statistic = list(
    all_continuous() ~ c("{mean} ({min},{max},{sd})",
                         "{N_miss} ({p_miss}%)"),
    all_categorical() ~ "{n} / {N} ({p}%)"
    ),
    type = all_continuous() ~ "continuous2"
    ) %>%
  add_p() %>%
  as_flex_table() %>%
  flextable::set_caption(caption=" BPM externalizing and attention problems answered by child self (Stra
TobExp_df %>%
  dplyr::select(c('bpm_att', 'bpm_ext',  'mom_smoke_post')) %>%
  tbl_summary(
    by = mom_smoke_post,
    missing = "no",
    statistic = list(
    all_continuous() ~ c("{mean} ({min},{max},{sd})",
                         "{N_miss} ({p_miss}%)"),
    all_categorical() ~ "{n} / {N} ({p}%)"
    ),
    type = all_continuous() ~ "continuous2"
    ) %>%
  add_p() %>%
  as_flex_table() %>%
  flextable::set_caption(caption=" BPM externalizing and attention problems answered by child self (Stra
pairs.panels(TobExp_df %>% select(preg_smk_severity, post_smk_severity) %>% na.omit())
TobExp_df %>% select(preg_smk_severity, post_smk_severity, swan_inattentive, swan_hyperactive) %>% na.o
TobExp_df %>% select(preg_smk_severity, post_smk_severity, bpm_att, bpm_ext) %>% na.omit() %>% ggpairs(
TobExp_df %>% select(smoke_exposure_6mo, smoke_exposure_12mo, smoke_exposure_2yr, smoke_exposure_3yr, s
TobExp_df %>% select(smoke_exposure_6mo, smoke_exposure_12mo, smoke_exposure_2yr, smoke_exposure_3yr, s
```