Project 1
Github Link

Zihan Zhou

2023-10-08

1. Introduction

Exposure to smoking during pregnancy (SDP) and environmental tobacco smoke (ETS) are two of the major public health concerns. These exposures are associated with an increased risk of externalizing behaviors, such as Attention-Deficit/Hyperactivity Disorder, rates of substance use problems and self-regulation problems, which involve maintaining physiological, emotional, behavioral, and cognitive control. The original study (Risica et al. 2017) evaluated the efficacy of a tailored video intervention versus usual care of reducing ETS exposure among 738 pregnant individuals, who are either smokers or non-smokers. Newsletters and videos focused on smoking cessation were assigned to intervention participants, while content focusing on general healthy pregnancy topics were assigned to comparison participants. Based on this study, Dr. Micalizzi re-recruited around 100 mothers and children from the original study and collected data. The purpose of this EDA (Exploratory Data Analysis) is to investigate the effects of SDP/ETS on adolescent self-regulation, substance use, and externalizing behaviors.

The data used in this EDA comes from both the original and current studies. It includes baseline data along with two longitudinal follow-up assessments scheduled for 6 and 12 months post-baseline. The dataset contains 49 observations, representing 49 mothers and their 49 children. There are a total of 79 variables, including demographic characteristics for both mothers and children, smoke exposure information, computerized self-regulatory assessments, self-reported substance use, and self-regulation based on both parent-reported and child-reported behaviors.

2. Data Quality Check

2.1. Data Type and Recording

Before further investigating the data, it is important to conduct a data quality check. Firstly, I checked the data type of each variable. There are 9 character variables, 54 integer variables, and 15 numeric variables. However, according to the codebook of the data, some variables have incorrect types. For instance, among the demographic variables, 'sex' (where 0 = Male, 1 = Female, 2 = Intersex) should be categorical (character in R) instead of numeric. The 'income' variable, which should be a continuous numeric variable, is coded as a character type. Other variables like smoke exposure and substance use, which use numbers to represent different responses (1 = Yes, 2 = No), should also be categorized as categorical rather than numeric.

In addition, there are some data recording discrepancies. For 'income', one entry is recorded as the character string '250, 000' instead of the numeric value 250000, leading to the incorrect data type mentioned earlier. In the 'mom_numcig' variable, which logs the number of cigarettes a mother consumes daily, there are two strange records: "2 black and miles a day" and "44989". The former value likely should be recorded as '2', while the latter is obviously an outlier by common knowledge; I converted it to NA. Another entry in the same variable shows a range "20-25" instead of a exact number. I imputed this as the mean, 22.5, for convenience

in subsequent analysis. Furthermore, there are some blank entries and the character string "none" in the data frame and all such entries should be converted to NAs. It's also worth noting the inconsistency in response recording between the two studies. In the original study, a response was captured with 1 for 'Yes' and 2 for 'No', whereas in the current study, 0 represents 'No' and 1 represents 'Yes'. To maintain consistency, I converted all responses from the original study to the 0 and 1 format to match the current study.

2.2. Missing Data

One of the most common issue in data analysis is missing values. In our dataset, 63 out of the total 79 variables have missing values, thus making it essential to find out the missing pattern. The extend of missingness varies significantly across variables. Variables related to the mother's smoking status during pregnancy, such as 'mom_smoke_16wk', have around 18.37% missing data. Demographic details about mother, such as 'page', 'psex' and 'pedu', collectively show a missingness rate of approximately 16.33%. Variable representing the number of days a child smoked in the past 30 days ('num_cigs_30') has the highest missingness at 97.96%. Following are variables related to e-cigarette usage and marijuana consumption in the past 30 days, 'num_e_cigs_30' and 'num_mj_30', having missing data of 95.92% and 93.88% respectively. This kind of high missingness is due to that these questions are posed only if the participant has previously reported substance use. Variable indicating self-reported smoking status at 16 weeks into the pregnancy (mom_smoke_16wk) has the least missingness at only 2.04%. Moreover, some variables are missing together such as the mother's demographic information like 'page' and 'employ', substance use and exposure variables like 'nidaalc' and 'nidatob', smoke exposure data for the third and fourth years ('smoke_exposure_3yr' and 'smoke_exposure_4yr'), and children's demographic details, substance use indicators, and brief problem monitor responses, such as 'tage', 'cig_ever', and 'bpm_att'. Such patterns of missingness could result from participants' personal preferences or the study design.

Table 1: Summary of Missing Values

Variable	Number	Pct	Variable	Number	Pct	Variable	Number	Pct
num_cigs_30	48	97.96%	ppmq_parental_knowledge	12	24.49%	nidaill	10	20.41%
num_e_cigs_30	47	95.92%	ppmq_child_disclosure	12	24.49%	momcig	10	20.41%
num_mj_30	46	93.88%	ppmq_parental_control	12	24.49%	bpm_int_p	10	20.41%
num_alc_30	45	91.84%	tage	12	24.49%	smoke_exposure_6mo	10	20.41%
mom_smoke_pp1	39	79.59%	language	12	24.49%	smoke_exposure_12mo	10	20.41%
childasd	28	57.14%	tethnic	12	24.49%	smoke_exposure_2yr	10	20.41%
mom_smoke_pp2	20	40.82%	cig_ever	12	24.49%	smoke_exposure_5yr	10	20.41%
pmq_parental_control	16	32.65%	e_cig_ever	12	24.49%	bpm_int_a	10	20.41%
ppmq_parental_solicitation	15	30.61%	mj_ever	12	24.49%	erq_cog_a	10	20.41%
bpm_int	14	28.57%	bpm_att	12	24.49%	erq_exp_a	10	20.41%
pmq_parental_knowledge	14	28.57%	bpm_ext	12	24.49%	mom_smoke_16wk	1	2.04%
pmq_parental_solicitation	14	28.57%	nidapres	11	22.45%	mom_smoke_32wk	9	18.37%
bpm_att_p	13	26.53%	cotimean_34wk	11	22.45%	mom_smoke_pp6mo	9	18.37%
tsex	13	26.53%	cotimean_pp6mo_baby	11	22.45%	page	8	16.33%
alc_ever	13	26.53%	cotimean_pp6mo	11	22.45%	psex	8	16.33%
erq_cog	13	26.53%	smoke_exposure_3yr	11	22.45%	plang	8	16.33%
erq_exp	13	26.53%	smoke_exposure_4yr	11	22.45%	pethnic	8	16.33%
pmq_child_disclosure	13	26.53%	bpm_att_a	11	22.45%	employ	8	16.33%
income	12	24.49%	bpm_ext_a	11	22.45%	pedu	8	16.33%
mom_numcig	12	24.49%	nidaalc	10	20.41%	mom_smoke_22wk	7	14.29%
bpm_ext_p	12	24.49%	nidatob	10	20.41%	mom_smoke_pp12wk	7	14.29%

3. Demographic Characteristics

There are some common demographic variables for both mother and child. The children median age is 14, while for parents, it is 37. Gender distribution shows 64% of children are male, whereas 98% of the parents are female. In terms of language spoken at home, 70% of children and 63% of parents do not speak another

language at home. The dataset shows diversity in race and ethnity with 52% of children identifying as Hispanic/Latino and 32% of parents, but it's worth noting that future studies might benefit from including a broader array of racial backgrounds to better generalize results for the population. The detailed information are presented in Table 2.

There are also some socio-economic variables unique to parents, which are 'employ', 'pedu' and 'income'. The data indicates that 54% of parents work full-time, 17% work part-time, and 29% are currently not employed. While the predominant education level is 'some college' at 37%, postgraduate degrees are the least common, accounting for only 4.9%. The median parental income is $46,848.

Table 2: Summary for Demographic Variables

| Characteristic	Child	Parent
Age	14 (13, 15)	37 (35, 39)
Unknown	12	8
Sex		
Male	23 (64%)	1 (2.4%)
Female	13 (36%)	40 (98%)
Unknown	13	8
Language Spoken at Home	11 (30%)	15 (37%)
Unknown	12	8
Race/Ethnicity		
American Indian/Alaska Native	3 (10%)	4 (9.8%)
Hispanic/Latino	15 (52%)	13 (32%)
Native Hawaiian/Pacific Islander	0 (0%)	4 (9.8%)
White	11 (38%)	19 (46%)
Other	0 (0%)	1 (2.4%)
Unknown	20	8

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

4. Impact of Smoking During Pregnancy (SDP) and Environmental Tobacco Smoke (ETS)

In our dataset, there are a series of variables that are related to SDP and ETS. By combining both timing and dosage effects, we can derive two independent variables (IVs).

4.1. SDP

To evaluate the influence of SDP, we must consider both timing and dosage effects from the data variables. In our dataset, 4 variables pertain to SDP: 'mom_smoke_16wk', 'mom_smoke_22wk', 'mom_smoke_32wk', 'cotimean_34wk'. These were assessed in the original study. The first three variables represent self-reported current smoking status, with 1 indicating "Yes" and 0 indicating "No", which is asked at week 16, 22 and 32 during pregnancy respectively. In contrast, the last variable measures mothers' urine cotinine levels tested at 34 weeks of gestation. The following plot shows the distribution of the smoke status during pregnancy.

Distribution of Smoke Status During Pregnancy across Different Weeks
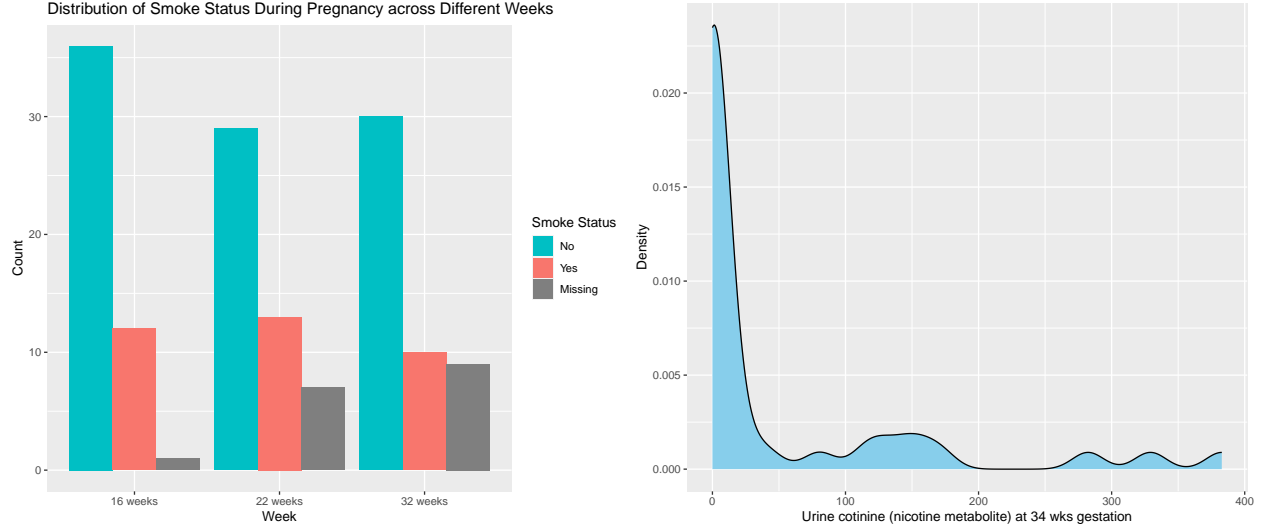
In order to create a new SDP variable, I first aim to assess the accuracy of mothers' self-reported smoking status by comparing their responses at week 32 to the urine cotinine levels at week 34. Typically, non-smoker is defined as a urine cotinine level of 10 ng/mL or less, while active smokers have cotinine levels ranging from 10-500 ng/mL(Pirkle 1996). Although there are 11 missing entries, I was able to identify 2 discrepancies in self-reporting, which could be due to either dishonest reporting or exposure to secondhand smoke. So when creating the new SDP variable, I would prioritize using 'cotimean_34wk' over 'mom_smoke_32wk' if 'cotimean_34wk' data is available. At the four time points, a smoker is identified either by a 'yes' response or by cotinine levels exceeding 10. Some mothers reported smoking at week 16 but ceased later, while others reported being non-smokers at week 16 but resumed at week 22 before quitting again.

Consequently, the mothers can be classified into three categories based on their smoking frequency: non-smoker, light-smoker, and heavy-smoker. A non-smoker is defined as a mother who reported as such across all three time points and exhibited urine cotinine levels below 10 at week 34. A light-smoker is a mother identified as a non-smoker at least once but not all time during the four time points. A heavy smoker, on the other hand, is consistently identified as a smoker throughout. For NA values of certain observations at less than 2 certain time points, I use the available information to categorize them. Finally, I identified 10 heavy-smoker, 5 light-smoker and 30 non-smoker.

Then I use this new SDP variable to assess the impact of SDP on the outcome we interested.

4.1.1. Adolescent Self-regulation

According to description of the project, self-regulation problems encompass the maintenance of physiological, emotional, behavioral, and cognitive control. I have selected variables 'bpm_att', 'erq_cog', 'erq_exp', 'bpm_att_p', 'erq_cog_a', 'erq_exp_a', and 'childasd'. The 'bpm' variables come from the Brief Problem Monitor, where both adolescents and parents report the accuracy of specific statements about themselves or their children with questions to rate Internalizing, Attention, and Externalizing problems. The 'erq' variables come from the Emotion Regulation Questionnaire, representing cognitive reappraisal and expressive suppression. The 'childasd' variable is a metric that reflects suspicions or diagnoses of Autism Spectrum Disorder (ASD) in the child. I include this variable because both externalizing and internalizing problems can be observed in individuals with ASD.

Table 3 presents the results, showing the differences in the self-regulation outcomes based on the SDP categories. Adolescents from the "heavy-smoker" group, on average, show highest median values in the 'bpm_att' and 'bpm_att_p', indicating potential greater attention problems. However, they exhibit moderate levels in both cognitive reappraisal and expressive suppression. Adolescents from "non-smoker" group have the lowest median values in the 'erq_exp' and 'erq_exp_a', while having higher median values in the 'erq_cog'

and 'erq_cog_a'. It is also interesting to observe one child from the "non-smoker" group have diagonosed ASD, while none were reported in the other two categories.

Table 3: Comparation of Self-regulation Metrics Across SDP Categories

Characteristic	heavy-smoker, N = 10	light-smoker, N = 5	non-smoker, N = 30
bpm_att	5 (2, 7)	2 (0, 4)	2 (0, 5)
erq_cog	3.00 (3.00, 3.42)	3.42 (3.00, 3.92)	3.33 (2.83, 3.92)
erq_exp	2.63 (2.50, 3.56)	3.25 (3.25, 3.88)	2.50 (2.00, 3.00)
bpm_att_p	4 (1, 6)	1 (1, 2)	1 (0, 2)
erq_cog_a	5.25 (4.33, 5.75)	5.50 (4.08, 6.63)	5.67 (4.79, 6.67)
erq_exp_a	4.00 (2.38, 5.88)	6.00 (4.63, 6.81)	2.88 (2.25, 3.75)
childasd	0 (0%)	0 (0%)	1 (8.3%)

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

The following heatmap shows the interrelationships among self-regulation variables. Notably, 'bpm_att' (adolescents' self-assessment of attention problems) demonstrates a strong positive correlation of 0.50 with 'bpm_att_p' (parents' perspective on their child's attention problems), indicating a consistency between adolescent self-reports and parental observations. Moreover, 'bpm_att_p' also shows a high positive correlation of 0.53 with 'erq_exp_a', which means that greater attention problems might be associated with higher expressive suppression. The variable 'childasd' also shows moderate positive correlations with 'bpm_att_p' and 'erq_exp_a', suggesting potential relations between ASD and attention and expressive issues. Other variables only present low correlations, which are not high enough to deduce the interrelationships.

Correlation Heatmap among Self−regulation Variables
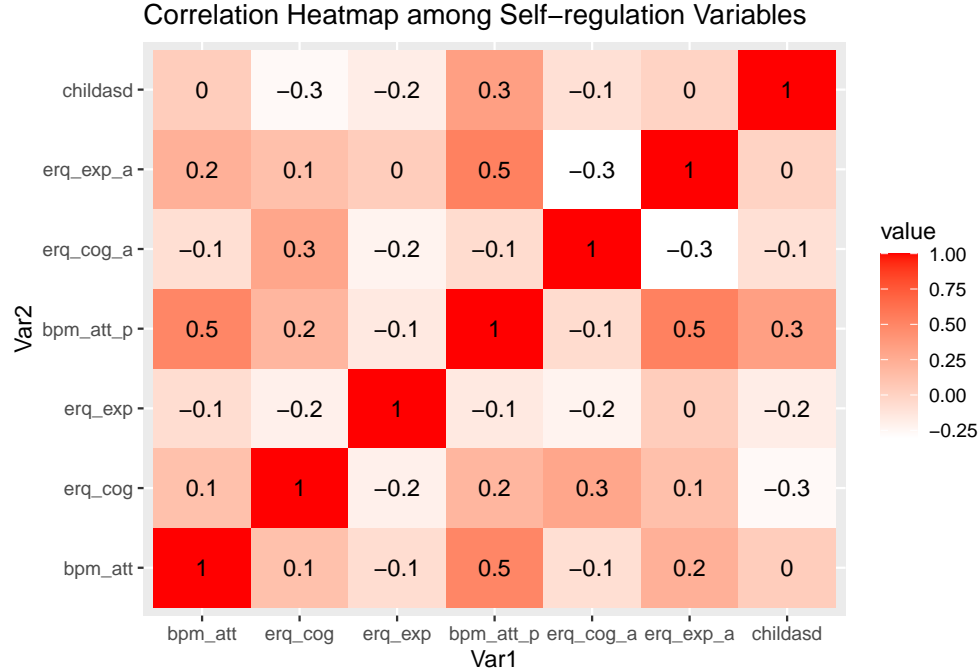
4.1.2. Adolescent Substance Use

The following bar plot shows the proportion of adolescents with substance use across the three SDP categories. Adolescents whose mothers were categorized as "heavy-smokers" during pregnancy show the highest rates of

cigarette and marijuana use, with 13% having tried cigarettes and 25% having tried with marijuana. None of the adolescent form the "non-smoker" category reported cigarette use. E-cigarette use occur across the categories, with adolescents from the "light-smoker" group showing the highest proportion at 25%. In the "non-smoker" category, e-cigarette and marijuana use was reported by a relatively small proportion, at 4.3% each. These children also have the lowest proportion in the alcohol use at 8.7%. In contrast, adolescents from the "heavy-smoker" and "light-smoker" categories reported considerably higher alcohol consumption rates, at 29% and 25%, respectively. In terms of the frequency of substance use over the past 30 days—which is asked only if substance use is responded as "Yes" —one adolescent in the "heavy-smoker" group disclosed using e-cigarettes twice. Additionally, two children in the same group mentioned using marijuana 3 and 12 times, respectively and one adolescent in the "non-smoker" group reported using marijuana 18 times. For alcohol, two individuals from the "heavy-smoker" category consumed it once and 10 times respectively. Data regarding number of cigarette consumption was unavailable. It seems that the "heavy-smoker" group tends to report substance use in the past 30 days at a higher frequency.

Smoking during pregnancy appears to have varying influences on adolescent substance use. The data suggests that adolescents with a history of heavier maternal smoking during pregnancy may be more susceptible to substance use, particularly cigarettes and marijuana. However, given the small sample sizes, further research with larger cohorts is needed to further validate these findings and draw more accurate conclusions.

Proportion of Adolescents with Substance Use Across SDP Categories
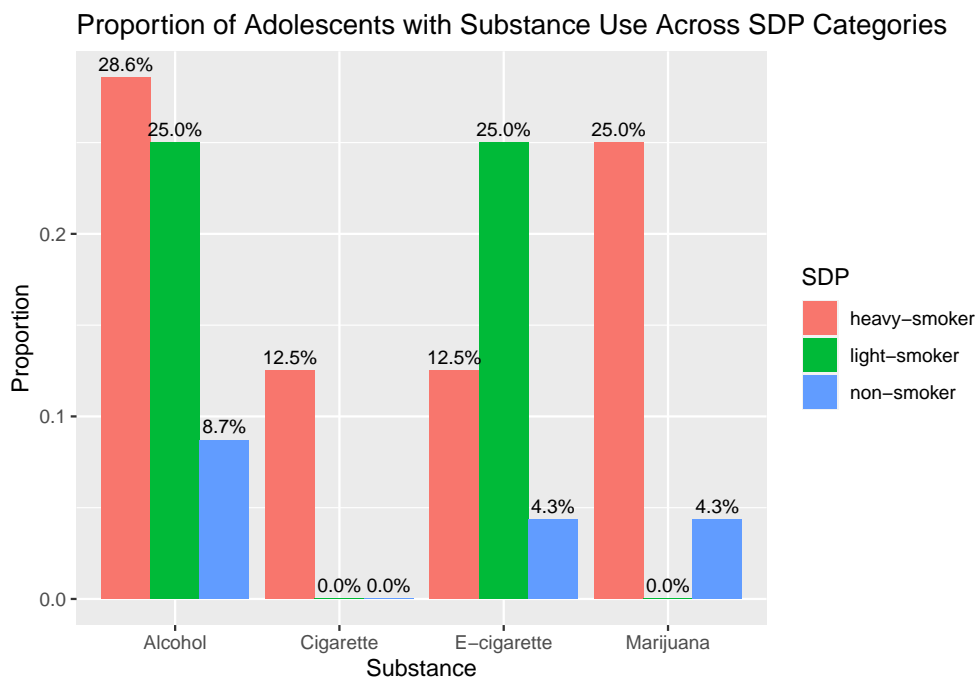
4.1.3. Adolescent Externalizing Problems

Externalizing behaviors in children include Attention-Deficit/Hyperactivity Disorder, so when analyzing the effect of SDP on adolescent externalizing problems, in addition to 'bpm_ext', 'bpm_ext_p' and 'childasd', I included 'swan_hyperactive' and 'swan_inattentive'. These two variables are the sum of responses on SWAN Rating Scale Qs 1-9: a score of 6 or greater indicates the child is likely ADHD- Hyperactive/Impulsive type and Inattentive type respectively.

Table 4 shows the summary statistics for these five variables stratified by the three SDP categories. In the "non-smoker" group, adolescents consistently show the lowest median scores across 'bpm_ext', 'bpm_ext_p', hyperactivity, and inattentiveness, suggesting fewer perceived externalizing issues. In contrast, children in the "heavy-smoker" group have the highest median scores in 'bpm_ext', 'bpm_ext_p' and inattentiveness, which

is much higher than children with non-smoker mother during pregnancy. For the "light-smoker" group, the data indicates moderate levels of externalizing problems. It is notable that in this group, 'swan_hyperactive' has the highest median. All these results appear to suggest that heavier smoking during pregnancy might be linked to increased externalizing problems in offspring.

Table 4: Comparation of Externalizing Problems Across SDP Categories

Characteristic	heavy-smoker, N = 10	light-smoker, N = 5	non-smoker, N = 30
bpm_ext	4 (3, 5)	3 (2, 5)	2 (1, 4)
bpm_ext_p	3 (0, 4)	1 (0, 4)	1 (0, 2)
childasd	0 (0%)	0 (0%)	1 (8.3%)
swan_hyperactive	10 (3, 18)	12 (6, 13)	4 (0, 7)
swan_inattentive	14 (3, 17)	10 (9, 12)	8 (4, 12)

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

4.2. ETS

Similarly, to evaluate the impact of ETS, we must consider both the frequency of ETS exposure and urine cotinine levels from both mother and baby. This data was collected in both the original and current studies. In the original study, mothers self-reported their smoking status at 12 weeks and 6 months postpartum. Urine cotinine levels were also tested for both mother and baby at 6 months postpartum. In the current study, mothers were asked to recall postpartum smoke exposure from both themselves and their partners during several time intervals: 0-6 months, 7-12 months, and the 2nd to the 5th year. It is important to note that the 6-month postpartum time point overlaps in reports from both the original and current studies. For this specific time point, if there was any reported smoking or if the urine cotinine level of mom or baby exceeded 10 ng/mL, I would code it as 1, indicating that there was smoke exposure at 6 months postpartum. The following plot shows the distribution of postnatal smoke exposure across different time points. The density plot of the urine cotinine levels in mothers and babies at 6 months postpartum reveals a broader distribution for mothers and mothers tend to have significantly higher cotinine levels compared to babies. The distribution of mothers' cotinine levels at 6 months is similar to that at 34 weeks gestation.
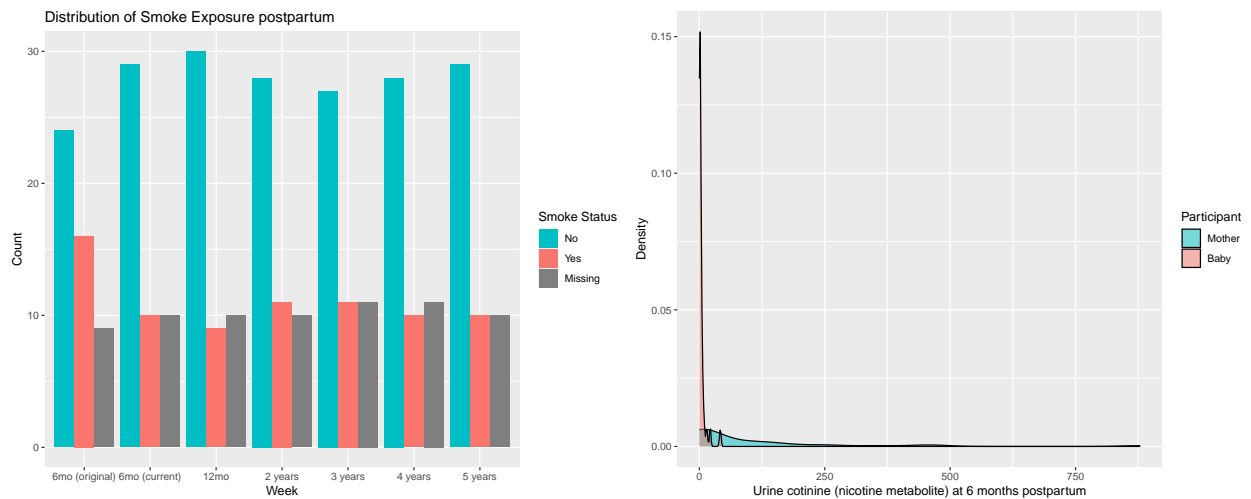
Following the same logic used when creating the SDP variable, ETS has three levels: no-exposure, light-exposure, and heavy-exposure. A child is classified as having no-exposure if no smoke exposures were reported

across all six time points. Light-exposure is when there were 1-2 time points reported smoke exposure. Heavy-exposure is classified when more than 2 time points have reported smoke exposure. Finally, I identified 10 heavy-exposure, 10 light-exposure and 13 no-exposure.

4.2.1. Adolescent Self-regulation

Table 5 shows the results of self-regulation variables across the three ETS categories. Except for variables associated with cognitive reappraisal ('erq_cog' and 'erq_cog_a') and 'childasd', adolescents in the heavy-exposure category consistently exhibit the highest median values for the other variables. Adolescents in the category "light-exposure" have the highest median in 'erq_cog' and share highest median with the "no-exposure" category for 'erq_cog_a'. Only one child in the category "no-exposure" was suspected with ASD. This table suggests a potential positive association between heavier smoke exposure and self-regulation problems regarding attention and expressive suppression.

Table 5: Comparison of Self-regulation Metrics Across ETS Categories

Characteristic	heavy-exposure, N = 10	light-exposure, N = 10	no-exposure, N = 13
bpm_att	4 (2, 7)	2 (1, 5)	2 (0, 3)
erq_cog	3.00 (3.00, 3.50)	3.33 (2.83, 4.33)	3.25 (2.71, 4.00)
erq_exp	3.50 (2.50, 3.75)	2.50 (2.25, 3.00)	2.50 (2.00, 3.00)
bpm_att_p	2 (1, 5)	2 (1, 2)	0 (0, 1)
erq_cog_a	5.08 (4.37, 6.21)	5.67 (4.87, 5.96)	5.67 (4.67, 6.67)
erq_exp_a	4.25 (2.56, 6.13)	3.13 (2.25, 3.75)	3.25 (2.75, 4.25)
childasd			
0	4 (100%)	7 (100%)	3 (75%)
2	0 (0%)	0 (0%)	1 (25%)

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

4.2.2. Adolescent Substance Use

The following bar plot illustrates the substance use of adolescents categorized by their smoke exposure levels. It is worth noting that across all types of substances, adolescents in the "no-exposure" group have the lowest proportion of reported substance use. There is only one child reported e-cigarette use, leading to a proportion of 8.3% in the "no-exposure" category. In contrast, the "light-exposure" group has the highest proportions of substance use across all categories. For alcohol consumption, both the heavy-exposure and light-exposure groups reported an identical usage rate of 11%. In the case of traditional cigarettes, only adolescents in the "light-exposure" group reported any use, accounting for 11%, while no use was reported in the other groups. For e-cigarettes, the "heavy-exposure" and "light-exposure" groups have very close proportions, with 10% and 11% respectively. Marijuana consumption was only reported in the "light-exposure" group with a proportion of 33%. These figures suggest that while substance use rates appear to be influenced by environmental smoking exposure, heavier exposure does not necessarily result in increased usage.

Proportion of Adolescents with Substance Use Across ETS Categories
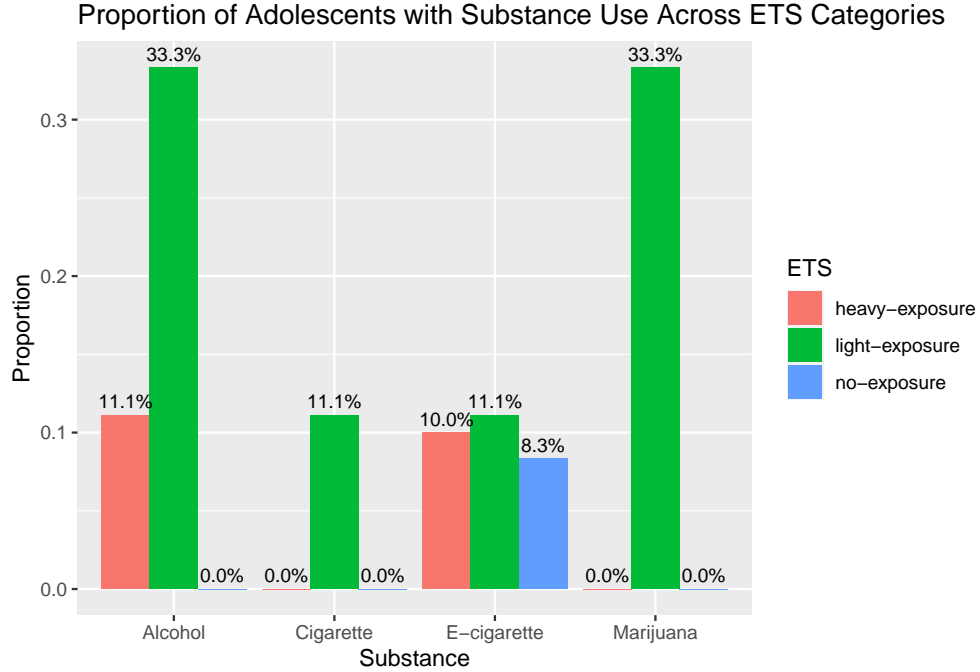
4.2.3. Adolescent Externalizing Problems

The following Table 6 presents a comparison of externalizing problems ETS exposure categories. For the Brief Problem Monitor externalizing scale (`bpm_ext` and `bpm_ext_p`), adolescents in the "heavy-exposure" group consistently exhibit higher median scores, which suggests a possible higher degree of externalizing issues. The "no-exposure" group has the lowest median scores, indicating the fewest externalizing problems among the three categories. In terms of the two variables about ADHD ('swan_hyperactive' and 'swan_inattentive'), they shows a different pattern. Regarding the Hyperactive/Impulsive type ADHD, the data indicates that adolescents from the "light-exposure" group have the highest median score, while the "no-exposure" group still exhibits the lowest median score. In terms of Inattentive type, the "heavy-exposure" group presents the highest median value and the "no-exposure" group reports lowest median value. This table highlights that adolescents with heavier smoke exposure tend to show more pronounced externalizing problems, while those with no smoke exposure show fewer issues.

Table 6: Comparation of Externalizing Problems Across ETS Categories

Characteristic	heavy-exposure, N = 10	light-exposure, N = 10	no-exposure, N = 13
bpm_ext	4 (3, 4)	3 (2, 4)	2 (1, 3)
bpm_ext_p	3 (0, 5)	1 (1, 2)	0 (0, 0)
childasd			
0	4 (100%)	7 (100%)	3 (75%)
2	0 (0%)	0 (0%)	1 (25%)
swan_hyperactive	10 (6, 17)	12 (5, 14)	4 (1, 5)
swan_inattentive	14 (11, 17)	13 (10, 16)	8 (6, 9)

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

4.3. Interrelatedness of prenatal and postnatal exposure

Table 7 presents the distribution of postnatal environmental tobacco smoke (ETS) exposure across categories of smoking during pregnancy (SDP). For mothers identified as "heavy-smokers" during pregnancy, a significant 75% of their children were found to have "heavy-exposure" to ETS after birth. 25% of these children experienced "light-exposure" ETS after birth, while no child in this category was found to have "no-exposure" ETS. This suggests that the mother's smoking behavior continued after the child was born. In the category of "light-smoker", half of the children in this group have "heavy-exposure", while 25% experience "light-exposure" and another 25% have "no-exposure". Lastly, for children whose mothers were "nonsmokers" during pregnancy, the majority (55%) had been unexposed to ETS postpartum. However, there were still 10% of these children who had been heavily exposed to ETS postpartum and 35% in the category of "light-exposure", which might be due to either the mothers picking up smoking after childbirth or the smoking from their partners. There seems to be a correlation between prenatal smoking behaviors of mothers and the postnatal ETS exposure. Children with mothers who smoked heavily during pregnancy are more likely to experience higher levels of postnatal ETS.

Table 7: Correlation between SDP and ETS

Characteristic	heavy-smoker, N = 10	light-smoker, N = 5	non-smoker, N = 30
ets			
heavy-exposure	6 (75%)	2 (50%)	2 (10%)
light-exposure	2 (25%)	1 (25%)	7 (35%)
no-exposure	0 (0%)	1 (25%)	11 (55%)
Unknown	2	1	10

[1] n (%)

We can also check mother's substance use history. The variables 'nidaalc', 'nidatob', 'nidapres', and 'nidaill' are realted to substance use behaviors. They represent the mother's consumption response for alcohol, tobacco, prescription drugs, and illegal drugs respectively. The categorization for these variables is 0 = Never, 1 = Once or Twice, 2 = Monthly, 3 = Weekly, 4 = Daily or almost daily. The variable 'momcig' is about the number of days the past 30 days the mother smoke, and 'mom_numcig' is related the average number of cigarettes mother consumes daily.

The following Table 8 then shows the comparison of these characteristics across both SDP and ETS categories. There are some similar patterns observed in SDP and ETS. For instance, in the 'nidalc' and 'nidatob' variables, both heavy-smoker and heavy-exposure groups have higher frequencies of substance usage, suggesting that heavier prenatal and postnatal smoke exposures are possibly associated with increased recent substance use. The trend is also observed in 'nidaill', but no so obvious. It is also found that while some mothers in the "light-smoker" category during pregnancy report no cigarette usage ('momcig' and 'mom_numcig' variables), a few have picked up the habit post-pregnancy. This indicates that while some mothers may quit smoking during pregnancy, they might resume or initiate the habit postnatally, posing an ETS risk for their children.

Table 8: Parental Substance Usage and Frequency

Characteristic	SDP			ETS		
	heavy-smoker, N = 10	light-smoker, N = 5	non-smoker, N = 30	heavy-exposure, N = 10	light-exposure, N = 10	no-exposure, N = 13
nidaalc						
0	5 (63%)	1 (25%)	13 (54%)	4 (40%)	6 (60%)	8 (62%)
1	2 (25%)	1 (25%)	4 (17%)	2 (20%)	3 (30%)	2 (15%)
2	0 (0%)	0 (0%)	3 (13%)	0 (0%)	0 (0%)	2 (15%)
3	1 (13%)	2 (50%)	3 (13%)	4 (40%)	0 (0%)	1 (7.7%)
4	0 (0%)	0 (0%)	1 (4.2%)	0 (0%)	1 (10%)	0 (0%)
nidatob						
0	1 (13%)	1 (25%)	18 (75%)	1 (10%)	5 (50%)	12 (92%)
1	0 (0%)	2 (50%)	1 (4.2%)	1 (10%)	2 (20%)	0 (0%)
3	0 (0%)	1 (25%)	3 (13%)	3 (30%)	1 (10%)	0 (0%)
4	7 (88%)	0 (0%)	2 (8.3%)	5 (50%)	2 (20%)	1 (7.7%)
nidapres						
0	8 (100%)	3 (100%)	22 (92%)	10 (100%)	10 (100%)	10 (83%)
1	0 (0%)	0 (0%)	1 (4.2%)			
3	0 (0%)	0 (0%)	1 (4.2%)	0 (0%)	0 (0%)	1 (8.3%)
4				0 (0%)	0 (0%)	1 (8.3%)
nidaill						
0	7 (88%)	3 (75%)	24 (100%)	8 (80%)	10 (100%)	13 (100%)
3	0 (0%)	1 (25%)	0 (0%)	1 (10%)	0 (0%)	0 (0%)
4	1 (13%)	0 (0%)	0 (0%)	1 (10%)	0 (0%)	0 (0%)
momcig	30 (24, 30)	1 (0, 9)	0 (0, 0)	30 (25, 30)	0 (0, 2)	0 (0, 0)
mom_numcig	7.5 (4.5, 8.5)	2.0 (1.0, 3.5)	0.0 (0.0, 0.0)	7.00 (3.00, 8.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)

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

5. Recommendations and Future Study

Our current explanatory data analysis is limited due to several reasons. Firstly, the sample size of our dataset is very small with only 49 observations, which might limit the generalization of our findings as they may not accurately reflect the variety of the population. Additionally, when estimating the intensity of SDP and ETS, our two new independent variables (IVs) heavily rely on self-reported data, which sometimes could be unreliable due to recall bias or participants' reluctance to disclose certain information. Furthermore, the study has a large number of missing values. Imputation methods might not be applicable given the small sample size and the nature of our variables, which are mostly categorically nominal. The missing values can challenge the robustness of our results.

For future studies, it is recommended to re-recruit more participants to increase the sample size, which is helpful to improve the statistical power and generality of the study. Researchers could also improve study design or refine the administration process. For example, arranging more urine cotinine tests might be more reliable than depending solely on self-report. Moreover, there are missing values in some basic demographic characteristics, which should be avoided when documenting participants' information. Extending the follow-up period might also help in reducing memory bias. If these issues are handled, future study could provide more conclusive and comprehensive insights into the impact of SDP and ETS.

6. Conclusion

Based on our previous analysis, it can be found that there is a potential positive trend linking prenatal and postnatal smoke exposure to various adolescent issues. Both Smoking During Pregnancy (SDP) and Environmental Tobacco Smoke (ETS) exposure seem to play a significant role in influencing self-regulation metrics, rates of substance use, and externalizing problems among adolescents. Specifically, as the intensity of SDP and ETS exposure increases, it can be observed a corresponding rise in median values across self-regulation scores, ratings from questionnaires, and the proportion of adolescents reporting substance use. But there are still several variables that did not follow this pattern, such as variables related cognitive reappraisal and ASD. This findings suggests= that smoke exposure both during and after pregnancy, could

have negative effects on child development. However, the conclusions drawn here are limited by the small sample size of the study and the lack of statistical validations from various tests.

References

Pirkle, James L. 1996. "Exposure of the US Population to Environmental Tobacco Smoke: The Third National Health and Nutrition Examination Survey, 1988 to 1991." *JAMA* 275 (16): 1233. https://doi.org/10.1001/jama.1996.03530400021033.

Risica, Patricia Markham, Adam Gavarkovs, Donna R. Parker, Ernestine Jennings, and Maureen Phipps. 2017. "A Tailored Video Intervention to Reduce Smoking and Environmental Tobacco Exposure During and After Pregnancy: Rationale, Design and Methods of Baby's Breath." *Contemporary Clinical Trials* 52 (January): 1–9. https://doi.org/10.1016/j.cct.2016.10.010.