Project 1: Exploring the Effects of SDP/ETS on Self-Regulation, Externalizing Behavior, and Substance Use

Mavis Liang

1 Introduction

Maternal Smoking during Pregnancy (SDP) and environmental tobacco smoke (ETS) are two of the most ubiquitous and hazardous of children's environmental exposures. Early smoke-exposure increases rates of externalizing behaviors in children.The changes in the brain are associated with adverse neurodevelopmental outcomes. Epidemiologic studies suggest maternal active smoking (i.e. direct intake of tobacco through cigarettes) during pregnancy is associated with child behavior problems, such as attention deficit and hyperactivity disorder (ADHD) as well as adult criminal behavior (Liu et al. 2012).

A subset of 100 mothers from a previous study on smoke avoidance intervention to reduce low-income women's (N=738) smoking and ETS exposure during pregnancy and children's exposure to ETS in the immediate postpartum period are randomly selected for this study. This dataset is provided by Lauren Micalizzi, Ph.D. from Brown university, and this project is within the large NIH project of Prenatal Tobacco Exposure: Sel-regulatory Pathways to Externalizing Behaviours.

With this dataset, we aim at examining the effects of SDP/ETS on adolescent self-regulation, substance use, and externalizing.

The code for this analysis can be found in Github: https://github.com/Mavis-Liang/pda_project1.

2 Data

The dataset comprises data from 49 individuals, including demographic information, maternal smoking behavior, socio-economic status, and mental health indicators for both mother and child. Notably, all participants had records in ADHD-related variables (swan_hyperactive and

swan_inattentive), although a significant portion of the data contained missing values in other variables.

The variables in the dataset can be summarized into three categories, according to the study aim:

- **Outcome**:
  - Self-regulation (which can also be a mediator) - `erq_cog`, `erq_exp`
  - Substance use (SU) - cig/e_cig/mj/alc ever/num
  - Externalizing (EXT) - `bpm_ext_p`, `bpm_ext`, ADHD(`swan_hyperactive` and `swan_inattentive`)
- **Predictors**:
  - Smoking during pregnancy (SDP) - `mom_smoke_16`, `mom_smoke_22`, `mom_smoke_32wk`, `cotimean_34week`
  - Environmental tobacco exposure (ETS) - `mom_smoke_pp1`, `mom_smoke_pp2`, `mom_smoke_pp12wk`, `mom_smoke_pp6mo`, `smoke_exposure_6mo`, `cotimean_pp6mo`, `cotimean_pp6mo_baby`
- **Confounders**:
  - age, race, language, ect.

3 Methods

3.1 Data Preprocessing

The preprocessing code provided by Dr. Alice Paul performs several steps to clean, transform, and prepare both child and parent data for analysis. Here's a summary of what each section of the code accomplishes:

Child Data Processing

Initial Data Loading and Selection: Child data (child_df) is loaded from a CSV file and filtered for specific records (child_baseline_arm_1). Only a subset of columns is retained.

Demographic Variable Selection: Removes various demographic variables and renames some race-related variables for clarity.

Dropping Brief YSR Variables: Excludes a range of variables related to the Brief Youth Self-Report due to scoring difficulties.

Substance Use Variables Summarization: Creates new variables summarizing cigarette, e-cigarette, marijuana, and alcohol usage (cig_ever, num_cigs_30, etc.) and drops related detailed variables.

Dropping Other Drugs and Norms Variables: Omits variables related to other drugs and perceived norms.

Brief Problem Monitor Scoring: Calculates scores for attention, externalizing, and internalizing problems based on the Brief Problem Monitor (BPM) questionnaire.

Emotional Regulation: Computes cognitive and expressive emotional regulation scores and drops related detailed variables.

Dropping Physical and Life Stress Variables: Excludes variables related to physical development and life stress.

Parental Monitoring Scoring: Computes scores for parental knowledge, child disclosure, parental solicitation, and parental control based on the Parental Monitoring questionnaire.

Dropping Dysregulation and Early Adolescent Temperament Variables: Removes variables related to dysregulation inventory and early adolescent temperament.

Alcohol and Substance Abuse: Excludes alcohol and substance abuse-related variables due to few observations.

Diet Questions Removal: Drops remaining diet-related questions.

Parent Data Processing

Loading and Filtering Parent Data: Loads parent data from the same CSV file and filters it for a specific record (parent_baseline_arm_2). Selects a subset of columns.

Demographics Processing: Similar to child data, processes demographic variables for parents.

Dropping Brief Parent Variables: Removes variables related to the Brief Parent questionnaire.

SWAN Processing: Calculates scores for inattentive and hyperactive symptoms based on the SWAN questionnaire.

Dropping Connors Variables: Excludes Connors questionnaire variables.

Parental Brief Problem Monitor: Calculates scores for parental BPM related to the child.

Alcohol and Drug Use: Processes variables related to alcohol and drug use, especially focusing on smoke exposure at different child ages.

Dropping Additional Variables: Removes variables related to other assessments like CHAOS, adult BPM, adult emotional regulation, adult temperament, and ETQ.

Combining Data: Joins the processed child and parent datasets on parent_id.

Final Data Export: The cleaned and combined dataset is exported as a CSV file for further analysis.

Overall, the preprocessing focuses on refining the dataset to include only relevant variables for the study, summarizing complex questionnaire data into actionable metrics, and aligning child and parent data for comprehensive analysis.

For our analysis we began with meticulous data preprocessing. Key steps included:

Assignment of Zero Values: For variables related to substance use (SU) such as num_cigs_30, num_e_cigs_30, num_mj_30, and num_alc_30, we assigned a value of zero where corresponding 'ever used' variables (e.g., cig_ever, e_cig_ever, mj_ever, alc_ever) indicated no usage. This approach clarified the dataset by aligning numerical usage data with categorical usage indicators.

Missing Data Visualization: We visualized the missing data pattern using a heatmap. Each row represented an individual participant, and each column corresponded to a variable. Blocks in the heatmap were color-coded—black for missing values and green for non-missing values. This visualization provided a clear overview of the missing data patterns and their associations across different variables.

3.2 Summary Statistics

Substance Use (SU): Focused on variables like num_cigs_30, num_e_cigs_30, num_mj_30, and num_alc_30.

Self-regulation: Analyzed variables such as erq_cog and erq_exp.

Externalizing Behaviors: Included variables like bpm_ext, bpm_ext_p, swan_hyperactive, and swan_inattentive.

Maternal Smoking During Pregnancy (SDP): Examined variables indicating maternal smoking at different pregnancy stages and postpartum periods.

Environmental Tobacco Smoke (ETS): Focused on variables representing environmental smoke exposure.

3.3 Analyzing Outcomes in Relation to SDP and ETS

Each outcome was assessed for the effects of SDP or ETS, both separately and in combination.

Self-regulation We stratified summary tables of self-regulation variables by missingness to explore the impact of SDP/ETS on missing data patterns, given the substantial proportion of missing values in these outcomes. Additionally, we utilized Lasagna plots, boxplots, and scatterplots to delve deeper into these relationships.

Substance Use (SU) Given the limited number of records available for SU outcomes, we conducted a detailed analysis of the few individuals with relevant data.

Externalizing Behaviors For Externalizing outcomes, we produced correlation plots to examine the relationships between different variables, including parent and child-reported externalizing behaviors and ADHD-related scores.

3.4 Visualizations and Statistical Analysis

Our analysis employed a range of visualizations:

Lasagna Plots: These plots provided a temporal view of maternal smoking behavior during and after pregnancy.

Boxplots and Scatterplots: Used to explore associations between SDP/ETS and various outcomes like self-regulation and externalizing behaviors.

Correlation Plots (ggpairs): Employed to visualize pairwise relationships among variables related to externalizing behaviors.

4 Results

4.1 Overall missing pattern

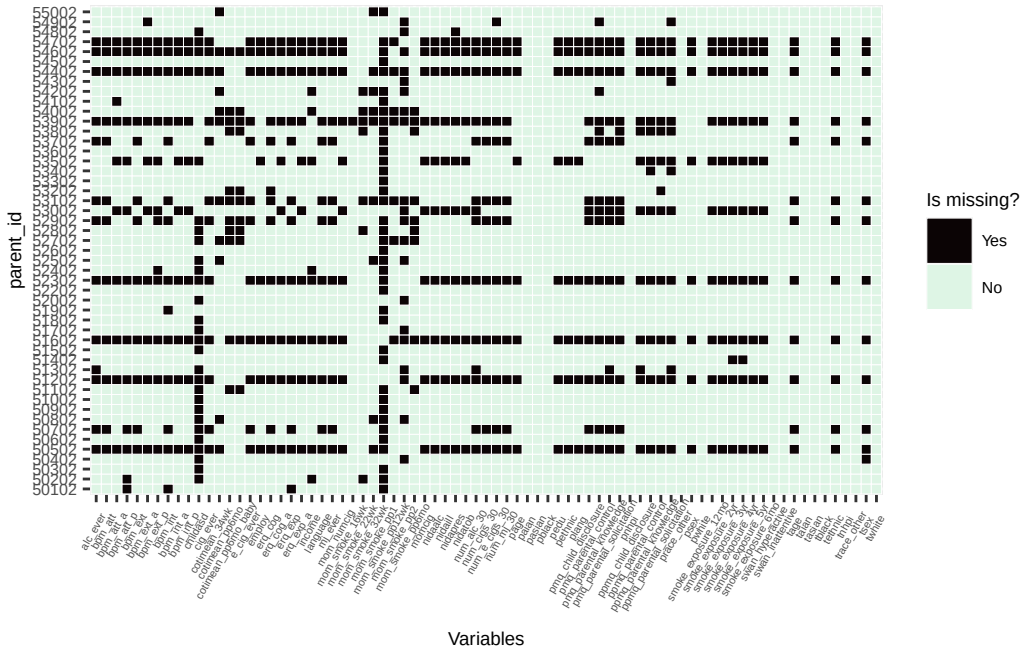

Figure 1: Missingness in each variable for each individual

Figure 1 shows the missing values for each individuals. Those with missing values usually have missing values in multiple variables. Looking at individual data, those who have missingness in many variables (eg. `parent_id` = 54702) have a relatively larger value in `cotimean_pp6mo` and `cotimean_6mo`, but this pattern is not substantial. Also, these individuals usually have values of 0 in `swan_hyperactive` and `swan_inattentive`; These two variables measure the externalizing behaviors of the child.

4.2 Summary for Variables of Interest

Summary statistics of the 5 groups of variables belongs to outcome and predictors are made into Table 1 to Table 5 below.

Table 1: Summary of Adolescents Substance Use Outcomes

Substance use	**N = 49**
num_cigs_30	0 (0%)
missing	12
num_e_cigs_30	
0	35 (97%)
2	1 (2.8%)
missing	13
num_mj_30	
0	34 (92%)
3	1 (2.7%)
12	1 (2.7%)
18	1 (2.7%)
missing	12
num_alc_30	
0	33 (94%)
1	1 (2.9%)
10	1 (2.9%)
missing	14

[1] n (%)

Table 2: Summary of Adolescents Externalizing Outcomes

Externalizing	**N = 49**
bpm_ext	2.81 (2.01)
missing	12
bpm_ext_p	1.68 (2.50)
missing	12
swan_hyperactive	6 (7)
swan_inattentive	9 (7)

[1] Mean (SD)

Table 3: Summary of SDP variables

SDP	N = 49
mom_smoke_16wk	
1=Yes	12 (25%)
2=No	36 (75%)
missing	1
mom_smoke_22wk	
1=Yes	13 (31%)
2=No	29 (69%)
missing	7
mom_smoke_32wk	
1=Yes	10 (25%)
2=No	30 (75%)
missing	9
cotimean_34wk	50 (98)
missing	11

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

Table 4: Summary of SDP variables

ETS	N = 49
mom_smoke_pp1	
1=Yes	3 (30%)
2=No	7 (70%)
missing	39
mom_smoke_pp2	
1=Yes	7 (24%)
2=No	22 (76%)
missing	20
mom_smoke_pp12wk	
1=Yes	12 (29%)
2=No	30 (71%)
missing	7
mom_smoke_pp6mo	
1=Yes	16 (40%)
2=No	24 (60%)
missing	9
smoke_exposure_6mo	10 (26%)
missing	10
cotimean_pp6mo	100 (179)
missing	11
cotimean_pp6mo_baby	4.0 (7.6)
missing	11

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

4.3 The effect of SDP/ETS on the outcome of Self-regulation

Among all the 49 individuals, 12 misses values in both `erq_cog` and `erq_exp`, which indicate the child's self-regulation behaviors. 2 individuals miss values in on of `erq_cog` or `erq_exp`. To explore if these missingness are related to SDP/ETS, in **Table 6**, we divide the 49 individuals into two groups with or without values in `erq_cog` and `erq_exp`. We notice that the individuals with missing values in these two variables tend to have a higher value in `cotimean_pp6mo` and `cotimean_pp6mo_baby`, but due to the small sample size, we can't get statistical significant conclusions.

Table 5: SDP/ETS grouped by missing in self-regulation

SDP/ETS	**With self-regulation**, N = 37	**Without self-regulation**, N = 12	p-value
mom_smoke_16wk			>0.9
1=Yes	9 (24%)	3 (27%)	
2=No	28 (76%)	8 (73%)	
Unknown	0	1	
mom_smoke_22wk			>0.9
1=Yes	10 (31%)	3 (30%)	
2=No	22 (69%)	7 (70%)	
Unknown	5	2	
mom_smoke_32wk			>0.9
1=Yes	8 (25%)	2 (25%)	

2=No	24 (75%)	6 (75%)	
Unknown	5	4	
cotimean_34wk	1 (0, 32)	1 (0, 73)	>0.9
Unknown	6	5	
mom_smoke_pp1			>0.9
1=Yes	2 (25%)	1 (50%)	
2=No	6 (75%)	1 (50%)	
Unknown	29	10	
mom_smoke_pp2			0.6
1=Yes	5 (21%)	2 (40%)	
2=No	19 (79%)	3 (60%)	
Unknown	13	7	
mom_smoke_pp12wk			0.4
1=Yes	9 (26%)	3 (43%)	
2=No	26 (74%)	4 (57%)	
Unknown	2	5	
mom_smoke_pp6mo			0.7
1=Yes	12 (38%)	4 (50%)	
2=No	20 (63%)	4 (50%)	
Unknown	5	4	
smoke_exposure_6mo	10 (29%)	0 (0%)	0.6
Unknown	2	8	
cotimean_pp6mo	4 (1, 111)	83 (15, 140)	0.2
Unknown	6	5	
cotimean_pp6mo_baby	1.4 (0.5, 3.2)	3.2 (1.7, 5.6)	0.090
Unknown	6	5	

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

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

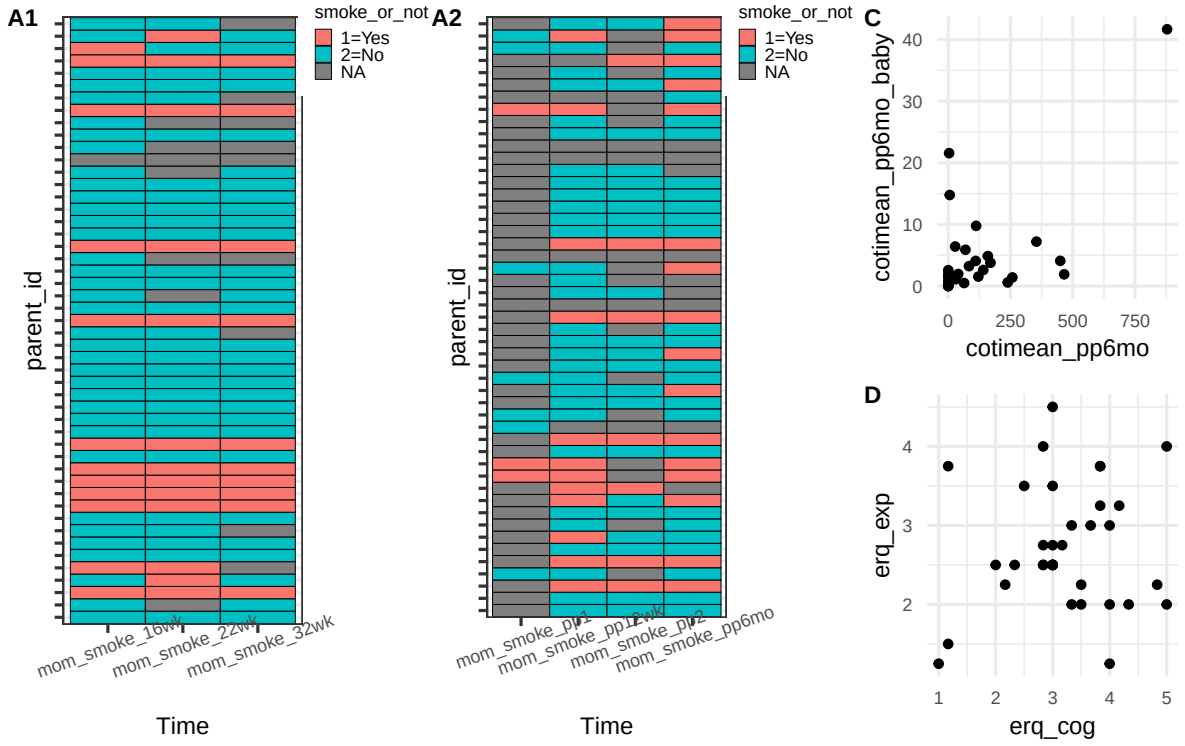

Figure 2: Relationships within SDP/ETS variables and within self-regulation variables. A1 and A2 are the Lasagna plots for SDP and for ETS related to mom smoking over time. A1 shows if the moms have been smoking during pregnancy in the three time points; and A2 shows the results during postpartum. C shows the relationship between the Urine cotinine (nicotine metabolite) at 6 months postpartum from baby (Y-axis) and from mom (X-axis). D shows the relationship between the two variables representing self-regulation behavious in children, in which Expressive Suppression on Y axis and Cognitive Reappraisal on X-axis.

In **Figure 2**, we observed that, (1) (**Figure 2 A1 and A2**)The missingness increased over time in the visits during pregnancy. There is a substantial missingness in `mom_smoke_pp1`. Missing in smoking values during pregnant usually results in missing in postpartum. (2) (**Figure 2 A1 and A2**)The smoking behaviors within individuals tend to persist, both during pregnancy and postpartum. (3) (**Figure 2 A1 and A2**)If an individual has smoking behaviors during pregnancy, she will persist the same behavior. (4) (**Figure 2 B**)The correlation between the Urine cotinine detected in mom and in baby is not obvious. We believe that there is other sources of tabbaco exposure for the babies. (5) (**Figure 2 C**)The relationship between exp and cog is not obvious.

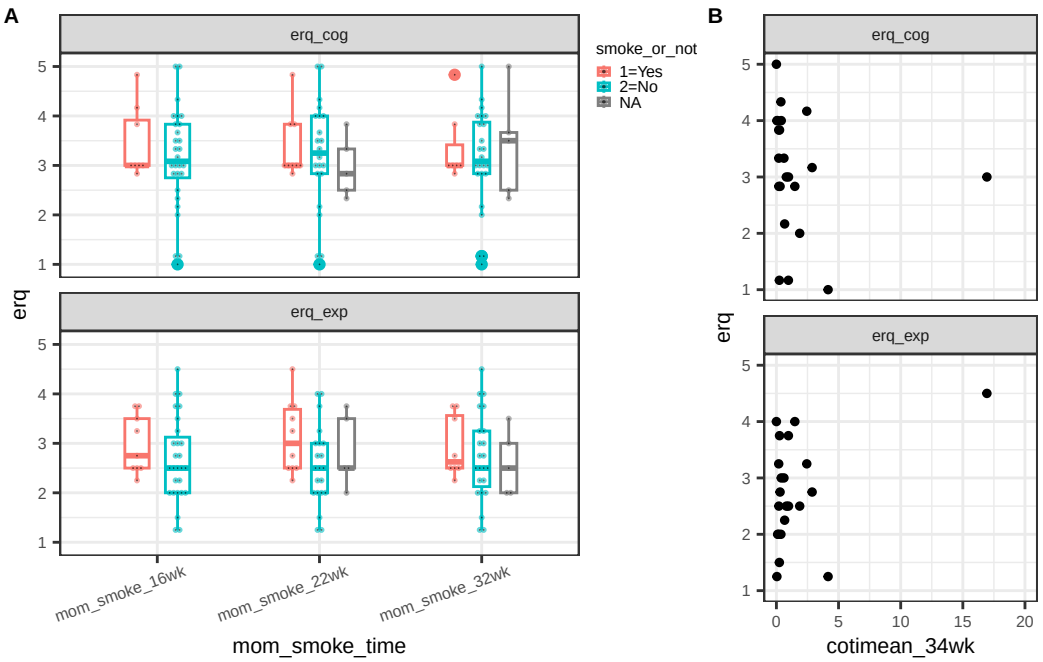

Figure 3: Association between self-regulation and SDP. `cotimean_34wk` > 20 is not shown in plot B for better display.

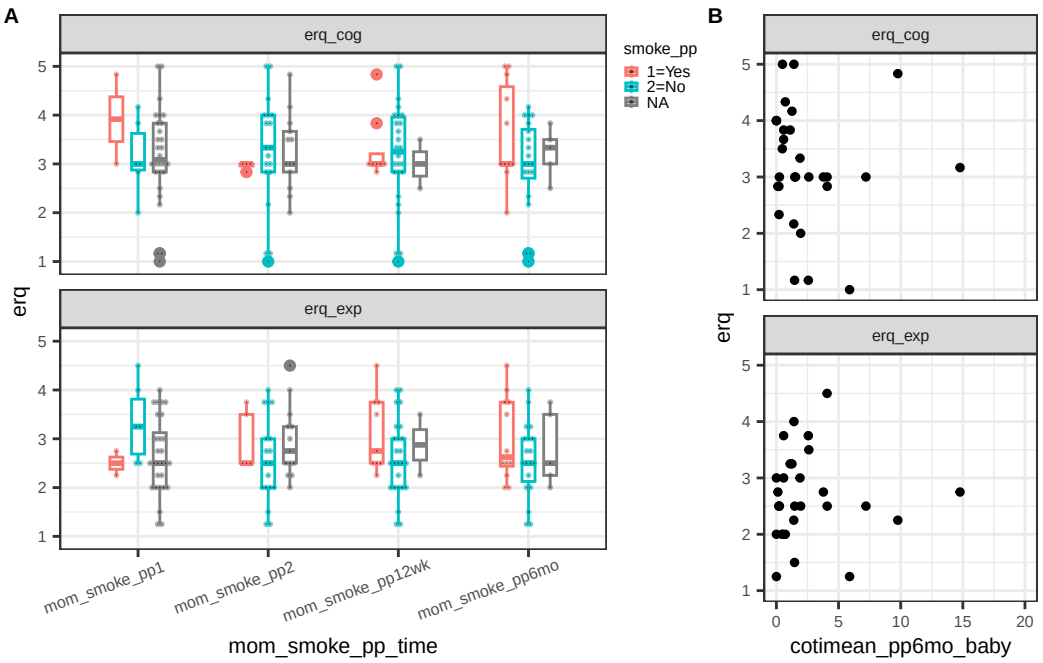

Figure 4: Association between self-regulation and ETS

Figure 3 shows the association between self-regulation and SDP and **Figure 4** shows the association between self-regulation and ETS. We observed that: (1) The box plots shows that not smoking during pregnancy and postpartum results in a lower score in child's self-regulation issues, compring the mean values in the blue boxes and the red boxes. (2) On the other hand, the dots on the boxes indicate that the data is evenly distributed within groups of smoking or not smoking or those have missing values. However, outliers exist in all three groups. (3) The scattter plots shows that the association between cog/exp and urine cotinine is weak, which indicates that the self-regulation is not related to SDP or ETS.

4.4 The effect of SDP/ETS on substance use (SU)

Based on **Table 2** in the **Data overview** section, we know that only a few (0 to 3) of the adolescents have outcomes in substance use that are not missing values or not equal 0.

Table 7 below shows the three individuals that have non-Zero/NA records in substance use. The child from all three of them have been highly exposed to SDP or ETS.

r} \end{table}

Table 6: Individuals have non-zero/NA outcome in substance use

parent_id	num_cigs_30	num_e_cigs_30	num_mj_30	num_alc_30
51902	0	0	18	0
52602	0	2	3	10
54302	0	0	12	1

parent_id	mom_smoke_16wk	mom_smoke_22wk	mom_smoke_32wk	cotimean_34wk	mom_smoke_pp1	mom_smoke_pp2
51902	2=No	2=No	2=No	0.362795	NA	2=No
52602	1=Yes	1=Yes	1=Yes	282.280000	NA	1=Yes
54302	1=Yes	1=Yes	1=Yes	171.218375	1=Yes	NA

parent_id	mom_smoke_pp12wk	mom_smoke_pp6mo	smoke_exposure_6mo	cotimean_pp6mo	cotimean_pp6mo_baby
51902	2=No	1=Yes	0	2.000525	0.7167082
52602	1=Yes	1=Yes	0	449.830000	4.0871285
54302	1=Yes	1=Yes	0	112.017535	9.7688382

4.5 The effect of SDP/ETS on Externalizing

Exploring variables of externalizing behaviors, we notice that the ADHD variable range from 1 to 20. We observed some missing values in the `bpm_ext_p` and `bpm_ext`, which are the answers of externalizing behaviors of the child from mom and from the child themselves. Those have missing values in `bpm_ext_p` tend to also have missing values in `bpm_ext`.

missing in exp answered by mom also missing in child

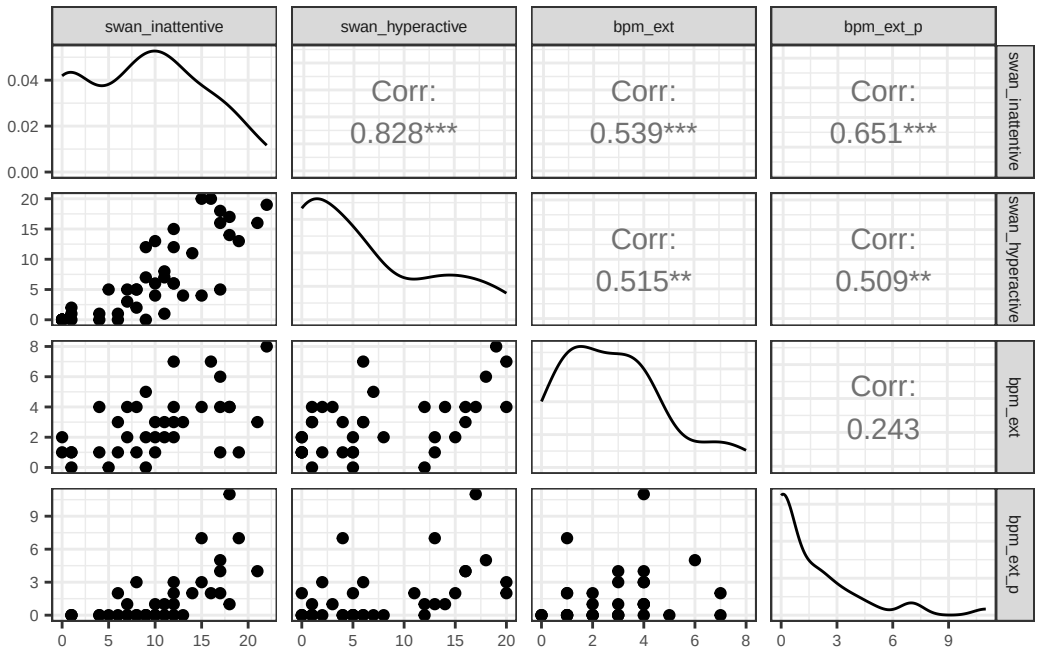

Figure 5: Pairwise relationships within Externalizing variables. Sub-plots at the lower triangle shows the scatterplot of the pairwise variables. Numbers in the upper triangle are the correlations of corresponding variables in pairs. Density curve in the diagnol is the density distribution of the variables in the top facet label. Note that both bpm_ext_p and bpm_ext are scores describing externalizing problems in child. bpm_ext_p are answered by parents, and bpm_ext are answered by children themselves.

In **Figure 5**, what we observed is that: (1) In the sub-figure at the 4th row and the 3rd column, although both `bpm_ext_p` and `bpm_ext` are scores describing externalizing problems in child, parents tend to underestimate the externalizing problems in child. We believe that the answers from the children themselves are more reliable. Thus in the following analysis, we only show the effect of SDP/ETS on `bpm_ext`. (2) In the first plot of the first column, we see that higher inattentive means higher hyperactive, they have a corr of more than 0.8.

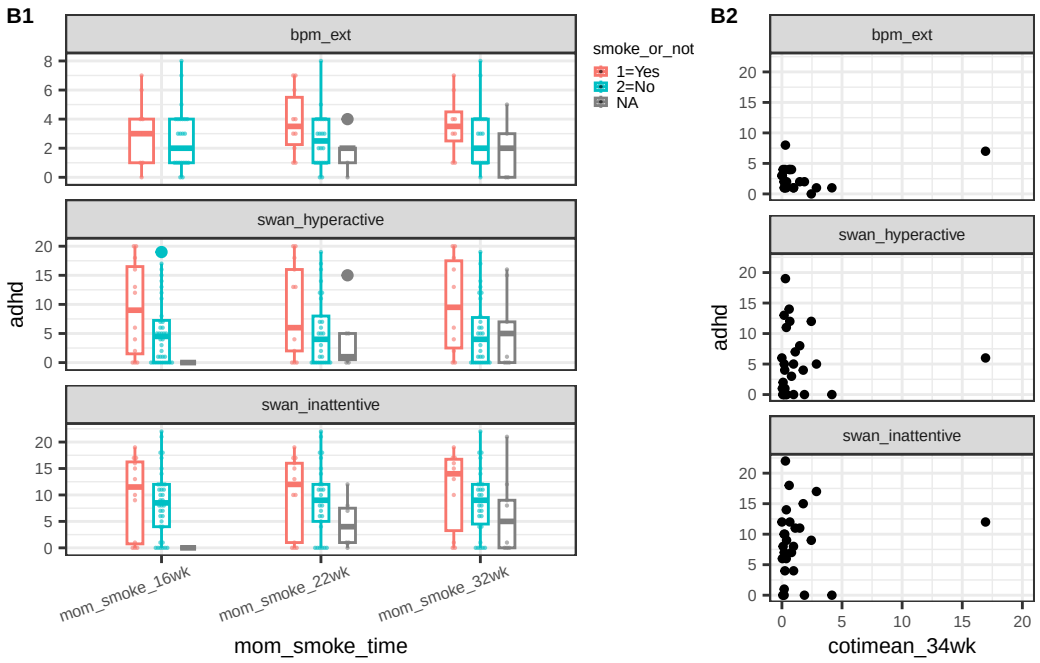

Figure 6: Association between externalizing and SDP

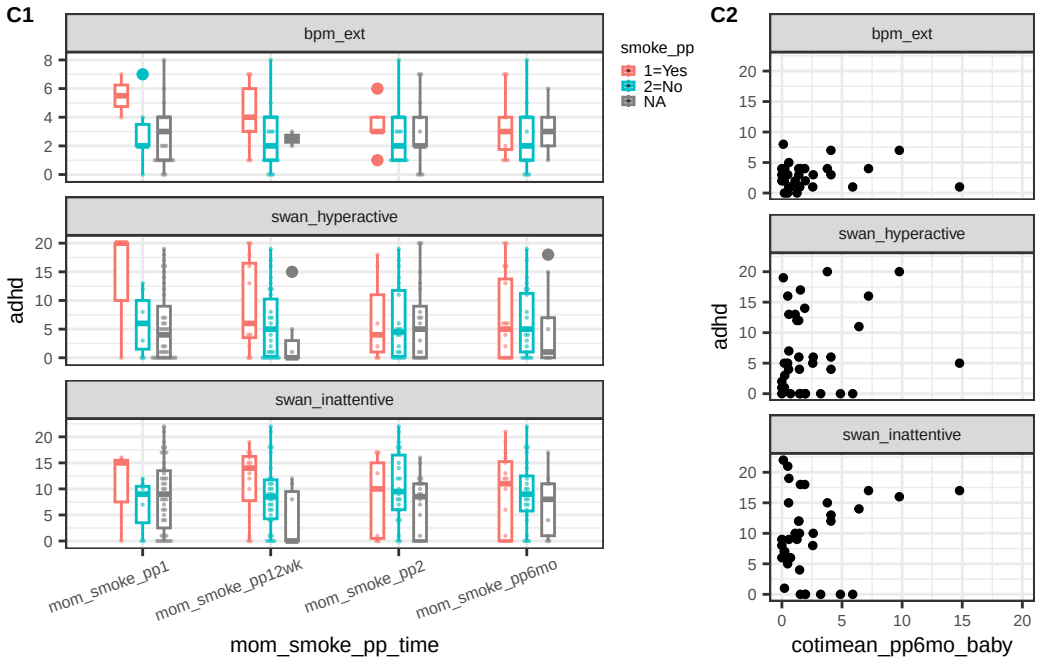

Figure 7: Association between externalizing and ETS

- Similar to the result for self-regulation, the association between externalizing and SDP/ETS is not strong.

5 Limitations

The data has a small size, and within the data, many individuals have missing values. This low quality of data result in the difficulty of finding significant relationships between SDP/ETS on the child's behavial problems. It also make imputation to the missingness difficult. On the other hand, the variation (noise) within the same group (smoke or not) is large.

On the other hand, there are many confoundings we can adjust. For example, the ADHD or other mental/behavial history can be genetically passed to the child. Also, the social-economic status can cause the difference of a child's development of behavior issues. However, due to the limitation of the quality of the data, we are not able to control all these confoundings when we explore the relationships between SDP/ETS and behavior problems.

6 Reference

Liu, Jianghong, Patrick Leung, Linda Mccauley, Yuexian Ai, and Jennifer Pinto-Martin. 2012. "Mother's Environmental Tobacco Smoke Exposure During Pregnancy and Externalizing Behavior Problems in Children." *Neurotoxicology* 34 (November). https://doi.org/10.1016/j.neuro.2012.11.005.

7 Code Appendix

```
packages_to_check <- c("gtsummary", "gt", "tidyverse", "kableExtra", "mice",
                       "viridis", "gridExtra", "cowplot", "xtable", "knitr",
                       "GGally")
# Check if each package is installed and load it if available;
# otherwise, install and load it
for (pkg in packages_to_check) {
  if (!require(pkg, character.only = TRUE, quietly = TRUE)) {
    # If the package is not installed, install it
    install.packages(pkg)

    # Now, load the package
    library(pkg, character.only = TRUE)
  }
}

knitr::opts_chunk$set(warning = FALSE, message = FALSE,
                      echo = FALSE, fig.align = "center")
df <- read.csv("../project1.csv")
df$num_cigs_30[df$cig_ever==0]=0
df$num_e_cigs_30[df$e_cig_ever==0]=0
df$num_mj_30[df$mj_ever==0]=0
df$num_alc_30[df$alc_ever==0]=0

df_filled <-
df %>%
  mutate_all(~ifelse(.=="", NA, .)) %>%
  mutate(bpm_ext = as.numeric(bpm_ext),
         bpm_ext_p = as.numeric(bpm_ext_p))
## put numeric variables into numeric
df_missing <-
  apply(df_filled, 2, function(x) ifelse(is.na(x), "Yes", "No")) %>%
  as.data.frame()

df_missing %>%
  mutate(parent_id = as.factor(df_filled$parent_id)) %>%
  pivot_longer(cols = names(df_filled)[2:78],
               names_to = "Variables",
               values_to = "is_missing") %>%
  mutate(is_missing = factor(is_missing, levels= c("Yes", "No")))%>%
```

```
  ggplot(aes(Variables, parent_id, fill=is_missing, )) +
  geom_tile(colour = "white") +
  scale_fill_viridis(discrete = TRUE, option = "G") +
  theme(text = element_text(size = 7),
        axis.text.x = element_text(size = 4,
                                   angle = 60, hjust = 0.8, vjust = .9))+
  labs(fill = "Is missing?")
## Table showing partial variables (those we are interested in)
## Outcome
## Sel-reg
df_test <- df_filled %>%
  select(erq_cog, erq_exp) %>%
  tbl_summary(statistic = list(all_continuous() ~ "{mean} ({sd})"),
                  missing_text = "missing") %>%
  modify_header(label = "**Self-regulation**") %>%
  as_kable_extra(booktabs = TRUE,
                     format = "latex",
                     longtable = TRUE,
                     caption = "Summary of Adolescents Self-regulation Outcomes") %>%
      kable_styling(latex_options = "hold_position", font_size = 7)

## SU
tb_su <-
df_filled %>%
  select(num_cigs_30, num_e_cigs_30, num_mj_30, num_alc_30) %>%
  tbl_summary(statistic = list(all_continuous() ~ "{mean} ({sd})"),
                  missing_text = "missing") %>%
  modify_header(label = "**Substance use**") %>%
  as_kable_extra(booktabs = TRUE,
                     format = "latex",
                     longtable = TRUE,
                     caption = "Summary of Adolescents Substance Use Outcomes") %>%
      kable_styling(latex_options = "hold_position", font_size = 7)

## EXT
tb_ext <-
df_filled %>%
  select(bpm_ext, bpm_ext_p, swan_hyperactive, swan_inattentive) %>%
  tbl_summary(type = list(bpm_ext ~ "continuous", bpm_ext_p ~ "continuous"),
              statistic = list(all_continuous() ~ "{mean} ({sd})"),
              missing_text = "missing") %>%
```

```
  modify_header(label = "**Externalizing**") %>%
  as_kable_extra(booktabs = TRUE,
                       format = "latex",
                       longtable = TRUE,
                       caption = "Summary of Adolescents Externalizing Outcomes") %>%
      kable_styling(latex_options = "hold_position", font_size = 7)

## Placing tables side by side
cat(c("\\begin{table}[!htb]
    \\begin{minipage}{.5\\linewidth}
      \\centering",
        tb_su,
    "\\end{minipage}
    \\begin{minipage}{.5\\linewidth}
      \\centering",
        tb_ext,
    "\\end{minipage}
\\end{table}"
))
## SDP/ETS
tb_sdp <-
df_filled %>%
  select(mom_smoke_16wk, mom_smoke_22wk, mom_smoke_32wk,
         cotimean_34wk) %>%
  tbl_summary(statistic = list(all_continuous() ~ "{mean} ({sd})"),
              missing_text = "missing") %>%
  modify_header(label = "**SDP**") %>%
  as_kable_extra(booktabs = TRUE,
                       format = "latex",
                       longtable = TRUE,
                  caption = "Summary of SDP variables") %>%
      kable_styling(latex_options = "hold_position", font_size = 7)

tb_ets <-
df_filled %>%
  select(mom_smoke_pp1, mom_smoke_pp2, mom_smoke_pp12wk,
         mom_smoke_pp6mo, smoke_exposure_6mo, cotimean_pp6mo,
         cotimean_pp6mo_baby) %>%
  tbl_summary(statistic = list(all_continuous() ~ "{mean} ({sd})"),
              missing_text = "missing") %>%
  modify_header(label = "**ETS**") %>%
```

```
  as_kable_extra(booktabs = TRUE,
                      format = "latex",
                      longtable = TRUE,
                  caption = "Summary of SDP variables") %>%
      kable_styling(latex_options = "hold_position", font_size = 7)

cat(c("\\begin{table}[H]
    \\begin{minipage}{.5\\linewidth}
      \\centering",
        tb_sdp,
    "\\end{minipage}%
    \\begin{minipage}{.5\\linewidth}
      \\centering",
        tb_ets,
    "\\end{minipage}
\\end{table}"
))
## SDP on self-regulation
# Select corresponding
df_sr <-
  df_filled %>%
  select(parent_id,erq_cog, erq_exp, mom_smoke_16wk, mom_smoke_22wk, mom_smoke_32wk,
         cotimean_34wk, mom_smoke_pp1, mom_smoke_pp2, mom_smoke_pp12wk,
         mom_smoke_pp6mo, smoke_exposure_6mo,
         cotimean_pp6mo, cotimean_pp6mo_baby)

df_sr %>%
  mutate(is_missing = ifelse(is.na(erq_cog)&is.na(erq_exp), 1, 0)) %>%
  select(-c(erq_cog, erq_exp, parent_id)) %>%
  tbl_summary(by = is_missing) %>%
  add_p() %>%
  modify_header(label = "**SDP/ETS**",
                stat_1 = "**With self-regulation**, N = 37",
                stat_2 = "**Without self-regulation**, N = 12") %>%
  as_kable_extra(booktabs = TRUE,
                format = "latex",
              longtable = TRUE,
              caption="SDP/ETS grouped by missing in self-regulation") %>%
  kable_styling(font_size = 7, latex_options = "hold_position")
# association between erq_cog and erq_exp
erq_cog <-
```

```r
df_sr %>%
  ggplot(aes(x=erq_cog, y=erq_exp)) +
  geom_point(size = 1)+
  theme_minimal() +
  theme(text = element_text(size = 8))

# Lasagna plot:
## mom_smoke over time during pregnancy
lasagna_smk <-
df_sr %>%
  pivot_longer(cols = c("mom_smoke_16wk", "mom_smoke_22wk",
                        "mom_smoke_32wk"),
               values_to = "smoke_or_not",
               names_to = "Time") %>%
  mutate(smoke_or_not = addNA(smoke_or_not)) %>%
  ggplot(aes(x = Time, y = factor(parent_id), fill = smoke_or_not))+
  geom_tile(colour = "black") +
  labs(y = "parent_id") +
  theme_bw()+
  theme(legend.key.size = unit(0.2, 'cm'),
        legend.title = element_text(size = 6),
        legend.box.margin=margin(-10,-10,-10,-10),
        legend.box.spacing = unit(0.3, 'cm'),
        legend.justification = "top",
        text = element_text(size = 8),
        axis.text.x = element_text(size = 6,
                                   angle = 20),
        axis.text.y = element_blank())

## mom_smoke postpartrum overtime
lasagna_smk_pp <-
df_sr %>%
  pivot_longer(cols = c("mom_smoke_pp1", "mom_smoke_pp2", "mom_smoke_pp12wk",
         "mom_smoke_pp6mo"),
               values_to = "smoke_or_not",
               names_to = "Time") %>%
  mutate(YesOrNo = addNA(smoke_or_not)) %>%
  ggplot(aes(x = Time, y = factor(parent_id), fill = smoke_or_not))+
  geom_tile(colour = "black") +
  labs(y = "parent_id") +
  theme_bw() +
```

```
  theme(legend.key.size = unit(0.2, 'cm'),
        legend.title = element_text(size = 6),
        legend.box.margin=margin(-10,-10,-10,-10),
        legend.box.spacing = unit(0.3, 'cm'),
        legend.justification = "top",
        text = element_text(size = 8),
        axis.text.x = element_text(size = 6,
                                   angle = 20),
        axis.text.y = element_blank())
coti_momVSbaby <-
df_sr %>%
  ggplot(aes(cotimean_pp6mo, cotimean_pp6mo_baby)) +
  geom_point(size = 1) +
  theme_minimal() +
  theme(text = element_text(size = 8))

# Arrange the plots into subfigures
p_base <- plot_grid(coti_momVSbaby, erq_cog,
                    cols = 1,
                    labels = c("C", "D"),
                    label_size = 8)

p_lasagna <- plot_grid(lasagna_smk, lasagna_smk_pp,
                       ncol = 2,
                       labels = c("A1", "A2"),
                       label_size = 8,
                       hjust = 0)

plot_grid(p_lasagna, p_base,
                      ncol = 2,
                    rel_widths = c(2.5, 1))
# slf-regulation vs SDP
box_slf_sdp <-
df_sr %>%
  pivot_longer(cols = c("mom_smoke_16wk", "mom_smoke_22wk",
                        "mom_smoke_32wk"),
               values_to = "smoke_or_not",
               names_to = "mom_smoke_time") %>%
  mutate(smoke_or_not = addNA(smoke_or_not)) %>%
  pivot_longer(cols = c(erq_cog, erq_exp),
               values_to = "erq",
```

```
              names_to = "cog_or_exp") %>%
  ggplot(aes(x = mom_smoke_time, y = erq, color = smoke_or_not)) +
  geom_boxplot(position = position_dodge(.5), width=0.3) +
  geom_dotplot(binaxis='y', stackdir='center', dotsize = 0.6, alpha = 0.6,
               position = position_dodge(.5))+
  facet_wrap(.~cog_or_exp, ncol = 1)+
  theme_bw() +
  theme(legend.key.size = unit(0.2, 'cm'),
        legend.title = element_text(size = 6),
        legend.box.margin=margin(-10,-10,-10,-10),
        legend.justification = "top",
        text = element_text(size = 8),
        axis.text.x = element_text(size = 6,
                                   angle = 20, hjust = 0.8, vjust = .9))

slf_coti <-
df_sr %>%
  pivot_longer(cols = c(erq_cog, erq_exp),
               values_to = "erq",
               names_to = "cog_or_erq") %>%
  ggplot(aes(x = cotimean_34wk, y = erq)) +
  geom_point(size = 1) +
  facet_wrap(cog_or_erq~., ncol = 1)+
  theme_bw() +
  theme(text = element_text(size = 8))+
  xlim(c(0, 20))

plot_grid(box_slf_sdp, slf_coti, ncol = 2,
          labels = c("A", "B"),
          label_size = 8,
          rel_widths = c(2, 1))
box_slf_ext <-
df_sr %>%
  pivot_longer(cols = c("mom_smoke_pp1", "mom_smoke_pp2", "mom_smoke_pp12wk",
                        "mom_smoke_pp6mo"),
               values_to = "smoke_pp",
               names_to = "mom_smoke_pp_time") %>%
  mutate(smoke_pp = addNA(smoke_pp),
         mom_smoke_pp_time = factor(mom_smoke_pp_time,
                                    levels = c("mom_smoke_pp1",
                                               "mom_smoke_pp2",
```

```
                                                  "mom_smoke_pp12wk",
                        "mom_smoke_pp6mo"))) %>%
  pivot_longer(cols = c(erq_cog, erq_exp),
               values_to = "erq",
               names_to = "cog_or_exp") %>%
  ggplot(aes(x = mom_smoke_pp_time, y = erq, color = smoke_pp)) +
  geom_boxplot(position = position_dodge(0.5), width=0.3) +
  geom_dotplot(binaxis='y', stackdir='center', dotsize = 0.6, alpha = 0.6,
               position = position_dodge(0.5))+
  facet_wrap(.~cog_or_exp, ncol = 1)+
  theme_bw() +
  theme(legend.key.size = unit(0.2, 'cm'),
        legend.title = element_text(size = 6),
        legend.box.margin=margin(-10,-10,-10,-10),
        legend.justification = "top",
        text = element_text(size = 8),
        axis.text.x = element_text(size = 6,
                                   angle = 20, hjust = 0.8, vjust = .9))

slf_coti_pp <-
df_sr %>%
  pivot_longer(cols = c(erq_cog, erq_exp),
               values_to = "erq",
               names_to = "cog_or_erq") %>%
  ggplot(aes(x = cotimean_pp6mo_baby, y = erq)) +
  geom_point(size = 1) +
  facet_wrap(cog_or_erq~., ncol = 1)+
  theme_bw() +
  theme(text = element_text(size = 8))+
  xlim(c(0, 20))


plot_grid(box_slf_ext, slf_coti_pp, ncol = 2,
                       labels = c("A", "B"),
                       label_size = 8,
          rel_widths = c(2, 1))
df_su <-
df_filled %>%
  select(parent_id, num_cigs_30, num_e_cigs_30, num_mj_30, num_alc_30,
         mom_smoke_16wk, mom_smoke_22wk, mom_smoke_32wk,
         cotimean_34wk, mom_smoke_pp1, mom_smoke_pp2, mom_smoke_pp12wk,
```

```
         mom_smoke_pp6mo, smoke_exposure_6mo,
         cotimean_pp6mo, cotimean_pp6mo_baby) %>%
  filter(num_e_cigs_30==2 | num_mj_30 %in% c(3, 12, 18) |
           num_alc_30 %in% c(1, 10))

df_su[,1:5] %>%
  knitr::kable(booktabs = TRUE,
                    format = "latex",
                    caption = "Individuals have non-zero/NA\\
               outcome in substance use") %>%
      kable_styling(font_size = 7, latex_options = "hold_position",
                    full_width = FALSE)
df_su[,c(1, 6:11)] %>%
  knitr::kable(booktabs = TRUE,
                    format = "latex",) %>%
      kable_styling(font_size = 6, latex_options = "hold_position")
df_su[,c(1, 12:16)] %>%
  knitr::kable(booktabs = TRUE,
                    format = "latex") %>%
      kable_styling(font_size = 6, latex_options = "hold_position",
                    full_width = F)

df_ext <-
df_filled %>%
  select(parent_id, bpm_ext, bpm_ext_p, swan_inattentive, swan_hyperactive,
         mom_smoke_16wk, mom_smoke_22wk, mom_smoke_32wk,
         cotimean_34wk, mom_smoke_pp1, mom_smoke_pp2, mom_smoke_pp12wk,
         mom_smoke_pp6mo, cotimean_pp6mo_baby)

df_ext %>%
  select(swan_inattentive, swan_hyperactive, bpm_ext, bpm_ext_p) %>%
  ggpairs() +
  theme_bw() +
  theme(text = element_text(size = 7))
p_adhd <-
df_ext %>%
  ggplot(aes(swan_inattentive, swan_hyperactive)) +
  geom_point(size = 1) +
  theme_bw()

p_ext_momVSbaby <-
```

```r
df_ext %>%
  ggplot(aes(bpm_ext, bpm_ext_p)) +
  geom_point(size = 1)+
  theme_bw()



plot_grid(p_adhd, p_ext_momVSbaby,
          ncol = 1,
          labels= c("A", "B"),
          label_size = 8)
# Externalizing vs SDP
box_ext_sdp <-
df_ext %>%
  pivot_longer(cols = c("mom_smoke_16wk", "mom_smoke_22wk",
                        "mom_smoke_32wk"),
               values_to = "smoke_or_not",
               names_to = "mom_smoke_time") %>%
  mutate(smoke_or_not = addNA(smoke_or_not)) %>%
  pivot_longer(cols = c(bpm_ext, swan_inattentive, swan_hyperactive),
               values_to = "adhd",
               names_to = "adhd_type") %>%
  ggplot(aes(x = mom_smoke_time, y = adhd, color = smoke_or_not)) +
  geom_boxplot(position = position_dodge(.5), width=0.3) +
  geom_dotplot(binaxis='y', stackdir='center', dotsize = 0.6, alpha = 0.6,
               position = position_dodge(.5))+
  facet_wrap(.~adhd_type, ncol = 1, scales = "free_y")+
  theme_bw() +
  theme(legend.key.size = unit(0.2, 'cm'),
        legend.title = element_text(size = 6),
        legend.box.margin=margin(-10,-10,-10,-10),
        legend.justification = "top",
        text = element_text(size = 8),
        axis.text.x = element_text(size = 6,
                                   angle = 20, hjust = 0.8, vjust = .9))

ext_coti <-
df_ext %>%
  pivot_longer(cols = c(bpm_ext, swan_inattentive, swan_hyperactive),
               values_to = "adhd",
               names_to = "adhd_type") %>%
```

```
  ggplot(aes(x = cotimean_34wk, y = adhd)) +
  geom_point(size = 1) +
  facet_wrap(adhd_type~., ncol = 1)+
  theme_bw() +
  theme(text = element_text(size = 8))+
  xlim(c(0, 20))

plot_grid(box_ext_sdp, ext_coti, ncol = 2,
          labels = c("B1", "B2"),
          label_size = 8,
          rel_widths = c(2, 1))
# Externalizing vs SDP
box_ext_sdp <-
df_ext %>%
  pivot_longer(cols = c("mom_smoke_pp1", "mom_smoke_pp2", "mom_smoke_pp12wk",
                        "mom_smoke_pp6mo"),
               values_to = "smoke_pp",
               names_to = "mom_smoke_pp_time") %>%
  mutate(smoke_pp = addNA(smoke_pp)) %>%
  pivot_longer(cols = c(bpm_ext, swan_inattentive, swan_hyperactive),
               values_to = "adhd",
               names_to = "adhd_type") %>%
  ggplot(aes(x = mom_smoke_pp_time, y = adhd, color = smoke_pp)) +
  geom_boxplot(position = position_dodge(.5), width=0.3) +
  geom_dotplot(binaxis='y', stackdir='center', dotsize = 0.6, alpha = 0.6,
               position = position_dodge(.5))+
  facet_wrap(.~adhd_type, ncol = 1, scales = "free_y")+
  theme_bw() +
  theme(legend.key.size = unit(0.2, 'cm'),
        legend.title = element_text(size = 6),
        legend.box.margin=margin(-10,-10,-10,-10),
        legend.justification = "top",
        text = element_text(size = 8),
        axis.text.x = element_text(size = 6,
                                   angle = 20, hjust = 0.8, vjust = .9))

ext_coti <-
df_ext %>%
  pivot_longer(cols = c(bpm_ext, swan_inattentive, swan_hyperactive),
               values_to = "adhd",
               names_to = "adhd_type") %>%
```

```
  ggplot(aes(x = cotimean_pp6mo_baby, y = adhd)) +
  geom_point(size = 1) +
  facet_wrap(adhd_type~., ncol = 1)+
  theme_bw() +
  theme(text = element_text(size = 8))+
  xlim(c(0, 20))

plot_grid(box_ext_sdp, ext_coti, ncol = 2,
          labels = c("C1", "C2"),
          label_size = 8,
          rel_widths = c(2, 1))
```