Examine effects of SDP on adolescent self-regulation, substance use, and externalizing

Caiwei Xiong

Abstract

Background: This study builds upon a prior investigation focused on smoke avoidance interventions aimed at reducing smoking and environmental tobacco smoke (ETS) exposure among low-income women (N=738) during pregnancy, as well as limiting ETS exposure to children in the immediate postpartum period. For the current study, a random selection of adolescents (N=100) and their mothers, originally participating in the previous study, were recruited. Participants will be evaluated using baseline data, with two subsequent longitudinal follow-up assessments scheduled at 6 and 12 months post-baseline.

Methods: Exploratory Data Analysis (EDA) will be conducted for both the original and the current studies. The EDA process will encompass an analysis of baseline data, supplemented by two longitudinal follow-up assessments scheduled at 6 and 12 months post-baseline. This approach aims to provide comprehensive insights into the progression and outcomes of the studies over time

Results: The data does not reveal a statistically significant correlation between the smoking status of parents during pregnancy and the diagnosis of Autism Spectrum Disorder (ASD), attention-related problems, or hyperactive and inattentive behaviors typically associated with ADHD. This suggests that prenatal exposure to smoking may not have a direct or singular influence on these particular developmental outcomes. Furthermore, our exploration into substance use revealed that adolescents whose parents are heavy smokers reported a higher incidence of lifetime substance use, particularly alcohol, which was statistically significant. However, current use of substances like cigarettes, electronic cigarettes, and marijuana in the past 30 days did not show significant differences across the groups. This indicates that while there is a notable association between parental smoking and adolescent experimentation with substances, this does not necessarily translate into ongoing substance use behaviors. Additionally, the lack of significant differences in self-reported externalizing problems among parents, as well as the behavioral manifestations in their children, suggests that factors beyond prenatal smoking exposure may contribute more significantly to these behaviors.

Conclusions: Our findings imply that while prenatal smoking exposure has some association with initial substance experimentation in adolescence, its impact on ongoing substance use and other aspects of self-regulation and externalizing behaviors may be limited. These outcomes underscore the complexity of the developmental trajectories impacted by prenatal smoking and highlight the need for a multifaceted approach to understand the full range of influences on adolescent development. It is essential to consider genetic, environmental, and social factors that may interact with prenatal exposure to smoking to affect adolescent well-being.

Introduction

During the period of fetal development, the nervous system exhibits heightened vulnerability to the adverse impacts of environmental toxicants, such as the components found in cigarettes.(Reeves and Bernstein 2008) Neurodevelopmental outcomes can be linked to alterations in brain structure and function, resulting in diminished psychomotor abilities and heightened susceptibility to behavioural difficulties. Tobacco smoke comprises several hazardous compounds, such as lead and cadmium, both of which have the potential to impact the maturing neurological system. Several studies have revealed the adverse effects of exposure to environmental tobacco smoke (ETS) during pregnancy on infant neurodevelopment, specifically highlighting the increased risk of developmental delay. Furthermore, some studies have indicated that exposure to mother passive smoking during pregnancy might potentially have an influence on the behavioural development of offspring, specifically in relation to aggressive and externalizing tendencies. (Polanska et al. 2017) This study aims to investigate the impact of self-regulation difficulties in the relationship between smoking during pregnancy (SDP) on the levels of substance use (SU) and severity of externalizing behaviour (EXT) over a period of time.

Motivation datasets

Data Preprocessing

The dataset comprises data from 49 pairs of mothers and children. It includes 9 character variables, 54 integer variables, and 15 numeric variables. However, we identified discrepancies in the data recording, such as an implausible record of an individual consuming 44,989 cigarettes in a day, likely due to errors during data collection. Consequently, we excluded this individual's data. Further investigation revealed that 'Black and Mild' is a brand of tobacco cigar. For simplicity, we will equate tobacco cigars with conventional cigarettes in our analysis. Additionally, when a participant reported smoking between 20 to 25 cigarettes daily, we used the median value for uniformity, resulting in an average of 23 cigarettes per day.

In the dataset, the variable `momcig` indicates the frequency of cigarette smoking over the past 30 days. However, an anomaly was noted where a participant reported smoking for 40 days, exceeding the defined time frame. This discrepancy is likely a result of data recording errors during collection. Consequently, this individual's data was excluded from the dataset. The dataset also contained a significant number of missing values. Nevertheless, some of these could be inferred from related variables. For example, if a participant reported never having tried cigarettes, e-cigarettes, or marijuana (a value of 0), we could confidently assign a value of 0 to the variables measuring usage of these substances in the past 30 days.

The variable `swan_hyperactive` in our dataset represents the total score from questions 1-9 of the SWAN Rating Scale. A score of 6 or higher on this scale suggests a potential diagnosis of Hyperactive/Impulsive ADHD in children. Accordingly, `swan_hyperactive` can be converted into a binary variable to indicate the likelihood of a child having Hyperactive/Impulsive ADHD. Similarly, 'swan_inattentive' sums up the responses to the same questions on the SWAN Rating Scale, with a score of 6 or more indicating a potential diagnosis of Inattentive ADHD. This variable can also be transformed into a binary format to signify the likelihood of a child being diagnosed with the Inattentive type of ADHD.

To better understand the baseline characteristics differentiating children at a higher risk of ADHD-Inattentive type from those without, we created the following table. The data in **Table 1** reveals a statistically significant difference in the `pedu` variable between the two groups being compared.

Table 1: Baseline characteristics of children with or without higher possiblity to have ADHD-Inattentive type

Characteristics	Likely Inattentive type	Not likely Inattentive type	p-value
page	36 (2)	38 (4)	0.3
MISSING	8	0	
psex	7 (100%)	31 (97%)	>0.9

MISSING	8	0	
smoke_exposure_6mo	0 (0%)	8 (25%)	0.6
MISSING	10	0	
pedu			0.030
0	2 (29%)	1 (3.1%)	
1	1 (14%)	2 (6.3%)	
2	0 (0%)	5 (16%)	
3	0 (0%)	13 (41%)	
4	1 (14%)	2 (6.3%)	
5	2 (29%)	8 (25%)	
6	1 (14%)	1 (3.1%)	
MISSING	8	0	
mom_smoke_16wk			>0.9
1	3 (21%)	8 (25%)	
2	11 (79%)	24 (75%)	
MISSING	1	0	
mom_smoke_22wk			>0.9
1	3 (27%)	9 (31%)	
2	8 (73%)	20 (69%)	
MISSING	4	3	
mom_smoke_32wk			>0.9
1	2 (20%)	7 (25%)	
2	8 (80%)	21 (75%)	
MISSING	5	4	
smoke_exposure_12mo	1 (20%)	6 (19%)	>0.9
MISSING	10	0	
smoke_exposure_2yr	1 (20%)	8 (25%)	>0.9
MISSING	10	0	
smoke_exposure_3yr	1 (20%)	8 (26%)	>0.9
MISSING	10	1	
smoke_exposure_4yr	1 (20%)	7 (23%)	>0.9
MISSING	10	1	
smoke_exposure_5yr	1 (20%)	7 (22%)	>0.9
MISSING	10	0	

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

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

This table delineates the baseline characteristics of children categorized by the likelihood of having ADHD-Inattentive type. Upon examination, the average age (`page`) shows no statistically significant difference between the two groups, with means of 36 and 38 months for the likely and not likely categories respectively (p-value: 0.3). Gender distribution (`psex`) is also not significantly different, with the majority of children in both groups being male (100% and 97%, p-value: >0.9). Notably, 'pedu', indicating parental education level, demonstrates a significant disparity (p-value: 0.030), suggesting that parental education might be associated with the inattentiveness in children. The data on smoke exposure at various age milestones (`smoke_exposure_6mo` through `smoke_exposure_5yr`) does not show a significant correlation with the likelihood of ADHD-Inattentive type (p-values: >0.9), except for the missing values at 6 months which cannot be evaluated due to data absence. Overall, while some variables show no significant association, the parental education level stands out as a potential factor related to the ADHD-Inattentive type in children.

The following **Table 2** displays baseline characteristics of children grouped by their likelihood of having ADHD-Hyperactive/Impulsive type. Analysis of Table 2 indicates that the variables `smoke_exposure_6mo`, `smoke_exposure_2yr`, `smoke_exposure_3yr`, and `smoke_exposure_5yr` show statistically significant differences between the two groups.

Table 2: Baseline characteristics of children with or without higher possiblity to have Hyperactive/Impulsive type.

Characteristics	Likely Hyperactive/Impulsive type	Not likely Hyperactive/Impulsive type	p-value
page	37 (3)	38 (4)	0.4
MISSING	8	0	

psex	19 (100%)	19 (95%)	>0.9
MISSING	8	0	
smoke_exposure_6mo	1 (5.9%)	7 (35%)	0.048
MISSING	10	0	
pedu			0.2
0	2 (11%)	1 (5.0%)	
1	2 (11%)	1 (5.0%)	
2	1 (5.3%)	4 (20%)	
3	4 (21%)	9 (45%)	
4	1 (5.3%)	2 (10%)	
5	8 (42%)	2 (10%)	
6	1 (5.3%)	1 (5.0%)	
MISSING	8	0	
mom_smoke_16wk			0.2
1	4 (15%)	7 (35%)	
2	22 (85%)	13 (65%)	
MISSING	1	0	
mom_smoke_22wk			0.11
1	4 (19%)	8 (42%)	
2	17 (81%)	11 (58%)	
MISSING	6	1	
mom_smoke_32wk			0.2
1	3 (14%)	6 (35%)	
2	18 (86%)	11 (65%)	
MISSING	6	3	
smoke_exposure_12mo	1 (5.9%)	6 (30%)	0.10
MISSING	10	0	
smoke_exposure_2yr	1 (5.9%)	8 (40%)	0.023
MISSING	10	0	
smoke_exposure_3yr	1 (6.3%)	8 (40%)	0.026
MISSING	11	0	
smoke_exposure_4yr	1 (6.3%)	7 (35%)	0.053
MISSING	11	0	
smoke_exposure_5yr	1 (5.9%)	7 (35%)	0.048
MISSING	10	0	

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

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

Table 2 outlines the baseline characteristics of children assessed for a higher likelihood of ADHD-Hyperactive/Impulsive type compared to those without. The average age (`page`) of children does not differ significantly between the likely and not likely groups, with means of 37 and 38 months respectively (p-value: >0.4). Gender distribution (`psex`) also does not show a significant difference, with a majority being male in both groups (100% in the likely group vs. 95% in the not likely group, p-value: >0.9). A notable finding is the statistical significance in smoke exposure at 6 months (`smoke_exposure_6mo`) with 5.9% in the likely group exposed versus 35% in the not likely group (p-value: 0.048), suggesting an association with a decreased risk of developing Hyperactive/Impulsive type ADHD. The parental education level (`pedu`), however, does not show a significant difference (p-value: >0.2). Additionally, smoke exposure at 2, 3, and 4 years (`smoke_exposure_2yr`, `smoke_exposure_3yr`, `smoke_exposure_4yr`) shows a pattern of significance (p-values: 0.023, 0.026, 0.053 respectively), with lower exposure correlating with a higher likelihood of ADHD diagnosis. The findings at 5 years (`smoke_exposure_5yr`) are also significant (p-value: 0.048), reinforcing the potential impact of early life smoke exposure on ADHD development. These results indicate that certain environmental factors, particularly smoke exposure at specific early ages, may be related to the likelihood of developing Hyperactive/Impulsive ADHD in children.

Missing values

In the data analysis, we encountered a common issue of missing values, with 63 variables in our dataset exhibiting gaps. The following **Table 3** presents the missing data percentages of variables within our dataset, highlighting the challenges inherent in comprehensive data analysis. It's critical to investigate the patterns

of these omissions. A notable 60 variables have more than 15\% of their data missing, with the percentage of missing information varying significantly across different variables. The variable `mom_smoke_pp1`, which represents self-reported smoking status at the first postpartum visit, displayed the highest percentage of missing values at 78%. On the other end of the spectrum, `mom_smoke_16wk`, indicating self-reported smoking status at 16 weeks of pregnancy, had the lowest at just 2.12%. Understanding and addressing these missing patterns is essential for accurate analysis and interpretation of the dataset. The **Table 3** shows a wide variation in missing data percentages, reflecting a significant range in the completeness of responses across different variables. This pattern necessitates a thorough evaluation of missing data handling methods to ensure the integrity and validity of any subsequent analysis conducted on this dataset.

Missing percentage of variables

variable_name	missing percentage
Self-reported current smoker at first postpartum visit	0.7872340
child participating diagnosed or suspected of having Autism Spectrum Disorder	0.5744681
Self-reported current smoker at second postpartum visit	0.4042553
Average response on the Parental Knowledge Questionnaire on items related to parental control	0.3404255
Average response on the Parental Knowledge Questionnaire on items related to parental solicitation	0.3191489
how many of the past 30 days did you drink alcohol?	0.2978723
Sum of responses on the Brief Problem Monitor on items related to internalizing problems on self	0.2978723
Average response on the Parental Knowledge Questionnaire on items related to parental knowledge	0.2978723
Average response on the Parental Knowledge Questionnaire on items related to parental solicitation	0.2978723
Sum Brief Problem Monitor on items related to attention problems on child	0.2765957
Child biological sex assigned at birth	0.2765957
In the past 30 days, on how many days did you use an e-cigarette or vape	0.2765957
Have you ever had alcohol not as part of a religious ceremony?	0.2765957
Average response on the Emotion Regulation Questionairre related to Expressive Suppression	0.2765957
Average response on the Parental Knowledge Questionnaire on items related to child disclosure	0.2765957
family's estimated annual household income	0.2553191
Sum of Brief Problem Monitor on items related to externalizing problems on child	0.2553191
Average response on the Parental Knowledge Questionnaire on items related to child disclosure	0.2553191
Average response on the Parental Knowledge Questionnaire on items related to parental control	0.2553191
Child age	0.2553191
Do you speak another language at home?	0.2553191
Are you Hispanic or Latino	0.2553191
Have you ever tried or experimented with cigarette smoking, even a few puff /drags?	0.2553191
On how many of the past 30 days did you smoke cigarettes	0.2553191
Have you ever used an e-cigarette or vaped?	0.2553191
Have you ever used marijuana (including in a vape containing marijuana)	0.2553191
On how many of the past 30 days did you use marijuana	0.2553191
Sum of Brief Problem Monitor on items related to attention problems on self	0.2553191
Sum of Brief Problem Monitor on items related to externalizing problems on self	0.2553191
Average response on the Emotion Regulation Questionairre related to Cognitive Reappraisal	0.2553191
In the past 6 months, how often have you used prescription drugs for non-medical reasons?	0.2340426
Self-reported smoke exposure from mom or partner in 3rd year	0.2340426
Self-reported smoke exposure from mom or partner in 4th year	0.2340426
Average response on the Parental Knowledge Questionnaire on items related to parental knowledge	0.2340426
Sum of Brief Problem Monitor on items related to attention problems on self	0.2340426
Sum of Brief Problem Monitor on items related to externalizing problems on self	0.2340426
In the past 6 months, how often have you used alcohol (4 or more drinks a day)?	0.2127660
In the past 6 months, how often have you used tobacco products?	0.2127660
In the past 6 months, how often have you used illegal drugs?	0.2127660
On how many of the past 30 days did you smoke cigarettes?	0.2127660

How many cigarettes per day do you usually smoke?	0.2127660
Urine cotinine (nicotine metabolite) at 34 wks gestation	0.2127660
Sum of Brief Problem Monitor on items related to internalizing problems on child	0.2127660
Self-reported smoke exposure from mom or partner from 0 to 6 months	0.2127660
Self-reported smoke exposure from mom or partner from 7 to 12 months	0.2127660
Self-reported smoke exposure from mom or partner in 2nd year	0.2127660
Self-reported smoke exposure from mom or partner in 5th year	0.2127660
Sum of Brief Problem Monitor on items related to internalizing problems on self	0.2127660
Average response on the Emotion Regulation Questionairre related to Cognitive Reappraisal	0.2127660
Average response on the Emotion Regulation Questionairre related to Expressive Suppression	0.2127660
Self-reported current smoker at 32 weeks pregnant	0.1914894
Urine cotinine (nicotine metabolite) at 6 months postpartum from baby	0.1914894
Urine cotinine (nicotine metabolite) at 6 months postpartum from mom	0.1914894
Parent age	0.1702128
Parent biological sex assigned at birth	0.1702128
Do you speak another language at home	0.1702128
Are you Hispanic or Latino	0.1702128
Do you currently have a paid job?	0.1702128
What is the highest level of education you have completed?	0.1702128
Self-reported current smoker at 6 months postpartum	0.1702128
Self-reported current smoker at 22 weeks pregnant	0.1489362
Self-reported current smoker at 12 weeks postpartum	0.1276596
Self-reported current smoker at 16 weeks pregnant	0.0212766

Demographic Characteristics

Table 4 provides a demographic breakdown of the children and parents in the study. The median age of the children is 14 years, with a range from 13 to 15 years (IQR), while the median age of the parents is 37, with a slightly broader range between 35 to 39 years (IQR). There are some cases where the age is unknown, affecting 12 children and 8 parents. In terms of gender, a majority of the children are male (62%), whereas the parents are predominantly female (97%). There are instances where gender has not been reported, which applies to 13 children and 8 parents. When looking at the language spoken at home, a substantial portion of children (31%) and parents (38%) did not have this data reported. For race and ethnicity, the children are primarily Hispanic/Latino (54%) and White (36%), while a smaller number are American Indian/Alaska Native (11%). The parents have a higher percentage identifying as White (46%), followed by Hispanic/Latino (33%), and there's a notable representation of Native Hawaiian/Pacific Islander (7.7%). A small minority falls under 'Other' or had their race/ethnicity unreported. These figures provide insight into the cultural and demographic landscape of the study's participants.

Table 4: Summary for Demographic Variables

Characteristic	Child	Parent
Age	14 (13, 15)	37 (35, 39)
Unknown	12	8
Sex		
Male	21 (62%)	1 (2.6%)
Female	13 (38%)	38 (97%)
Unknown	13	8
Language Spoken at Home	11 (31%)	15 (38%)
Unknown	12	8
Race/Ethnicity		
American Indian/Alaska Native	3 (11%)	4 (10%)
Hispanic/Latino	15 (54%)	13 (33%)
Native Hawaiian/Pacific Islander	0 (0%)	3 (7.7%)
White	10 (36%)	18 (46%)
Other	0 (0%)	1 (2.6%)
Unknown	19	8

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

Table 5 provides a snapshot of the educational background, income, and employment status of the 47 parents in the study. The highest level of educational attainment for the majority of parents is 'some college', comprising 33% of the sample. Ten parents, or 26%, hold a 4-year degree, while 3 (7.7%) have completed a 2-year degree, and an equivalent percentage have only some high school education. A General Education Diploma (GED) is held by 13% of the parents, and another 7.7% have completed high school. Postgraduate degrees are the least common, with only 5.1% of parents having attained this level of education. There are 8 instances where the education level was not reported. The average household income is approximately $64,460, with a notable variation, as indicated by a standard deviation of $61,158. However, income data for 12 parents is missing. Employment status shows that a slight majority of parents (51%) are employed full-time, 31% are not employed, and 18% are employed part-time. The employment status for 8 parents remains unreported. These figures reveal a diverse range of educational backgrounds and economic conditions among the parents involved in the study.

Table 5: Characteristics of Parents

Characteristics	N = 47
highest level of education	
2 year degree	3 (7.7%)
4 year degree	10 (26%)
GED	5 (13%)
high school	3 (7.7%)
postgraduate degree	2 (5.1%)
some college	13 (33%)
some high school	3 (7.7%)
MISSING	8
family's estimated annual household income	64,460 (61,158)
MISSING	12
Do you currently have a paid job?	
Full-time	20 (51%)
No	12 (31%)
Part-time	7 (18%)
MISSING	8

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

Effects of Smoking During Pregnancy on adolescent self-regulation, substance use, and externalizing

Effects of Smoking During Pregnancy on adolescent self-regulation

The study's code book specifies that parental smoking behavior during pregnancy was monitored at three key gestational periods: 16 weeks, 22 weeks, and 32 weeks. At each of these time points, parents self-reported their current smoking status, which was recorded using a binary variable: `1` denotes a current smoker, and `0` indicates a non-smoker during the respective time period. Additionally, the variable `cotimean_34_wk` was introduced to provide a biochemical assessment of smoking behavior, measuring the mean urine cotinine levels at 34 weeks of gestation. This biometric data offers a more objective measure of tobacco exposure during the later stage of pregnancy, complementing the self-reported data collected earlier.

In our study, we categorized parents into three distinct groups: non-smokers, moderate smokers, and heavy smokers. To classify parents as non-smokers, they must have reported no smoking at the 16, 22, and 32-week marks, and their `cotimean_34wk` values should be equal to 0. Parents who consistently reported smoking at all three time points — 16, 22, and 32 weeks — were classified as heavy smokers. Those who did not fit into either of these categories were designated as moderate smokers. This categorization was essential for our analysis to distinctly assess the impact of varying smoking habits.

Table 6: Effects of Smoking During Pregnancy on adolescent self-regulation

Characteristic	heavy smoker, N = 9	moderate smokers, N = 15	non smoker, N = 23	p-value
childasd				0.7
0	2 (100%)	6 (86%)	10 (91%)	
1	0 (0%)	0 (0%)	1 (9.1%)	
2	0 (0%)	1 (14%)	0 (0%)	
MISSING	7	8	12	
bpm_att_a				0.3
0	1 (14%)	5 (42%)	9 (53%)	
1	1 (14%)	3 (25%)	3 (18%)	
2	3 (43%)	3 (25%)	3 (18%)	
4	0 (0%)	0 (0%)	1 (5.9%)	
5	0 (0%)	0 (0%)	1 (5.9%)	
6	2 (29%)	0 (0%)	0 (0%)	
8	0 (0%)	1 (8.3%)	0 (0%)	
MISSING	2	3	6	
bpm_att_p				0.2
0	0 (0%)	3 (27%)	6 (38%)	
1	2 (29%)	4 (36%)	3 (19%)	
2	1 (14%)	2 (18%)	4 (25%)	
3	0 (0%)	1 (9.1%)	0 (0%)	
4	0 (0%)	1 (9.1%)	0 (0%)	
5	1 (14%)	0 (0%)	2 (13%)	
6	2 (29%)	0 (0%)	0 (0%)	
7	0 (0%)	0 (0%)	1 (6.3%)	
8	1 (14%)	0 (0%)	0 (0%)	
MISSING	2	4	7	
erq_cog_a	5.02 (0.70)	5.57 (1.36)	5.26 (1.46)	0.4
MISSING	2	3	5	
erq_cog	3.36 (0.73)	3.45 (0.77)	2.94 (1.15)	0.6
MISSING	2	4	6	
bpm_att				0.6
0	0 (0%)	4 (36%)	5 (29%)	
1	0 (0%)	1 (9.1%)	2 (12%)	
2	2 (29%)	3 (27%)	2 (12%)	
3	0 (0%)	1 (9.1%)	1 (5.9%)	
4	0 (0%)	1 (9.1%)	1 (5.9%)	
5	2 (29%)	0 (0%)	3 (18%)	
6	0 (0%)	0 (0%)	1 (5.9%)	
7	2 (29%)	1 (9.1%)	2 (12%)	

9	1 (14%)	0 (0%)	0 (0%)	
MISSING	2	4	6	
erq_exp	2.82 (0.57)	2.93 (0.81)	2.56 (0.87)	0.6
MISSING	2	4	7	
erq_exp_a	3.86 (2.01)	3.79 (1.46)	3.03 (1.32)	0.2
MISSING	2	3	5	

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

[2] Fisher's exact test; Kruskal-Wallis rank sum test

Across the heavy, moderate, and non-smoker parental groups, the reported lifetime cigarette usage (`cig_ever`) by their children shows that a larger percentage of children from heavy smoker parents have tried cigarettes at least once, though this finding is not statistically significant with a p-value of 0.2. For electronic cigarette usage (`e_cig_ever`), again children of heavy smokers report higher usage, but the difference across groups does not reach statistical significance (p-value of 0.2). Notably, the number of electronic cigarettes used in the past 30 days (`num_e_cigs_30`) is only reported by children of heavy smokers. The reported lifetime marijuana usage (`mj_ever`) is higher among children of heavy smokers, with 29% indicating usage compared to none from moderate smokers and a small percentage from non-smokers. However, this observation is not statistically significant (p-value of 0.12). No significant use of marijuana in the past 30 days (`num_mj_30`) is reported across all groups, with a p-value of 0.10. Lastly, alcohol consumption (`alc_ever`) is reported among children from all categories, with those from heavy smoking backgrounds reporting slightly higher lifetime consumption, which is statistically significant (p-value of 0.028). However, the number of alcohol uses in the past 30 days (`num_alc_30`) is negligible across all groups.

These data suggest that while children of heavy smokers are more likely to report lifetime substance use, the differences are not statistically significant except for alcohol, which shows a minor but significant correlation. This could indicate a potential influence of parental smoking on children's substance use behaviors, but further investigation with more detailed data would be needed to establish stronger evidence for these patterns.

Effects of Smoking During Pregnancy on adolescent substance use

To investigate the impact of smoking during pregnancy on substance use among adolescents, we have selected a range of variables for our analysis. These include `cig_ever`, which indicates whether the child has ever tried cigarettes; `num_cigs_30`, the number of cigarettes smoked in the past 30 days; `e_cig_ever`, denoting any electronic cigarette use; and `num_e_cigs_30`, the quantity of electronic cigarettes used in the past 30 days. Additionally, we are examining `mj_ever` for lifetime marijuana use; `num_mj_30` for marijuana use in the past 30 days; alc_ever to identify any history of alcohol consumption; and `num_alc_30` for the frequency of alcohol use in the past 30 days. These variables are critical for assessing the patterns and prevalence of substance use behaviors in the children of mothers who smoked during pregnancy.

Table 7: Effects of Smoking During Pregnancy on adolescent substance use

Characteristic	heavy smoker, N = 9	moderate smokers, N = 15	non smoker, N = 23	p-value
cig_ever				0.2
0	6 (86%)	11 (100%)	17 (100%)	
1	1 (14%)	0 (0%)	0 (0%)	
MISSING	2	4	6	
num_cigs_30	0 (0%)	0 (0%)	0 (0%)	
MISSING	2	4	6	
e_cig_ever				0.2
0	6 (86%)	9 (82%)	17 (100%)	
1	1 (14%)	2 (18%)	0 (0%)	
MISSING	2	4	6	
num_e_cigs_30				0.2
0	6 (86%)	10 (100%)	17 (100%)	
2	1 (14%)	0 (0%)	0 (0%)	
MISSING	2	5	6	
mj_ever				0.12

9

0	5 (71%)	11 (100%)	16 (94%)	
1	2 (29%)	0 (0%)	1 (5.9%)	
MISSING	2	4	6	
num_mj_30				0.10
0	5 (71%)	11 (100%)	16 (94%)	
3	1 (14%)	0 (0%)	0 (0%)	
12	1 (14%)	0 (0%)	0 (0%)	
18	0 (0%)	0 (0%)	1 (5.9%)	
MISSING	2	4	6	
alc_ever				0.4
0	4 (67%)	10 (91%)	15 (88%)	
1	2 (33%)	1 (9.1%)	2 (12%)	
MISSING	3	4	6	
num_alc_30				0.028
0	4 (67%)	11 (100%)	16 (100%)	
1	1 (17%)	0 (0%)	0 (0%)	
10	1 (17%)	0 (0%)	0 (0%)	
MISSING	3	4	7	

[1] n (%)

[2] Fisher's exact test

The above table delineates the self-reported substance use characteristics among three distinct parental groups—heavy smokers, moderate smokers, and non-smokers—in our study population. When examining lifetime cigarette usage (`cig_ever`), we observe that all moderate and non-smoking parents reported not having smoked, compared to 86% of heavy smokers, suggesting a correlation between self-identification as a heavy smoker and past cigarette use; however, this did not reach statistical significance (p-value of 0.2). The number of cigarettes smoked in the past 30 days (`num_cigs_30`) is not reported by moderate and non-smokers, which is consistent with their classification, and there are no reported instances in the heavy smoker group either, perhaps due to a data reporting issue, as indicated by missing responses. Similarly, for electronic cigarette usage (`e_cig_ever`), the majority of heavy smokers have used them, but again, the data is limited by a number of non-responses, and the difference is not statistically significant (p-value of 0.2). The pattern is repeated for the number of electronic cigarettes used in the past 30 days (`num_e_cigs_30`), with all moderate and non-smokers reporting zero usage. The usage of marijuana (`mj_ever`) shows a notable difference, with 29% of heavy smokers reporting usage compared to none in the moderate smokers and a very small percentage (5.9%) in non-smokers, although this difference is not significant (p-value of 0.12). For the number of marijuana uses in the past 30 days (`num_mj_30`), again, most moderate and non-smokers report no use, but the difference is not statistically significant (p-value of 0.10). When assessing alcohol consumption (`alc_ever`), we find a higher reported lifetime use among heavy smokers (33%) compared to moderate (9.1%) and non-smokers (12%), and this is the only variable in the table to show a statistically significant difference (p-value of 0.028). However, in the past 30 days (`num_alc_30`), the majority report no alcohol use, and missing data again complicates the analysis.

In conclusion, the data suggests that self-identified heavy smokers are more likely to report lifetime use of various substances, with a statistically significant association found for alcohol. Nevertheless, current use in the past 30 days does not appear to be significantly higher among heavy smokers, which may be due to underreporting or missing data. This analysis underscores the potential relationships between smoking status and other substance use, but also highlights the necessity for more comprehensive data collection to support these findings.

Effects of Smoking During Pregnancy on adolescent externalizing

In the study, externalizing issues among adolescents are quantified using the Brief Problem Monitor (`BPM`). This tool captures the sum of responses pertaining to externalizing problems as reported by the child (`bpm_ext_p`) and by the parent (`bpm_ext_a`). Additionally, the variable `childasd` is used to indicate whether the child has been diagnosed with or is suspected of having Autism Spectrum Disorder (ASD), which is considered an aspect of externalizing behaviors. Attention-related problems are further detailed through the SWAN

Rating Scale, where a score of 6 or higher suggests significant attention issues, typically associated with ADHD. This condition is categorized into two types: Hyperactive/Impulsive (`swan_hyperactive`) and Inattentive (`swan_inattentive`). For a comprehensive assessment, all five variables—`bpm_ext_p`, `bpm_ext_a`, `childasd`, `swan_hyperactive`, and `swan_inattentive`—are incorporated into the measurement of adolescent externalizing behaviors, thereby facilitating a nuanced understanding of these complex issues.

Table 8: Effects of Smoking During Pregnancy on adolescent externalizing

Characteristic	heavy smoker, N = 9	moderate smokers, N = 15	non smoker, N = 23	p-value
bpm_ext_p				0.3
0	2 (29%)	9 (75%)	6 (38%)	
1	0 (0%)	1 (8.3%)	3 (19%)	
2	1 (14%)	1 (8.3%)	3 (19%)	
3	1 (14%)	0 (0%)	2 (13%)	
4	1 (14%)	1 (8.3%)	0 (0%)	
5	1 (14%)	0 (0%)	0 (0%)	
7	1 (14%)	0 (0%)	1 (6.3%)	
11	0 (0%)	0 (0%)	1 (6.3%)	
MISSING	2	3	7	
bpm_ext_a				0.065
0	1 (14%)	5 (45%)	11 (61%)	
1	4 (57%)	3 (27%)	2 (11%)	
2	0 (0%)	1 (9.1%)	2 (11%)	
3	0 (0%)	1 (9.1%)	2 (11%)	
4	2 (29%)	0 (0%)	0 (0%)	
5	0 (0%)	1 (9.1%)	0 (0%)	
6	0 (0%)	0 (0%)	1 (5.6%)	
MISSING	2	4	5	
swan_hyperactive	11 (8)	6 (6)	5 (6)	0.2
swan_inattentive	12 (7)	7 (6)	9 (7)	0.2
childasd				0.7
0	2 (100%)	6 (86%)	10 (91%)	
1	0 (0%)	0 (0%)	1 (9.1%)	
2	0 (0%)	1 (14%)	0 (0%)	
MISSING	7	8	12	

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

[2] Fisher's exact test; Kruskal-Wallis rank sum test

For externalizing problems in children (`bpm_ext_p`), the data does not reveal a statistically significant difference across the groups of heavy smokers, moderate smokers, and non-smokers, as indicated by a p-value of 0.3. Although heavy smokers' children show a wider range of scores, the prevalence of externalizing behaviors does not significantly increase with the intensity of parental smoking. When examining self-reported externalizing problems (`bpm_ext_a`), there is a noticeable trend where non-smokers report fewer problems, but this difference approaches, yet does not reach, statistical significance with a p-value of 0.065. This suggests a potential pattern where smoking may correlate with self-perception of behavioral issues, warranting further investigation. The SWAN Rating Scale responses, which assess hyperactive (`swan_hyperactive`) and inattentive (`swan_inattentive`) tendencies, yield p-values of 0.2 for both, indicating no significant differences among the parental smoking groups in terms of these ADHD-related characteristics in children. Finally, the assessment of Autism Spectrum Disorder (`childasd`) shows no significant variation in diagnosis or suspicion rates among the children of the different smoking status groups (p-value of 0.7). This consistency suggests that, within this dataset, parental smoking habits do not influence the diagnosis of ASD.

Conclusion

In conclusion, our investigation into the effects of smoking during pregnancy on adolescent self-regulation, substance use, and externalizing behaviors has yielded nuanced insights. The data does not reveal a statistically significant correlation between the smoking status of parents during pregnancy and the diagnosis

of Autism Spectrum Disorder (ASD), attention-related problems, or hyperactive and inattentive behaviors typically associated with ADHD. This suggests that prenatal exposure to smoking may not have a direct or singular influence on these particular developmental outcomes. Furthermore, our exploration into substance use revealed that adolescents whose parents are heavy smokers reported a higher incidence of lifetime substance use, particularly alcohol, which was statistically significant. However, current use of substances like cigarettes, electronic cigarettes, and marijuana in the past 30 days did not show significant differences across the groups. This indicates that while there is a notable association between parental smoking and adolescent experimentation with substances, this does not necessarily translate into ongoing substance use behaviors. Additionally, the lack of significant differences in self-reported externalizing problems among parents, as well as the behavioral manifestations in their children, suggests that factors beyond prenatal smoking exposure may contribute more significantly to these behaviors.

Our findings imply that while prenatal smoking exposure has some association with initial substance experimentation in adolescence, its impact on ongoing substance use and other aspects of self-regulation and externalizing behaviors may be limited. These outcomes underscore the complexity of the developmental trajectories impacted by prenatal smoking and highlight the need for a multifaceted approach to understand the full range of influences on adolescent development. It is essential to consider genetic, environmental, and social factors that may interact with prenatal exposure to smoking to affect adolescent well-being.

Refereneces

Polanska, Kinga, Anna Krol, Dorota Merecz-Kot, Danuta Ligocka, Karolina Mikolajewska, Fiorino Mirabella, Flavia Chiarotti, Gemma Calamandrei, and Wojciech Hanke. 2017. "Environmental Tobacco Smoke Exposure During Pregnancy and Child Neurodevelopment." *International Journal of Environmental Research and Public Health* 14 (7): 796. https://doi.org/10.3390/ijerph14070796.

Reeves, Shane, and Ira Bernstein. 2008. "Effects of Maternal Tobacco-Smoke Exposure on Fetal Growth and Neonatal Size." *Expert Review of Obstetrics & Gynecology* 3 (6): 719–30. https://doi.org/10.1586/17474108.3.6.719.

Code appendix

```r
knitr::opts_chunk$set(echo = TRUE,warning = FALSE, message = FALSE)
library("tidyverse")
library("HDSinRdata")
library("gtsummary")
library("gt")
library("webshot2")
library("kableExtra")
library("mice")
library("mediation")
library("dplyr")
library("shiny")
project_one = read.csv("/Users/xiongcaiwei/Downloads/project1.csv")
# Data preprocessing
index_number = 0
get_extreme_number = function(data_set_name){
  for(i in (1:dim(data_set_name)[1])){
    if(data_set_name$mom_numcig[i] == "44989")
      index_number = i
  }
  return(index_number)
}
project_one = project_one[-get_extreme_number(project_one), ]
project_one$mom_numcig[project_one$mom_numcig == 'None'] = '0'
project_one$mom_numcig[project_one$mom_numcig == '2 black and miles a day'] = '2'
project_one$mom_numcig[project_one$mom_numcig == '20-25'] = '23'
project_one$mom_numcig = as.numeric(project_one$mom_numcig)
project_one = project_one[- which(project_one$momcig == 40), ]
project_one[project_one == ''] = NA
project_one$income[project_one$income == '250, 000'] = '250000'
project_one = project_one %>%
  mutate(num_cigs_30 = case_when(cig_ever == 0 ~ 0,
                                 cig_ever == 1 ~ 0,
                                 TRUE ~ NA))
project_one$num_e_cigs_30[which(project_one$e_cig_ever == 0)] = 0
project_one$num_mj_30[which(project_one$mj_ever == 0)] = 0
project_one$num_alc_30[which(project_one$alc_ever == 0)] = 0
project_one$mom_smoke_16wk = str_replace(project_one$mom_smoke_16wk, "2=No","2")
project_one$mom_smoke_16wk = str_replace(project_one$mom_smoke_16wk, "1=Yes","1")

project_one$mom_smoke_22wk = str_replace(project_one$mom_smoke_22wk, "2=No","2")
project_one$mom_smoke_22wk = str_replace(project_one$mom_smoke_22wk, "1=Yes","1")

project_one$mom_smoke_32wk = str_replace(project_one$mom_smoke_32wk, "2=No","2")
project_one$mom_smoke_32wk = str_replace(project_one$mom_smoke_32wk, "1=Yes","1")

project_one$mom_smoke_pp1 = str_replace(project_one$mom_smoke_pp1, "2=No","2")
project_one$mom_smoke_pp1 = str_replace(project_one$mom_smoke_pp1, "1=Yes","1")

project_one$mom_smoke_pp2 = str_replace(project_one$mom_smoke_pp2, "2=No","2")
project_one$mom_smoke_pp2 = str_replace(project_one$mom_smoke_pp2, "1=Yes","1")
```

```r
project_one$mom_smoke_pp12wk = str_replace(project_one$mom_smoke_pp12wk, "2=No","2")
project_one$mom_smoke_pp12wk = str_replace(project_one$mom_smoke_pp12wk, "1=Yes","1")

project_one$mom_smoke_pp6mo = str_replace(project_one$mom_smoke_pp6mo, "2=No","2")
project_one$mom_smoke_pp6mo = str_replace(project_one$mom_smoke_pp6mo, "1=Yes","1")
project_one = project_one %>%
  mutate(swan_hyperactive_new = case_when(swan_hyperactive < 6 ~ 0,
                                          swan_hyperactive >= 6 ~ 1))
project_one = project_one %>%
  mutate(swan_inattentive_new = case_when(swan_inattentive < 6 ~ 0,
                                          swan_inattentive >= 6 ~ 1))
project_one %>%
  dplyr::select(c(page, psex, swan_inattentive_new, smoke_exposure_6mo, pedu,
                                mom_smoke_16wk, mom_smoke_22wk,
                                mom_smoke_32wk,
                                smoke_exposure_6mo,
          smoke_exposure_12mo, smoke_exposure_2yr, smoke_exposure_3yr,
          smoke_exposure_4yr, smoke_exposure_5yr)) %>%
  tbl_summary(by = "swan_inattentive_new",
              statistic = list(all_continuous() ~ "{mean} ({sd})"),
              missing_text = "MISSING") %>%
  add_p() %>%
  modify_header(label = "**Characteristics**",
                stat_1 = "**Likely Inattentive type**",
                stat_2 = "**Not likely Inattentive type**") %>%
  as_kable_extra(booktabs = TRUE,
                 format = "latex",
                 longtable = TRUE,
                 caption = "Baseline characteristics of children with or
                 without higher possiblity to have ADHD-Inattentive type") %>%
  kable_styling(font_size = 7, latex_options = "hold_position")
project_one %>%
  dplyr::select(c(page, psex, swan_hyperactive_new, smoke_exposure_6mo, pedu,
                                mom_smoke_16wk, mom_smoke_22wk,
                                mom_smoke_32wk,
                                smoke_exposure_6mo,
                                smoke_exposure_12mo,
          smoke_exposure_2yr, smoke_exposure_3yr,
          smoke_exposure_4yr, smoke_exposure_5yr)) %>%
  tbl_summary(by = "swan_hyperactive_new",
              statistic = list(all_continuous() ~ "{mean} ({sd})"),
              missing_text = "MISSING") %>%
  add_p() %>%
  modify_header(label = "**Characteristics**",
                stat_1 = "**Likely Hyperactive/Impulsive type**",
                stat_2 = "**Not likely Hyperactive/Impulsive type**") %>%
  as_kable_extra(booktabs = TRUE,
                 format = "latex",
                 longtable = TRUE,
                 caption = "Baseline characteristics of children with or
                 without higher possiblity to have Hyperactive/Impulsive type.") %>%
  kable_styling(font_size = 7, latex_options = "hold_position")
project_one = project_one %>%
```

```r
  dplyr::select(-c(swan_hyperactive_new, swan_inattentive_new))
project_one$psex = as.factor(project_one$psex)
project_one$plang = as.factor(project_one$plang)
project_one$pethnic = as.factor(project_one$pethnic)
project_one$paian = as.factor(project_one$paian)
project_one$pasian = as.factor(project_one$pasian)
project_one$pnhpi = as.factor(project_one$pnhpi)
project_one$pblack = as.factor(project_one$pblack)
project_one$pwhite = as.factor(project_one$pwhite)
project_one$prace_other = as.factor(project_one$prace_other)
project_one$employ = as.factor(project_one$employ)
project_one$pedu = as.factor(project_one$pedu)
project_one$childasd = as.factor(project_one$childasd)
project_one$nidaalc = as.factor(project_one$nidaalc)
project_one$nidatob = as.factor(project_one$nidatob)
project_one$nidapres = as.factor(project_one$nidapres)
project_one$nidaill = as.factor(project_one$nidaill)

project_one$mom_smoke_16wk = as.factor(project_one$mom_smoke_16wk)
project_one$mom_smoke_22wk = as.factor(project_one$mom_smoke_22wk)
project_one$mom_smoke_32wk = as.factor(project_one$mom_smoke_32wk)
project_one$mom_smoke_pp1 = as.factor(project_one$mom_smoke_pp1)
project_one$mom_smoke_pp2 = as.factor(project_one$mom_smoke_pp2)
project_one$mom_smoke_pp12wk = as.factor(project_one$mom_smoke_pp12wk)

project_one$mom_smoke_pp6mo = as.factor(project_one$mom_smoke_pp6mo)
project_one$smoke_exposure_6mo = as.factor(project_one$smoke_exposure_6mo)
project_one$smoke_exposure_12mo = as.factor(project_one$smoke_exposure_12mo)
project_one$smoke_exposure_2yr = as.factor(project_one$smoke_exposure_2yr)
project_one$smoke_exposure_3yr = as.factor(project_one$smoke_exposure_3yr)
project_one$smoke_exposure_4yr = as.factor(project_one$smoke_exposure_4yr)

project_one$smoke_exposure_5yr = as.factor(project_one$smoke_exposure_5yr)
project_one$tsex = as.factor(project_one$tsex)
project_one$language = as.factor(project_one$language)

project_one$tethnic = as.factor(project_one$tethnic)
project_one$tsex = as.factor(project_one$tsex)
project_one$taian = as.factor(project_one$taian)
project_one$tasian = as.factor(project_one$tasian)
project_one$tnhpi = as.factor(project_one$tnhpi)
project_one$tblack = as.factor(project_one$tblack)

project_one$twhite = as.factor(project_one$twhite)
project_one$trace_other = as.factor(project_one$trace_other)
project_one$cig_ever = as.factor(project_one$cig_ever)
project_one$e_cig_ever = as.factor(project_one$e_cig_ever)
project_one$mj_ever = as.factor(project_one$mj_ever)
project_one$alc_ever = as.factor(project_one$alc_ever)
probability_missing = unlist(lapply(project_one, function(x) sum(is.na(x))))/nrow(project_one)

temp_missing_value = sort(probability_missing[probability_missing > 0], decreasing = TRUE)
temp_missing_value = as.data.frame(temp_missing_value)
```

```r
variable_names = c("Self-reported current smoker at first postpartum visit",
                   "child participating diagnosed or suspected
                   of having Autism Spectrum Disorder",
                   "Self-reported current smoker at second postpartum visit",
                   "Average response on the Parental Knowledge
                   Questionnaire on items related to parental control",
                   "Average response on the Parental Knowledge Questionnaire
                   on items related to parental solicitation",
                   "how many of the past 30 days did you drink alcohol?",
                   "Sum of responses on the Brief Problem Monitor on items
                   related to internalizing problems on self",
                   "Average response on the Parental Knowledge Questionnaire
                   on items related to parental knowledge",
                   "Average response on the Parental Knowledge Questionnaire
                   on items related to parental solicitation",
                   "Sum Brief Problem Monitor on items
                   related to attention problems on child",
                   "Child biological sex assigned at birth",
                   "In the past 30 days, on how many days did you use an e-cigarette or vape",
                   "Have you ever had alcohol not as part of a religious ceremony?",
                   "Average response on the Emotion Regulation Questionairre related
                   to Expressive Suppression",
                   "Average response on the Parental Knowledge Questionnaire on items
                   related to child disclosure",
                   "family's estimated annual household income",
                   "Sum of Brief Problem Monitor on items
                   related to externalizing problems on child",
                   "Average response on the Parental Knowledge Questionnaire on
                   items related to child disclosure",
                   "Average response on the Parental Knowledge Questionnaire on
                   items related to parental control",
                   "Child age",
                   "Do you speak another language at home?",
                   "Are you Hispanic or Latino",
                   "Have you ever tried or experimented with cigarette smoking,
                   even a few puff /drags?",
                   "On how many of the past 30 days did you smoke cigarettes",
                   "Have you ever used an e-cigarette or vaped?",
                   "Have you ever used marijuana (including in a vape containing marijuana)",
                   "On how many of the past 30 days did you use marijuana",
                   "Sum of Brief Problem Monitor on items related
                   to attention problems on self",
                   "Sum of Brief Problem Monitor on items related
                   to externalizing problems on self",
                   "Average response on the Emotion Regulation Questionairre related
                   to Cognitive Reappraisal",
                   "In the past 6 months, how often have you used prescription drugs
                   for non-medical reasons?",
                   "Self-reported smoke exposure from mom or partner in 3rd year",
                   "Self-reported smoke exposure from mom or partner in 4th year",
                   "Average response on the Parental Knowledge Questionnaire on
                   items related to parental knowledge",
                   "Sum of Brief Problem Monitor on items related
```

```r
                  to attention problems on self",
                  "Sum of Brief Problem Monitor on items related
                  to externalizing problems on self",
                  "In the past 6 months, how often have you used alcohol
                  (4 or more drinks a day)?",
                  "In the past 6 months, how often have you used tobacco products?",
                  "In the past 6 months, how often have you used illegal drugs?",
                  "On how many of the past 30 days did you smoke cigarettes?",
                  "How many cigarettes per day do you usually smoke?",
                  "Urine cotinine (nicotine metabolite) at 34 wks gestation",
                  "Sum of Brief Problem Monitor on items related
                  to internalizing problems on child",
                  "Self-reported smoke exposure from mom or partner from 0 to 6 months",
                  "Self-reported smoke exposure from mom or partner from 7 to 12 months",
                  "Self-reported smoke exposure from mom or partner in 2nd year",
                  "Self-reported smoke exposure from mom or partner in 5th year",
                  "Sum of Brief Problem Monitor on items related
                  to internalizing problems on self",
                  "Average response on the Emotion Regulation Questionairre related
                  to Cognitive Reappraisal",
                  "Average response on the Emotion Regulation Questionairre related
                  to Expressive Suppression",
                  "Self-reported current smoker at 32 weeks pregnant",
                  "Urine cotinine (nicotine metabolite) at 6 months postpartum from baby",
                  "Urine cotinine (nicotine metabolite) at 6 months postpartum from mom",
                  "Parent age",
                  "Parent biological sex assigned at birth",
                  "Do you speak another language at home",
                  "Are you Hispanic or Latino",
                  "Do you currently have a paid job?",
                  "What is the highest level of education you have completed?",
                  "Self-reported current smoker at 6 months postpartum",
                  "Self-reported current smoker at 22 weeks pregnant",
                  "Self-reported current smoker at 12 weeks postpartum",
                  "Self-reported current smoker at 16 weeks pregnant")

temp_missing_value$variable_name = variable_names

temp_missing_value = temp_missing_value %>% relocate(variable_name, .before = temp_missing_value)

temp_missing_value <- temp_missing_value %>%
      rename("missing percentage" = "temp_missing_value")

gt_tbls = gt(temp_missing_value)

gt_tbls <-
  gt_tbls |>
  tab_header(
    title = "Missing percentage of variables" )

gt_tbls

temp = project_one %>%
```

```r
  mutate(
    prace =  case_when(
      pethnic == 1 ~ "Hispanic/Latino",
      paian == 1 ~ "American Indian/Alaska Native",
      pasian == 1 ~ "Asian",
      pnhpi == 1 ~ "Native Hawaiian/Pacific Islander",
      pblack  == 1 ~ "Black/African American",
      pwhite == 1 ~ "White",
      prace_other == 1 ~ "Other",
      is.na(pethnic)|pethnic == 2 ~ NA
    ),
    trace = case_when(
      tethnic == 1 ~ "Hispanic/Latino",
      taian == 1 ~ "American Indian/Alaska Native",
      tasian == 1 ~ "Asian",
      tnhpi == 1 ~ "Native Hawaiian/Pacific Islander",
      tblack  == 1 ~ "Black/African American",
      twhite == 1 ~ "White",
      trace_other == 1 ~ "Other",
      is.na(tethnic)|tethnic == 2 ~ NA
    )
  ) %>%
  dplyr::select(c(parent_id, page, psex, plang, prace, employ, pedu, income, tage, tsex, language, trac
temp$page = as.numeric(temp$page)
temp$tage = as.numeric(temp$tage)
temp$prace = factor(temp$prace, levels = c("American Indian/Alaska Native",
                                           "Hispanic/Latino", "Native Hawaiian/Pacific Islan
temp$trace = factor(temp$trace, levels = c("American Indian/Alaska Native",
                                           "Hispanic/Latino", "Native Hawaiian/Pacific Islan

temp = temp %>%
  mutate(
    psex = factor(psex, levels = c(0, 1), labels = c("Male", "Female")),
    tsex = factor(tsex, levels = c(0, 1), labels = c("Male", "Female")),
    plang = factor(plang, levels = c(0, 1), labels = c("No", "Yes")),
    language = factor(language, levels = c(0, 1), labels = c("No", "Yes")),
    employ = factor(employ, levels = c(0, 1, 2), labels = c("No", "Part-Time", "Full-Time")),
    pedu = factor(pedu, levels = c(0, 1, 2, 3, 4, 5, 6), labels = c("Some high school", "High school",
                                                     "GED", "Some college", "2-year degr
                                                     "Postgraduate degree"))
  )

Demo_p = temp[, c("page", "psex", "plang", "prace", "employ", "pedu", "income")] %>%
  tbl_summary(label = list(page ~ "Age",
                           psex ~ "Sex",
                           plang ~ "Language",
                           prace ~ "Race",
                           employ ~ "Employ",
                           pedu ~ "Education",
                           income ~ "Income"))
Demo_t = temp[, c("tage", "tsex", "language", "trace")] %>%
  tbl_summary(type = list(tage ~ "continuous"),
              label = list(tage ~ "Age",
```

```r
                                   tsex ~ "Sex",
                                   language ~ "Language",
                                   trace ~ "Race"))

parent_demo = temp %>%
  dplyr::select(c(parent_id, page, psex, plang, prace)) %>%
  rename_with(~ sub("^p", "", .x), starts_with("p")) %>%
  mutate(participant = "Parent")
colnames(parent_demo)[c(1,4)] <- c("parent_id","language")

child_demo = temp %>%
  dplyr::select(c(parent_id, tage, tsex, language, trace)) %>%
  rename_with(~ sub("^t", "", .x), starts_with("t")) %>%
  mutate(participant = "Child")

data_demo_long = bind_rows(parent_demo, child_demo)

data_demo_long[,-1] %>% tbl_summary(by = participant,
                                    label = list(
                                      parent_id = "Parent ID",
                                      age = "Age",
                                      sex = "Sex",
                                      language = "Language Spoken at Home",
                                      race = "Race/Ethnicity")) %>%
  modify_header(all_stat_cols() ~ "**{level}**") %>%
  modify_caption("Summary for Demographic Variables") %>%
  bold_labels() %>%
  as_kable_extra(booktabs = T, escape = T) %>%
  kable_styling(latex_options = "HOLD_position")
project_one$income = as.numeric(project_one$income)
temp_temp = project_one %>%
  dplyr::select(c(pedu,income, employ)) %>%
  mutate(pedu = case_when(
    pedu == 0 ~ "some high school",
    pedu == 1 ~ "high school",
    pedu == 2 ~ "GED",
    pedu == 3 ~ "some college",
    pedu == 4 ~ "2 year degree",
    pedu == 5 ~ "4 year degree",
    pedu == 6 ~ "postgraduate degree"),
    employ = case_when(
      employ == 0 ~ "No",
      employ == 1 ~ "Part-time",
      employ == 2 ~ "Full-time"
    ))
temp_temp %>%
  dplyr::select(c(pedu,income, employ)) %>%
  tbl_summary(label = list(
    pedu = "highest level of education",
    employ = "Do you currently have a paid job?",
    income = "family's estimated annual household income"),
              statistic = list(all_continuous() ~ "{mean} ({sd})"),
              missing_text = "MISSING") %>%
```

```r
  modify_header(label = "**Characteristics**",
                stat_1 = "**Parent (N=47)**") %>%
  as_kable_extra(booktabs = TRUE,
                 format = "latex",
                 longtable = TRUE,
                 caption = "Characteristics of Parents") %>%
  kable_styling(font_size = 7, latex_options = "hold_position")
project_one = project_one %>%
  mutate(smoke_conditions = case_when(
    mom_smoke_16wk == 1 & mom_smoke_22wk == 1 & mom_smoke_32wk == 1 ~ "heavy smoker",
    mom_smoke_16wk == 2 & mom_smoke_22wk == 2 & mom_smoke_32wk == 2 ~ "non smoker",
    .default = "moderate smokers"
  ))
project_one %>%
  dplyr::select(c(smoke_conditions, childasd, bpm_att_a, bpm_att_p, erq_cog_a,
                  erq_cog, bpm_att, erq_exp, erq_exp_a)) %>%
  tbl_summary(by = "smoke_conditions",
              statistic = list(all_continuous() ~ "{mean} ({sd})"),
              missing_text = "MISSING") %>%
  add_p() %>%
  #modify_header(label = "**Characteristics**",
                #stat_1 = "**Likely Hyperactive/Impulsive type**",
                #stat_2 = "**Not likely Hyperactive/Impulsive type**") %>%
  as_kable_extra(booktabs = TRUE,
                 format = "latex",
                 longtable = TRUE,
                 caption = "Effects of Smoking During Pregnancy on adolescent self-regulation") %>%
  kable_styling(font_size = 7, latex_options = "hold_position")
project_one %>%
  dplyr::select(c(cig_ever, num_cigs_30, e_cig_ever, num_e_cigs_30, mj_ever,smoke_conditions,
                  num_mj_30, alc_ever, num_alc_30)) %>%
  tbl_summary(by = "smoke_conditions",
              statistic = list(all_continuous() ~ "{mean} ({sd})"),
              missing_text = "MISSING") %>%
  add_p() %>%
  #modify_header(label = "**Characteristics**",
                #stat_1 = "**Likely Hyperactive/Impulsive type**",
                #stat_2 = "**Not likely Hyperactive/Impulsive type**") %>%
  as_kable_extra(booktabs = TRUE,
                 format = "latex",
                 longtable = TRUE,
                 caption = "Effects of Smoking During Pregnancy on adolescent substance use") %>%
  kable_styling(font_size = 7, latex_options = "hold_position")
project_one %>%
  dplyr::select(c(bpm_ext_p, bpm_ext_a, swan_hyperactive, swan_inattentive, childasd,smoke_conditions))
  tbl_summary(by = "smoke_conditions",
              statistic = list(all_continuous() ~ "{mean} ({sd})"),
              missing_text = "MISSING") %>%
  add_p() %>%
  #modify_header(label = "**Characteristics**",
                #stat_1 = "**Likely Hyperactive/Impulsive type**",
                #stat_2 = "**Not likely Hyperactive/Impulsive type**") %>%
  as_kable_extra(booktabs = TRUE,
```

```
                format = "latex",
                longtable = TRUE,
                caption = "Effects of Smoking During Pregnancy on adolescent externalizing") %>%
  kable_styling(font_size = 7, latex_options = "hold_position")
```