# The Effects of Pre-Pregnancy BMI on the Relationship Between Maternal Smoking and Low Birth Weight

## BSTA 512-612: Winter 2021, OHSU-PSU School of Public Health

Matthew Hoctor, Bradley Hopkins

2/10/2021

## Contents

---

## Summary

## 1. Introduction

The relationship between maternal prenatal smoking and low birth weight (LBW) is a topic that has long been of interest to investigators, due to the association of LBW with poor health outcomes (Blencowe et al., 2019). An infant is considered to be LBW when they are born weighing less than 2500 grams, or less than about 5.5 lbs (WHO, 1993). LBW has been associated with many adverse effects, such as a higher mortality rate (Martin et al., 2013); various psychiatric and neurological problems (Hack et al., 2005); and even poor academic achievement (Hack et al., 2002). LBW has also been implicated as an economic impact, with Kowlessar et al. (2011) reporting an increased cost of hospital stays at $27,200 for LBW infants compared to $3,200 for all newborns. Maternal smoking during pregnancy has been of intense interest for decades because it has been implicated in LBW (Zhang & Yang, 2019; Parascandola, 2014; Yerushalmy, 1971).

However, although the association of maternal smoking with LBW is well-documented and consistent (Ventura et al., 2003), the mechanism by which it contributes to LBW and how this mechanism might be affected by other factors are not well understood. For example, the risk of LBW from prenatal smoking has been demonstrated to be significantly reduced among overweight and obese mothers (La Merrill et al., 2011). But it is unclear how these mechanisms interact with one another, as maternal pre-pregnancy BMI is itself implicated as a predictor for birth-weight (Gul et al., 2020). This relationship is further complicated by the fact that one of the primary outcomes that makes LBW a topic of interest for researchers, perinatal death, is not always associated with LBW in the way that it might seem that it should — it was demonstrated by Yerushalmy (1971) that among LBW infants, maternal smoking was actually associated with a decreased risk of perinatal death. Shah(2014) proposed that these seemingly contradictory findings, often referred to as the "Birth weight Paradox" (VanderWeele, 2014). This would reasonably explain Yerushalmy's finding if the association of this phenotype with maternal smoking were strong enough such that LBW infants with smoking mothers were much more likely to have this phenotype compared to others.

The mechanisms that might contribute to these "other" phenotypes of LBW are not entirely clear, although existing work does point toward some reasonable candidates to start with. Vanderweele (2014) proposes that the increase in mortality for LBW infants with non-smoking mothers compared to smoking mothers may be that there is a much more insidious root cause for LBW in non-smoking mothers, such as malnutrition. There is some support for malnutrition as a potential cause for LBW, as it has been demonstrated there is a higher risk of pregnancies resulting in LBW in situations of inadequate social support, such as in low-income households (Gould, Davey, & Leroy, 1989) and in teenage pregnancies (Fraser et al., 1995). As noted above, low pre-pregnancy BMI has been noted to be associated with LBW (Gul et al., 2020), as well. These findings suggest that evaluating the effects that BMI has on the relationship between maternal smoking and LBW may help to clarify potential directions for future work related to the mechanisms underlying LBW and infant mortality.

## 2. Study Background

The sample used for this analysis was taken from the Child Health and Development Studies (CHDS), a data set used in Yerushalmy's (1971) paper. Below is Yerushalmy's description of this data set (p. 445):

> The information is derived from our Child Health and Development Studies (CHDS)—a comprehensive investigation of all pregnancies that occurred between 1960-1967, among women in the Kaiser Foundation Health Plan in the San Francisco-East Bay area. The Kaiser Health Plan is a prepaid medical care program.

> The members represent a broadly based group which is not atypical of an employed population. It is deficient only in the two extremes: the very affluent and the very indigent portions of the population. The women were interviewed on a variety of medical, genetic, and environmental subjects, including behavior variables, such as smoking, drinking, use of contraceptive methods and the like. The interview took place early in pregnancy. The information was thus derived prospectively before the woman knew the outcome of the pregnancy. The child was followed to evaluate his physical and mental development including survival and the development of congenital anomalies.

> The interviewed group comprised some 15,000 pregnancies. This study is based on single live born infants among the whites and the blacks. The numbers of members in the other ethnic groups in the sample were too small and were therefore left out. The study is based on 9,793 pregnancies among the whites and 3,290 among the blacks

A key point to note from the description above is that the majority of pregnancies from the data set involve middle-class women who have established healthcare. This will be an important limitation going forward, because low social status has been shown to be associated with both LBW (Gould, Davey, & Leroy, 1989) and, to some extent, with BMI (Basto-Abreu et al., 2017). There is the possibility that there are aspects of the relationship between maternal smoking and LBW that would not be apparent from this sample simply

because it has poor representation of women with low socio-economic status or women who are not established with a regular medical provider. With that being said, however, this analysis is still expected to provide some value in the sense that it may provide insight into potential avenues that might be taken. Potentially, this analysis could reveal important attributes of the relationship between maternal smoking and factors such as pre-pregnancy BMI, maternal height, or maternal age, which could then be evaluated further in future work.

Due to limitations of scope, this analysis was performed with a reduced sample of 680 white women from the source data set. Many of the variables described by Yerushalmy in the quote above, such as alcohol use or contraceptive methods, were not examined. Instead, the focus of this analysis is on 6 variables from the 1971 Yerushalmy study: birth-weight as the response variable; and gestational age, maternal age, mean number of cigarettes smoked per day by the mother, maternal height, and maternal pre-pregnancy weight as the predictors. Additionally, pre-pregnancy BMI was derived from maternal height and pre-pregnancy weight and examined as well. Each of these variables is discussed briefly below:

**Birth weight** This is the the weight of the infant at the time of birth, in pounds. This is the response variable for this analysis and, as noted previously, this is of interest because LBW has been associated with poor health outcomes, such as infant mortality (Martin et al., 2013).

**Mean number of cigarettes smoked per day** is split into four categories, for levels of smoking: light ($<$ 10 cigarettes per day), moderate (10 - 20 cigarettes per day), and heavy ($>$ 20 cigarettes per day) smokers. This is the primary predictor of interest in this analysis, since, as described above, it has been demonstrated in multiple studies in the past that maternal smoking is associated with LBW (Zhang and Yang, 2019; Parascandola, 2014; Yerushalmy, 1971). These category thresholds were chosen somewhat arbitrarily, although it should be noted that consistency in definitions for smoking severity has been poor. For example, Husten (2009) reports that measures such as thresholds of $<4$, $<10$, $<15$, and $<20$ have all been used to classify "light" smoking. Husten (2009) further points out that cigarettes smoked per day may not necessarily be the best proxy for toxin exposure levels, due to reasons such as the differences in toxin levels in different tobacco products and the nonlinear relationship of cigarette consumption with disease risk in some cases, among others. For these reasons, levels of smoking severity were chosen out of convenience in providing separation of smoking groups, with the intention of potentially identifying general trends and without interpreting too heavily in what significant differences between these levels might mean. It should be noted that the effects of other potential exposures, such as paternal smoking during the pregnancy, were not assessed in this analysis.

**Gestational age** is a continuous variable, measured in weeks. This is a secondary predictor variable for this analysis and is of interest in the study because, as noted by Engle (2006), birth weight is often associated with gestational age but can differ in clinically-significant ways. This suggests that controlling for gestational age may reveal more about the biological mechanisms underlying the relationships of birth weight with the other variables in this study than if birth weight were examined without gestational age.

**Maternal age** is is a continuous variable, measured in years. This is also a secondary predictor of interest in this analysis, although the relationship that maternal age has with birth weight is more complex. Dennis and Mollborn (2014) discuss how pregnancies at both ends of the spectrum for maternal age are associated with an increased risk for LBW. However, they also report that there are a variety of other non-biological factors involved and that "LBW disparities by maternal age are a complex product of socioeconomic disadvantage and current social and behavioral factors, such that LBW risk does not operate uniformly by race/ethnicity or maternal age." (Mollborn & Dennis, 2014). Similarly, Wang et al. (2020) found that the relationship between maternal age and birthweight is non-linear and that the risk of LBW increased both with very young and very old maternal age. For these reasons, including maternal age in the analysis can serve to verify past results and also provide another means for observing how relationships with the response variable change when other predictors are present.

**Maternal height** is a continuous variable, measured in inches. This is a secondary variable of interest for this analysis because maternal height has been shown to have a direct association with LBW, with increasing risk for LBW as height decreases, even in cases where socio-economic status is thought not to be a factor (Inoue et al., 2016). This would suggest that there could be a biological component in LBW independent of malnutrition or social status. Including this variable in the study is useful because it serves to motivate the biological argument for LBW not explained by maternal smoking.

**Maternal pre-pregnancy weight** is a continuous variable, measured in pounds. Similar to maternal height, this is a secondary predictor of interest in this analysis due to the biological implications it might have about the mechanism for LBW. Also like maternal height, pre-pregnancy weight has been shown to have a direct association with LBW, with increasing risk of LBW as pre-pregnancy weight decreases (Zhang & Yang, 2019). Pre-pregnancy weight is useful to include in the analysis because it has been postulated that the association between pre-pregnancy weight and LBW can be explained by poor nutrition or malnutrition (Inoue et al., 2016), suggesting that pre-pregnancy weight might be a good proxy for the mother's pre-pregnancy nutritional status.

**Maternal pre-pregnancy BMI** is a continuous variable, derived from maternal height and pre-pregnancy maternal weight data from the study. This is another secondary predictor that is of interest for this analysis because it provides for a slightly different way to examine the relationship between maternal height and pre-pregnancy weight, since BMI is a measure of weight relative to height.This may make pre-pregnancy BMI a better proxy for nutritional status than pre-pregnancy weight alone.

There is a strong likelihood that the small size and relative uniformity of the sample used for this analysis will result in issues of generalization, particularly when the original data set the sample was itself fairly homogeneous. The purpose of this analysis for this reason is not to make generalizable conlusions about the population but instead to identify general trends towards significance that may inform future work.

# 3. Statistical Analysis

This analysis has two primary aims: (1) to evaluate the effects of BMI on the relationship between maternal smoking and birth weight; and (2) to develop a model to predict birth weight based on factors such as smoking status, maternal BMI, and age. The primary hypothesis that will be evaluated during this analysis is that pre-pregnancy BMI is a confounder for the relationship between birthweight and smoking status. A major reason BMI is of interest in this analysis is due to its usefulness as a diagnostic tool for evaluating malnutrition, which can be otherwise difficult to assess (Cederholm et al., 2015). If a high BMI were to significantly mitigate the risk of LBW from maternal smoking compared to low BMI, this may highlight a potential biological mechanism for LBW and infant mortality that could be evaluated in future work. The motivation behind the second aim in this analysis would be to provide validation for existing methods of predicting LBW based on maternal characteristics, such as maternal pre-pregnancy BMI (Gul et al., 2020), maternal height (Inoue et al., 2016), and smoking status (Knopik et al., 2016).

The analysis was accomplished in three major steps:

- A preliminary exploratory analysis was performed, in order to identify any basic relationships between the variables in the sample.
- An association model was built using birth weight as the primary response variable and cigarettes smoking as the primary predictor. Multiple relationships were examined during this step, and specifically the role of BMI as a confounder for the relationship between maternal smoking and birth weight was examined.
- A prediction model was then build, again with birth weight as the primary response variable and maternal smoking as the primary predictor.

## 3.1 Exploratory Analysis

For this section of the analysis, the primary goal was to gain a basic understanding of the characteristics of the variables used in this analysis. A summary of descriptive statistics is presented in Table 1. A brief analysis of each of the variables shown in Table 1 is described below:

**Birth weight** is a continuous ratio variable that is measured in lbs. From Figure B.2, it can be seen that the distribution of Birth weight is very symmetric and approximately normal as the response variable. There

Table 1: Descriptive statistics from sample of 680 pregnancies

| Variable | Mean | Standard deviation | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|
| Birth weight (lbs) | 7.52 | 1.09 | 3.30 | 6.80 | 7.60 | 8.20 | 11.4 |
| Gestation age (weeks) | 39.77 | 1.88 | 29.00 | 39.00 | 40.00 | 41.00 | 48.0 |
| Maternal age (years) | 25.86 | 5.46 | 15.00 | 21.00 | 25.00 | 29.00 | 42.0 |
| Number of cigarettes smoked daily | 7.43 | 11.27 | 0.00 | 0.00 | 0.00 | 12.00 | 50.0 |
| Maternal height (inches) | 64.43 | 2.48 | 57.00 | 63.00 | 64.00 | 66.00 | 71.0 |
| Maternal pre-pregnancy weight (lbs) | 126.90 | 17.88 | 85.00 | 115.00 | 125.00 | 135.00 | 246.0 |
| Maternal pre-pregnancy BMI | 21.47 | 2.63 | 15.55 | 19.74 | 21.12 | 22.59 | 39.7 |

would be no indication for transformation of this distribution, since the shape is already very reasonably normal and adding transformations to the analysis would therefore be of questionable benefit.

**Gestation age** is a continuous ratio variable that is measured in weeks. From Figure B.2, it can be seen that this variable is also approximately normal and fairly symmetric.

**Maternal age** is a continuous ratio variable that is measured in years. Figure B.2 shows that this variable is skewed to the right.

**Cigarettes smoked per day** is a continuous ratio variable (since it is an average). From Figure B.2, it can be seen that the distribution of this variable is very asymmetric and there appear to be multiple regions around which the data points tend to be concentrated. The most salient of these is around 0 cigarettes per day, which represents non-smokers. This was one of the primary reasons that cigarette-smoking was turned into a categorical variable, as there would be good reason to believe that there are significant differences between the very large non-smoker group and all of the other data points in the sample, particularly when considering the response variable, birth weight. As noted in the previous section, the thresholds chosen for this variable were 0-10, 10-20, and >20 cigarettes per day. These thresholds were chosen partially out of convenience, although Husten (2009) notes that both <10 and <20 cigarettes per day have been used to designate "light" smoking by researchers in the past (it may be helpful to think of the 0-10 and 10-20 thresholds as "light" and "medium-light" for this reason).

**Maternal height** is a continuous ratio variable, measured in inches. Figure B.2 shows that this variable is reasonably symmetric and approximately normal.

**Maternal pre-pregnancy weight** is a continuous ratio variable, measured in lbs. Figure B.2 shows that this variable is skewed to the right.

Figure B.1 shows a scatter-plot matrix for each of the variables described in this analysis. This representation shows no obvious departure from linearity for any of the variables in this sample. Some of the variables appear to lack any obvious association at all, such as between BMI and age. Other pairs do seem to more clearly show a possible association, such as birth weight vs gestation age or pre-pregnancy weight and BMI. However, neither of these findings are surprising, as it has been noted by authors such as Engle (2006) that birth weight and gestational age are associated with one another and BMI was calculated directly from pre-pregnancy weight. These findings suggest that if any associations are found in this analysis, they are likely to be weak.

The correlation matrix from Output A.1 shows similar findings: there are a few pairs, such as birth weight vs gestation age (0.4212) and pre-pregnancy weight vs maternal height (0.9925), that are highly correlated with one another, but otherwise the correlations are much weaker. Other notable associations outside of these more clear associations would be birth weight vs mean cigarettes per day (-0.1707) followed by birth weight

vs pre-pregnancy weight (0.0340), although these associations are much weaker.

## 3.2 Method

A simple linear regression analysis was performed to establish the association between birth weight as the primary response variable and cigarette use as the primary predictor. This is appropriate because the birth weight is a quantitative variable. Simple regression analysis was used to develop an association model and a prediction model with birth weight as the response variables and the remaining variables in this sample as the predictors, namely: gestation age, maternal age, average cigarettes smoked per day, maternal height, and maternal pre-pregnancy weight. Additionally, a new variable BMI was derived from existing maternal height and pre-pregnancy weight data and was included in the analysis, as well. The association model was developed using the "backward-elimination" method for efficiency purposes. The prediction model was developed using the parsimonious framework.

The advantages of this strategy is that simple regression analysis is relatively easy to perform and is not hardware intensive. The findings from this analysis should also be easier to communicate and easier for the reader to interpret, compared to some methods. One potential limitation of this method would be that if the relationships between the variables in this sample are truly non-linear, this would diminish the value of the models produced by this analysis greatly.

## 3.3 Association Model

## 3.4 Prediction Model

# 4. Discussion

# Appendix

## Appendix A: R Output

### Output A.1: Correlation Matrix

```
##             bwt gestwks     age  mnocig mheight    mppwt      BMI
## bwt      1.0000  0.4212 -0.0132 -0.1707 -0.0070   0.0340 -0.0235
## gestwks  0.4212  1.0000  0.0074  0.0096  0.0407  -0.0467  0.0444
## age     -0.0132  0.0074  1.0000  0.0476 -0.0035   0.0068  0.0029
## mnocig  -0.1707  0.0096  0.0476  1.0000 -0.0240   0.0318 -0.0337
## mheight -0.0070  0.0407 -0.0035 -0.0240  1.0000   0.9925 -0.9904
## mppwt    0.0340 -0.0467  0.0068  0.0318  0.9925   1.0000  0.9970
## BMI     -0.0235  0.0444  0.0029 -0.0337 -0.9904   0.9970  1.0000
```

## Appendix B: Plots and Figures
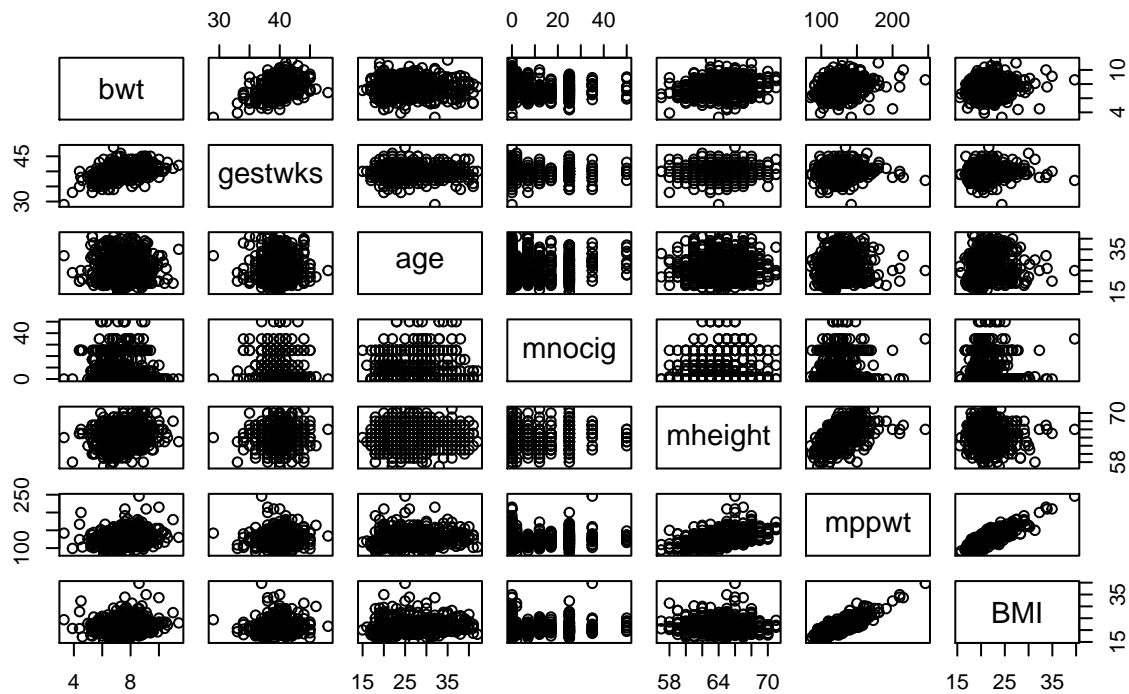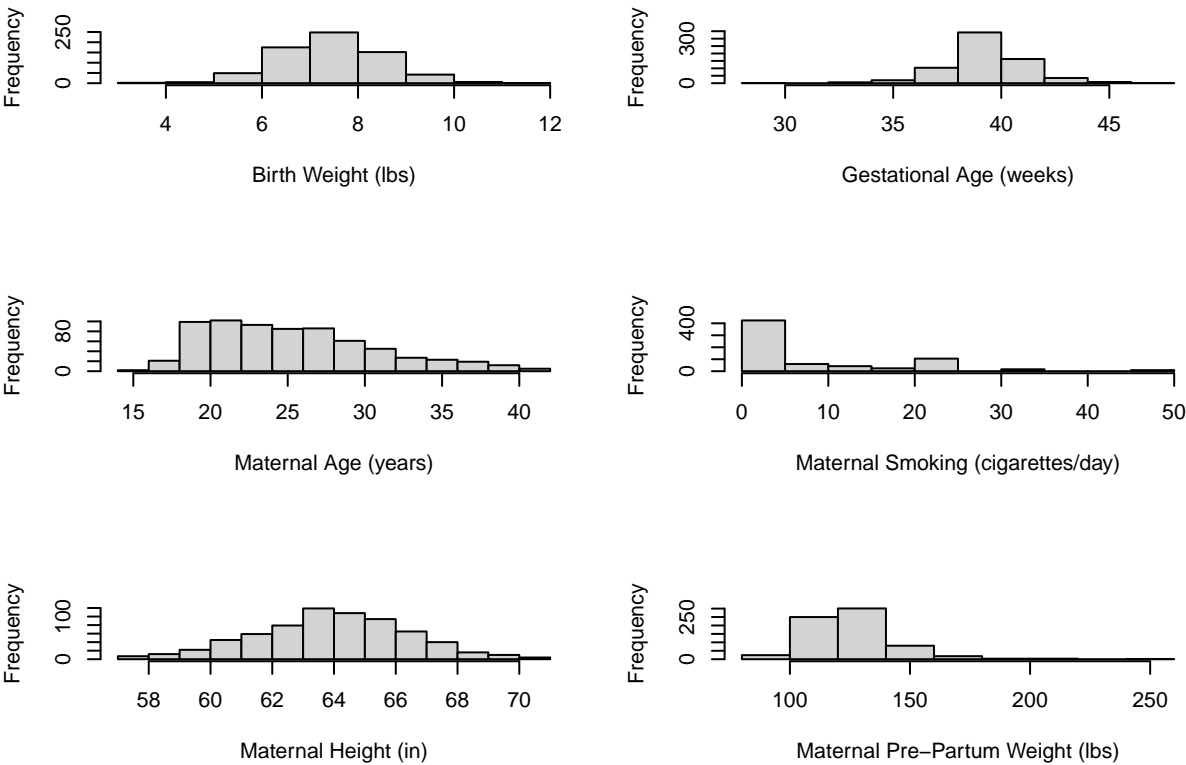
### Figure B.1 Scatterplot Matrix

**Figure B.2 Distributions of Sample Variables**



## Appendix C: Model Diagnostics

# References