Low Birth Weight Data

Aditya Badola
r0768685@student.kuleuven.be

Danie Daaboul
r0775655@student.kuleuven.be

Matthew Watson
r0766802@student.kuleuven.be

Kristen Nader
r0771801@student.kuleuven.be

Prabhat Juyal
r0772775@student.kuleuven.be

Magdalena Zielonka
r0769926@student.kuleuven.be

Michael Neilson
r0770285@student.kuleuven.be

James O'Reilly
r0773125@student.kuleuven.be

March 13, 2020

1 Introduction

Low birth weight (LBW) is defined by the World Health Organisation as a birth of an infant weighing less than 2500 grams. Low birth weight is closely associated with fetal and perinatal mortality, inhibited growth and cognitive development, and chronic diseases later in life. Physicians are therefore interested in understanding the effects of different behavioural and environmental variables on the likelihood of having a low birth weight.

The variables identified in Table 1 have been shown to be associated with low birth weight in the obstetrical literature. The goal of this study is to ascertain if these variables were important in the population being served in the medical center where this data were collected. We then build, compare, and discuss different models for predicting the likelihood of a low weight birth given these variables.

Variable	Abbreviation
Low Birth Weight (0 or 1)	LOW
Age of the Mother in Years	AGE
Weight in pounds at the last menstrual period	LWT
Race (1 - White, 2 = Black, 3 = Other)	RACE
Smoking status during pregnancy (Yes = 1, No = 0)	SMOKE
Number of previous premature labours	PTL
History of hypertension	HT
Presence of uterine irritability	UI
Number of visits to physician during first trimester	FTV
Birth weight in grams	BWT

Table 1: Variables included in the dataset.

Review of the Obstetrical Literature

Reviewing the obstetrical literature allows us to understand how these variables and their interactions should influence the likelihood of a low birth weight. This information can then be compared with the results from an exploratory analysis of the data, allowing us to determine how our data differs from what would be expected and to make informed decisions when building predictive models.

1.1 Effects of variables on low birth weight

A brief description of the effect of each variable on low birth weight is given below.

- **AGE:** Fraser and Brockert [1995] showed that teenage mothers have a significantly higher

risk than mothers who were 20 to 24 years of age of delivering an infant who had low birth weight.[8] However, advancing maternal age is associated with a decreased potential for fetal growth and it has been shown that the adjusted risk for low birth weight at term is the lowest in younger mothers and increases with advancing maternal age.[14]

- **SMOKE:** Smoking during pregnancy inhibits full fetal development and is closely associated with low birth weight.[3] However, the degree to which smoking affects low birth weight varies significantly with the average number of cigarettes smoked per day, with birth weight decreasing as the category of cigarette number per day increases.[11] This study uses a binary SMOKE variable - a more detailed categorisation of smokers would most likely improve the quality of the models.

- **RACE:** There is a higher rate of LBW among black women when compared with white women.[9] Race effectively acts as a proxy variable - encoding information about sociodemographic, health-related, and behavioural differences between different racial groups.[10]

- **LWT:** Progressive increase in pre-pregnancy weight was paralleled by progressive increase in mean birth weight and decrease in the incidence of low weight births.[5]

- **PTL:** The risk of the birth of a subsequent low birthweight infant is 2 to 5 times higher than average for mothers who have had a previous low birthweight delivery and increases with the number of prior low-weight births.[1]

- **HT:** Both chronic and pregnancy induced hypertension are strongly associated with retarded fetal growth and pre-term birth. As a result, maternal hypertension is strongly associated with low infant birthweight. [16][6]

- **FTV:** Increasing the number of physician visits can lead to earlier and more accurate diagnosis of LBW risk factors. Many studies have established a link between date of the first visit, total number of visits and length of pregnancy and LBW, a relation that is stronger if the first visit is delayed or if the number of visits is smaller than normal.[12][4][2]

- **UI:** The incidence of preterm labour in women with uterine irritability is not as frequent as in patients with other high-risk factors. However, preterm labour does occur in patients with uterine irritability at a rate higher than that in the general obstetric population.[13]

1.2 Interactions between variables

An interaction describes a situation in which the effect of one non-dependent variable on 'LOW' is dependant on the state of a second non-dependent variable. This is equivalent to saying the effects of the the non-dependent variables are not additive. Listed below are the interactions which are most relevant according to the obstetrical literature.

- **SMOKE/AGE -** Older mothers, who are already at risk of giving birth to low birthweight infants, might be more susceptible to the effects of maternal smoking. [reference] determined that maternal age has a modifying effect on the association between maternal smoking and birthweight. The association between low birthweight and age increases with maternal age.[17]

- **SMOKE/RACE -** The LBW risk difference associated with smoking is greater among black women. This might be due to the fact that even if black women smoked significantly fewer cigarettes per day, they might have higher cotinine levels compared to white women. [7]

- **AGE/FTV -** For adult first-time mothers, fewer than 10 prenatal care visits was the only significant factor affecting LBW. Therefore, health education programs or prenatal care aimed at preventing LBW should be tailored according to the age of pregnant mothers. The effect of FTV on LBW is stronger for younger mothers.[15]

- **UI/PTL -**

2 Structure of the Dataset

A total of 189 subjects took part in the study. There are no missing data. In this study, there is a single binary response variable "LOW", where

$$\text{LOW} = \begin{cases} 1, & \text{if Birth weight} < 2500\text{g} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

There are three continuous predictor variables (AGE, LWT, FTV), and five categorical predictor variables (RACE, SMOKE, PTL, HT, and UI).

3 Exploratory Analysis

An exploratory analysis is performed, first investigating the correlation between the predictor variables and the response variable, and then checking for meaningful interactions between predictor variables. The reasons for these analyses are twofold:

- Investigate whether the associations and correlations in our data are consistent with the obstetrical literature.

- As a guideline for which main effects and interactions should be included in the candidate models.

Note that when we calculate correlation, we're not making any inferences about causality. We are instead using it to compare the associations in our data with those in the obstetrical literature to see if our data is consistent. Also a significant correlation with low birthweight gives us reason to strongly consider these variables as main effects in our candidate models.

3.1 Correlation with LBW

The Spearman correlation coefficient was calculated between the dependant variable (BWT) and each of the continuous variables - AGE, LTW, FTV, and PTL. The Spearman correlation was chosen due to the non-normal distribution of each of these variables, as determined by a series of Shapiro-Wilks tests. The Spearman correlation coefficients and related p-values are given in Table 2. Both LWT and PTL are significantly correlated with BWT, with p-values of 0.001 and 0.005 respectively. They are therefore strong candidates to consider as parameters in our candidate models.

	AGE	LWT	FTV	PTL
ρ	0.061	0.248	0.07	-0.204
p-value	0.404	0.001	0.338	0.005

Table 2: Spearman correlation results and their respective p-values.
The highlighted cells indicate statistically significant results.

Box plots were used to identify how the non-continuous variables in the dataset were associated birth weight and are therefore strong candidates to consider as parameters in our candidate models (See Section 1 and Section 2).
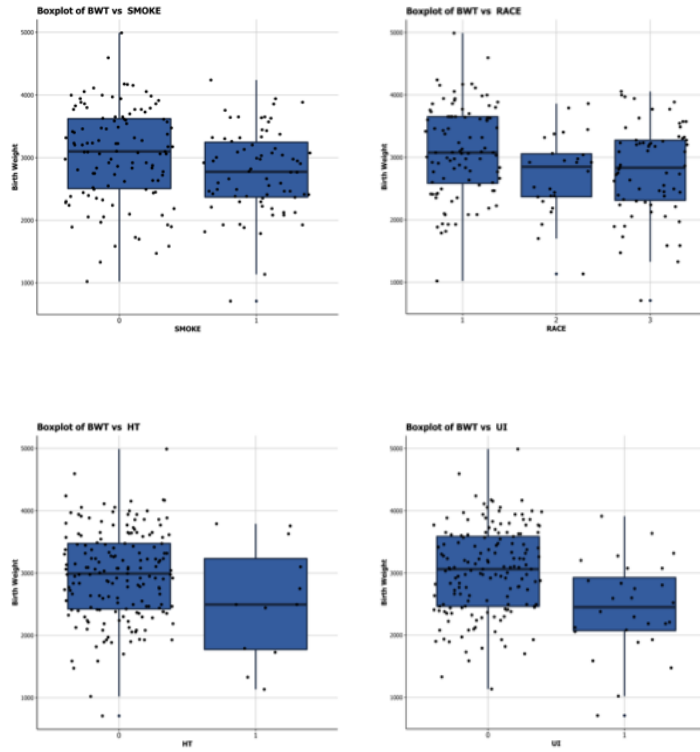

Figure 1: Box plots for each of the non-continuous variables against birthweight.

BWT vs RACE: The mean birth weight for black women and the 'other' group are both slightly lower than for white women (200 g). These results are consistent with the obstetrical literature.

BWT vs SMOKE: Note that the mean birth weight for smokers is 2700g and the mean for non-smokers is 3100g, showing the negative association between smoking and birth weight in our dataset.

BWT vs HT: The mean birth weight for pregnant women with hypertension is lower than the

mean birth weight of women with no hypertension history. It is important to note that only a few women in our dataset have a history of hypertension, nonetheless the association is consistent with the literature.

BWT vs UI: Note the 600g difference in mean birth weight between women without UI (3100g) and those with UI (2500g). This is again consistent with the literature and suggests that UI should be strongly considered as a main effect in the candidate models.

3.2 Interactions Between Sets of Variables

Drawing from the interactions found in the obstetrical literature, we investigate whether these interactions are also relevant in our dataset. The interactions are visualised int he plots below.
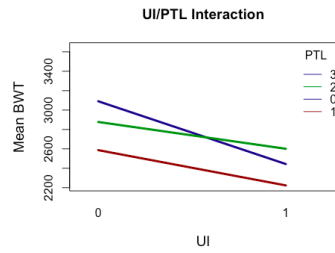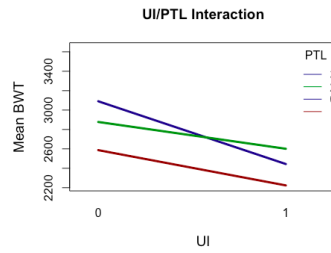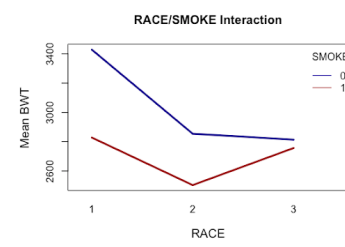

Figure 2

Figure 3

Figure 4

INSERT INTERACTION EXPLANATION FROM DOCS HERE!!

4 Question 1

The four variables which are considered the most clinically important are the AGE, LWT, RACE, and FTV. Using these variables, a model building exercise is carried out to determine the best model to describe the data using the AIC and AIC weights.

4.1 Model Building

There is a trade-off between accuracy and interpretability when deciding on the number of interaction terms to include in these models. The principle of constructing hierarchical models says that if we include an interaction term, the main effect variables should also be included in that interaction. The interaction terms included in the candidate model were the salient interactions determined by the literature review and exploratory analysis.

First we consider a model with the main effects only:

$$RACE + FTV + AGE + LWT \tag{2}$$

Then a model with all interaction terms:

$$RACE + FTV + AGE + LWT + LWT/FTV + AGE/FTV \tag{3}$$

Then Models with each singular interaction term:

$$RACE + FTV + AGE + LWT + AGE/FTV \tag{4}$$

$$RACE + FTV + AGE + LWT + LWT/FTV \tag{5}$$

The final two candidate models retain AGE/FTV as an interaction term because we have seen that this interaction term is most likely significant through our exploratory data analysis and through model testing. These models examine the effect of adding both LWT and RACE individually as main effects.

$$RACE + FTV + AGE + LWT + AGE/FTV \tag{6}$$

$$FTV + AGE + LWT + AGE/FTV \tag{7}$$

4.2   Comparing the Models

The models are evaluated against each other using the Akaike Information Criterion (AIC). The results for the models are shown in Table 3.

Model	AIC Score	ΔAic	AIC Weight
$LOW \sim AGE + RACE + FTV + LWT$	235.03	9.34	0.01
$LOW \sim AGE + RACE + FTV + LWT + AGE : FTV + LWT : FTV$	228.79	3.09	0.12
$LOW \sim AGE + RACE + FTV + LWT + AGE : FTV$	227.24	1.54	0.26
$LOW \sim AGE + RACE + FTV + LWT + LWT : FTV$	236.64	10.94	0.00
$LOW \sim AGE + FTV + LWT + AGE : FTV$	225.70	0.00	0.57
$LOW \sim AGE + RACE + FTV + AGE : FTV$	231.34	5.65	0.03

Table 3: Comparison of candidate models using AIC. The best performing model is highlighted.

Based on the AIC weights, we determine that the model 5 is the best model from the list of candidate models. The final model including coefficients is given by:

$$\mathbb{E}(LOW) = -0.166 + 0.0658 \star AGE + 2.844 \star FTV - 0.0155 \star LWT - 0.131 \star Age : FTV \tag{8}$$

A final model was also constructed using the backward elimination procedure. Starting with the full model given by equation 3, the variable with the highest p-value controlling the type 1 error

at $\alpha = 0.1$ was iteratively removed. Both $FTV : LWT$ and $RACE$ were removed using this procedure, resulting in the final model:

$$\mathbb{E}(LOW) = -0.166 + 0.0658 \star AGE + 2.844 \star FTV - 0.0155 \star LWT - 0.131 \star Age : FTV \quad (9)$$

Both the AIC and backward elimination procedures give the same final model.

4.3 Interpreting the Final Models

5 Question Two

5.1 Determining Interaction Terms

5.2 Model Building

5.3 Comparing the Models

5.4 Interpreting the Final Models

References

[1] BAKKETEIG, L. S., HOFFMAN, H. J., AND HARLEY, E. E. The tendency to repeat gestational age and birth weight in successive births. *American journal of obstetrics and gynecology 135*, 8 (1979), 1086–1103.

[2] BRATTON, S. L., SHOULTZ, D. A., AND WILLIAMS, M. A. Recurrence risk of low birthweight deliveries among women with a prior very low birthweight delivery. *American journal of perinatology 13*, 03 (1996), 147–150.

[3] CHIOLERO, A., BOVET, P., AND PACCAUD, F. Association between maternal smoking and low birth weight in switzerland: the eden study. *Swiss medical weekly 135*, 35-36 (2005), 525–530.

[4] DESJARDINS, E., AND HARDWICK, D. How many visits by health professionals are needed to make a difference in low birthweight? a dose-response study of the toronto healthiest babies possible program. *Canadian journal of public health 90*, 4 (1999), 224–228.

[5] EASTMAN, N. J., AND JACKSON, E. Weight relationships in pregnancy: I. the bearing of maternal weight gain and pre-pregnancy weight on birth weight in full term pregnancies. *Obstetrical & gynecological survey 23*, 11 (1968), 1003–1025.

[6] FANG, J., MADHAVAN, S., AND ALDERMAN, M. The influence of maternal hypertension on low birth weight: differences among ethnic populations. *Ethnicity & disease 9*, 3 (1999), 369–376.

[7] FICHTNER, R. R., SULLIVAN, K. M., ZYRKOWSKI, M. C. L., AND TROWBRIDGE, R. F. L. Reports on selected racial/ethnic groups special focus: Maternal and child health racial/ethnic differences in smoking, other risk factors, and low birth weight among low-income pregnant women, 1978-1988.

[8] FRASER, A. M., BROCKERT, J. E., AND WARD, R. H. Association of young maternal age with adverse reproductive outcomes. *New England journal of medicine 332*, 17 (1995), 1113–1118.

[9] HESSOL, N. A., FUENTES-AFFLICK, E., AND BACCHETTI, P. Risk of low birth weight infants among black and white parents. *Obstetrics & Gynecology 92*, 5 (1998), 814–822.

[10] KALLAN, J. E. Race, intervening variables, and two components of low birth weight. *Demography 30*, 3 (1993), 489–506.

[11] KATAOKA, M. C., CARVALHEIRA, A. P. P., FERRARI, A. P., MALTA, M. B., CARVALHAES, M. A. D. B. L., AND DE LIMA PARADA, C. M. G. Smoking during pregnancy and harm reduction in birth weight: a cross-sectional study. *BMC pregnancy and childbirth 18*, 1 (2018), 67.

[12] LETAMO, G., AND MAJELANTLE, R. Factors influencing low birth weight and prematurity in botswana. *Journal of Biosocial science 33*, 3 (2001), 391–403.

[13] ROBERTS, W. F., PERRY JR, K. G., NAEF III, R. W., WASHBURNE, J. F., AND MORRISON, J. C. The irritable uterus: A risk factor for preterm birth? *American journal of obstetrics and gynecology 172*, 1 (1995), 138–142.

[14] SUN LEE, K., FERGUSON, R. M., CORPUZ, M., AND GARTNER, L. M. Maternal age and incidence of low birth weight at term: A population study. *American Journal of Obstetrics*

and Gynecology 158, 1 (1988), 84 – 89.

[15] Wang, C.-S., and Chou, P. Risk factors for low birth weight among first-time mothers in southern taiwan. *Journal of the Formosan Medical Association 100*, 3 (2001), 168–172.

[16] Xiong, X., Mayes, D., Demianczuk, N., Olson, D. M., Davidge, S. T., Newburn-Cook, C., and Saunders, L. D. Impact of pregnancy-induced hypertension on fetal growth. *American journal of obstetrics and gynecology 180*, 1 (1999), 207–213.

[17] Zheng, W., Suzuki, K., Tanaka, T., Kohama, M., Yamagata, Z., Group, O. C. H. S., et al. Association between maternal smoking during pregnancy and low birthweight: effects by maternal age. *PloS one 11*, 1 (2016).