# The harm of maternal smoking during pregnancy to the health of newborns*

## A propensity score matching case study on the effect of smoking during pregnancy on birthweight and length of pregnancy

Ruoyun Wang

27 April 2022

**Abstract**

This paper studies the impact of maternal smoking on newborns' health. A property score matching method is used to march these observations to eliminate the difference between smoking and non-smoking mothers in the sample. The matching data is used to study the effect of maternal smoking on babies' birth weight and premature birth. The results showed that smoking would decrease birth weight and shorten the length of pregnancy. Thus, smoking during pregnancy is harmful to a baby's health..

**Keywords**: Smoking, birth weight, propensity score matching, health

# 1 Introduction

Smoking is harmful to yourself and adversely affects the health of those around you. Studies have shown that lung cancer in non-smoking women may be caused by inhaling fumes from their husbands (Smith 2003). Passive smoking is also harmful to the human body. The fetus gets Nutrition through the mother, and the mother's habits will affect the fetus's health. Babies of smoking mothers are more likely to develop the respiratory disease than other infants (Stick et al. 1996). Birth weight is one of the essential indicators to measure the health of babies. Low birth weight will have many adverse effects on infants. In the United States, 65% of infant events occur in low birth weight (LBW) infants (<2500). The causes of LBW are unknown. Environment, Nutrition, genetics and so on may cause LBW (Wang et al. 2002).

In addition to the newborn's weight, the length of gestation is another indicator to measure the newborn's health. The average length of human pregnancy is 280 days or 40 weeks ("A.d.a.m. Medical Encyclopedia [Internet]. Johns Creek (GA)" 2020). If the gestation period is less than 37 weeks, it will be considered premature. In general, after 34 weeks of gestation, the survival rate of the baby increases substantially. Many factors can lead to premature birth, such as drinking or smoking during pregnancy, low maternal weight, insufficient weight gain, prenatal care, mood swings, diabetes and other diseases. However, a long pregnancy cycle can also have adverse effects, so the baby needs to be born by labour induction beyond 42 weeks of pregnancy.

Both the mother's habits and weight impact the baby's health. Mothers generally gain weight due to careful family care during pregnancy and better nutrition. The baby will also cause the mother to gain weight, and most women gain 25 to 35 lbs during pregnancy. Pregnant women are not advised to diet or lose weight during pregnancy, and they should eat healthy food and do more exercise. Dieting may cause a baby to be underweight, while overeating will cause a baby to be overweight, making it challenging to produce. Therefore, antenatal care is particularly important for pregnant mothers, especially maternal weight management.

From the previous analysis, it can be seen that the mother's bad habits may have adverse effects on the baby's health, such as reducing the baby's weight, shortening the pregnancy cycle, causing premature birth outcomes,

*Code and data are available at: https://github.com/royintot/sta304-final and affecting the mother's weight gain. We generally use the regression method to study the effect of smoking. Due to selection bias, the regression results could not reflect the causal relationship between smoking and infant health indicators. We want the smoking group to be similar to the no-smoking group, i.e. have similar ages, genders, etc. However, we cannot achieve a perfect match between the control and treatment groups when there are many explanatory variables. The probability score match method is proposed to solve the unbalance problem of the smoking and no smoking groups.

This paper mainly includes the following parts: 1. The introduction of the survey data. 2. The regression model and propensity score method. 3. The analysis results for the given survey data 4. Discussion of the importance of the results and the possible problems in this study

# 2 Data

In this paper, the dataset randomly sampled the birth information of infants in the United States in 2014 (Çetinkaya-Rundel et al. 2021). There are 1000 observations and 13 variables. These variables include father's age, mother's age, maturity status, length of pregnancy, premature status, number of hospital visits during pregnancy, weight gained by mother during pregnancy, birth weight, sex of baby, habit, marital status and race. In this dataset, length of gestation, premature status, weight gained by mother during pregnancy and birthweight are the response variables, and habit is the variable of interest. The other variables like father's age, mother's age, marital status and race are the variables that need to be balanced between the control and treatment groups.

Since the variable "visit to hospital" has many missing values, I excluded this variable from the dataset and then removed the observations with missing values. After the data cleaning, 851 observations are kept. Table 1 shows the distribution of the categorical variables. From Table 1, we find that the number of smokers in the sample is 92, while the number of nonsmokers is 743, which is much higher than the number of smokers. There are 60 low birth weight babies in the sample, accounting for 7.2%. Even now, there is a large proportion of low birth weight. It is necessary to investigate those factors that lead to low birth weight. The gender distribution of babies is relatively even, close to half and half. In the United States, unmarried pregnancies are also high, reaching 33.4%. White mothers accounted for 79% of pregnant mothers, far higher than other races.

Table 2 shows the distribution of categorical variables in the smoker, and Table 3 shows the distribution of categorical variables in the nonsmoker. Comparing Table 2 and Table 3, we find that 17.4% of the children of smoker moms have low birth weight, while the probability of nonsmokers moms is only 7.2%. In addition, the proportion of smoker moms giving birth to boys was 54.3, which was significantly higher than that of girls. Among nonsmoker moms, the difference between males and females is only 0.2%. It can be seen that smokers may cause sex disproportion in infants. There were also significant differences in marital status between the Smoker group and the nonsmoker group. Smoking mothers were married at 45.7%, while nonsmokers were married at 69.2%, higher than smoker moms. There were also considerable differences in age and race among pregnant mothers between smokers and nonsmokers. From the above analysis that the smoker group and the nonsmoker group have obvious unbalance problems.

Table 4 shows the distribution of the numerical variables in the sample. The average father's age is 31.2, and the minimum is 15 while the maximum is 85. The average mother's age is 28.9, a little lower than that of the father, the minimum is 14, and the maximum is 46. It can be seen that the reproductive age range of males is much more extensive than that of females. The peak reproductive age for men or women occurs around the age of 30, and only a few men still have children over the age of 60. As shown in Figure 1, in nonsmokers, father's age has a slight positive skewness. In contrast, in smoker father, except for one age reaching 85 years old, the other ages are kept around 30 years old, and approximate symmetry. For the mother's age, in the nonsmoker group, the distribution of mother's age is approximately symmetric. In the smoker mom, the age distribution is more scattered, and there is no apparent symmetric relationship.

By comparing the length of pregnancy, in the nonsmoker group, the number of weeks of pregnancy is concentrated in 39 weeks. In the smoker group, there are several moms whose pregnancy weeks are less than 30 weeks, particularly premature babies. Also, nonsmoker and smoker groups show positive skewness features

Table 1: categorical variable in sample

| | | N | % |
|---|---|---|---|
| mature | mature mom | 142 | 17.0 |
| | younger mom | 693 | 83.0 |
| habit | nonsmoker | 743 | 89.0 |
| | smoker | 92 | 11.0 |
| lowbirthweight | low | 60 | 7.2 |
| | not low | 775 | 92.8 |
| sex | female | 413 | 49.5 |
| | male | 422 | 50.5 |
| marital | married | 556 | 66.6 |
| | not married | 279 | 33.4 |
| whitemom | not white | 175 | 21.0 |
| | white | 660 | 79.0 |

Table 2: categorical variable in smoker

| | | N | % |
|---|---|---|---|
| mature | mature mom | 10 | 10.9 |
| | younger mom | 82 | 89.1 |
| lowbirthweight | low | 16 | 17.4 |
| | not low | 76 | 82.6 |
| sex | female | 42 | 45.7 |
| | male | 50 | 54.3 |
| marital | married | 42 | 45.7 |
| | not married | 50 | 54.3 |
| whitemom | not white | 14 | 15.2 |
| | white | 78 | 84.8 |

for mothers gain weight. Comparing the birth weight, we can see that the birth weight of the nonsmoker group obeys the approximately normal distribution, and the mean value is around 7.5. In contrast, the smoker group shows apparent negative skewness.

# 3 Model

From the above analysis, we know that length of gestation, premature status, weight gained by the mother during pregnancy and birthweight are the response variables as infant health indicators. The premature status is a categorical variable among the four response variables, which is analyzed by the logistic regression method. The other three numerical variables are analyzed by the multiple linear regression method. Considering the significant difference between smoking and non-smoking group variables, we need to use the property score matching method to match the smoking group's observations to achieve similar results in the control and treatment groups. This paragraph will introduce the principles and assumptions of MLR, logistic regression, and the property score matching method.

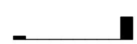
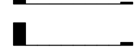
## 3.1 Multiple linear regression

Multiple linear regression refers to a statistical method that uses two or more independent variables to predict the response variable, which can determine the scale and ability of each independent variable to explain the response variable. MLR assumes a linear relationship between the response variable and other independent variables, and its model is (Anselin 2003):

Table 3: categorical variable in nonsmoker

| | | N | % |
|---|---|---|---|
| mature | mature mom | 132 | 17.8 |
| | younger mom | 611 | 82.2 |
| lowbirthweight | low | 44 | 5.9 |
| | not low | 699 | 94.1 |
| sex | female | 371 | 49.9 |
| | male | 372 | 50.1 |
| marital | married | 514 | 69.2 |
| | not married | 229 | 30.8 |
| whitemom | not white | 161 | 21.7 |
| | white | 582 | 78.3 |

Table 4: numerical variable in sample

| | Unique (#) | Missing (%) | Mean | SD | Min | Median | Max | |
|---|---|---|---|---|---|---|---|---|
| fage | 38 | 0 | 31.2 | 7.1 | 15.0 | 31.0 | 85.0 |  |
| mage | 30 | 0 | 28.9 | 5.6 | 14.0 | 29.0 | 47.0 |  |
| weeks | 22 | 0 | 38.7 | 2.4 | 23.0 | 39.0 | 46.0 |  |
| premie | 2 | 0 | 0.9 | 0.3 | 0.0 | 1.0 | 1.0 |  |
| gained | 78 | 0 | 30.8 | 15.2 | 0.0 | 30.0 | 98.0 |  |
| weight | 323 | 0 | 7.3 | 1.2 | 1.1 | 7.4 | 10.4 |  |
| lbw | 2 | 0 | 0.1 | 0.3 | 0.0 | 0.0 | 1.0 |  |
| smoker | 2 | 0 | 0.1 | 0.3 | 0.0 | 0.0 | 1.0 |  |

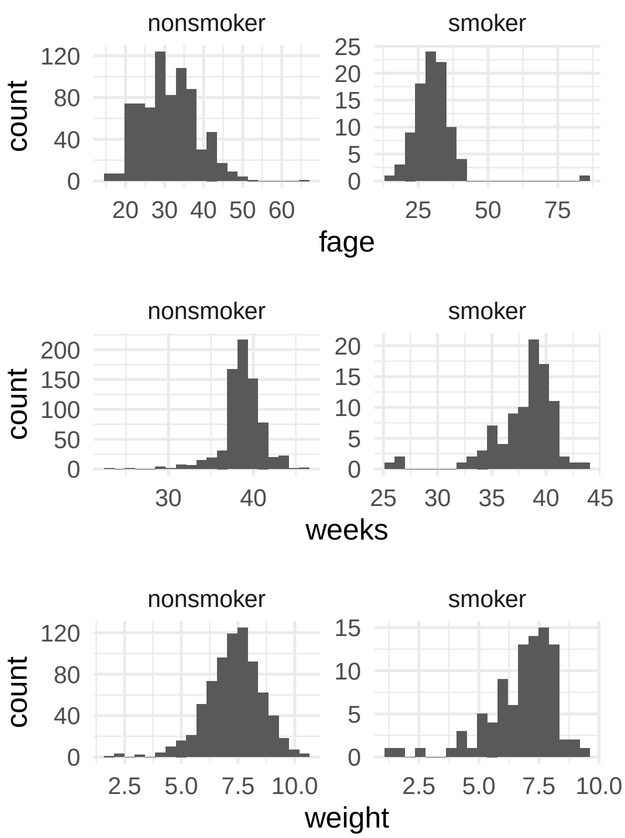



Figure 1: Distributions of numerical variables

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} + \epsilon_i, \quad i = 1, ..., n$$

Where, $y_i$ is the response variable, $x_i$ is the value of the predictor variable of observation i; $\beta_0, \beta_1, ..., \beta_p$ are the unknown coefficients need to be estimated; $\epsilon_i$ is the error term of observation i.

In order to maker the regression results are valid, there are some assumptions need to be satisfied:

A1. $Y$ should be linearly related to X;

A2. The errors terms are uncorrelated, $cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$;

A3. The variance of the are constant errors $Var(\epsilon_i) = \sigma^2, for\ i = 1, ..., n$

A4. The errors are normally distributed

In order to estimated the coefficients of $\beta$, we can write the MLR model into matrix form:

$$Y = X'\beta + \epsilon$$

and the $\epsilon \sim N(0, \sigma^2 I)$. If these assumption holds, we can use the least square method to estimate $\beta_s$, the estimated formula is:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

After the regression, we can use residual plot to check if these assumptions hold for the given model.

## 3.2 logit regression

The output of low birth weight is a category variable, we can code it as a binary response variable. Logistic regression is a statistical tool that uses study the predictors on the binary response variable. The regression model is:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_{smoker} + \beta_2 x_{fage} + \beta_3 x_{mage} + \beta_4 x_{mature} + \beta_5 x_{white} + \beta_6 x_{marital} + \beta_7 x_{sex}$$

Where:

$p$ is the propability to be low birth weight

$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$ are the coefficients to be estimated.

The assumption for logistic regression model is (Z, 2020):

Assumption 1: the response variable is binary

Assumption 2: The observations are independent

Assumption 3: There is no multicollinearity among explanatory variables

Assumption 4: There are no extreme outliers

Assumption 5: There is a Linear Relationship Between Explanatory Variables and the Logit of the Response Variable

Assumption 6: The Sample Size is Sufficiently Large

After we get the estimation of the regression model by using the glm function, we can get the probability of a baby to be low birth weight:

$$\hat{p} = \frac{exp(\beta_0 + \beta_1 x_{smoker} + \beta_2 x_{fage} + \beta_3 x_{mage} + \beta_4 x_{mature} + \beta_5 x_{white} + \beta_6 x_{marital} + \beta_7 x_{sex})}{1 + exp(\beta_0 + \beta_1 x_{smoker} + \beta_2 x_{fage} + \beta_3 x_{mage} + \beta_4 x_{mature} + \beta_5 x_{white} + \beta_6 x_{marital} + \beta_7 x_{sex})}$$

We can test the coefficient of $\beta_1$ to study the effect of smoking on babies' health.

### 3.3 Propensity score matching

From the data analysis, we have seen that the distribution of predictor variables are not balance in smoking and nonsmoking groups. This may cause the regression results biased to the true effect of smoking. Thus, we need to use propensity score matching method to match the observations in smoking group and the observations in the non-smoking group. The propensity score method has 5 key steps:

1. collect data
2. estimate propensity score
3. match records
4. evaluate matching
5. evaluate treatment effect on the outcome.

In the estimate propensity score step, we use a logistic regression model to estimate the propensity score and match the most similar control observations with the observations in the treatment group. We often use a one-to-one matching method, and the probability score is a balanced score, which does not mean that each variable will be completely matched. And the property score method can only match the variables that can be observed but cannot match those variables that cannot be observed. Therefore, even using the property score method, it is still difficult for us to obtain the causal relationship between smoking and babies' health.

## 4 Results

### 4.1 Results without propensity score matching

Table 7 shows the effect of smoking on babies' health without property score matching. Column 1 studies the impact of length of pregnancy, the response variable of Column 2 is baby birth weight, and column 3 shows the effect of smoking on mother weight gain in pregnancy. Column 4 is the result of logistic regression, reflecting the impact of tobacco on whether it causes low birth weight. Data in parentheses are the standard derivation of estimated parameters (Demin 2022).

Column 1 shows that the coefficient of a smoker is -0.738. The p-value is less than 0.01, so we have significant solid evidence that smoking will affect the length of pregnancy. The estimated value tells us that if we hold other variables constant, the length of pregnancy of a smoker mother is 0.74 weeks less than that of a nonsmoker mother. And the baby of a smoker mother may have more chance of being premature. In addition to this, we found that white moms will have longer pregnant length. Other explanatory variables did not show a significant effect on length of pregnancy.

Column 2 studies the effect of smoking on birth weight. The coefficient of a smoker is -0.556, and the p-value is less than 0.001, so we have very strong significant evidence to say that smoking will lose birth weight. If we hold other variables constant, the birth weight of the smoker mother is 0.56 pounds less than the nonsmoker mother. In addition to the effect of smoking on birth weight, we also found that a white mom's birth weight teaches nonwhite mom's birth weight to be 0.412 pounds higher, while a male baby's birth weight is 0.382 pounds more elevated than a female baby's.

Column 3 examines the effect of smoking on maternal weight gain during pregnancy. The results show that all variables are significant, and the R square of the corresponding model is less than 1%. We, therefore, believe that smoking does not affect maternal weight gain.

Finally, column 4 studies whether smoking will cause low birth weight. The response variable is the probability of low birth weight. The coefficient of the smoker is 1.318, so we conclude that we have very strong evidence to show that a smoker mother will have more chance of getting a low birthweight baby.

### 4.2 Results with propensity score matching

First, we got the matching dataset with the help of the property match function in the matchit library (Ho et al. 2011). There are 184 groups of data in the matching dataset, and the one-to-one matching relationship between smoking and nonsmoking group observations is realized. The matched data summary is shown in Table 6 and Table 7 (Arel-Bundock 2022). After matching, the proportion of mature moms in the smoker group is 89.1%, similar to 87% in the nonsmoker group. At the same time, in terms of marital status, the married ratio is 45.7% in the smoker group, while 47.8% in the nonsmoker group. The white mom ratio is 84.8% in the smoker group while 87% in the non-smoker group. Overall, the matched samples are more similar. However, Figure 2 observes the distribution of numerical variables after matching, and we find that there is still a big difference between the smoker group and the nonsmoker group (Auguie 2017). While matching realizes the matching of each smoker group observation, it is impossible to distinguish that there is a more obvious outlier in the smoker group.

I repeat these regressions in Table 5 on the matched data, and the results are shown in Table 8. After matching, the coefficient of smoker in column 1 is -0.988, which is less than the one in Table 5. Thus, smoking effect on the length of pregnancy becomes more pronounced. In addition, comparing the R square results in column 1, it is found that the R square of the model after matching is 8.9%, while the R square before matching is 1.7%. Thus, the model after matching can explain more variation in the length of pregnancy.

In the second column of results, the coefficient of the smoker is -0.492, whose scale is less than the one in Table 5. Thus, the effect of smoking on birth weight in the OLS method is overestimated. We can also find that the R square in column 2 is more extensive than in Table 5.

The results in Column 3 show that no predictor has significance in the mother's weight gain, the same as the conclusion in Table 5.

The effect of smoking on the probability of low birth weight in column 4 is 1.295, while the coefficient is 1.318 in Table 5. However, the p-value of the coefficient is less than 0.001 in Table 5, while the p-value in Table 8 is between 0.01 and 0.05.

## 5 Discussion

### 5.1 Smoking effect on Length of pregnancy

From Table 5 and Table 8, we can see that the length of pregnancy of a smoking mom is significantly lower than that of a non-smoking mom. Without matching, the effect of Smoking on the Length of pregnancy is lower than the result after matching. The shortening of the Length of pregnancy will make the fetus have a greater chance of premature birth. The harm caused by premature birth is not only the weight loss of the baby, but the bigger problem is that premature birth may lead to the baby's birth in an immature condition, which will bring many side effects, such as organs not fully functioning developed. Preterm birth may also lead to higher mortality rates, which is not conducive to population growth.

As seen from the distribution map of Length of pregnancy, a small fraction of smoking moms have a Length of pregnancy is less than 30 weeks. For babies born before 30 weeks, even if the baby survives, additional care will be required, adding to the social expenditure. And premature babies may also be accompanied by some diseases and the baby's future development. There is evidence that Smoking reduces pregnancy cycles, so we should urge mothers to minimize Smoking during pregnancy.

### 5.2 Smoking effect on birth weight

The results in Table 5 and Table 8 show that smoking will bring about weight loss in the newborn, and there is a higher probability that the baby will be born with a risk of low birth weight. When considering infant birth weight, we found that in addition to smoking affecting infant weight, infant gender and mother race also had an impact on infant birth weight. The results show that babies of white moms had higher birth weights than other non-white moms and that male babies had significantly higher birth weights than female

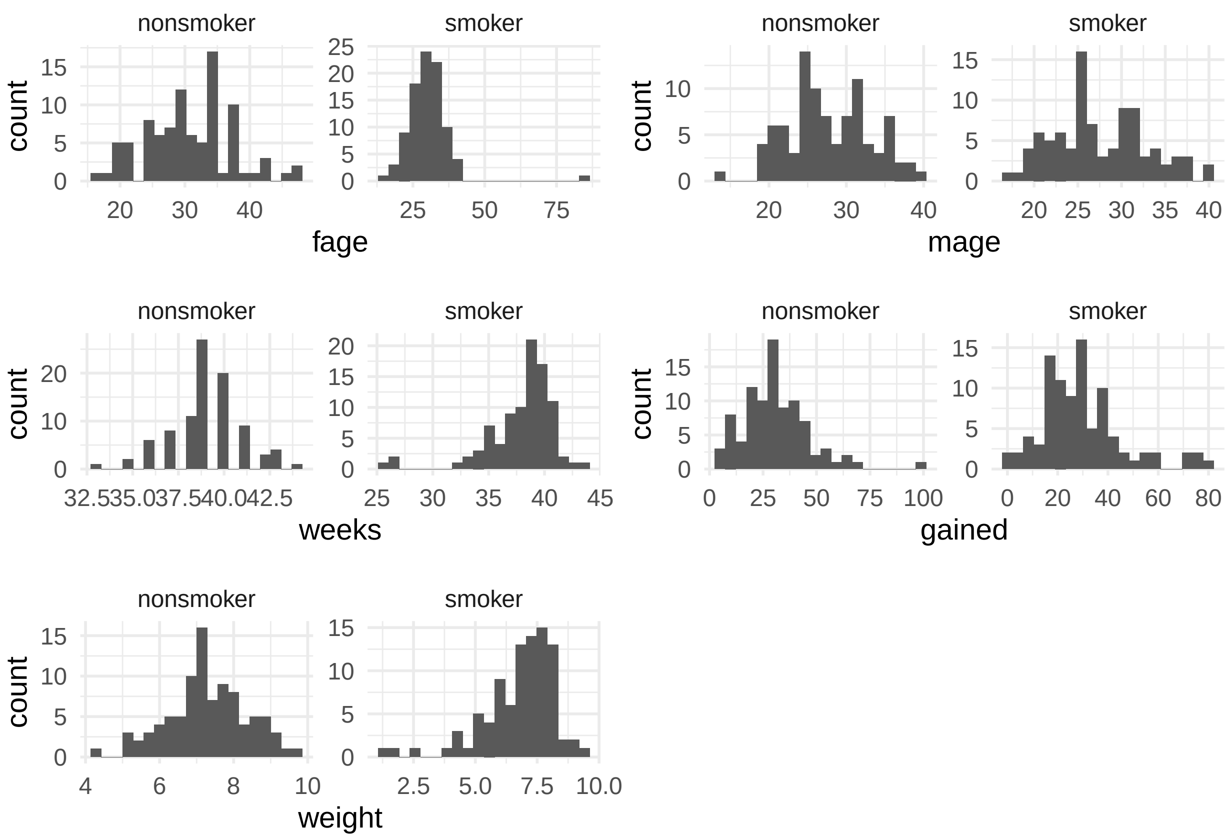



Figure 2: Distributions of numerical variables after matching

babies. Therefore, we can conclude that the newborn's weight will be related to the baby's genes. Caucasians are taller and heavier than Asians, so Caucasian babies also reflect this difference at birth weight.

Although the data can explain about 10% of birth weight variation after matching, the R square is still tiny. This also means that we may have overlooked some essential factors when estimating the variable of infant birth weight, limiting the model's explanatory power for birth weight. In addition, the MLR and logistic regression we used in the analysis process both need to satisfy assumptions for the results to be credible. However, we did not conduct research on whether the above assumptions are met, which may be why our results are not significant and have low explanatory power. In the last column of Table 5 and Table 8, the logistic regression method is used to study the effect of smoking on low birth weight. The results show that smoking increases the risk of low birth weight. So the result of column 2 and column 4 is consistent.

## 5.3 Smoking effect on mother's weight gain during pregnant

The final output of my study is to analyze the effect of smoking on a mother's weight gain during pregnancy. Mothers' weight gain reflects the baby's weight and the mother's health. Gaining too much weight during pregnancy can lead to maternal obesity, leading to side effects such as high blood pressure and diabetes. Therefore, weight control is something that pregnant mothers should pay special attention to. Our results showed that smoking did not affect maternal weight.

We, therefore, concluded that smoking during pregnancy only had adverse effects on the baby, not on the mother. Because of the small impact on mothers, many people ignore the harm caused by smoking. And this role in the harm of the newborn may affect the baby's development. We should call on people, especially mothers, to cut back on smoking and other unhealthy activities during pregnancy.

## 5.4 Weaknesses and next steps

We use the probability score matching method to match the variables of smoking and non-smoking groups and then compare the results of OLS regression before and after matching. Although we can find that smoking will adversely affect the baby's health, the magnitude of this effect is difficult to obtain through the above regression. Because for the observation study, we can only get the association relationship between the variables, but not the causal effect between them. First, the sampling method may produce selection bias in the observation study. For example, in the data studied in this paper, The proportion of white moms is much more significant than that of non-white moms. This paper explores the effect of smoking on infant weight, white moms, and other races. Therefore, there is obvious selection bias in this paper, so the results of our research may not reflect reality.

Secondly, we use MLR and the logistics model to study the relationship between variables. However, the above two methods require data to satisfy strong assumptions, and the observation dataset is challenging to meet these assumptions simultaneously, limiting the application of the above model. Whether the result of the above model is valid still needs to be diagnosed. Even if the above model assumptions are valid, we will still miss relevant variables, resulting in biased estimation results or low accuracy.

To overcome the above drawbacks, we can use, for example, RCT methods for research. However, for smoking or not smoking, RCT studies are not suitable to study the effect of smoking, considering that the experiment cannot be enforced on people. Another method is using the natural experiment method of difference in difference or instrumental variable estimation. Although these methods cannot obtain the causal effect of RCT, they can help us get more accurate conclusions. In addition, since the proportion of white moms in the sample is too large, we should expand the sampling scope and randomly collect people of different races for research.

## References


"A.d.a.m. Medical Encyclopedia [Internet]. Johns Creek (GA)." 2020.%20https://medlineplus.gov/ency/article/003247.htm.

Anselin, Luc. 2003. "An Introduction to Spatial Regression Analysis in r." *University of Illinois, Urbana-Champaign* 25.

Arel-Bundock, Vincent. 2022. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready.*

Auguie, Baptiste. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics.*

Çetinkaya-Rundel, Mine, David Diez, Andrew Bray, Albert Y. Kim, Ben Baumer, Chester Ismay, Nick Paterno, and Christopher Barr. 2021. *Openintro: Data Sets and Supplemental Functions from 'OpenIntro' Textbooks and Labs.*

Demin, Gregory. 2022. *Expss: Tables, Labels and Some Useful Functions from Spreadsheets and 'SPSS' Statistics.*

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2011. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software* 42 (8): 1–28.

Smith, George Davey. 2003. *Effect of Passive Smoking on Health. Bmj.* Vol. 326. 7398.

Stick, Stephen M, Paul R Burton, L Gurrin, PD Sly, and PN LeSouef. 1996. *Effects of Maternal Smoking During Pregnancy and a Family History of Asthma on Respiratory Function in Newborn Infants. The Lancet.* Vol. 348. 9034. Elsevier.

Wang, Xiaobin, Barry Zuckerman, Colleen Pearson, Gary Kaufman, Changzhong Chen, Guoying Wang, Tianhua Niu, Paul H Wise, Howard Bauchner, and Xiping Xu. 2002. "Maternal Cigarette Smoking, Metabolic Gene Polymorphism, and Infant Birth Weight." *Jama* 287 (2): 195–202.

Table 5: effect of smoking before mathching

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| (Intercept) | 38.349 *** | 6.784 *** | 34.372 *** | -3.002 * |
| | (0.880) | (0.440) | (5.544) | (1.418) |
| smoker | -0.738 ** | -0.556 *** | -1.893 | 1.318 *** |
| | (0.271) | (0.136) | (1.709) | (0.331) |
| fage | -0.011 | -0.003 | 0.020 | 0.022 |
| | (0.017) | (0.008) | (0.105) | (0.023) |
| mage | 0.013 | 0.010 | -0.202 | -0.000 |
| | (0.026) | (0.013) | (0.162) | (0.040) |
| matureyounger mom | 0.166 | -0.200 | -0.016 | 0.190 |
| | (0.304) | (0.152) | (1.912) | (0.490) |
| whitemomwhite | 0.475 * | 0.412 *** | 1.870 | -0.600 * |
| | (0.208) | (0.104) | (1.312) | (0.305) |
| maritalnot married | -0.018 | 0.009 | -0.344 | -0.160 |
| | (0.199) | (0.099) | (1.250) | (0.323) |
| sexmale | -0.179 | 0.382 *** | 0.877 | -0.297 |
| | (0.168) | (0.084) | (1.057) | (0.276) |
| N | 835 | 835 | 835 | 835 |
| R2 | 0.017 | 0.068 | 0.009 | |
| logLik | -1917.291 | -1338.483 | -3453.945 | -205.949 |
| AIC | 3852.582 | 2694.966 | 6925.890 | 427.898 |

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 6: matched data summary for smoker

| | | N | % |
|---|---|---|---|
| mature | mature mom | 10 | 10.9 |
| | younger mom | 82 | 89.1 |
| | smoker | 92 | 100.0 |
| lowbirthweight | low | 16 | 17.4 |
| | not low | 76 | 82.6 |
| sex | female | 42 | 45.7 |
| | male | 50 | 54.3 |
| marital | married | 42 | 45.7 |
| | not married | 50 | 54.3 |
| whitemom | not white | 14 | 15.2 |
| | white | 78 | 84.8 |

Table 7: matched data summary for nonsmoker

| | | N | % |
|---|---|---|---|
| mature | mature mom | 12 | 13.0 |
| | younger mom | 80 | 87.0 |
| | nonsmoker | 92 | 100.0 |
| lowbirthweight | low | 5 | 5.4 |
| | not low | 87 | 94.6 |
| sex | female | 45 | 48.9 |
| | male | 47 | 51.1 |
| marital | married | 44 | 47.8 |
| | not married | 48 | 52.2 |
| whitemom | not white | 12 | 13.0 |
| | white | 80 | 87.0 |

Table 8: Effect of smoking before mathching after Propensity score estiamtion

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| (Intercept) | 38.519 *** | 7.596 *** | 39.488 ** | -5.523 * |
| | (1.940) | (0.951) | (12.025) | (2.513) |
| smoker | -0.988 ** | -0.492 ** | -1.721 | 1.295 * |
| | (0.379) | (0.186) | (2.350) | (0.545) |
| fage | -0.050 | -0.010 | 0.129 | 0.007 |
| | (0.032) | (0.015) | (0.196) | (0.032) |
| mage | 0.036 | -0.002 | -0.308 | 0.088 |
| | (0.053) | (0.026) | (0.331) | (0.067) |
| matureyounger mom | 0.239 | -0.632 | -3.364 | 0.915 |
| | (0.759) | (0.372) | (4.705) | (0.996) |
| whitemomwhite | 1.363 * | 0.564 * | -0.272 | -0.845 |
| | (0.549) | (0.269) | (3.402) | (0.607) |
| maritalnot married | -0.397 | -0.182 | -0.927 | 0.417 |
| | (0.424) | (0.208) | (2.627) | (0.527) |
| sexmale | -0.187 | 0.377 * | -0.190 | -0.779 |
| | (0.388) | (0.190) | (2.404) | (0.513) |
| N | 184 | 184 | 184 | 184 |
| R2 | 0.089 | 0.110 | 0.010 | |
| logLik | -430.452 | -299.283 | -766.158 | -58.491 |
| AIC | 878.904 | 616.567 | 1550.315 | 132.982 |

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.