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

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

Matthew Hoctor, Bradley Hopkins

2/21/2021

Contents

Summary

The relationship between birth weight and smoking during pregnancy (SDP) has long been of interest to public health investigators due to the association of low birth weight (LBW) with adverse health outcomes, such as infant mortality in the first year of life. However, it has remained unclear how secondary factors, such as pre-pregnancy BMI, might affect the underlying mechanisms of this relationship. This analysis demonstrates that both pre-pregnancy BMI and gestational age can have a confounding and/or mediating effect on the relationship between birth weight and SDP. These findings suggest that the mechanisms underlying LBW are complex and multi-factorial, warranting further investigation.

1. Introduction

The relationship between maternal prenatal smoking and 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 Global Nutrition Targets 2025: Low Birth Weight Policy Brief" n.d.). LBW has been associated with many adverse effects, such as a higher mortality rate (Martin et al. 2015); 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, Jiang, and Steiner (2006) reporting an increased cost of hospital stays at $27,200 for LBW infants compared to $3,200 for all newborns. SDP during pregnancy has been of intense interest for decades because it has been implicated in LBW (Zhang and Yang 2019; Parascandola 2014; J. Yerushalmy 2014).

However, although the association of SDP 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 J. Yerushalmy (2014) that among LBW infants, SDP was actually associated with a decreased risk of perinatal death. Ebrahim (2014) proposed that these seemingly contradictory findings, often referred to as the "Birth weight Paradox" (Tyler J. VanderWeele 2014), could be explained by differences in LBW phenotypes. This would reasonably explain Yerushalmy's findings if the association of this phenotype with SDP were strong enough such that LBW infants with smoking mothers were much more likely to have this phenotype compared to others and that the phenotype associated with SDP was somehow more resistant to outcomes such as perinatal mortality.

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. Tyler J. 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, and LeRoy 1989) and in teenage pregnancies (Fraser, Brockert, and Ward 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 SDP 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 by J. Yerushalmy (2014). Below is Yerushalmy's description of this data set:

> 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, and LeRoy 1989) and, to some extent, with BMI (Basto-Abreu et al. 2018). There is the possibility that there are aspects of the relationship between SDP 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 SDP 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 J. Yerushalmy (2014) 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. Each of these variables is discussed briefly below:

Birth weight is the the weight of the infant at the time of SDP 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. 2015).

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 and, as described above, has been demonstrated in multiple studies in the past to be associated with LBW (Zhang and Yang 2019; Parascandola 2014; J. Yerushalmy 2014). 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 thresholds of <4, <10, <15, and <20 cigarettes per day 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. 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 the 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 (2013) 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." (Dennis and Mollborn 2013) 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 SDP.

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 and 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 accounts for both height and weight, it may make for 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 SDP 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 SDP 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 four major steps:

- A preliminary exploratory analysis was performed, in order to identify any basic relationships between the variables in the sample.

Table 1: Descriptive statistics from sample of 680 pregnancies from CHDS Study

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

- 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 SDP and birth weight was examined.
- A prediction model was then built, again with birth weight as the primary response variable and SDP as the primary predictor.
- Last, a mediation analysis was performed to determine whether any of the secondary predictors might have a mediating effect on the relationship between birth weight and SDP.

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 would be no indication for transformation of this distribution, since the shape is already very reasonably normal and adding transformations to the analysis would 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 noticeable of these is around 0 cigarettes per day, for non-smokers. This was one of the primary reasons that cigarette-smoking was turned into a categorical variable, as there is 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 pairs of variables appear to lack any obvious association at all, such as BMI vs age. Other pairs do seem to more clearly show a possible association, such as birth weight vs gestation age or pre-pregnancy weight vs BMI. However, neither of these findings are surprising, as BMI was calculated directly from pre-pregnancy weight and it has been noted by authors such as Engle (2006) that birth weight and gestational age are associated with one another. These findings suggest that if any other 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 most other correlations are much weaker. Other notable associations include birth weight vs mean cigarettes per day (-0.1707) and 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 method 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 by starting with a crude model of birth weight vs SDP and then checking each of the secondary predictors for effect modification and confounding. The prediction model was developed using a backward-elimination approach.

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.

BMI and categorical variables for SDP were derived according to the following equations:

$$\mathrm{BMI} = 703 * \mathrm{mppwt}/\mathrm{mheight}^2$$

$$SMK_1 = \begin{cases} 1 & \text{if } \; 0 < \text{mnocig} \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

$$SMK_2 = \begin{cases} 1 & \text{if } \; 10 < \text{mnocig} \leq 20 \\ 0 & \text{otherwise} \end{cases}$$

$$SMK_3 = \begin{cases} 1 & \text{if } \; \text{mnocig} > 20 \\ 0 & \text{otherwise} \end{cases}$$

where *mmpwt* is mean pre-pregnancy weight, *mheight* is maternal height, and *mnocig* is the mean number of cigarettes smoked per day.

3.3 Association Model

First, a crude association model was developed for the response variable *birth weight* and the primary predictor *mean number of cigarettes per day.* There was highly statistically significant evidence ($P < 0.001$, see Output A.2) that this model provides for a better fit of the data compared to the mean value for *birth weight* alone. These findings suggest that this crude model is appropriate for this sample.

Next, an analysis was performed for the remaining variables to determine if any qualify as effect modifiers or confounders. This process was conducted step-by-step with each variable evaluated against the crude model, shown below:

Gestational Age There was no significant evidence to support the inclusion of gestational age as an effect modifier for the relationship between birth weight and SDP ($p > 0.10$ for all three interaction terms; see Output A.3), but there was sufficient support to include gestational age as a confounder (>10% change in slope for all three smoking terms; see Output A.4).

Maternal Age There was no evidence supporting the inclusion of maternal age as an effect modifier for the relationship between birth weight and SDP ($p > 0.10$ for all three interaction terms; see Output A.5) and insufficient support to include it as a confounder (<10% change in slope for all smoking terms, see Output A.6).

Maternal Height There was no evidence to support the inclusion of maternal height as an effect modifier ($p > 0.10$ for all three interaction terms; see Output A.7) and no support for maternal height as a confounder in the model (<10% change in slope for all smoking terms; see Output A.8).

Maternal Pre-Pregnancy BMI There was no evidence to support the inclusion of maternal pre-pregnancy BMI as an effect modifier ($P > 0.10$ for all three interaction terms; see Output A.9), but there was support for the inclusion of BMI as a confounder for the relationship between SDP and birth weight (>10% change in one slope term for smoking with inclusion of BMI; see Output A.10).

Maternal Pre-Pregnancy Weight was not included at this stage of the model due to concerns about the high degree of colinearity it might introduce.

The full association model can be described by the following equation:

$$B\hat{W}T = -2.90 - 0.21 * \mathrm{SM\hat{K}1} - 0.61 * \mathrm{SM\hat{K}2} - 0.39 * \mathrm{SM\hat{K}3} + 0.25 * \mathrm{ges\hat{t}wks} + 0.04 * \mathrm{B\hat{M}I}$$

This model satisfies all of the assumptions for simple linear regression. Examination of model diagnostic Figures C.1 and C.2 reveals no overt change in variance with increasing $\hat{Y}$, suggesting that the homoscedasticity assumption is met. Furthermore, the residuals do not seem to be asymmetrically distributed, skewed, or multi-modal in any way, suggesting that the normality assumption is also met. Examination of Q-Q plots shown in Figures C.3, C.4, and C.5 also support the normality assumption, with the majority of points consistent with a straight line and only a handful of points at the extremes possibly signaling a deviation from normality (Shapiro-Wilk testing would not be appropriate in this situation because of the large sample size). The reference cell coding approach for creating dummy variables for smoking categories ensures that the linearity assumption is met. There is also no clear non-linear relationship between birth weight and any of the the variables used in the study (see Figure B.1).

Outliers An assessment of potential outliers was assessed using Cook's Distance, which is shown in Output C.6. None of the data points are implausible, with the exception of one with 48 weeks gestational age. According to Hoffman et al. (2008), a gestational age of 48 weeks is in the 99.9th percentile of births. A new data set with this outlier removed was created, with summary statistics shown in Output A.12. In comparison of this model with the original full model (Output A.11), it can be seen that the removal of this outlier only resulted in a slight change of slope coefficients.

Colinearity A variance inflation factor (VIF) assessment was performed to check for colinearity, shown in Output C.7. This revealed no factors over 10, suggest there is no significant colinearity present in the model.

Confidence Intervals Output A.13 shows confidence intervals for the full vs reduced association models produced by this analysis. The confidence intervals appear slightly more narrow for the full model compared to the reduced model.

From comparison of the crude model in Output A.2 and the adjusted full model from Output A.12, it can be seen that the adjusted R^2 has increased from 0.05432 for the crude model to 0.234 for the full model. This suggests that there is improvement in fit when includding pre-pregnancy BMI and gestational age as confounders in the full model compared to the crude model. These findings suggest that at least some of the association between SDP and birth weight is distorted by pre-pregancy BMI and gestational age.

This model was build by starting with a crude model and then separately assessing whether additional predictors should be considered as effect modifiers or confounders. The main advantage of this method was that it is simple and time-efficient. A potential drawback from this, however, is that it neglects many of the possible interactions and relationships between the different variables in the study that might not be apparent when only checked against the crude model. However, such an analysis would require extensive examination of interactions between variables and would likely be too time-consuming for the scope of this project.

3.4 Prediction Model

A prediction model was built using a backward-elimination procedure, starting with a complete model containing all variables from the study (except for maternal pre-pregnancy weight, again due to concerns about colinearity with BMI). This model was built over two iterations, shown in Outputs A.14 and A.15. Maternal age was removed during the first step, since it was the least significant term in the model and had a non-significant slope coefficient ($P > 0.10$). During the second step, no variables were removed since all slope coefficients were significant ($P < 0.10$). Although the coefficient for the first smoking term had the highest P-value during this step, it is still considered significant ($0.01 < P < 0.10$). Additionally, this variable will be left in the model due to its role as a primary predictor variable.

This procedure resulted the following model:

$$B\hat{W}T = -8.16 - 0.18 * \mathrm{SM\hat{K}1} - 0.55 * \mathrm{SM\hat{K}2} - 0.41 * \mathrm{SM\hat{K}3} + 0.24 * \mathrm{gest\hat{w}ks} + 0.04 * \mathrm{B\hat{M}I} + 0.08\mathrm{mhe\hat{i}ght}$$

By inspection, it can be seen that this procedure has produced almost the same model as that produced from the association model-building method, with the exception being that maternal height has been included in this model and was not added to the association model.

This model meets all of the basic assumptions of simple linear regression. Analysis of residuals in Figures C.8 and C.9 shows that there is no clear change in variance as the value of birth weight changes, suggesting that this model meets the assumption of homoscedasticity. There is also no clear departure from normality in either of these plots. VIF assessment does not suggest there is any significant degree of colinearity (output C.7). Confidence intervals for this model are shown in Output A.15 — the confidence intervals for this model appear to be about the same width as those for the association model. Again, since smoking status was split into dummy variables, the linearity assumption has also been met.

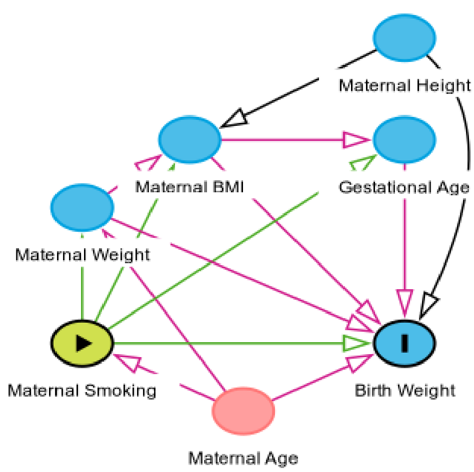
These findings suggest that a baby born after an average gestation time of 39.77 weeks to a non-smoking mother of average height (64.43") and BMI (21.47) is predicted to be $-8.16+39.77*0.24+0.04*21.47+0.08*64.43 = 7.36$ lbs. Gestation age has by far the most profound effect on this factor, with a coefficient of 0.24 lbs per week gestation. This is reasonable since it has been well established that premature babies have a higher risk of LBW. Smaller effects are seen from pre-pregnancy BMI and maternal height, which result in a change of 0.04

lbs per unit of pre-pregnancy BMI and 0.08 lbs per inch in the mother's height. When the mother smokes, this model predicts a loss of about 0.18 lbs birth weight if the mother is a light smoker (< 10 cigarettes per day), 0.55 lbs if the mother is a moderate smoker, and 0.41 lbs if the mother is a heavy smoker. Since the threshold for LBW is about 5.5 lbs, this model predicts that the mother would have to be below average in either BMI or height to have a baby that is below the LBW threshold, even when accounting for smoking.

The R^2 value for this prediction model is 0.2675 (Output A.15), which is very slightly higher than the R^2 of 0.234 produced by the association model. This suggests that this prediction model explains slightly more of the variance in the response variable *birth weight* compared to the association model. However, these differences are not large enough to be of particular concern that the association model would make for a significantly worse fit of the data compared to this model, and, since the association model is made up of one fewer variable, it may be reasonable to choose that model over this one simply for ease of interpretation. It may be that the effect of maternal height is just not significant enough that it would be noticeable as an effect modifier/confounder but was significant enough not to be removed from the backward elimination procedure used to create the prediction model. This is reasonable from a logical standpoint, since the association model already includes BMI, which does account for height to some degree. The slight increase in R^2 value in the prediction model compared to the association model could then be explained by the small degree of variance that is explained better by the direct relationship with height rather than what is already accounted for by BMI.

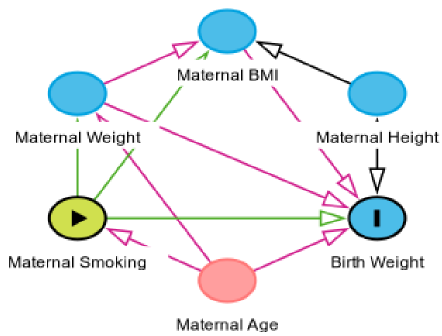
3.5 Mediation Analysis

Based on preliminary research, the possible relations between the variables in the dataset are depicted below:

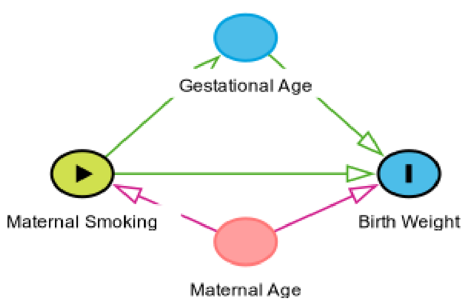


Regression of gestational age on BMI provides no significant evidence for an association ($p > 0.10$; Output D.1), suggesting that "intertwined" mediation can be ignored for this analysis (e.g. propensity matching) and that mediation analysis can be performed separately for each of the two potential mediators (gestational age and BMI). The following DAGs depict the hypothesized relations:

Mediation by BMI:



Mediation by Gestational Age:



Pre-Pregnancy BMI Linear regression of SDP on pre-pregnancy BMI and of birth weight on pre-pregnancy BMI shows significant relationships ($p < 0.01$ in all cases, see Outputs D.2 and D.3), suggesting the mediation of the smoking-gestational weight relation by BMI is plausible.

Simulations from Outputs D.4, D.5, and D.6 show evidence that ~23%, ~6%, and ~5% of the effect on birth weight from light smoking, moderate smoking, and heavy smoking, respectively, is mediated by pre-pregnancy BMI (slightly different values may be obtained upon subsequent simulations). Estimates for mediated effect and proportion mediated met significance thresholds ($P > 0.10$).

Gestational Age Output D.7 shows evidence that light and moderate smoking categories have a significant association with gestational age ($P < 0.10$) and Output D.8 shows evidence for a significant association between birth weight and gestational age. These findings suggest that the mediation of the smoking-gestational weight relation by BMI is plausible.

Simulations from D.9, D.10, and D.11 show evidence that ~34%, ~18%, and ~13% of the effect on birth weight from light smoking, moderate smoking, and heavy smoking, respectively, is mediated by gestational age (slightly different values may be obtained upon subsequent simulations). Estimates for mediated effect size met the significance threshold ($P < 0.10$) for light and moderate smokers, but not for heavy smokers. Estimates for proportion mediated met significance thresholds ($P < 0.10$) for moderate smokers only.

These findings provide evidence that both BMI and gestational age are independent mediators for the relationship between SDP and birth weight.

4. Discussion

The purpose of this analysis was to build association models and prediction models for the relationship of birth weight with SDP, as well as other predictor variables. It was found that there was a significant linear relationship between birth weight and SDP, at all three designated smoking severity levels. It was also found that pre-pregnancy BMI and gestational age could both qualify as confounders for the relationship between SDP and birth weight. The prediction model for birth weight contained one more variable than the association model, maternal height, suggesting that maternal height does have predictive power for birth weight, although there was not evidence to support maternal height as either an effect modifier or a confounder for this relationship. Mediation analysis provided support for both pre-pregnancy BMI and gestational age as mediators for the relationship between SDP and birth weight, suggesting that these variables could partially explain the relationship between birth weight and SDP.

These findings provide support for the hypothesis that BMI is a confounder (and possibly a mediator) for the relationship between SDP and birthweight. These findings might mean that there is some facet of the biological mechanism for low birth weight to which smoking mothers are more vulnerable to and mothers with high BMI are more resistant to.

This analysis had some fairly substantial limitations for evaluation of this hypothesis, as discussed earlier. Previous work has discussed how malnutrition may be a significant component in LBW and that the relationship that low pre-pregnancy BMI has with LBW is because BMI can serve as a reasonable proxy for nutritional status in situations of inadequate social support. However, the sample used for this analysis contained primarily middle-class white women who all had insurance and regular healthcare. This sample would have very poor representation of individuals with poor nutrition that would be of interest in testing the hypothesis that malnutrition has a significant effect on birth weight.

Another potential issue with the sample data used in this analysis, likely due to the age of the study it was taken from, is that there is a large amount of medical data that was missing from this sample that might have been useful for evaluating the relationships in the study. For example, it may have been helpful to know information about the mother's health status, such as whether she had chronic diseases like diabetes or heart disease. This information could have helped to broaden the picture and inform whether the relationships seen in this study were extraneous assocations only seen because of how little was known about the sample.

There were also limitations in the procedure used for this analysis, as well. A limited method of association-model building was performed where only the crude model was used to check each variable for effect modification and confounding. Although this method still produced statistically significant findings, it is possible that a more rigorous method for model-building could have produced more appropriate results and possibly helped to identify relationships that were not obvious from this relatively basic approach. Similarly, although a conventional backwards-elimination procedure was performed for the prediction model, the model was not verified against test data. Performing more rigorous model-building procedures in future work would likely provide some improvement for the conclusions made in this analysis.

For these reasons, this analysis should be interpreted with caution and is meant to only identify trends that might be of interest for future studies rather than be used as conclusive evidence for the relationships studied here. The value in this analysis is that it has identified some trends that may be apparent in more generalized populations and could be useful in future work. Conversely, however, there may be general trends present in the population that were not apparent in this sample but may have become apparent if a more representative sample had been used. Future work could examine whether these relationships are changed when mothers in situations of poor social support are included in the study, controlling for BMI. This may help to differentiate between physical stature as a cause for LBW vs malnutrition. Overall, these findings are promising in the sense that they show that it might be possible to understand LBW and infant mortality better, as more information about the factors surrounding these outcomes becomes available.

Appendices

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
```

Output A.2: Crude Model (Birth Weight and SDP)

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3, data = CHDS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4328 -0.7328 -0.0328  0.7214  3.6672
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.73281    0.05442 142.091  < 2e-16 ***
## SMK1        -0.35223    0.11797  -2.986  0.00293 **
## SMK2        -0.76008    0.14163  -5.367 1.10e-07 ***
## SMK3        -0.46665    0.10790  -4.325 1.76e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.062 on 676 degrees of freedom
## Multiple R-squared:  0.0585, Adjusted R-squared:  0.05432
## F-statistic:    14 on 3 and 676 DF,  p-value: 7.294e-09
```

Output A.3: Gestational Age as an Effect Modifier for Birth Weight and SDP

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + gestwks + gestwks * SMK1 +
##     gestwks * SMK2 + gestwks * SMK3, data = CHDS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9507 -0.6687  0.0313  0.5922  3.2100
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.10650    1.06362  -1.040    0.299
```

```
## SMK1         -2.39214    2.25385  -1.061    0.289
## SMK2         -0.34265    3.16583  -0.108    0.914
## SMK3         -1.76955    2.01751  -0.877    0.381
## gestwks       0.22135    0.02661   8.320 4.91e-16 ***
## SMK1:gestwks  0.05384    0.05682   0.948    0.344
## SMK2:gestwks -0.00756    0.08017  -0.094    0.925
## SMK3:gestwks  0.03432    0.05070   0.677    0.499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9678 on 672 degrees of freedom
## Multiple R-squared:  0.2231, Adjusted R-squared:  0.215
## F-statistic: 27.56 on 7 and 672 DF,  p-value: < 2.2e-16
```

Output A.4: Gestational Age as a Confounder for Birth Weight and SDP

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + gestwks, data = CHDS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9444 -0.6799  0.0219  0.5883  3.1785
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.71575    0.79583  -2.156   0.0314 *
## SMK1        -0.25749    0.10764  -2.392   0.0170 *
## SMK2        -0.63221    0.12932  -4.889 1.27e-06 ***
## SMK3        -0.40392    0.09832  -4.108 4.47e-05 ***
## gestwks      0.23660    0.01989  11.896  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9666 on 675 degrees of freedom
## Multiple R-squared:  0.2217, Adjusted R-squared:  0.2171
## F-statistic: 48.06 on 4 and 675 DF,  p-value: < 2.2e-16
```

```
## % change
```

```
##     SMK1     SMK2     SMK3
## 26.89641 16.82322 13.44370
```

Output A.5: Maternal Age as an Effect Modifier for Birth Weight and SDP

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + age + age * SMK1 + age *
##     SMK2 + age * SMK3, data = CHDS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3704 -0.7152 -0.0242  0.7164  3.7605
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.999695   0.264100  30.290  < 2e-16 ***
## SMK1        -0.636015   0.531049  -1.198  0.23147
## SMK2        -1.084608   0.695967  -1.558  0.11960
## SMK3        -1.520342   0.553027  -2.749  0.00614 **
## age         -0.010291   0.009965  -1.033  0.30211
## SMK1:age     0.010955   0.020247   0.541  0.58865
## SMK2:age     0.012542   0.026563   0.472  0.63696
## SMK3:age     0.040463   0.020826   1.943  0.05244 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.062 on 672 degrees of freedom
## Multiple R-squared:  0.06379,    Adjusted R-squared:  0.05403
## F-statistic: 6.541 on 7 and 672 DF,  p-value: 1.743e-07
```

Output A.6: Maternal Age as a Confounder for Birth Weight and SDP

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + age, data = CHDS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4302 -0.7311 -0.0319  0.7220  3.6710
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.7437494  0.2013039  38.468  < 2e-16 ***
## SMK1        -0.3524276  0.1181119  -2.984  0.00295 **
## SMK2        -0.7602196  0.1417548  -5.363 1.13e-07 ***
## SMK3        -0.4665944  0.1079818  -4.321 1.79e-05 ***
## age         -0.0004219  0.0074726  -0.056  0.95500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.063 on 675 degrees of freedom
## Multiple R-squared:  0.0585, Adjusted R-squared:  0.05292
## F-statistic: 10.49 on 4 and 675 DF,  p-value: 3.029e-08

## % change

##        SMK1        SMK2        SMK3
## -0.05725988 -0.01822295  0.01288602
```

Output A.7: Maternal Age as an Effect Modifier for Birth Weight and SDP

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + mheight + mheight * SMK1 +
##     mheight * SMK2 + mheight * SMK3, data = CHDS)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -4.3861 -0.6604  0.0013  0.6276  3.8070
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.72466    1.36754   1.261    0.208
## SMK1          0.16650    3.10800   0.054    0.957
## SMK2         -2.57609    3.52455  -0.731    0.465
## SMK3          3.13840    2.85958   1.098    0.273
## mheight       0.09315    0.02119   4.397 1.28e-05 ***
## SMK1:mheight -0.00757    0.04837  -0.157    0.876
## SMK2:mheight  0.02940    0.05507   0.534    0.594
## SMK3:mheight -0.05604    0.04417  -1.269    0.205
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.042 on 672 degrees of freedom
## Multiple R-squared:  0.09864,    Adjusted R-squared:  0.08925
## F-statistic: 10.51 on 7 and 672 DF,  p-value: 1.445e-12
```

Output A.8: Maternal Age as a Confounder for Birth Weight and SDP

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + mheight, data = CHDS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3901 -0.6779 -0.0048  0.6394  3.7952
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.23166    1.04577   2.134  0.03320 *
## SMK1        -0.32189    0.11585  -2.779  0.00561 **
## SMK2        -0.70440    0.13931  -5.056 5.51e-07 ***
## SMK3        -0.48885    0.10591  -4.616 4.69e-06 ***
## mheight      0.08529    0.01619   5.267 1.86e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.042 on 675 degrees of freedom
## Multiple R-squared:  0.09567,    Adjusted R-squared:  0.09031
## F-statistic: 17.85 on 4 and 675 DF,  p-value: 6.072e-14
```

```
## % change
```

```
##      SMK1      SMK2      SMK3
##  8.612399  7.325258 -4.755978
```

Output A.9: Maternal Age as an Effect Modifier for Birth Weight and SDP

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + BMI + BMI * SMK1 + BMI *
```

```
##     SMK2 + BMI * SMK3, data = CHDS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5371 -0.6979  0.0077  0.6899  3.6178
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.84268    0.44593  15.345   <2e-16 ***
## SMK1        -0.83185    1.09138  -0.762   0.4462
## SMK2        -0.34462    1.60278  -0.215   0.8298
## SMK3        -0.31620    0.81426  -0.388   0.6979
## BMI          0.04080    0.02029   2.011   0.0447 *
## SMK1:BMI     0.02574    0.05223   0.493   0.6223
## SMK2:BMI    -0.01826    0.07563  -0.241   0.8093
## SMK3:BMI    -0.00617    0.03755  -0.164   0.8695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.06 on 672 degrees of freedom
## Multiple R-squared:  0.06855,    Adjusted R-squared:  0.05885
## F-statistic: 7.065 on 7 and 672 DF,  p-value: 3.732e-08
```

Output A.10: Maternal Age as a Confounder for Birth Weight and SDP

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + BMI, data = CHDS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5381 -0.6864  0.0140  0.6775  3.6173
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.83464    0.34636  19.733  < 2e-16 ***
## SMK1        -0.30162    0.11903  -2.534  0.01150 *
## SMK2        -0.72880    0.14152  -5.150 3.42e-07 ***
## SMK3        -0.44781    0.10767  -4.159 3.61e-05 ***
## BMI          0.04117    0.01568   2.626  0.00885 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.058 on 675 degrees of freedom
## Multiple R-squared:  0.06801,    Adjusted R-squared:  0.06249
## F-statistic: 12.31 on 4 and 675 DF,  p-value: 1.136e-09

## % change

##      SMK1      SMK2      SMK3
## 14.367561  4.115376  4.038391
```

Output A.11: Full Association Model

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + gestwks + BMI, data = CHDS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1911 -0.6411  0.0124  0.5925  3.1337
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.51900    0.84650  -2.976  0.00303 **
## SMK1        -0.21063    0.10855  -1.940  0.05274 .
## SMK2        -0.60350    0.12916  -4.672 3.60e-06 ***
## SMK3        -0.38657    0.09808  -3.941 8.94e-05 ***
## gestwks      0.23573    0.01980  11.905  < 2e-16 ***
## BMI          0.03841    0.01427   2.692  0.00728 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9621 on 674 degrees of freedom
## Multiple R-squared:  0.2299, Adjusted R-squared:  0.2242
## F-statistic: 40.25 on 5 and 674 DF,  p-value: < 2.2e-16
```

Output A.12: Creation of New Full Model with Outlier Removed (See C.6)

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + gestwks + BMI, data = CHDS_60)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1929 -0.6312 -0.0012  0.5793  3.1058
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.90394    0.85123  -3.411 0.000685 ***
## SMK1        -0.21458    0.10792  -1.988 0.047169 *
## SMK2        -0.60595    0.12840  -4.719 2.88e-06 ***
## SMK3        -0.39169    0.09751  -4.017 6.57e-05 ***
## gestwks      0.24570    0.01996  12.308  < 2e-16 ***
## BMI          0.03817    0.01418   2.691 0.007294 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9564 on 673 degrees of freedom
## Multiple R-squared:  0.2396, Adjusted R-squared:  0.234
## F-statistic: 42.42 on 5 and 673 DF,  p-value: < 2.2e-16
```

Output A.13: Confidence Intervals for Full vs Reduced Models — Association

```
## For the crude model:
```

```
##                  2.5 %     97.5 %
## (Intercept)  7.6259530  7.8396639
## SMK1        -0.5838591 -0.1205927
## SMK2        -1.0381668 -0.4819955
## SMK3        -0.6785080 -0.2548011
```

```
## For the adjusted model:
```

```
##                   2.5 %       97.5 %
## (Intercept) -4.57533007 -1.232559664
## SMK1        -0.42647420 -0.002690799
## SMK2        -0.85806889 -0.353825765
## SMK3        -0.58315619 -0.200217851
## gestwks      0.20650008  0.284891594
## BMI          0.01032253  0.066019190
```

Output A.14: Prediction Model Backwards Elimination Procedure Step 1 — All variables

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + gestwks + age + BMI +
##     mheight, data = CHDS_60)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2771 -0.6106  0.0142  0.5561  3.2634
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.100887   1.256253  -6.448 2.16e-10 ***
## SMK1        -0.180649   0.105766  -1.708   0.0881 .
## SMK2        -0.550760   0.126018  -4.370 1.44e-05 ***
## SMK3        -0.410902   0.095487  -4.303 1.93e-05 ***
## gestwks      0.241351   0.019546  12.348  < 2e-16 ***
## age         -0.003674   0.006626  -0.554   0.5795
## BMI          0.044959   0.014016   3.208   0.0014 **
## mheight      0.082446   0.014603   5.646 2.43e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9358 on 671 degrees of freedom
## Multiple R-squared:  0.2743, Adjusted R-squared:  0.2667
## F-statistic: 36.23 on 7 and 671 DF,  p-value: < 2.2e-16
```

Output A.15: Prediction Model Backwards Elimination Procedure Step 2 — Maternal Age Removed

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + gestwks + BMI + mheight,
##     data = CHDS_60)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -3.2487 -0.6078  0.0073  0.5562  3.2309
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.16421    1.25041  -6.529 1.30e-10 ***
## SMK1        -0.18009    0.10571  -1.704  0.08890 .
## SMK2        -0.55037    0.12595  -4.370 1.44e-05 ***
## SMK3        -0.41179    0.09542  -4.315 1.83e-05 ***
## gestwks      0.24136    0.01954  12.355  < 2e-16 ***
## BMI          0.04403    0.01391   3.166  0.00162 **
## mheight      0.08226    0.01459   5.638 2.54e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9353 on 672 degrees of freedom
## Multiple R-squared:  0.274,  Adjusted R-squared:  0.2675
## F-statistic: 42.26 on 6 and 672 DF,  p-value: < 2.2e-16
```

Output A.16: Confidence Intervals — Prediction

```
##                    2.5 %      97.5 %
## (Intercept) -10.61938216 -5.70903683
## SMK1         -0.38764413  0.02746481
## SMK2         -0.79767379 -0.30306260
## SMK3         -0.59915690 -0.22442405
## gestwks       0.20300269  0.27972030
## BMI           0.01671973  0.07133750
## mheight       0.05360913  0.11090837
```

Appendix B: Plots and Figures

Figure B.1 Scatterplot Matrix

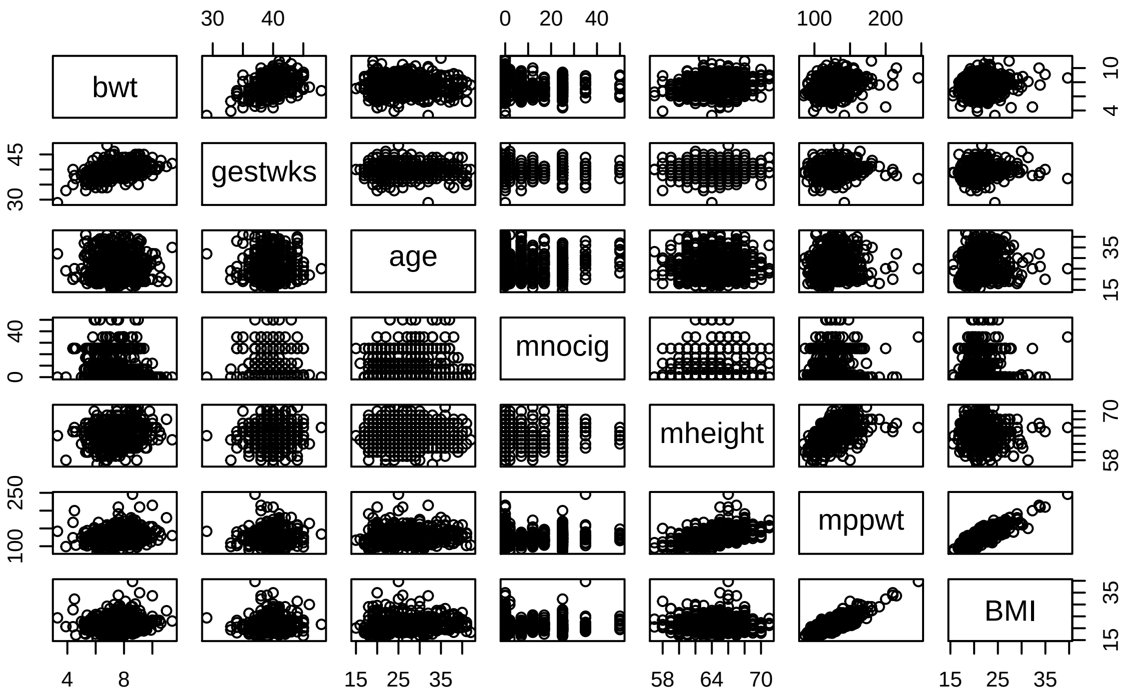
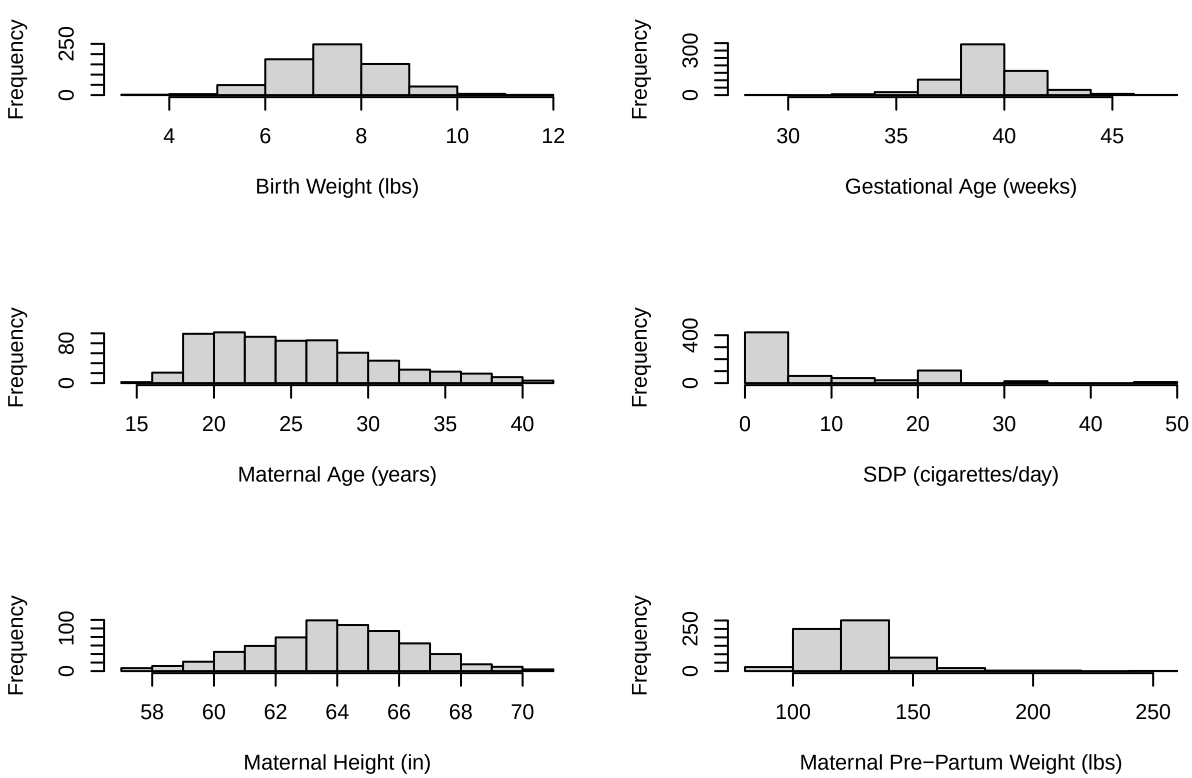


Figure B.2 Distributions of Sample Variables



Appendix C: Model Diagnostics

Figure C.1: Plot of Standardized Model Residuals as a Function of Fitted Values for Birth Weight — Association

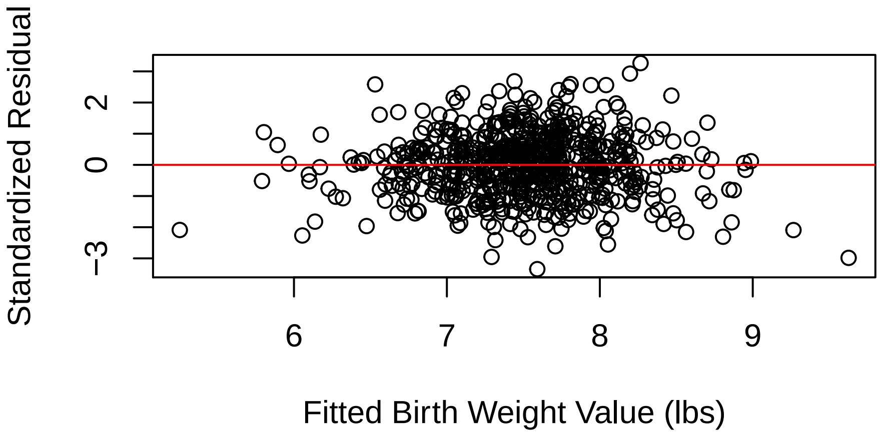


Figure C.2: Plot of Studentized Model Residuals as a Function of Fitted Values for Birth Weight — Association

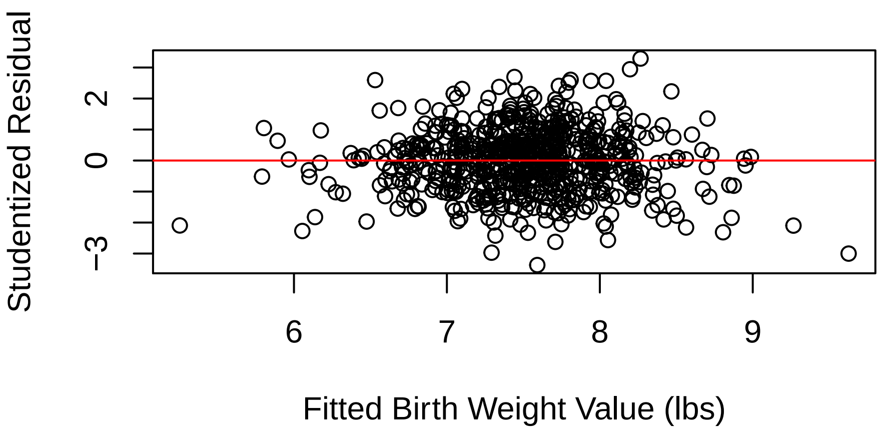


Figure C.3: Q-Q Plot of Residuals

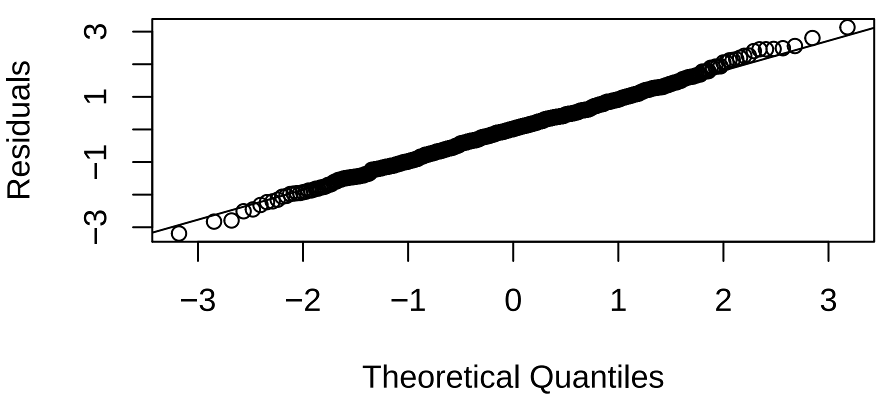


Figure C.4: Q-Q Plot of Standardized Residuals

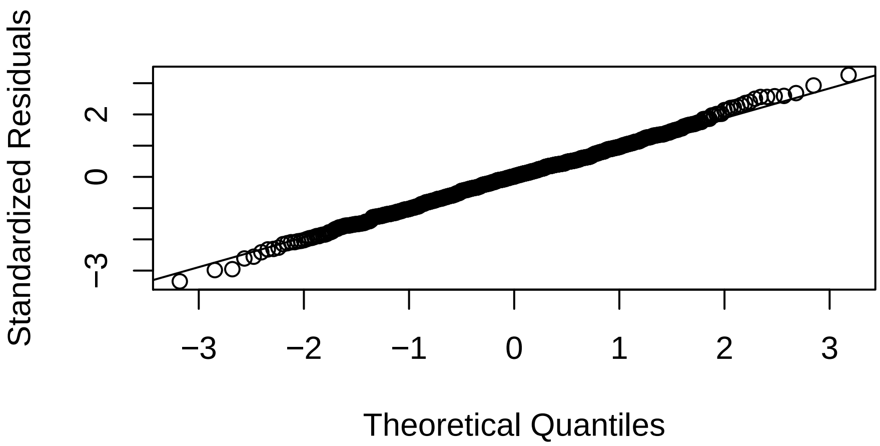
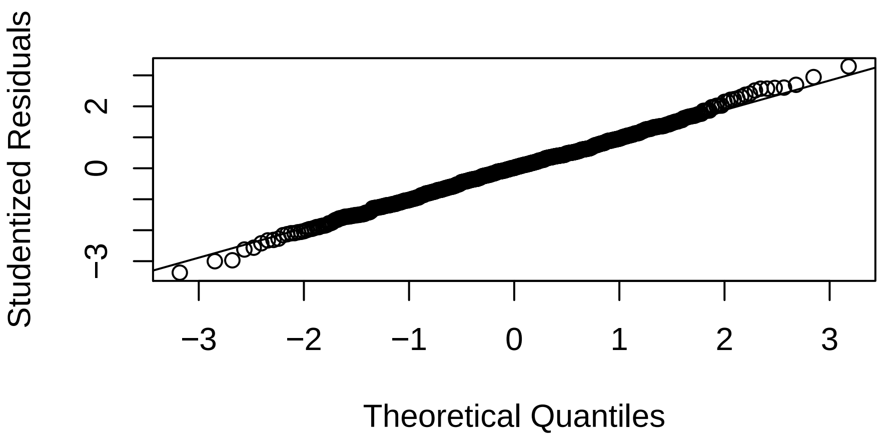


Figure C.5: Q-Q Plot of Studentized Residuals



Output C.6: Influential Outlier Detection with Cook's Distance

```
## The output below shows all potential outliers from this data

## [1] bwt     gestwks age     mnocig  mheight mppwt   BMI     h       c
## <0 rows> (or 0-length row.names)

##      bwt gestwks age mnocig mheight mppwt     BMI      h      c
## 9    3.3      29  32      0      64   142 24.3716 0.0550 0.0423
## 60   6.8      48  25      0      66   134 21.6258 0.0302 0.0462
## 166  5.3      33  20      7      63   109 19.3064 0.0281 0.0013
## 167  8.6      37  25     35      66   246 39.7011 0.0852 0.0290
## 356  8.1      44  30      2      58   150 31.3466 0.0431 0.0047
## 400 10.0      38  32      0      67   215 33.6701 0.0353 0.0352
## 404  7.6      39  26      0      66   210 33.8912 0.0352 0.0010
## 445  7.3      46  22      2      62   164 29.9927 0.0463 0.0354
## 506  4.5      38  25     25      66   200 32.2773 0.0353 0.0532
## 610  9.1      40  20      0      65   210 34.9420 0.0405 0.0057
```

Output C.7: Collinearity Assessment — Association

```
##      SMK1     SMK2     SMK3  gestwks      BMI
## 1.112337 1.073923 1.092602 1.010782 1.030826
```

```
## [1] 1.064094
```

Figure C.8: Plot of Standardized Model Residuals as a Function of Fitted Values for Birth Weight — Prediction

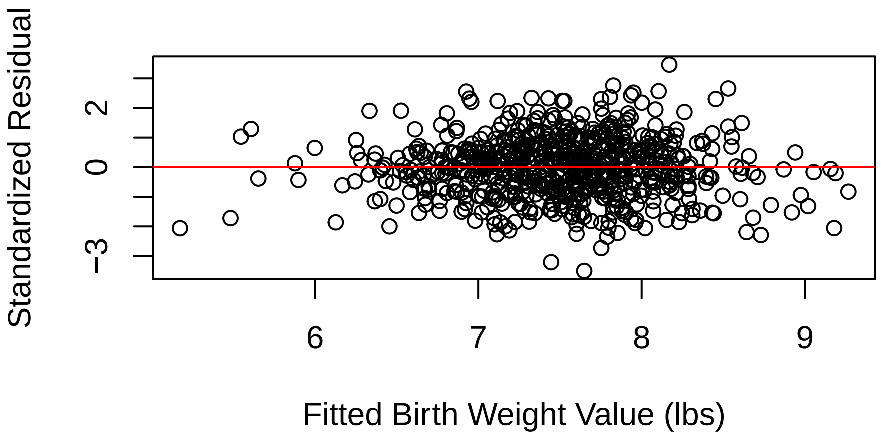
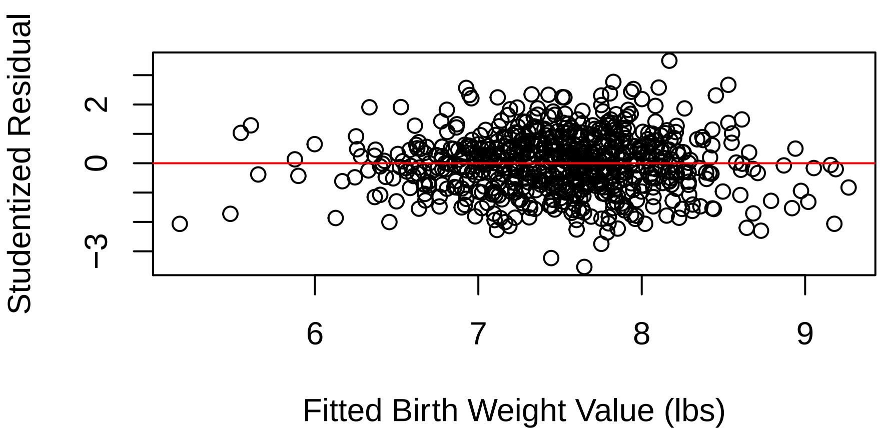


Figure C.9: Plot of Studentized Model Residuals as a Function of Fitted Values for Birth Weight — Prediction



Output C.7: Collinearity Assessment — Prediction

```
##      SMK1     SMK2     SMK3  gestwks      BMI  mheight
## 1.116076 1.080543 1.094130 1.012350 1.036611 1.018422
```

```
## [1] 1.059689
```

Appendix D: Mediation Analysis

Output D.1: Gestational Age and BMI

```
##
## Call:
## lm(formula = gestwks ~ BMI + SMK1 + SMK2 + SMK3, data = CHDS_60)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9440  -0.9178   0.0979   1.0855   6.3526
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 39.65007    0.60440  65.602   <2e-16 ***
## BMI          0.01206    0.02736   0.441   0.6595
## SMK1        -0.36435    0.20776  -1.754   0.0799 .
## SMK2        -0.51005    0.24699  -2.065   0.0393 *
## SMK3        -0.23840    0.18794  -1.269   0.2051
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.846 on 674 degrees of freedom
## Multiple R-squared:  0.01067,    Adjusted R-squared:  0.004795
## F-statistic: 1.817 on 4 and 674 DF,  p-value: 0.1238
```

Output D.2: Pre-Pregnancy BMI and Smoking

```
##
## Call:
## lm(formula = BMI ~ SMK1 + SMK2 + SMK3 + age, data = CHDS_60)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.917 -1.617 -0.351  1.069 18.404
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 20.36065    0.48860  41.672  < 2e-16 ***
## SMK1        -1.20267    0.28670  -4.195  3.1e-05 ***
## SMK2        -0.74168    0.34406  -2.156  0.03146 *
## SMK3        -0.46607    0.26212  -1.778  0.07584 .
## age          0.05609    0.01813   3.093  0.00206 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.58 on 674 degrees of freedom
## Multiple R-squared:  0.0432, Adjusted R-squared:  0.03753
## F-statistic: 7.609 on 4 and 674 DF,  p-value: 5.347e-06
```

Output D.3: Pre-Pregnancy BMI and Birth Weight

```
##
```

```
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + gestwks + BMI, data = CHDS_60)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1929 -0.6312 -0.0012  0.5793  3.1058
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.90394    0.85123  -3.411 0.000685 ***
## SMK1        -0.21458    0.10792  -1.988 0.047169 *
## SMK2        -0.60595    0.12840  -4.719 2.88e-06 ***
## SMK3        -0.39169    0.09751  -4.017 6.57e-05 ***
## gestwks      0.24570    0.01996  12.308  < 2e-16 ***
## BMI          0.03817    0.01418   2.691 0.007294 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9564 on 673 degrees of freedom
## Multiple R-squared:  0.2396, Adjusted R-squared:  0.234
## F-statistic: 42.42 on 5 and 673 DF,  p-value: < 2.2e-16
```

Output D.4: Light Smoking-BMI Mediation Analysis

```
## Running nonparametric bootstrap
```

```
##
## Causal Mediation Analysis
##
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
##
##                Estimate 95% CI Lower 95% CI Upper p-value
## ACME            -0.0459      -0.0971        -0.01    0.02 *
## ADE             -0.2146      -0.4283         0.01    0.06 .
## Total Effect    -0.2605      -0.4764        -0.05    0.02 *
## Prop. Mediated   0.1762       0.0107         0.87    0.04 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 679
##
##
## Simulations: 500
```

Output D.5: Moderate Smoking-BMI Mediation Analysis

```
## Running nonparametric bootstrap
```

```
##
## Causal Mediation Analysis
##
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
##
##                Estimate 95% CI Lower 95% CI Upper p-value
```

```
## ACME           -0.02831     -0.06688         0.00   0.008 **
## ADE            -0.60595     -0.83345        -0.35  <2e-16 ***
## Total Effect   -0.63426     -0.86513        -0.39  <2e-16 ***
## Prop. Mediated  0.04464      0.00495         0.13   0.008 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 679
##
##
## Simulations: 500
```

Output D.6: Heavy Smoking-BMI Mediation Analysis

```
## Running nonparametric bootstrap
```

```
##
## Causal Mediation Analysis
##
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
##
##                Estimate 95% CI Lower 95% CI Upper p-value
## ACME           -0.01779     -0.05365         0.00    0.12
## ADE            -0.39169     -0.58644        -0.21  <2e-16 ***
## Total Effect   -0.40948     -0.59622        -0.23  <2e-16 ***
## Prop. Mediated  0.04345     -0.00966         0.15    0.12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 679
##
##
## Simulations: 500
```

Output D.7: Gestational Age and SDP

```
##
## Call:
## lm(formula = gestwks ~ SMK1 + SMK2 + SMK3 + age, data = CHDS_60)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9172  -0.9125   0.0881   1.0861   6.4683
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 39.8960237  0.3495997 114.119   <2e-16 ***
## SMK1        -0.3788599  0.2051386  -1.847   0.0652 .
## SMK2        -0.5190000  0.2461802  -2.108   0.0354 *
## SMK3        -0.2440197  0.1875536  -1.301   0.1937
## age          0.0006606  0.0129751   0.051   0.9594
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.846 on 674 degrees of freedom
## Multiple R-squared:  0.01039,    Adjusted R-squared:  0.004512
## F-statistic: 1.768 on 4 and 674 DF,  p-value: 0.1334
```

Output D.8: Gestational Age and Birth Weight

```
##
## Call:
## lm(formula = bwt ~ SMK1 + SMK2 + SMK3 + gestwks + BMI, data = CHDS_60)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1929 -0.6312 -0.0012  0.5793  3.1058
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.90394    0.85123  -3.411 0.000685 ***
## SMK1        -0.21458    0.10792  -1.988 0.047169 *
## SMK2        -0.60595    0.12840  -4.719 2.88e-06 ***
## SMK3        -0.39169    0.09751  -4.017 6.57e-05 ***
## gestwks      0.24570    0.01996  12.308  < 2e-16 ***
## BMI          0.03817    0.01418   2.691 0.007294 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9564 on 673 degrees of freedom
## Multiple R-squared:  0.2396, Adjusted R-squared:  0.234
## F-statistic: 42.42 on 5 and 673 DF,  p-value: < 2.2e-16
```

Output D.9: Light Smoking-Gestational Age Mediation Analysis

```
## Running nonparametric bootstrap

##
## Causal Mediation Analysis
##
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
##
##                 Estimate 95% CI Lower 95% CI Upper p-value
## ACME            -0.09308     -0.19933         0.00   0.052 .
## ADE             -0.21458     -0.43320         0.01   0.064 .
## Total Effect    -0.30767     -0.55474        -0.08   0.004 **
## Prop. Mediated  0.30255     -0.00818         1.06   0.056 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 679
##
##
## Simulations: 500
```

Output D.10: Moderate Smoking-Gestational Age Mediation Analysis

```
## Running nonparametric bootstrap
##
## Causal Mediation Analysis
##
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
##
##                Estimate 95% CI Lower 95% CI Upper p-value
## ACME            -0.1275      -0.2375        -0.03   0.024 *
## ADE             -0.6059      -0.8233        -0.38  <2e-16 ***
## Total Effect    -0.7335      -0.9568        -0.49  <2e-16 ***
## Prop. Mediated   0.1739       0.0352         0.31   0.024 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 679
##
##
## Simulations: 500
```

Output D.11: Heavy Smoking-Gestational Age Mediation Analysis

```
## Running nonparametric bootstrap
##
## Causal Mediation Analysis
##
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
##
##                Estimate 95% CI Lower 95% CI Upper p-value
## ACME            -0.0600      -0.1513         0.03     0.2
## ADE             -0.3917      -0.5909        -0.21  <2e-16 ***
## Total Effect    -0.4516      -0.6696        -0.24  <2e-16 ***
## Prop. Mediated   0.1327      -0.0955         0.31     0.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 679
##
##
## Simulations: 500
```

References

Basto-Abreu, Ana, Tonatiuh Barrientos-Gutiérrez, Rodrigo Zepeda-Tello, Vanessa Camacho, David Gimeno Ruiz de Porras, and Mauricio Hernández-Ávila. 2018. "The Relationship of Socioeconomic Status with Body Mass Index Depends on the Socioeconomic Measure Used: Socioeconomic Associations of Body Mass Index." *Obesity* 26 (1): 176–84. https://doi.org/10.1002/oby.22042.

Blencowe, Hannah, Julia Krasevec, Mercedes de Onis, Robert E Black, Xiaoyi An, Gretchen A Stevens, Elaine Borghi, et al. 2019. "National, Regional, and Worldwide Estimates of Low Birthweight in 2015, with Trends from 2000: A Systematic Analysis." *The Lancet. Global Health* 7 (7): e849–60. https://doi.org/10.1016/S2214-109X(18)30565-5.

Cederholm, T., I. Bosaeus, R. Barazzoni, J. Bauer, A. Van Gossum, S. Klek, M. Muscaritoli, et al. 2015. "Diagnostic Criteria for Malnutrition – An ESPEN Consensus Statement." *Clinical Nutrition* 34 (3): 335–40. https://doi.org/10.1016/j.clnu.2015.03.001.

"Committee Opinion No. 591: Challenges for Overweight and Obese Women: Correction." 2016. *Obstetrics and Gynecology* 127 (1): 166. https://doi.org/10.1097/AOG.0000000000001227.

Dennis, Jeff A., and Stefanie Mollborn. 2013. "Young Maternal Age and Low Birth Weight Risk: An Exploration of Racial/Ethnic Disparities in the Birth Outcomes of Mothers in the United States." *The Social Science Journal* 50 (4): 625–34. https://doi.org/10.1016/j.soscij.2013.09.008.

Ebrahim, Shah. 2014. "Yerushalmy and the Problems of Causal Inference." *International Journal of Epidemiology* 43 (5): 1349–51. https://doi.org/10.1093/ije/dyu186.

Engle, William A. 2006. "A Recommendation for the Definition of 'Late Preterm' (Near-Term) and the Birth Weight–Gestational Age Classification System." *Seminars in Perinatology* 30 (1): 2–7. https://doi.org/10.1053/j.semperi.2006.01.007.

Fox, John, and Sanford Weisberg. 2018. *An r Companion to Applied Regression*. 3rd edition. Los Angeles: SAGE Publications, Inc.

Fox, John, Sanford Weisberg, Brad Price, Daniel Adler, Douglas Bates, Gabriel Baud-Bovy, Ben Bolker, et al. 2020. *Car: Companion to Applied Regression* (version 3.0-10). https://CRAN.R-project.org/package=car.

Fraser, Alison M., John E. Brockert, and R. H. Ward. 1995. "Association of Young Maternal Age with Adverse Reproductive Outcomes." *New England Journal of Medicine* 332 (17): 1113–18. https://doi.org/10.1056/NEJM199504273321701.

Goldstein, Harvey. 2014. "Commentary: Smoking in Pregnancy and Neonatal Mortality." *International Journal of Epidemiology* 43 (5): 1366–68. https://doi.org/10.1093/ije/dyu161.

Gould, J. B., B. Davey, and S. LeRoy. 1989. "Socioeconomic Differentials and Neonatal Mortality: Racial Comparison of California Singletons." *Pediatrics* 83 (2): 181–86.

Gul, Rafia, Samar Iqbal, Zahid Anwar, Saher Gul Ahdi, Syed Hamza Ali, and Saima Pirzada. 2020. "Pre-Pregnancy Maternal BMI as Predictor of Neonatal Birth Weight." *PLoS ONE* 15 (10). https://doi.org/10.1371/journal.pone.0240748.

Hack, Maureen, Daniel J. Flannery, Mark Schluchter, Lydia Cartar, Elaine Borawski, and Nancy Klein. 2002. "Outcomes in Young Adulthood for Very-Low-Birth-Weight Infants." *New England Journal of Medicine* 346 (3): 149–57. https://doi.org/10.1056/NEJMoa010856.

Hack, Maureen, H. Gerry Taylor, Dennis Drotar, Mark Schluchter, Lydia Cartar, Laura Andreias, Deanne Wilson-Costello, and Nancy Klein. 2005. "Chronic Conditions, Functional Limitations, and Special Health Care Needs of School-Aged Children Born with Extremely Low-Birth-Weight in the 1990s." *JAMA* 294 (3): 318–25. https://doi.org/10.1001/jama.294.3.318.

Hickey, R. J., R. C. Clelland, and E. J. Bowers. 1978. "Maternal Smoking, Birth Weight, Infant Death, and the Self-Selection Problem." *American Journal of Obstetrics and Gynecology* 131 (7): 805–11.

https://doi.org/10.1016/0002-9378(78)90253-3.

Hoffman, Caroline S., Lynne C. Messer, Pauline Mendola, David A. Savitz, Amy H. Herring, and Katherine E. Hartmann. 2008. "Comparison of Gestational Age at Birth Based on Last Menstrual Period and Ultrasound During the First Trimester." *Paediatric and Perinatal Epidemiology* 22 (6): 587–96. https://doi.org/10.1111/j.1365-3016.2008.00965.x.

Husten, C. G. 2009. "How Should We Define Light or Intermittent Smoking? Does It Matter?" *Nicotine & Tobacco Research* 11 (2): 111–21. https://doi.org/10.1093/ntr/ntp010.

Inoue, Sachiko, Hiroo Naruse, Takashi Yorifuji, Tsuguhiko Kato, Takeshi Murakoshi, Hiroyuki Doi, and SV Subramanian. 2016. "Association Between Short Maternal Height and Low Birth Weight: A Hospital-Based Study in Japan." *Journal of Korean Medical Science* 31 (3): 353–59. https://doi.org/10.3346/jkms.2016.31.3.353.

Keyes, Katherine M, George Davey Smith, and Ezra Susser. 2014. "Commentary: Smoking in Pregnancy and Offspring Health: Early Insights into Family-Based and 'Negative Control' Studies?" *International Journal of Epidemiology* 43 (5): 1381–88. https://doi.org/10.1093/ije/dyu166.

Kleinbaum, David G., Lawrence L. Kupper, Azhar Nizam, and Eli S. Rosenberg. 2013. *Applied Regression Analysis and Other Multivariable Methods*. 5th edition. Boston, MA: Cengage Learning.

Knopik, Valerie S., Kristine Marceau, Rohan H. C. Palmer, Taylor F. Smith, and Andrew C. Heath. 2016. "Maternal Smoking During Pregnancy and Offspring Birth Weight: A Genetically-Informed Approach Comparing Multiple Raters." *Behavior Genetics* 46 (3): 353–64. https://doi.org/10.1007/s10519-015-9750-6.

Kowlessar, Niranjana M., H. Joanna Jiang, and Claudia Steiner. 2006. "Hospital Stays for Newborns, 2011: Statistical Brief #163." In *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Rockville (MD): Agency for Healthcare Research; Quality (US). http://www.ncbi.nlm.nih.gov/books/NBK173954/.

Kramer, Michael S., Xun Zhang, and Robert W. Platt. 2014. "Commentary: Yerushalmy, Maternal Cigarette Smoking and the Perinatal Mortality Crossover Paradox." *International Journal of Epidemiology* 43 (5): 1378–81. https://doi.org/10.1093/ije/dyu164.

La Merrill, Michele A., Cheryl R. Stein, Philip Landrigan, Stephanie M. Engel, and David A. Savitz. 2011. "Prepregnancy Body Mass Index, Smoking During Pregnancy, and Infant Birth Weight." *Annals of Epidemiology* 21 (6): 413–20. https://doi.org/10.1016/j.annepidem.2010.11.012.

Martin, Joyce A., Brady E. Hamilton, Michelle Jk Osterman, Sally C. Curtin, and T. J. Matthews. 2015. "Births: Final Data for 2013." *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System* 64 (1): 1–65.

McGowan, Craig. 2019. *Describedata: Miscellaneous Descriptive Functions* (version 0.1.0). https://CRAN.R-project.org/package=describedata.

National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. 2014. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. Reports of the Surgeon General. Atlanta (GA): Centers for Disease Control; Prevention (US). http://www.ncbi.nlm.nih.gov/books/NBK179276/.

"National Vital Statistics Reports, Volume 64, Number 5, (06/01/2015)." n.d., 20.

"NCHS Data Brief No. 305 February 2018." 2018, no. 305: 8.

Parascandola, Mark. 2014. "Commentary: Smoking, Birthweight and Mortality: Jacob Yerushalmy on Self-Selection and the Pitfalls of Causal Inference." *International Journal of Epidemiology* 43 (5): 1373–77. https://doi.org/10.1093/ije/dyu163.

Pickett, Kate E., Barbara Abrams, and Steve Selvin. 2000. "Maternal Height, Pregnancy Weight Gain, and Birthweight." *American Journal of Human Biology: The Official Journal of the Human Biology Council* 12 (5): 682–87. https://doi.org/10.1002/1520-6300(200009/10)12:5%3C682::AID-AJHB13%3E3.0.CO;2-X.

Rp, Britto, Florêncio Tm, Benedito Silva Aa, Sesso R, Cavalcante Jc, and Sawaya Al. 2013. "Influence of Maternal Height and Weight on Low Birth Weight: A Cross-Sectional Study in Poor Communities of Northeastern Brazil." *Plos One* 8 (11): e80159–59. https://doi.org/10.1371/journal.pone.0080159.

SNOWDEN, Jonathan M, and Olga BASSO. 2018. "Causal Inference in Studies of Preterm Babies: A Simulation Study." *BJOG : An International Journal of Obstetrics and Gynaecology* 125 (6): 686–92. https://doi.org/10.1111/1471-0528.14942.

Srinivas, Sindhu K., Jamie Bastek, Meghan A. McShea, Markley N. Foreman, and Michal A. Elovitz. 2011. "526: Unraveling the Racial Disparity in Preterm Birth: Experiences of Discrimination Are Associated with Preterm Birth and Low Birth Weight in Black Women." *American Journal of Obstetrics & Gynecology* 204 (1): S209–10. https://doi.org/10.1016/j.ajog.2010.10.545.

Textor, Johannes, Benito van der Zander, and Ankur Ankan. 2021. *Dagitty: Graphical Analysis of Structural Causal Models* (version 0.3-1). https://CRAN.R-project.org/package=dagitty.

Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele, Kosuke Imai, Minh Trinh, and Weihuang Wong. 2019. *Mediation: Causal Mediation Analysis* (version 4.5.0). https://CRAN.R-project.org/package=mediation.

"Tobacco and Nicotine Cessation During Pregnancy: ACOG Committee Opinion, Number 807." 2020. *Obstetrics and Gynecology* 135 (5): e221–29. https://doi.org/10.1097/AOG.0000000000003822.

VanderWeele, Tyler J. 2014. "Commentary: Resolutions of the Birthweight Paradox: Competing Explanations and Analytical Insights." *International Journal of Epidemiology* 43 (5): 1368–73. https://doi.org/10.1093/ije/dyu162.

VanderWeele, Tyler J. 2016. "Mediation Analysis: A Practitioner's Guide." *Annual Review of Public Health* 37: 17–32. https://doi.org/10.1146/annurev-publhealth-032315-021402.

Ventura, Stephanie J., Brady E. Hamilton, T. J. Mathews, and Anjani Chandra. 2003. "Trends and Variations in Smoking During Pregnancy and Low Birth Weight: Evidence from the Birth Certificate, 1990-2000." *Pediatrics* 111 (5): 1176–80.

Vohr, Betty R., Jon E. Tyson, Linda L Wright, Rebecca L. Perritt, Lei Li, and W. Kenneth Poole. 2009. "Maternal Age, Multiple Birth and Extremely Low Birth Weight Infants." *The Journal of Pediatrics* 154 (4): 498–503.e2. https://doi.org/10.1016/j.jpeds.2008.10.044.

Wang, Shanshan, Liren Yang, Li Shang, Wenfang Yang, Cuifang Qi, Liyan Huang, Guilan Xie, Ruiqi Wang, and Mei Chun Chung. 2020. "Changing Trends of Birth Weight with Maternal Age: A Cross-Sectional Study in Xi'an City of Northwestern China." *BMC Pregnancy and Childbirth* 20 (1): 744. https://doi.org/10.1186/s12884-020-03445-2.

"WHO Global Nutrition Targets 2025: Low Birth Weight Policy Brief." n.d. *WHO*. Accessed February 19, 2021. http://www.who.int/nutrition/publications/globaltargets2025_policybrief_lbw/en/.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and RStudio. 2021. *Dplyr: A Grammar of Data Manipulation* (version 1.0.4). https://CRAN.R-project.org/package=dplyr.

Witter, Frank R., and Barbara Luke. 1991. "The Effect of Maternal Height on Birth Weight and Birth Length." *Early Human Development* 25 (3): 181–86. https://doi.org/10.1016/0378-3782(91)90114-I.

Yerushalmy, J. 2014. "The Relationship of Parents' Cigarette Smoking to Outcome of Pregnancy–Implications as to the Problem of Inferring Causation from Observed Associations." *International Journal of Epidemiology* 43 (5): 1355–66. https://doi.org/10.1093/ije/dyu160.

Yerushalmy, J. n.d. "THE RELATIONSHIP OF PARENTS' CIGARETTE SMOKING TO OUTCOME OF PREGNANCY—IMPLICATIONS AS TO THE PROBLEM OF INFERRING CAUSATION FROM OBSERVED ASSOCIATIONS," 14.

Zhang, Weihui, and Tse-Chuan Yang. 2019. "Maternal Smoking and Infant Low Birth Weight: Exploring the Biological Mechanism Through the Mother's Pre-Pregnancy Weight Status." *Population Research and*

Policy Review, October. https://doi.org/10.1007/s11113-019-09554-x.