Regression Analysis: Studying the Effect Of Maternal Age on Birthweight

Group Members:

Cody Cronin, Stefano Illanes, Daniela Martinez, Dylan Suda

Abstract

The United States has one of the highest infant mortality rates in the developed world, low birthweight being one of the many contributing factors to infant mortality. Babies can be born with a low birthweight for several reasons. This study looks to model and identify the effect of maternal age on the birthweight of a child. To examine this effect, we collected a random sample of 1000 birth cases from a dataset released by the state of North Carolina in 2004. Reviewing the literature, we found that race/ethnicity, socioeconomic status, and health habits all contributed towards birthweight. In our linear regression model, maternal age is the independent variable of interest and birthweight is the dependent variable. Because of the elaborate nature of the relationship under study, we included several control variables to offset some potential biases with our model. The findings showed that there was a positive, albeit small and inconsistent relationship between maternal age and birthweight. This implies that there are more factors to consider when studying the effect of maternal age on birthweight.

Introduction

Infant mortality remains an issue in the United States. According to the *Committee to Study the Prevention of Low Birthweight,* low birthweight is a growing contributing factor to infant mortality in the United States, and it accounts for 20 to 40 percent of post neonatal deaths in developing countries (1985). The aim of our study is to answer the question, “What role does maternal age play in determining a child’s birthweight?” There are many factors that contribute to birthweight, and while there are many studies that examine the effects of said factors, there are few that attempt to isolate maternal age and its impact on birthweight, ceteris paribus. Successfully answering our research question can provide direction on where the United States should more greatly focus healthcare policy regarding prenatal care and education. As discussed in the literature review section, many of the referenced studies analyze the influence of factors other than maternal age but oftentimes will include age and investigate its interaction with the variables of interest. We limit the scope of our study to the relationship between maternal age and birthweight, controlling for omitted variables to avoid bias in our estimates and to get a more accurate picture of the regression coefficient. We hypothesize that there is a negative relationship between maternal age and birthweight, saying that as maternal age increases, birthweight will decrease. Although we opt to study the effect of maternal age in isolation, we recognize that there are numerous factors that go into determining birthweight, and that many of them interact with each other and with maternal age. We will discuss this and devise how our study could be improved were it to be replicated in the future. Acknowledging our study's limitations will help us craft a more complete answer to our research question and suggest where healthcare policy is best (most productively) targeted regarding prenatal care and education.

Literature Review

Upon reviewing the literature surrounding our topic of interest, we found articles that examine the effects of maternal age alongside other related factors, such as socioeconomic status, race/ethnicity, healthcare access, health habits, and more.

The first article, “*Diverging associations of maternal age with low birthweight for black and white mothers”* (Rich-Edwards et al., 2003) evaluated the relationship between maternal age and race/ethnicity with the risk of low birthweight. They found a divergent trend between the two groups; as maternal age increased, the risk of low birthweight increased more for black mothers than it did for white mothers. Using birth certificate data from the city of Chicago, they developed an interaction model that would control for other variables not included in the dataset. Controlling for and defining the interactions between marital status, level of neighborhood poverty, adequacy of prenatal care, smoking habits, and maternal age and ethnicity would go on to diminish the divergent gap of the risk of low birthweight between black and white mothers. They ultimately found that “the risk of delivering a low birthweight child rose with maternal age for *both* white and black women who were unmarried, lived in poor neighborhoods, smoked cigarettes, or received inadequate prenatal care” (Rich-Edwards et al., 2003).

The second article, “*Association between maternal smoking during pregnancy and low birthweight: Effects by maternal age”* (Zheng et al., 2016) evaluates the association with maternal smoking and low birthweight by maternal age. The authors identify a U-shaped distribution between maternal age and birthweight, stating that younger mothers and older mothers share an increased likelihood to give birth to low birthweight infants. Although they largely focused on the effects of maternal smoking on low birthweight, they show that this association increased with maternal age in both adjusted and unadjusted models.

The last article we review, *“Changing trends of birth weight with maternal age: A cross-sectional study in Xi’an City of Northwestern China”* (Wang et al., 2020), offers a more forthright look into the trends of birthweight in relation to maternal age. Extracting from abundant birth records (N=536993) in Xi’an city of Shaanxi, a province in China, the authors would exclude the extremes of the data as these outliers had more complications and too many confounding factors that would introduce bias in their measurements. This is important to note because it reveals greater insight into the relationship between the typical maternal age range (20 to 40 years as examined in this study) and birthweight, something that was unclear and elusive at best in prior studies. They found the relationship between maternal age and birthweight/risk of low birthweight to be nonlinear. The data showed that birthweight increased at a decreasing rate from 20 to 34 years old then marginally decreased thereafter. The risk of low birthweight decreased at a decreasing rate from 20 to 36 years old and would thereafter significantly increase. While these findings seem to be credible, they acknowledge potentially confounding factors that were not adjusted for and that could affect the validity of the study.

The results of the aforementioned studies further our suspicions that determining the variables involved in birthweight is complex and cannot be explained by any individual indicator, though an attempt will be made to isolate the effect of maternal age.

<u>*Data*</u>

For our data, we use a random sample provided by OpenIntro Statistics on 1000 birth cases from a dataset released by the state of North Carolina in 2004 containing information on births recorded in the state. The data frame is categorized with information provided on 13 different variables, many of which we make use of in our regression models. The variables are described as follows: fage (father's age in years), mage (mother's age in years), mature (maturity status of mother), weeks (length of pregnancy in weeks), premie (whether the birth was classified as premature or full-term), visits (number of hospital visits during pregnancy), gained (weight gained by mother during pregnancy in pounds), weight (birthweight of baby in pounds), low birthweight (whether the baby is classified as low birthweight or not), gender (gender of baby, classified as male or female), habit (whether the mother is a smoker or nonsmoker), marital (whether the mother is married or unmarried at delivery) and white mom (whether the mother is white or not white). Seven of the thirteen variables are categorical, of which only five are used. We create dummy variables to capture their effects within the regression models, allowing for the other 6 discrete and nonbinary variables to remain as is (we exclude father's age from the regression). The summary statistics of our data is shown below:

Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
fage	829	30.3	6.8	14	55
mage	1,000	27.0	6.2	13	50
weeks	998	38.3	2.9	20	45
visits	991	12.1	4.0	0	30
gained	973	30.3	14.2	0	85
weight	1,000	7.1	1.5	1.0	11.8

Data that was missing from the sample were automatically excluded from the set, as noted by N<1000 for 4 of the 6 variables.

Statistical Methods

The association between maternal age and birthweight was examined using a multivariable linear regression analysis of the data in the statistical package R. We set up birthweight as the dependent variable and maternal age as the primary independent variable of interest. We constructed 9 linear models to test our predictions with, listed below:

- Model 1: $Weight_i = \beta_0 + \beta_1$ MothersAge $+ \mu_i$
- Model 2 (*BASE SPECIFICATION*): $Weight_i = \beta_0 + \beta_1$ MothersAge $+ \beta_2$ TermLength $+ \mu_i$
- Model 3: $Weight_i = \beta_0 + \beta_1$ MothersAge $+ \beta_2$ TermLength $+ \beta_3$ Smoker $+ \mu_i$
- Model 4: $Weight_i = \beta_0 + \beta_1$ MothersAge $+ \beta_2$ TermLength $+ \beta_3$ Smoker $+ \beta_4$ NotWhite $+ \mu_i$

- Model 5: $Weight_i = \beta_0 + \beta_1$ MothersAge + β_2 TermLength + β_3 Smoker + β_4 NotWhite + β_5 BabyGender + μ_i
- Model 6: $Weight_i = \beta_0 + \beta_1$ MothersAge + β_2 TermLength + β_3 Smoker + β_4 NotWhite + β_5 BabyGender + β_6 Premature + μ_i
- Model 7: $Weight_i = \beta_0 + \beta_1$ MothersAge + β_2 TermLength + β_3 Smoker + β_4 NotWhite + β_5 BabyGender + β_6 Premature + β_7 MaritalStatus + μ_i
- Model 8: $Weight_i = \beta_0 + \beta_1$ MothersAge + β_2 TermLength + β_3 Smoker + β_4 NotWhite + β_5 BabyGender + β_6 Premature + β_7 MaritalStatus + β_8 HospitalVisits + μ_i
- Model 9: $Weight_i = \beta_0 + \beta_1$ MothersAge + β_2 TermLength + β_3 Smoker + β_4 NotWhite + β_5 BabyGender + β_6 Premature + β_7 MaritalStatus + β_8 HospitalVisits + β_9 WeightGain + μ_i

We decided to make model 2 our base specification model because the coefficient in model 1 did not reach the desired significance level of 5% although it was statistically significant at the 10% level. In subsequent models, we would continue to add control variables to more accurately assess the MothersAge coefficient. The dummy variables (β_3 - β_7) were constructed using ifelse functions in R, where the value of Smoker was 1 if the mother was a smoker and 0 otherwise; the value of NotWhite was 1 if the mother was nonwhite and 0 otherwise; the value of BabyGender was 1 if the sex of the baby was male and 0 if female; the value of Premature was 1 if the baby was classified as premature and 0 if full-term; the value of MaritalStatus was 1 if the mother was married at the time of birth and 0 otherwise. The variables TermLength (length of pregnancy in weeks), HospitalVisits, and WeightGain (weight gained by mother during pregnancy) are discrete nonbinary control variables. We have purposefully excluded the father's age (fage), maturity status of mother (mature), and classification of birthweight (lowbirthweight) from the regression models. Previous studies are unclear on whether advanced paternal age (> 40 years of age) has any influence on birthweight, there is however an increase in genetic risk as men age (Wang et al., 2020). The maturity status of the mother was determined via age threshold, with mothers greater than or equal to the age of 35 classified as mature, and not mature otherwise. Seeing as how this is just a grouping of age, the variable would just overlap with the existing mage variable and would provide no further insight into the variable of interest. Similar logic can be applied to the lowbirthweight variable.

Results

Regression of Maternal Age on Birthweight

	Dependent variable:				
	weight				
	(1)	(2)	(3)	(4)	(5)
mage	0.0134^{*}	0.0199^{***}	0.0181^{***}	0.0136^{**}	0.0135^{**}
	(0.0077)	(0.0057)	(0.0057)	(0.0058)	(0.0057)
weeks		0.3457^{***}	0.3461^{***}	0.3409^{***}	0.3396^{***}
		(0.0120)	(0.0120)	(0.0120)	(0.0118)
smoker			-0.3230^{***}	-0.3471^{***}	-0.3699^{***}
			(0.1062)	(0.1056)	(0.1042)
notwhite				-0.2982^{***}	-0.3130^{***}
				(0.0788)	(0.0778)
babymale					0.3741^{***}
					(0.0688)
Constant	6.7400^{***}	-6.6855^{***}	-6.6104^{***}	-6.2031^{***}	-6.3284^{***}
	(0.2126)	(0.4917)	(0.4903)	(0.4971)	(0.4906)
Observations	1,000	998	998	996	996
R^2	0.0030	0.4558	0.4608	0.4683	0.4837
Adjusted R^2	0.0020	0.4547	0.4592	0.4661	0.4811
Residual Std. Error	1.5073 (df = 998)	1.1124 (df = 995)	1.1078 (df = 994)	1.0996 (df = 991)	1.0840 (df = 990)
F Statistic	3.0354^{*} (df = 1; 998)	416.6438^{***} (df = 2; 995)	283.1504^{***} (df = 3; 994)	218.1735^{***} (df = 4; 991)	185.4817^{***} (df = 5; 990)
Note:			*p<0.1; **p<0.05; ***p<0.01		

Regression of Maternal Age on Birthweight

	Dependent variable:			
	weight			
	(6)	(7)	(8)	(9)
mage	0.0125**	0.0060	0.0054	0.0063
	(0.0056)	(0.0062)	(0.0062)	(0.0063)
weeks	0.2844***	0.2836***	0.2806***	0.2807***
	(0.0172)	(0.0172)	(0.0178)	(0.0181)
smoker	-0.3674***	-0.3437***	-0.3395***	-0.3730***
	(0.1033)	(0.1034)	(0.1041)	(0.1051)
notwhite	-0.3132***	-0.2541***	-0.2522***	-0.2128***
	(0.0771)	(0.0803)	(0.0809)	(0.0816)
babymale	0.3935***	0.3895***	0.3947***	0.4019***
	(0.0683)	(0.0682)	(0.0686)	(0.0691)
premature	-0.6104***	-0.5967***	-0.6063***	-0.5309***
	(0.1402)	(0.1399)	(0.1410)	(0.1435)
married		-0.2091**	-0.2226***	-0.2271***
		(0.0818)	(0.0834)	(0.0840)
visits			-0.0028	-0.0051
			(0.0090)	(0.0092)
gained				0.0083***
				(0.0025)
Constant	-4.0992***	-3.8342***	-3.6677***	-3.9399***
	(0.7061)	(0.7117)	(0.7314)	(0.7510)
Observations	996	996	988	962
R^2	0.4934	0.4967	0.4835	0.4803
Adjusted R^2	0.4903	0.4932	0.4793	0.4754
Residual Std. Error	1.0743 (df = 989)	1.0714 (df = 988)	1.0733 (df = 979)	1.0660 (df = 952)
F Statistic	160.5317*** (df = 6; 989)	139.3018*** (df = 7; 988)	114.5539*** (df = 8; 979)	97.7484*** (df = 9; 952)
Note:				*p<0.1; **p<0.05; ***p<0.01

Confidence Interval of Base Specification Model (Model 2) at 95% Significance Level: (0.0088, 0.0311)

Discussion

There are many takeaways that can be drawn from our regression models. For starters, the small but positive coefficient of maternal age (mage) trended downwards from our base specification model (0.0199) to near-zero (0.0063), losing statistical significance from model 7 onward. This can be explained in several ways. In preceding studies, the relationship between maternal age and birthweight was observed to be non-linear, where the ends above and below the "typical" maternal age range (20-40) suffered from lower birthweights, forming a U-shaped distribution. Because we are using a linear regression model to estimate an ostensibly nonlinear relationship, the regression line attempts to best fit the data and thus the observed coefficients one would see in a non-linear model effectively cancel out to near-zero in the linear model. It could also be that as more control variables are added to regression, the coefficient of maternal age converges to its true value of 0, indicating that maternal age has little effect of birthweight, but this is very unlikely considering the findings of other studies and the lack of statistical significance. It should also be noted that the adjusted R^2 reaches a maximum of 0.4932, after the inclusion of every dummy variable of interest but before the 8^{th} and 9^{th} models' inclusion of number of hospital visits and weight gained during pregnancy. Furthermore, adding the number of hospital visits to the regression seemed to have no statistical relevance in helping to explain variation in our regression model. This comes as a surprise, because it has been shown that access to prenatal care reduces the likelihood of low birthweight among all women (Committee to Study the Prevention of Low Birthweight, 1985) and so we would expect to see a strong positive relationship between the regressor and the regressand. The intercept in our model does not follow a logical interpretation as it is a negative number (indicating a negative birthweight, holding all variables constant), which doesn't make sense here. Seeing as how the regression coefficient on maternal age loses significance as more variables are introduced, we construct a confidence level with 95% significance using the base specification model. In answering our research question, "What role does maternal age play in determining a child's birthweight?" We cannot provide a sound answer using the results from our analysis due to the complexity of the relationship as it cannot be properly estimated using a linear model. More testing is needed to further investigate this issue.

Conclusion

This paper attempts to study the relationship between maternal age and birthweight to better understand how it plays a role in low birthweight and by extension, risk of infant mortality. Our research question, "What role does maternal age play in determining a child's birthweight?" guides us in the design and analysis of our data. Using multivariable linear regression, we found a small but positive relationship between maternal age and birthweight. The magnitude of this finding is relatively low, but statistically significant in 5 of 9 of our regression models, at a p-value < .05. In designing our statistical model, we used several control variables to improve the adjusted R-squared value and improve the accuracy of the coefficient on the variable of interest. Still, there was a maximum adjusted R-squared value of 0.4932, indicating that there are omitted variables that could be included to better explain the variation in birthweight. Because of our limited data set, we were unable to account for omitted dummy variables such as quality of prenatal care, socioeconomic or medical risk factors, which could help draw a more complete picture. Furthermore, due to the sample size of the population and geographic setting, the statistical analysis has reduced external validity. In addition, the methodology that we used to estimate the relationship is inherently flawed: we used a linear regression to model a non-linear relationship, and thus our study has reduced internal validity as well. Given the sophistication involved with explaining variation of birthweight, we suggest that the United States continue to focus on healthcare policy involving prenatal care and education on the risk factors that are known to reduce birthweight, such as smoking and inadequate access to quality prenatal care. Birthweight is a complex subject that consists of various moving parts and interactions, and it will take many iterations of modeling to best capture the effects of any singular variable.

References

Committee to Study the Prevention of Low Birthweight. (1985, January 1). *The significance of low birthweight*. Preventing Low Birthweight. https://www.ncbi.nlm.nih.gov/books/NBK214473/

North Carolina births, 1000 cases. Data Sets. (n.d.). https://www.openintro.org/data/index.php?data=ncbirths

Rich-Edwards, J. W., Buka, S. L., Brennan, R. T., & Earls, F. (2003). Diverging associations of maternal age with low birthweight for black and white mothers. *International Journal of Epidemiology*, *32*(1), 83–90. https://doi.org/10.1093/ije/dyg008

Wang, S., Yang, L., Shang, L., Yang, W., Qi, C., Huang, L., Xie, G., Wang, R., & Chung, M. C. (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). https://doi.org/10.1186/s12884-020-03445-2

Zheng, W., Suzuki, K., Tanaka, T., Kohama, M., & Yamagata, Z. (2016). Association between maternal smoking during pregnancy and low birthweight: Effects by maternal age. *PLOS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0146241