

Exploratory Analysis on Effects of SDP/ETS on Adolescent Self-regulation, Substance Use, and Externalizing

Project 1: Exploratory Data Analysis*

Haiyue Song

2023-12-15

Abstract

Background: Exposure to smoking during pregnancy and environmental tobacco smoke are two of the most ubiquitous and hazardous of children’s environmental exposures (Micalizzi 2023; Paul 2023). Early smoke exposure is linked to externalizing behaviors, self-regulation problems, and substance use. This project aims to conduct exploratory data analysis to examine the associations between smoking during pregnancy, environmental tobacco smoke exposure, and self-regulation, externalizing behavior, and substance use in adolescents.

Methods: Population in this study were recruited from a previous study on smoke avoidance intervention, aimed at reducing low-income women’s smoking and environmental tobacco smoke exposure during pregnancy, as well as children’s exposure to environmental tobacco smoke in the immediate postpartum period. A subset of adolescents (N=100) and their mothers were randomly selected for recruitment into this current study. After preprocessing, we described the missing data pattern, conducted univariate analyses of demographic variables, smoking exposure variables, and variables related to externalizing, self-regulation, and substance use; intercorrelations were then analyzed. Bivariate analysis was conducted to meet the project’s research goals.

Results: The data accessed is the baseline data for the current study, including self-regulatory assessments, self-reported substance use, self-regulation, and EKG. After preprocessing, 49 observations of 78 variables were available. Significant missing data affected the exploratory data analysis. In terms of interrelationships, we found that prenatal and postnatal smoke exposure are mildly positively correlated. For variables related to self-regulation, externalizing, and substance use, some showed strong intercorrelations, respectively. Bivariate associations indicated that smoke exposure during the 2nd to 5th year is significantly associated with expressive suppression. For externalizing problems, smoke exposure from 0 to 6 months was associated with attention problems, externalizing problems, and two types of ADHD; prenatal smoke exposure was associated with attention problems evaluated by parents and ADHD - Hyperactive/Impulsive type.

Conclusions: Based on bivariate analysis, different stages of early smoke exposure are associated with self-regulation, externalizing problems, and substance use, respectively. Further studies should consider improving the sample size and the quality of the data to obtain more powerful statistical associations.

Keywords: Smoking during Pregnancy, Environmental Tobacco Smoke, Externalizing Behaviors, Self-regulation, Substance Use

1. Introduction

Exposure to smoking during pregnancy (SDP) and environmental tobacco smoke (ETS) are two of the most ubiquitous and hazardous of children’s environmental exposures (Micalizzi 2023; Paul 2023). Early smoke-exposure increases rates of externalizing behaviors in children, including Attention-Deficit/Hyperactivity

*This project is a collaboration with Dr. Lauren Micalizzi from the Department Behavioral and Social Sciences, Brown University School of Public Health. The instructor is Dr. Alice Paul from Department of Biostatistics, Brown University School of Public Health.

Table 1: Summary Statistics for Demographic Information

Demographic Variable	Child, N = 49	Parent, N = 49
age		
Mean (SD); Median (Range)	13.62 (1.21); 14.00 (12.00, 16.00)	37.54 (3.57); 37.00 (32.00, 45.00)
(Missing)	12	8
sex	13 / 36 (36%)	40 / 41 (98%)
(Missing)	13	8
language	11 / 37 (30%)	15 / 41 (37%)
(Missing)	12	8
race		
American Indian/Alaska Native	3 / 37 (8.1%)	4 / 41 (9.8%)
Black/African American	8 / 37 (22%)	0 / 41 (0%)
Hispanic/Latino	15 / 37 (41%)	13 / 41 (32%)
Native Hawaiian/Pacific Islander	0 / 37 (0%)	4 / 41 (9.8%)
Other	0 / 37 (0%)	1 / 41 (2.4%)
White	11 / 37 (30%)	19 / 41 (46%)
(Missing)	12	8

¹ n / N (%)

Disorder (ADHD), and rates of substance use problems, all of which have clear public health implications. Early smoke exposure is also linked to self-regulation problems, such as emotional regulation problems.

In this project, our goal is to conduct exploratory data analysis to examine the association between smoking during pregnancy (SDP) and environmental tobacco smoke (ETS) exposure and self-regulation, externalizing behavior, and substance use on adolescent. The women in this study were recruited 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 (Risica et al. 2017). A subset of adolescents (N=100) and their mothers are randomly selected for recruitment into this current study.

2. Data Analysis

2.1. Data Description

The data includes a part of original 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 (Risica et al. 2017). A subset of adolescents (N=100) and their mothers are randomly selected for recruitment into current study, and the data we access is the baseline data for the current study. At this baseline, self-regulatory assessments, self-reported substance use, self-regulation and EKG are recorded.

The data is filtered to the part of baseline data and is partially pre-processed with dropping and computation for comprehension variables; 49 observations of 78 variables are available. Firstly, the data includes the participant ID of parent, the demographic characteristics of the child and the parent, Autism Spectrum Disorder (ASD) information for the child from his/her parent. The summary statistics for demographic characteristics of child and parent are shown in Table 1.

For parents, we find that the age of parent has mean 37.54 with 3.57 standard deviation; 40 parents reported female sex; 15 parents speak another language at home; 13 parents identified as Hispanic or Latino; 26 parents are white; a few parents belong to American Indian/ Alaska Native/ Native Hawaiian or Pacific Islander and others; no Asian or black is reported from parents; 22 parents have full-time paid jobs, with

7 parents having part-time paid jobs and 12 parents not having paid jobs; 15 parents finished some college and 10 parents finished 4-year degree; the family’s annual income has mean 63138.05 with a great standard deviation of 59884.90. For children, their current age has mean 13.62 with standard deviation of 1.21; 23 children are male and 13 children are female; 11 children speak another language at home; 21 children are not Hispanic or Latino, while 15 children belongs to those, and 1 child preferred not to answer; 19 children are white and 15 children are black, as well as a few children are American Indian/ Alaska Native and others; no Asian and Native Hawaiian or Pacific Islander is reported from children.

The data also includes the self-reported smoking during pregnancy and in the immediate postpartum period from mom from original study, the self-reported smoking exposure from mom or partner from 0 months to 5 years postnatal from current study, the urine nicotine metabolite levels at 34 weeks gestation/6 months postpartum from mom and baby; these variables are mainly self-reported answers and fewer lab results, which are representing for SDP and ETS. SDP is based on the self-reported smoke status during pregnancy; ETS is based on the exposure to smoke during pregnancy and also postpartum.

Besides, the data includes the results of emotional regulation from parents and children, which are related to self-regulation; it also includes the attention and externalizing problems based on Brief Problem Monitor from the child and from the parent on self and on child, and also SWAN Rating Scale for ADHD from parent on child, which are related to externalizing problems; it finally includes current substance use of children and also their parents. All of these are evaluated in current study.

2.2. Additional Pre-processing

We will firstly conduct some steps to clean the data and modify the data format. For the parent income variable, there is a record of “250,000”, which is not identified as numeric data automatically, so we edited it to 250000. For the self-reported number of cigarettes smoked per day by mom, there is an answer of “2 black and miles a day”, and we simply change it to 2. However, in a single black and mild, the tobacco is as harmful as 10 to 15 cigarettes and even more, we cannot make sure which number to replace, thus changing to 2 simply is part of limitations and problems of data quality. Besides, there is an answer of “20-25” in the same variable, we use 23 to simply replace it. And there is also an answer of “None” in the same variable, we use 0 to simply replace it. We fill all the blank in the data set as NA and then set the continuous variables as numeric variables and set the categorical ones as factors.

There are four variables from four questions about intensity of child substance use, which are linked to their front questions about whether they have related substance use. If the front question has negative answer of no substance use, then the intensity question will not be asked. Thus there are some NAs related to this. Assume they are consistent, we fill the intensity of child substance use by 0 if they stated never had substance use. Additionally, for two SWAN rates of ADHD, they are some NAs wrongly processed as 0 in the previous pre-processing part, we use NA to replace them. The parent IDs are ‘50502’, ‘51202’, ‘51602’, ‘52302’, ‘53002’, ‘53502’, ‘53902’, ‘54402’, ‘54602’, ‘54702’, ‘55102’ from checking of the raw data. We replace the 0s by NAs.

2.3. Outliers

After exploring the data, firstly there is no duplicated data. Then, I found that there is an record of 44989 in the variable of `mom_numcig`, representing for the self-reported number of cigarettes smoked per day by mom. This is identified as an outlier.

2.4. Missing Data

There is much missing data after pre-processing as Section 2.2. There is no complete entry in this data set and there are 8 records with at least 57 variables missing. Table 2 gives an overview of missingness of each variable. For 5 variables with the first five greatest percentages of missing, there is 39 (79.59%) records of missing in `mom_smoke_pp1` representing for self-reported smoker status at first postpartum visit. 28 (57.14%) records has `childasd` missing, which is the variable for the child ASD status. Then, 20 (40.82%) records has `mom_smoke_pp2` missing, which is the variable for the self-reported smoker status at the second postpartum visit. 16 (32.65%) records are missing for `pmq_parental_control`, which is the average response on the

Parental Knowledge Questionnaire on items related to parental control. 15 (30.61%) records are missing for `ppmq_parental_solicitation`, which is the average response on the Parental Knowledge Questionnaire on items related to parental solicitation.

Table 2: Missing Data in Each Variable

Variable	Number	Percentage(%)	Variable	Number	Percentage(%)
<code>mom_smoke_pp1</code>	39	79.59	<code>nidaalc</code>	10	20.41
<code>childasd</code>	28	57.14	<code>nidatob</code>	10	20.41
<code>mom_smoke_pp2</code>	20	40.82	<code>nidaill</code>	10	20.41
<code>pmq_parental_control</code>	16	32.65	<code>momcig</code>	10	20.41
<code>ppmq_parental_solicitation</code>	15	30.61	<code>mom_numcig</code>	10	20.41
<code>num_alc_30</code>	14	28.57	<code>swan_inattentive</code>	10	20.41
<code>bpm_int</code>	14	28.57	<code>swan_hyperactive</code>	10	20.41
<code>pmq_parental_knowledge</code>	14	28.57	<code>bpm_int_p</code>	10	20.41
<code>pmq_parental_solicitation</code>	14	28.57	<code>smoke_exposure_6mo</code>	10	20.41
<code>bpm_att_p</code>	13	26.53	<code>smoke_exposure_12mo</code>	10	20.41
<code>tsex</code>	13	26.53	<code>smoke_exposure_2yr</code>	10	20.41
<code>num_e_cigs_30</code>	13	26.53	<code>smoke_exposure_5yr</code>	10	20.41
<code>alc_ever</code>	13	26.53	<code>bpm_int_a</code>	10	20.41
<code>erq_cog</code>	13	26.53	<code>erq_cog_a</code>	10	20.41
<code>erq_exp</code>	13	26.53	<code>erq_exp_a</code>	10	20.41
<code>pmq_child_disclosure</code>	13	26.53	<code>mom_smoke_32wk</code>	9	18.37
<code>income</code>	12	24.49	<code>mom_smoke_pp6mo</code>	9	18.37
<code>bpm_ext_p</code>	12	24.49	<code>page</code>	8	16.33
<code>ppmq_parental_knowledge</code>	12	24.49	<code>psex</code>	8	16.33
<code>ppmq_child_disclosure</code>	12	24.49	<code>plang</code>	8	16.33
<code>ppmq_parental_control</code>	12	24.49	<code>pethnic</code>	8	16.33
<code>tage</code>	12	24.49	<code>employ</code>	8	16.33
<code>language</code>	12	24.49	<code>pedu</code>	8	16.33
<code>tethnic</code>	12	24.49	<code>mom_smoke_22wk</code>	7	14.29
<code>cig_ever</code>	12	24.49	<code>mom_smoke_pp12wk</code>	7	14.29
<code>num_cigs_30</code>	12	24.49	<code>mom_smoke_16wk</code>	1	2.04
<code>e_cig_ever</code>	12	24.49	<code>parent_id</code>	0	0.00
<code>mj_ever</code>	12	24.49	<code>paian</code>	0	0.00
<code>num_mj_30</code>	12	24.49	<code>pasian</code>	0	0.00
<code>bpm_att</code>	12	24.49	<code>pnhpi</code>	0	0.00
<code>bpm_ext</code>	12	24.49	<code>pblack</code>	0	0.00
<code>nidapres</code>	11	22.45	<code>pwhite</code>	0	0.00
<code>cotimean_34wk</code>	11	22.45	<code>prace_other</code>	0	0.00
<code>cotimean_pp6mo_baby</code>	11	22.45	<code>taian</code>	0	0.00
<code>cotimean_pp6mo</code>	11	22.45	<code>tasian</code>	0	0.00
<code>smoke_exposure_3yr</code>	11	22.45	<code>tnhpi</code>	0	0.00
<code>smoke_exposure_4yr</code>	11	22.45	<code>tblack</code>	0	0.00
<code>bpm_att_a</code>	11	22.45	<code>twhite</code>	0	0.00
<code>bpm_ext_a</code>	11	22.45	<code>trace_other</code>	0	0.00

Then, we explored the missing pattern across variables. For demographic characteristics for children, as in Table 1, there are 1 record with missing of sex, and 12 records with missing of age, another language, ethnic, and also sex. For the demographic characteristics for parent, there are 4 records with missing of income, as well as 8 records with missing of age, sex, another language, ethnic, employment status, education level and income.

For the variables related to SDP and ETS, we are going to explore the variables related to the self-reported smoking during pregnancy and in the immediate postpartum period from mom, the self-reported smoking exposure from mom or partner from 0 months to 5 years postnatal, the urine nicotine metabolite levels at 34 weeks gestation/6 months postpartum from mom and baby. There is a record with all of these variables missing. Then, separately, for the self-reported smoking during pregnancy and in the immediate postpartum period from mom, from Figure 1 (left), we can see that mainly there are 22 cases with only missingness at first postpartum visit, and there are 9 cases with only missingness at the second postpartum visit; there are some cases with different combinations of missingness here; 1 case shows missingness across all time; most

of missing data is happened at first postpartum visit, as we displayed in Table 2. Then, for the self-reported smoke exposure, from Figure 1 (right) below, we can see that there are 10 cases having all missing data of self-reported exposure, as well as 1 case having missing in the 3rd and 4th year.

Figure 1: The combinations of missingness for self-reported smoker (left) or smoke exposure (right)

For variables related to self-regulation of child and parent, there are 8 cases with all 4 variables missing, 4 cases with missing from child's answers, 2 cases with missing from parent's answers, both cognitive reappraisal and expressive suppression answers from child also have 1 record missing. For variables related to externalizing problems, there are 8 cases with all missing of Brief Problem Monitor results and SWAN rating results, and there are 2 cases with missing from Brief Problem Monitor of child as well as 2 cases with only Brief Problem Monitor of child is complete. Finally, for variables related to substance use, we find that there are 8 cases with all missing of substance use from child and parent, 4 cases with missing of substance use from child, 1 case with missing of substance use from parent, as well as other cases with different combinations.

2.5. Univariate Analysis

Then, Table 3 shows the summary of statistics for variables related to the self-regulation. We can see the difference of mean and standard deviation of average response between parent and the child. The first two rows are for the average response on the Emotion Regulation Questionnaire related to Cognitive Reappraisal from child and the average response on the Emotion Regulation Questionnaire related to Expressive Suppression from child. The last two rows are from parent. We can see that the average responses on the Emotion Regulation Questionnaire related to Cognitive Reappraisal from parents are generally higher than

those from children; but for the average responses on the Emotion Regulation Questionnaire related to Expressive Suppression, we can see the values from parents have a greater range; this can also be observed in Figure 2.

Table 3: Summary statistics for variables related to self regulation

vars	n	mean	sd	median	min	max	range
erq_cog	36	3.19	0.97	3.00	1.00	5.0	4.00
erq_exp	36	2.75	0.80	2.50	1.25	4.5	3.25
erq_cog_a	39	5.38	1.30	5.67	1.00	7.0	6.00
erq_exp_a	39	3.46	1.58	3.25	1.00	7.0	6.00

Note:

Four variables above refer to (1) Average response on ERQ related to Cognitive Reappraisal (child), (2) Average response on ERQ related to Expressive Suppression (child), (3) Average response on ERQ related to Cognitive Reappraisal (parent), (4) Average response on ERQ related to Expressive Suppression (parent).

Table 4 shows the summary of statistics for variables related to externalizing problems. The first two row are from children, the 3rd and 4th rows are from parent on child, the 5th and 6th row are from parent on self. These are about sum of responses on the Brief Problem Monitor (BPM) on items related to attention problems or externalizing problems. Besides, the last two rows are for SWAN rates related to ADHD. We can see that the sum of responses from BPM related to attention problems or externalizing problems from children has higher mean, median, standard deviation and range.

Table 4: Summary statistics for variables related to externalizing problems

vars	n	mean	sd	median	min	max	range
bpm_att	37	3.00	2.62	2	0	9	9
bpm_ext	37	2.81	2.01	3	0	8	8
bpm_att_a	38	1.47	1.96	1	0	8	8
bpm_ext_a	38	1.24	1.57	1	0	6	6
bpm_att_p	36	2.06	2.22	1	0	8	8
bpm_ext_p	37	1.68	2.50	1	0	11	11
swan_hyperactive	39	7.79	6.38	6	0	20	20
swan_inattentive	39	10.92	5.44	11	1	22	21

Note:

Eight variables above refer to (1) the sum of responses on the Brief Problem Monitor on items related to attention problems from child on self, (2) the sum of responses on the Brief Problem Monitor on items related to externalizing problems from child on self, (3) the sum of responses on the Brief Problem Monitor on items related to attention problems from parent on self, (4) the sum of responses on the Brief Problem Monitor on items related to externalizing problems from parent on self, (5) the sum of responses on the Brief Problem Monitor on items related to attention problems from parent on child, (6) the sum of responses on the Brief Problem Monitor on items related to externalizing problems from parent on child, (7) SWAN rating for ADHD- Hyperactive/Impulsive type, (8) SWAN rating for ADHD- Inattentive type.

From Figure 2 (right), we can see that SWAN rates have different scale. As we know a score of 6 or greater indicates the child is likely ADHD, we see that many children are likely ADHD Hyperactive/Impulsive type or ADHD Inattentive type.

2.6. Prenatal and Postnatal Smoke Exposure Intensity Calculation

For the next step, before we analyzing the bivariate relationships, we are going to create two variable for intensity of prenatal and postnatal respectively. For intensity of prenatal, which can also be treated as the intensity of SDP, we are going to sum up the self-reported smoker status at 16, 22 and 32 weeks pregnant,

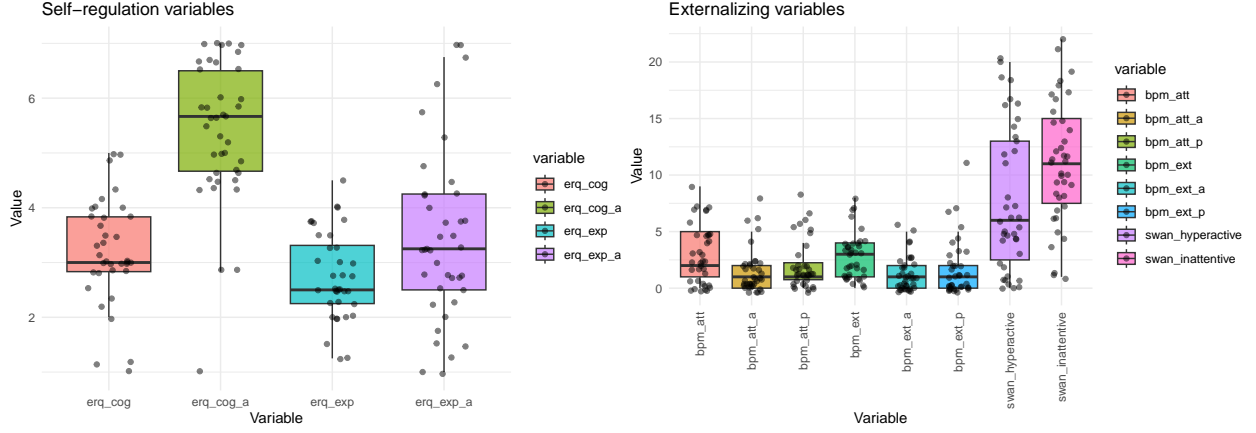


Figure 2: Descriptive Plot

which shows the intensity of smoking during pregnancy. If there is any missing data for the self-reported smoker status at 16, 22 and 32 weeks pregnant, the intensity of prenatal smoke exposure is missing as NA.

As we have the urine cotinean at 34 week gestation from lab, we can refer to the lab results to see if smoker status at 32 weeks pregnant is reliable. Cotinine levels in a nonsmoker are generally less than 10 ng/mL; cotinine levels in a light smoker or someone exposed to secondhand smoke are 11 ng/mL to 30 ng/mL; cotinine levels in a heavy smoker may be more than 500 ng/mL. Most of cotimean results at week 34 pregnant meet our expectation when comparing to self-reported smoker status at week 32 pregnant. However, a parent that had 22.10 ng/mL cotinine level reported non-smoker at 32 weeks pregnant; another parent has 16.94 ng/mL cotinine level reported non-smoker as well; we do not know the real situation whether the parent was a light smoker or was exposed to secondhand smoke, this also might be a data quality problem. While the correlation between the urine cotimean at 34 week gestation and the SDP we created for smoking intensity during pregnancy is 0.81, which makes us think SDP is to some extent reliable.

For the postnatal exposure to smoke, we have the smoke exposure variables from 0 to 6 months, from 7 to 12 months, and in each year from 2nd to 5th years. While for the smoke exposure from 0 to 6 months, there are also 6 other variables related, including the self-reported current smoker at first/ second postpartum visit and at 12 weeks/ 6 months postpartum, as well as urine cotimean at 6 months postpartum from mom and baby. As we explored before, from Table 1 and Figure 1, there are great missing values among the self-reported current smoker at first/ second postpartum visit. We leverage mainly from the self-reported current smoker at 12 weeks/ 6 months postpartum and urine cotimean at 6 months postpartum from mom and baby to revise the smoke exposure from 0 to 6 months, as if one variable shows smoke exposure from 0 to 6 months, we think that there was smoke exposure from 0 to 6 months.

Again, for intensity of postpartum smoke exposure, we sum the smoke exposure variables from 0 to 6 months, from 7 to 12 months, and in each year from 2nd to 5th years, which shows the intensity of smoke exposure of child. If there is any missing data during this period, the intensity of postpartum smoke exposure is missing as NA.

The intensity of SDP refers to the prenatal exposure intensity, while the intensity of ETS refers to both the prenatal exposure intensity and the postnatal intensity.

2.7. Intercorrelations

After checking the correlation between prenatal and postnatal smoke exposure, we find the Pearson's correlation is 0.50, and we can see positive relationship obviously from Figure 3. We guess that they are not very strongly correlated because the smoke exposure postpartum also includes the smoking information from other family members living together; besides, the prenatal data is from the original study, but mainly the postnatal data is from the current study. As in Figure 3, we can see as there are few complete data, the

correlation is not a strong result and may be biased, which is also part of data quality problems.

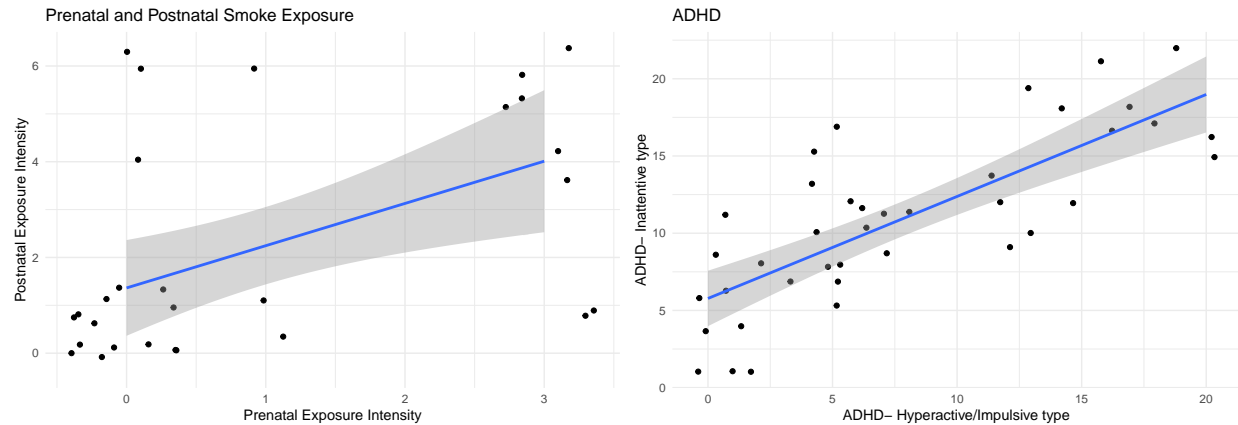


Figure 3: Interrelationship Plots

For the self regulation variables, after calculating correlation coefficients as they are continuous variables, we find that these correlation coefficients are close to 0, with most of them having absolute values less than 0.2. While the average response on the Emotion Regulation Questionnaires related to Expressive Suppression has 0.26 correlation coefficient between parent and child.

For the externalizing variables, we have that the SWAN ratings between ADHD- Hyperactive/Impulsive type and ADHD- Inattentive type show highest correlation of 0.78 significantly, as shown in Figure 3. Two SWAN ratings for ADHD shows significant positive correlation with the other measures for attention and externalizing problems from BPM. Besides, the measures for attention problems from BPM are strongly correlated between from child on self and from parent on child significantly ($\rho = 0.66$), while the measures for externalizing problems are slightly correlated between from child on self and from parent on child and not significant ($\rho = 0.24$).

For the substance use variables, we can see that different categories of substance use are highly positive correlated, significantly. However, as the cases of substance use are limited, those correlations might be biased.

2.8. Bivariate Analysis

Next, we are going to analysis bivariate relationships for to meet our research goal to examine the association between SDP/ETS and self-regulation, externalizing behavior, and substance use on adolescent.

2.8.1. Association between SDP/ETS and Self-regulation

We firstly check the association between prenatal smoke exposure and self-regulation problems on adolescent. We use both the self-reported smoker status from mom at 16/ 22/ 32 weeks pregnant and also the prenatal exposure intensity we created before. From correlation matrix, the prenatal smoke exposure has a little positive relationship with self-regulation problems on ERQ related to cognitive reappraisal ($\rho = 0.090$) and ERQ related to expressive suppression ($\rho = 0.109$) on adolescent.

Then, we check the association between postnatal smoke exposure and self-regulation problems on adolescent. We find that the correlation between postnatal exposure intensity and the average response on the Emotion Regulation Questionnaires related to Expressive Suppression from child is 0.424, which is relatively high and significant. Furthermore, we use the time-series smoke exposure variables, and find the correlation coefficients between the smoke exposure in 2nd, 3rd, 4th and 5th year and the average response on the Emotion Regulation Questionnaires related to Expressive Suppression from child is around 0.48 to 0.50 significantly. And we can also see the positive relationship in Figure 4. Children with smoke exposure in 2nd, 3rd, 4th and 5th year has more cases of higher expressive suppression. Referred to the questions related

to Expressive Suppression from Emotion Regulation Questionnaires (Gross and John 2003), we hold the opinion that a higher score on expressive suppression subscale indicates a greater tendency to use expressive suppression as an emotional regulation strategy, such as inhibiting or masking outward emotional expressions, which means worse situation of self regulation. Thus, I think that expressive suppression is associated with smoke exposure during 2nd, 3rd, 4th and 5th year.

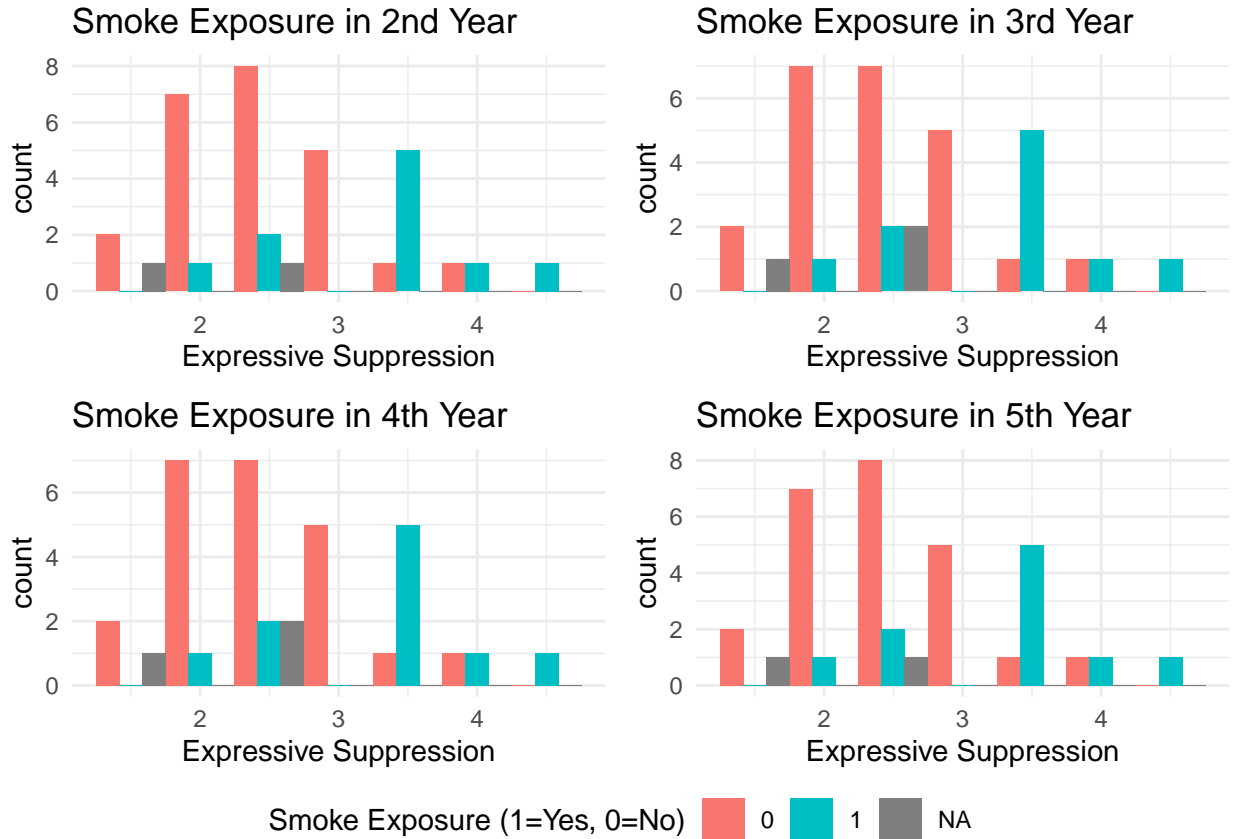


Figure 4: Smoke exposure during 2nd, 3rd, 4th and 5th year versus expressive suppression

2.8.2. Association between SDP/ETS and Externalizing Problems

We firstly check the association between prenatal smoke exposure and externalizing problems on adolescent. We use both the self-reported smoker status from mom at 16/ 22/ 32 weeks pregnant and also the prenatal exposure intensity we created before. We do not find useful pattern from 3 self-reported smoker status, while there are some pattern using the prenatal intensity variable. We find that the correlation between prenatal exposure intensity and the sum of responses on the Brief Problem Monitor on items related to attention problems from parent on child is 0.51 significantly, and the correlation between prenatal exposure intensity and the SWAN rating for ADHD- Hyperactive/Impulsive type is 0.58 significantly. We can also see the relationships from Figure 5. Higer intensity of prenatal smoke exposure means possibly more serious attention problems and ADHD problems for adolescents. However, as the number of cases is limited, the result is not very strong.

Next, for the postnatal smoke exposure, though the intensity does not show strong relationship with externalizing problems, we further use time-series variables of smoke exposure and find some useful pattern. Especially for the smoke exposure during 0 to 6 months, it is strongly correlated with sum of responses on the Brief Problem Monitor for attention problems from child on self ($\rho = 0.39$) and from parent on child ($\rho = 0.47$), and sum of responses on the Brief Problem Monitor for externalizing problems from child on self ($\rho = 0.40$) and from parent on child ($\rho = 0.43$), as well as SWAN rating related to ADHD- Hyper-

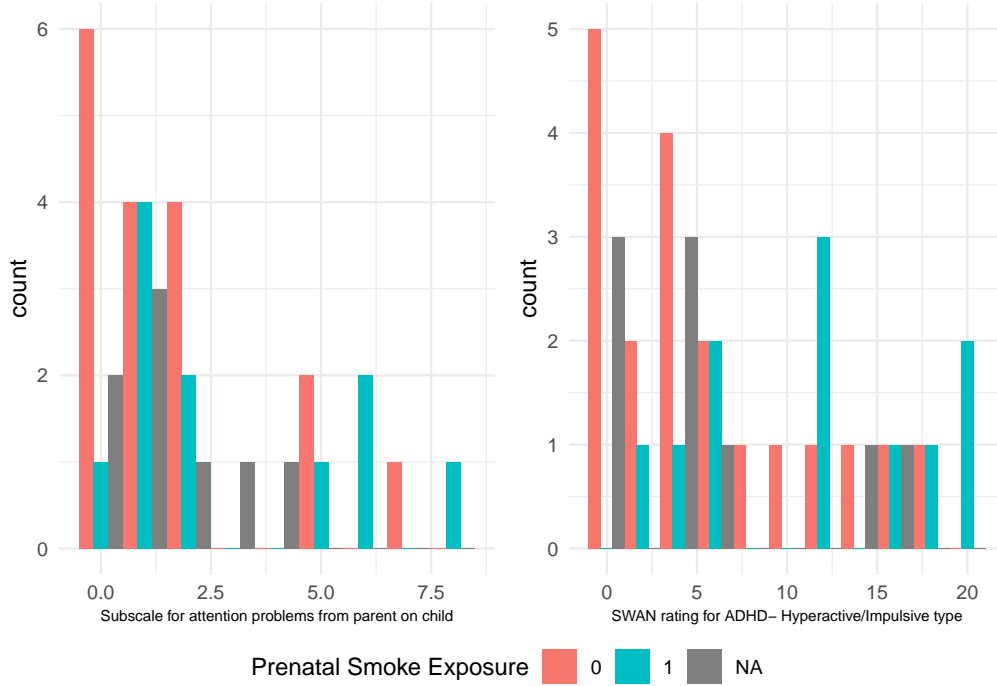


Figure 5: Prenatal Exposure Intensity and Externalizing Problems

active/Impulsive type ($\rho = 0.45$) and ADHD- Inattentive type ($\rho = 0.56$). Scatter plots are displayed in Figure 10 and we can see some positive relation to some extent. Besides, the sum of responses on the Brief Problem Monitor for externalizing problems from parent on child is also significantly correlated with smoke exposure in 3rd year ($\rho = 0.39$) and in 4th year ($\rho = 0.39$); the sum of responses on the Brief Problem Monitor for attention problems from parent on child is also significantly correlated with smoke exposure in 5th year ($\rho = 0.46$).

Referred to the BRP (ASEBA 2019) and also SWAN rating scales, we know that the higher response in BRP and the higher SWAN ratings means greater externalizing problems. Based on analysis above, we find that the smoke exposure is associated with externalizing problems, especially the smoke exposure from 0 to 6 months.

2.8.3. Association between SDP/ETS and Substance Use

As Section 2.2. discussed, we filled the number of substance use in last 30 days as 0 if the child states that they never had such substance use. After exploring the pairwise relationship, we find prenatal smoke exposure has mildly positive correlation with the substance use on adolescent, while postnatal smoke exposure shows slight negative correlation with the substance use on adolescent. As limited cases of substance use on adolescent from the data set, it is hard to find useful patterns here, and the correlation may be biased, which is part of data quality problem.

3. Conclusions and Limitations

Exposure to smoking during pregnancy (SDP) and environmental tobacco smoke (ETS) are common and pose serious hazards as environmental factors affecting children. The exposure to smoke during early stages of development has been found to be associated with higher prevalence of externalizing behaviors in children, such as Attention-Deficit/Hyperactivity Disorder (ADHD), as well as increased rates of substance abuse problems (Micalizzi 2023; Paul 2023). In this exploratory data analysis project, our goal is to examine the

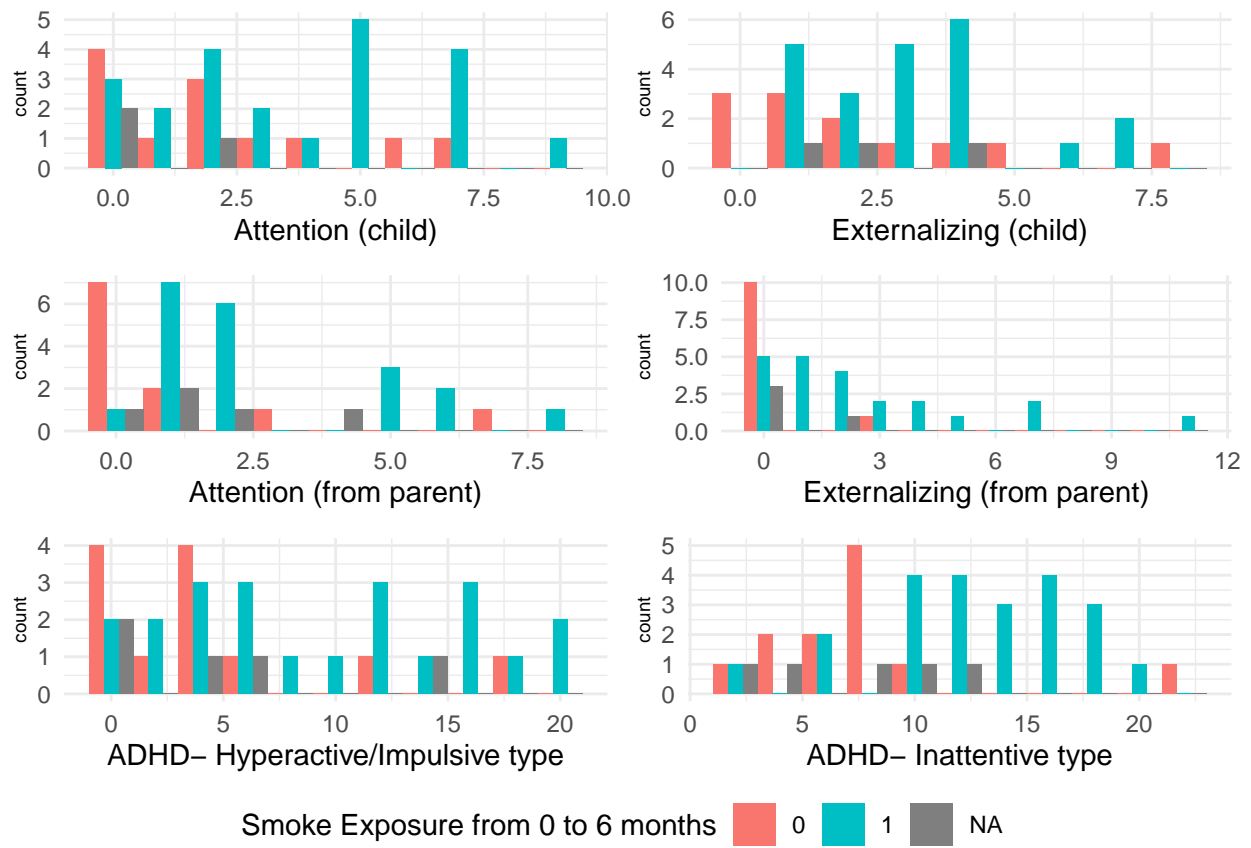


Figure 6: Smoke exposure from 0 to 6 months versus externalizing problems

association between smoking during pregnancy (SDP) and environmental tobacco smoke (ETS) exposure and self-regulation, externalizing behavior, and substance use on adolescent. Based on bivariate analysis, we find different stages of early smoke exposure are associated with self-regulation, externalizing problems and substance uses, respectively.

For interrelationship, we find that the prenatal and postnatal smoke exposure are mildly positive correlated ($\rho = 0.50$). They are not strongly correlated as the reason might be the prenatal smoke exposure is only from self-reported mom smoker status, but the postnatal smoke exposure combines the information from other family members; and the prenatal data is from the original study, but the postnatal data is mainly from current study. Besides, as the complete cases of smoking exposure across the original study and current study of whole timeline is rare, the analysis is limited and may be biased. For variables related to self-regulation, externalizing, and substance use, we find some of them show strongly intercorrelations, respectively. To some extent, this also shows the data meets consistency.

For bivariate associations, firstly, for self-regulation, we find that the smoke exposure during 2nd to 5th year is associated with expressive suppression, significantly. For externalizing problems, the smoke exposure from 0 to 6 months is associated with attention problems, externalizing problems, and two types of ADHD; the prenatal smoke exposure is associated with attention problems evaluated by parent on child and ADHD-Hyperactive/Impulsive type. Finally, for substance use, we find prenatal smoke exposure has mildly positive correlation with the substance use on adolescent, while postnatal smoke exposure shows slight negative correlation with the substance use on adolescent. Due the limited sample size, the association might be spurious.

For the limitations, firstly, as mentioned before, there are only 49 observations at baseline and there are many missing values in the data set. The sample size is quite small when we analyze some variables and require this part of data complete. This may lead to biased and insignificant results, as some special data records may totally change the analytics results. Secondly, some parts of data are self-reported and some are from reflection, which are part of data quality problems. Besides, we focus on analyze univariate and bivariate relationships. Without adjusting for the other covariates, the results is biased and not comparable among different groups of smoking exposure. For further study, we need control the other covariates properly and try to fit proper models to measure the association between SDP/ETS and self-regulation, externalizing behavior, and substance use on adolescent.

References

- ASEBA (2019), “[ASEBA BPM/6-18™ - ASEBA](#).”
- Gross, J. J., and John, O. P. (2003), “Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being,” *Journal of Personality and Social Psychology*, 85, 348–362. <https://doi.org/10.1037/0022-3514.85.2.348>.
- Micalizzi, L. (2023), “Prenatal Tobacco Exposure & Child Outcomes.”
- Paul, A. (2023), “[Project 1: Exploratory Data Analysis](#).”
- Risica, P. M., Gavarkovs, A., Parker, D. R., Jennings, E., and 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>.

Appendix

Table 5: Summary Statistics for SDP

Variable	N = 49
mom_smoke_16wk	
0	36 / 48 (75%)
1	12 / 48 (25%)
(Missing)	1
mom_smoke_22wk	
0	29 / 42 (69%)
1	13 / 42 (31%)
(Missing)	7
mom_smoke_32wk	
0	30 / 40 (75%)
1	10 / 40 (25%)
(Missing)	9
mom_smoke_pp1	
0	7 / 10 (70%)
1	3 / 10 (30%)
(Missing)	39
mom_smoke_pp2	
0	22 / 29 (76%)
1	7 / 29 (24%)
(Missing)	20
mom_smoke_pp12wk	
0	30 / 42 (71%)
1	12 / 42 (29%)
(Missing)	7
mom_smoke_pp6mo	
0	24 / 40 (60%)
1	16 / 40 (40%)
(Missing)	9
cotimean_34wk	
Mean (SD); Median (Range)	49.75 (97.69); 1.04 (0.00, 382.77)
(Missing)	11
cotimean_pp6mo_baby	
Mean (SD); Median (Range)	4.04 (7.62); 1.50 (0.00, 41.68)
(Missing)	11
cotimean_pp6mo	
Mean (SD); Median (Range)	100.49 (179.34); 14.79 (0.00, 878.96)
(Missing)	11

¹ n / N (%)

Table 6: Summary Statistics for ETS

Variable	N = 49
smoke_exposure_6mo	
0	29 / 39 (74%)
1	10 / 39 (26%)
(Missing)	10
smoke_exposure_12mo	
0	30 / 39 (77%)
1	9 / 39 (23%)
(Missing)	10
smoke_exposure_2yr	
0	28 / 39 (72%)
1	11 / 39 (28%)
(Missing)	10
smoke_exposure_3yr	
0	27 / 38 (71%)
1	11 / 38 (29%)
(Missing)	11
smoke_exposure_4yr	
0	28 / 38 (74%)
1	10 / 38 (26%)
(Missing)	11
smoke_exposure_5yr	
0	29 / 39 (74%)
1	10 / 39 (26%)
(Missing)	10

¹ n / N (%)

Code Appendix

All code is compatible with R version 4.1.2, and can be found in the *Project 1 Exploratory Data Analysis* folder at [Github](#).