Project 1

Han Ji

2023-10-08

Data Description

The data set is a survey study with 49 pairs of children and their mothers at birth. The main interest of this study is to examine how smoking during pregnancy (SDP) and environmental tobacco smoke (ETS) exposure will affect self-regulation, externalizing behavior, and substance use on children. This study is based on a previous study that recorded the smoking exposure from pregnancy to 6 months postpartum, and these 49 pairs of children and mothers are randomly and recruited to the current study. In addition to the previous records, this data set contains demographic information of mothers and children, their substance use habits, and the evaluations of their self-regulation, externalizing behavior, and parental knowledge. Also, the survey asks mothers to reflect whether the children were exposed to smoking from 0 to 6 months until 5 years postpartum.

Data Processing

We change outliers to the maximum possible or observed values. For example, in the question "on how many of the past 30 days did you smoke cigarettes?", we change all values above 30 into 30 days. For the questionnaire responses, if there are some responses missing, we use the non-missing values approximate the participant's answer (observed mean for average response; observed mean times the number of questions for sum). Otherwise, the number of complete data within each type of questionnaire would be too small.

Data Quality

Missing Data

For survey data, it's common to have many missing values. In this survey, the demographic variable race is presented as checkbox for participants where 0 means the participant didn't check the corresponding option, and thus this variable doesn't have missingness. We then use the number of races checked as indicator to explore missing patterns. There are 13 children didn't check any race options, and 12 of them miss all child-related information (Fig. 1). Moreover, one of them misses all information from the previous study. There are 8 parents didn't check any race options, and all of them miss all parent-related information in the current survey. 7 observations even miss all information related to the current survey. This causes the actual data size to be even smaller.

Besides, there are missing patterns within each category of variables For example, some observations miss all substance use information instead of only one or two responses missing, implying an occurrence pattern (Fig. 2).

Consistency

We also checked the consistency between related variables to see if the quality of this survey data is good. There are two questions related to the smoking habit of mothers: "In the past 6 months, how often have

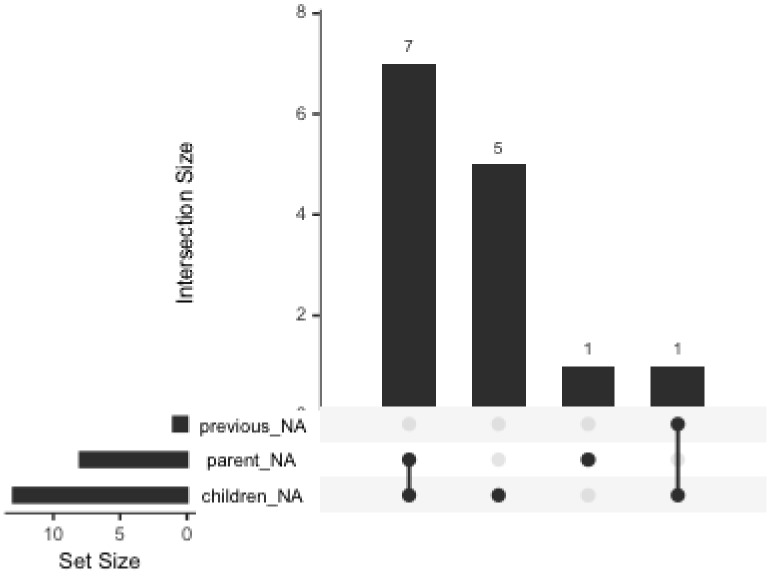

Figure 1: Missing patterns observed in the data. The rows represent missing in the all information related tp the previous study, parent information, and children information in the current study, respectively.

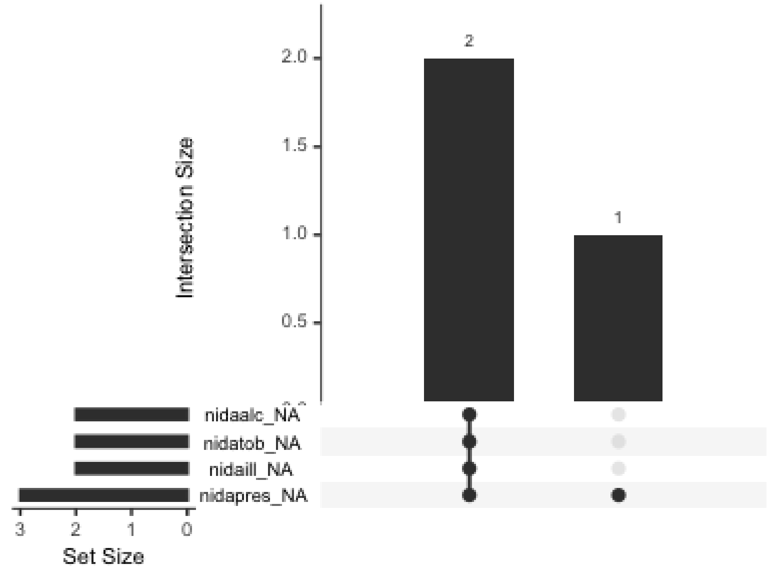

Figure 2: Missing patterns observed within substance use of mothers.

you used tobacco products?” and “On how many of the past 30 days did you smoke cigarettes?” We assume the responses are correlated, and they should answer yes on the first question if the number of cigarettes is more than 0. However, we observe a strong inconsistency between the responses (Fig. 3). Most participants who responded “never” or “once or twice” reported 0 days of smoking in the past month, but 2 participants responded 30 days. This may imply that they responded one question incorrectly or misinterpreted the question.

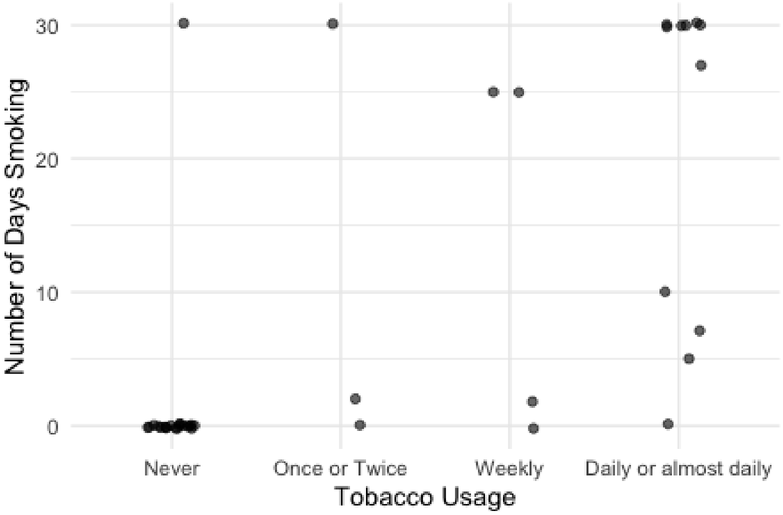

Figure 3: Number of days smoking in the past 30 days versus the tobacco usage of mothers in the past 6 months. No participants responded monthly so this option is removed from the plot.

In the previous study, the urine cotinine, a marker for nicotine metabolite, was measured for mothers at 34 weeks gestation and measured for both mothers and babies at 6 months postpartum. We see the nicotine at 34 weeks gestation is a good indicator when comparing the number of smoking exposure responded from 16 to 32 weeks at pregnancy (Fig. 4).

We combined all responses from 0 to 6 months in the previous and current studies into a new exposure variable, and we want to check if cotinine is still a good indicator for exposure (Fig. 5). A few babies have relatively high cotinine values even if no smoking exposure reported and their mothers had low cotinine values. Also, we only see expected correlation between maternal and child cotinine values in the subset with smoking exposure reported. This result is counterintuitive and may imply some inconsistency or entry errors in our data.

Questionnaire Evaluation

In this survey, there are four questionnaires presented to the participants: Brief Problem Monitor, Emotion Regulation, SWAN Rating for ADHD, and Parental Knowledge. We are mainly interested in the first three and if they are associated with smoking exposure. We observe strong correlations within Brief Problem Monitor and SWAN Rating (Fig. 6, 7). Especially, we see many children are suspected to be ADHD based on the SWAN Rating responded by parents. This becomes susceptible whether this prevalence of ADHD relates to the smoking exposure since the we saw different levels of exposure. It is possible that he parents’ responses don’t reflect the actual behavior of their children, or the children in this data set are likely to be ADHD at the baseline. Also, the number of ASD diagnosized and suspected cases is too small, so we decided not to explore more on that variable. In addition, we also look at the correlations between answers from children and parents on child for the same survey, but none of them are high.

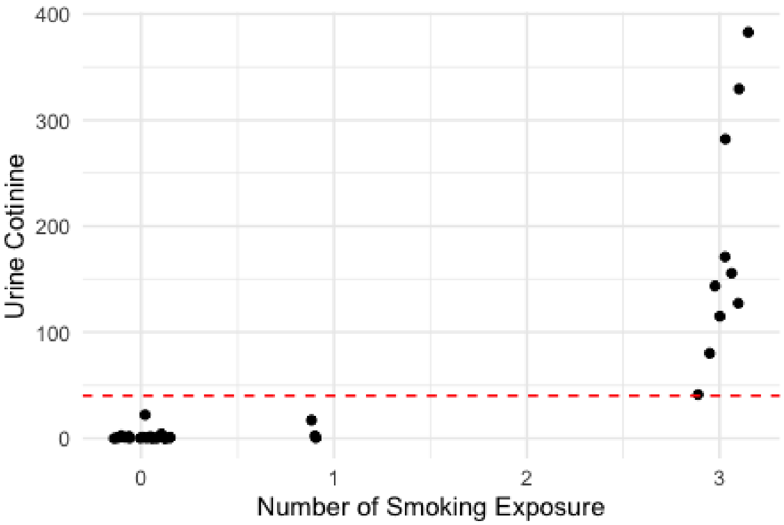

Figure 4: Urine cotinine from mothers at 34 weeks gestation versus the number of smoking exposure reported . The red dashed line proposed at 40 ug/mL can separate the data clearly.

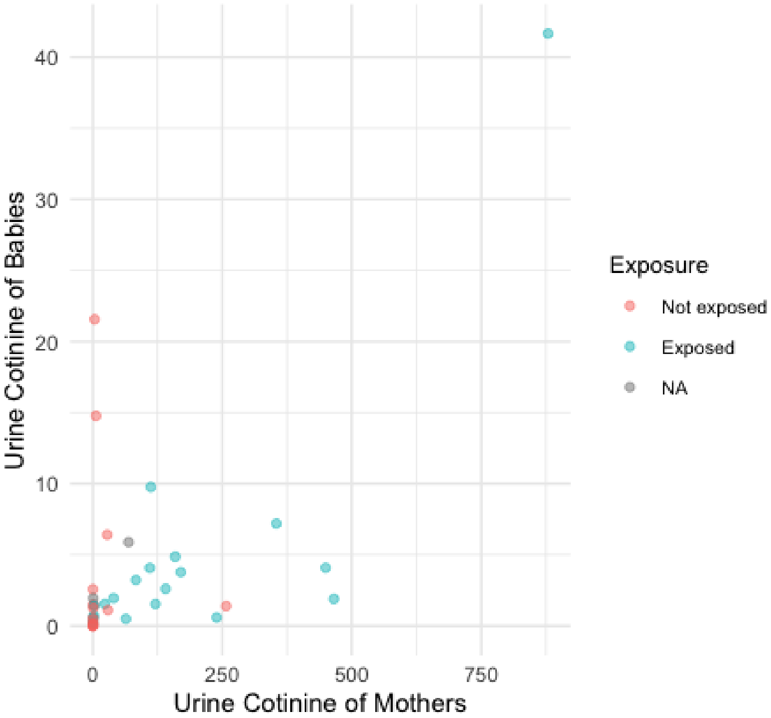

Figure 5: Urine cotinine from babies versus from mothers at 6 months postpartum. Points are colored by whether the babies were exposed to smoking from 0 to 6 months.

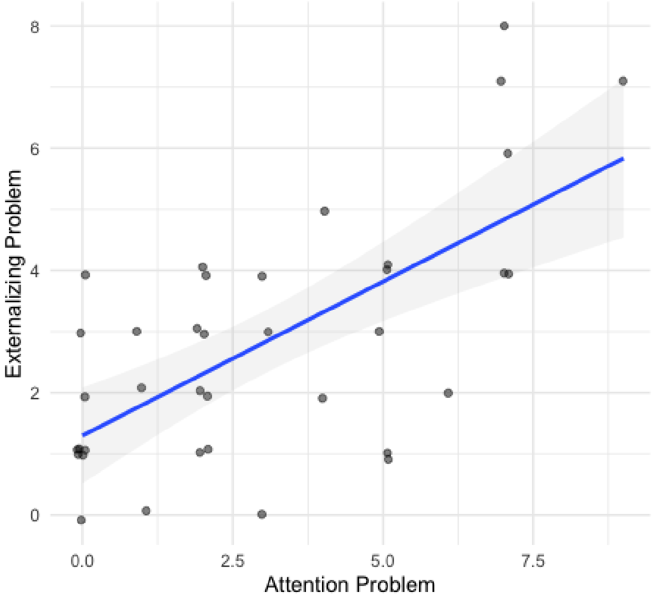

Figure 6: Sum of responses related to attention problems and externalizing problems responded by children.

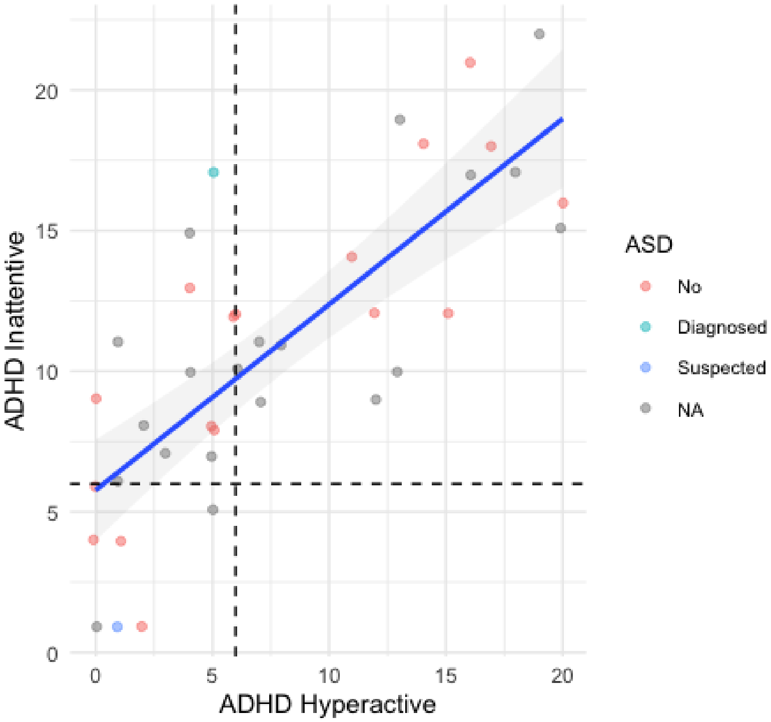

Figure 7: Sum of responses on SWAN Rating responded by parents. A score of 6 or greater indicates that the child is likely ADHD (dashed lines), and the type is on the axes.

Prenatal/Postnatal Exposure with Self-regulation

We focus on the counts of prenatal and postnatal exposures as the severity of exposures, cotinine values, and indicators whether mothers smoked during pregnancy and at 6 month postpartum combined from previous and the current studies. We want check their associations with the questionnarie responses. Among all pairwise comparisons, we see that the the indicator whether mothers smoked at 6 month postpartum can explain the attention problems somewhat (Fig. 8). Also, it also correlated with the SWAN inattentive scores (Fig. 9). Most observations are likely inattentive ADHD if they were exposed to smoking at 6 month postpartum. This may imply that smoking exposure at 6 month postpartum is related to attention problems on children overall.

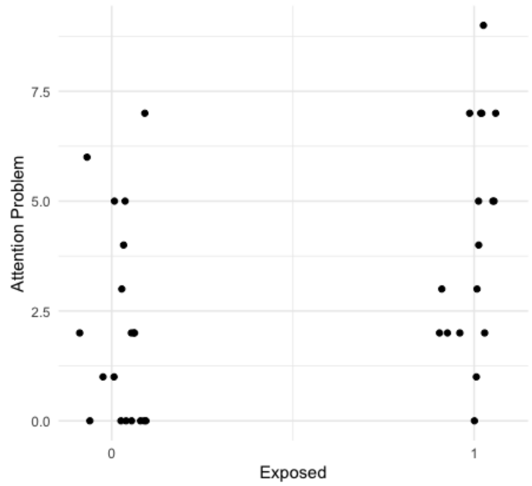

Figure 8: Sum of attention problems reported by children versus exposure to smoking at 6 month postpartum.

We also look at whether smoking exposure is related to substance use of children. However, the number of observations with substance use is too small, so it is not easy to find a clear pattern or draw a conclusion (Fig. 10).

Conclusion and Limitations

Based on our exploratory analysis, we create some indicator variables for smoking exposure and severity based on the previous and current studies, and the exposure is correlated with the marker urine cotinine from mothers at least. We also see that whether the child exposed to smoking at 6 months postpartum is correlated with the attention problems in Brief Problem Monitor and SWAN Rating.

However, the quality of this data is not good, especially that the number of complete cases with both parent and child information about smoking and self-regulation is too small. Thus, we shouldn't be confident in any findings because it only represents a small subset of the population. Also, the inconsistency in the data makes the answers of the survey less trustworthy, and we may doubt whether the relationship we find reflect the truth or not.

Data Privacy

The data is provided by Dr. Lauren Micalizzi from the Department Behavioral and Social Sciences at Brown. The original data cannot be shared directly for privacy. For data-related question, please email lauren_micalizzi@brown.edu.

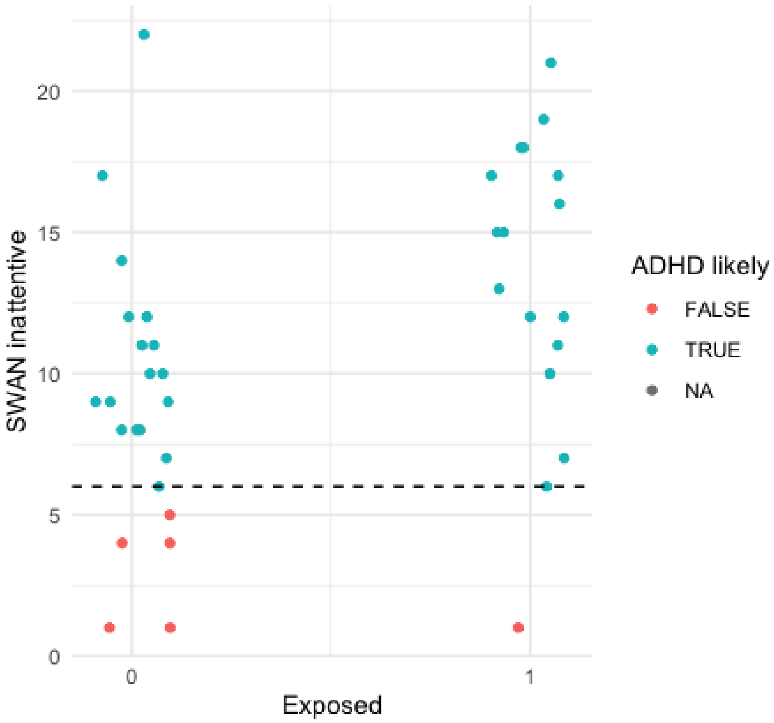

Figure 9: SWAN inattentive scores reported by mothers versus exposure to smoking at 6 month postpartum.

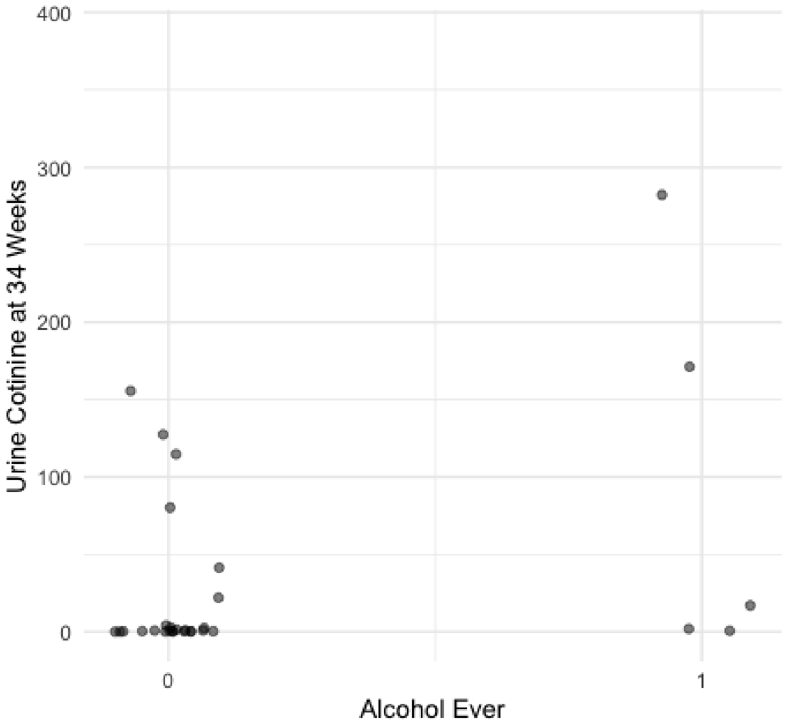

Figure 10: Urine Cotinine at 34 weeks versus children drinking alcogol ever. The number of observations with using alcohol is too low to draw a clear pattern.

Code Availability

The code used to analyze this data can be found on: https://github.com/RuBBiT-hj/PHP2550_Project1. Although the data cannot be accessed, the code contains the steps of processing and how we explore among different values with the detailed comments.