Archit Gubiligari
Data Analysis Project

Evaluating The Relationship Between Health Risk During Pregnancy and Blood Glucose Levels and Blood Pressure

Introduction:

In recent years, there has been an increasing emphasis on prioritizing health during pregnancy to ensure positive outcomes for both parents and their infants. Pregnancy health risks pose significant challenges, making it crucial to identify and address potential complications proactively.

High blood pressure and uncontrolled glucose levels can have a significant impact on parental health during pregnancy. It is crucial for expectant parents to monitor and manage their blood pressure and glucose levels to ensure a healthy pregnancy and reduce potential risks.

Blood glucose levels have been recognized to have a great effect on the chance of health risks during pregnancy and "staying in your target range during pregnancy, which may be different than when you aren't pregnant, is also important" (NIH). This occurs due to high blood sugar harming the baby in the first couple of weeks of pregnancy before even knowing you are pregnant. This variable in association to risk level is important to analyze in this dataset.

In addition, blood pressure can be detrimental as well to the fetus during pregnancy since hypertension during pregnancy can lead to "severe or uncontrolled high blood pressure in both the pregnant individual and the fetus" (ACOG).

The purpose of this data analysis report is to examine and prioritize the key parental health risks faced by pregnant individuals. By leveraging data sources and advanced analytical techniques, I aimed to uncover patterns, correlations, and risk factors associated with adverse health outcomes.

In order to look at two relationships with risk involved, I will be focusing on the blood glucose levels of the parents and the systolic blood pressure, since I believe that measuring changes in these variables will give us the greatest accuracy in our association of health risk chance and assessing the importance of these variables in this relationship, while adjusting for age as a potential confounder.

Dataset Background:

This dataset describes the parental health risk predicted during pregnancy considering information about the other variables in the dataset. This data has been collected directly from healthcare IoT devices from Daffodil International University. These healthcare IoT devices collected data from different hospitals, community clinics, and other parental health care organizations through the IoT risk monitoring system. This dataset was then released to the public on Kaggle.com by Csafrit.

These Iot devices can record and transmit data that include the variables of this dataset and much more in the recording and measurement functioning department. "Iot devices are pieces of hardware, such as sensors, actuators, gadgets, appliances, or machines, that are

programmed for certain applications and can transmit data over the internet or other networks" (Arm).

The people studied are the patients from these different healthcare institutions who are pregnant and are measured for their blood glucose levels and blood pressure as well as the other variables. This study would be classified as an observational study. This data was collected in 2020 in Bangladesh and Singapore hospitals and contains 1014 patients. These clinics came to use these IoT devices in order to transmit more data to researchers that can try to use their analysis to reduce these health risks experienced.

Overall Scientific Goal:

To use the measurements of systolic blood pressure and blood glucose levels of pregnant individuals to assess the chance of health risks during pregnancy and determine the importance of these variables when determining chance of health risks.

Scientific Questions:
1.How do changes in the blood glucose levels affect the chance of health risks during pregnancy?
2.How do changes in systolic blood pressure affect the chance of health risks during pregnancy?

Statistical Questions:
1. Are blood glucose levels during pregnancy, when adjusting for age, associated with the odds of high risk of health risks during pregnancy?
2. Are systolic blood pressure measurements during pregnancy, when adjusting for age, associated with the odds of high risk of health risks during pregnancy?

Variables:

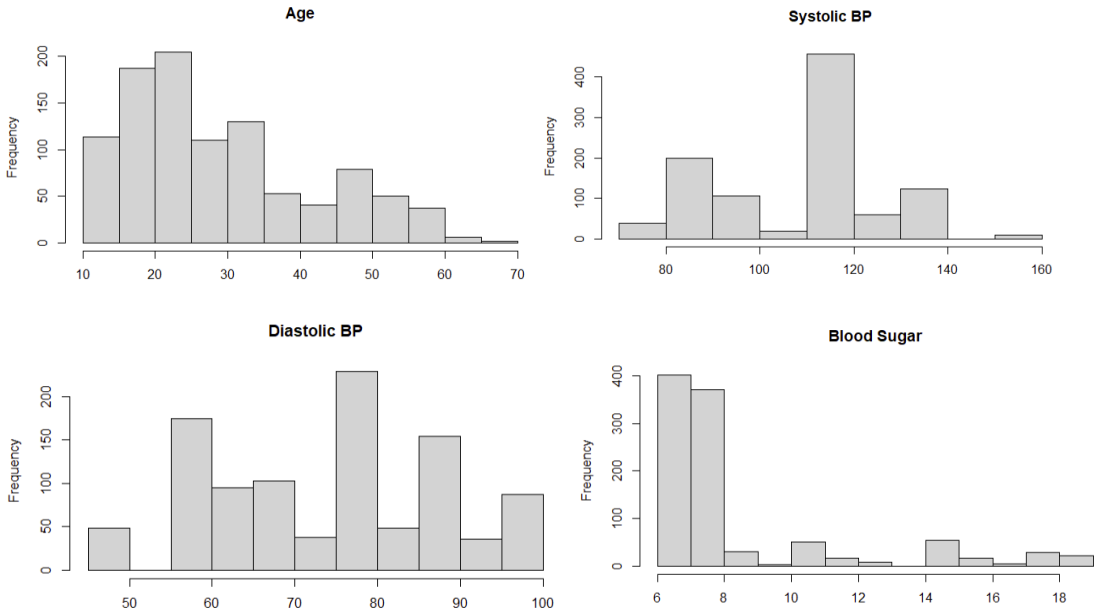
Name	Description	Type	Units	Meaning	Role
Age	Age of the patient	Quantitative	Years	Demographic	Potential Confounder
SystolicBP	Systolic Blood Pressure of patient at time of pregnancy	Quantitative	mmHg	Baseline physiology	Predictor of Interest
DiastolicBP	Diastolic Blood Pressure of patient at time of	Quantitative	mmHg	Baseline physiology	Predictor of Interest

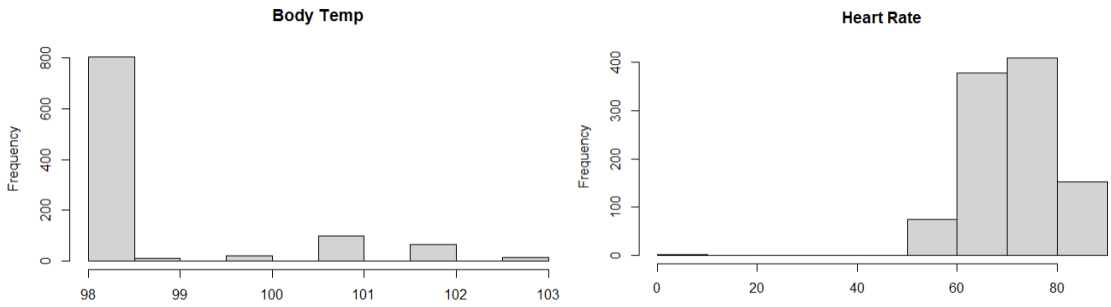
	pregnancy				
BS	Blood sugar of the patient at time of pregnancy	Quantitative	mmol/L	Baseline physiology	Predictor of Interest
BodyTemp	Body Temperature	Quantitative	Fahrenheit	Baseline physiology	Potential Confounder
Heart Rate	Heart Rate of the patient at time of pregnancy	Quantitative	Beats Per Minute	Baseline physiology	Potential Confounder
RiskLevel	Risk Level of the Patient during pregnancy for Health Risks	Categorical -> Made into Binary	High Risk or Low/Mid Level High Risk Or Not High Risk	Baseline risk factors	Outcome/Exposure

Methods:

Descriptive Analysis:

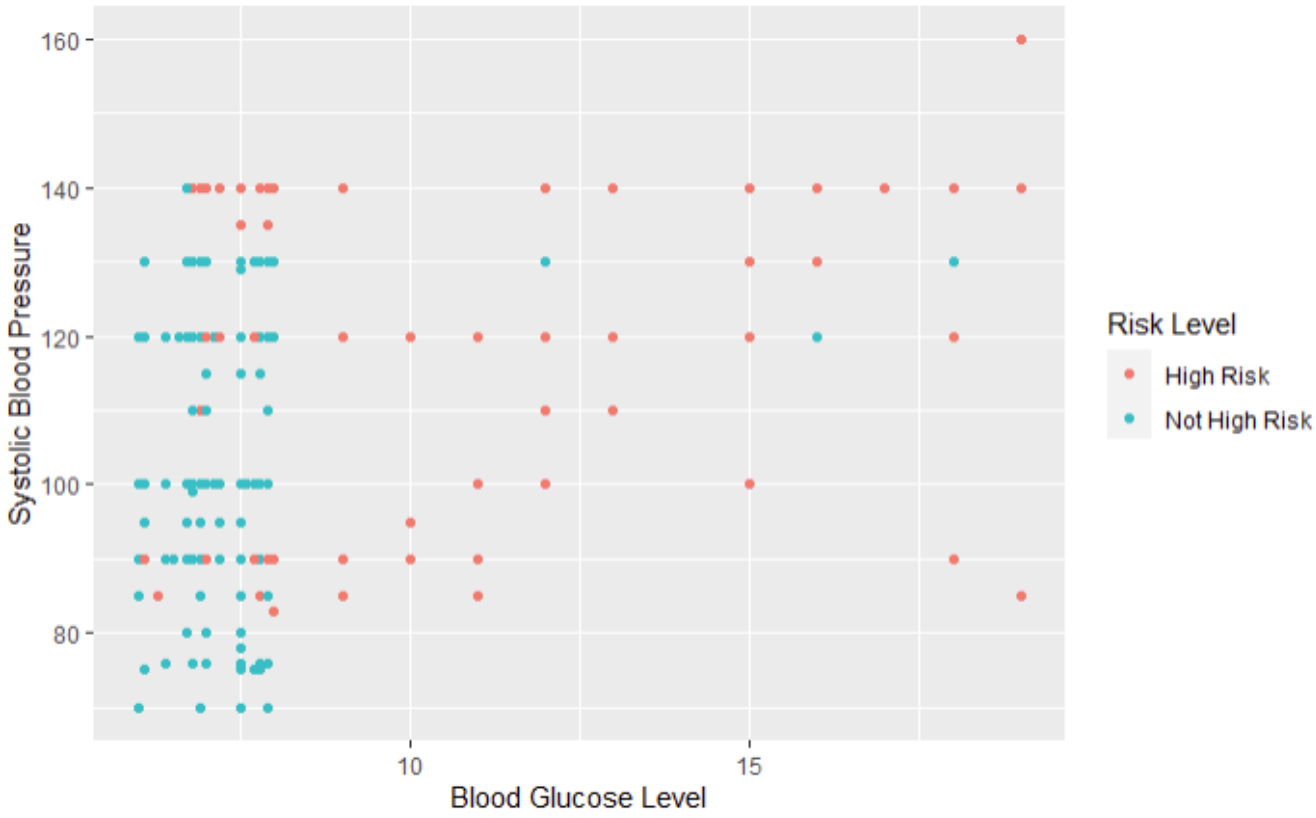
Figure 1:





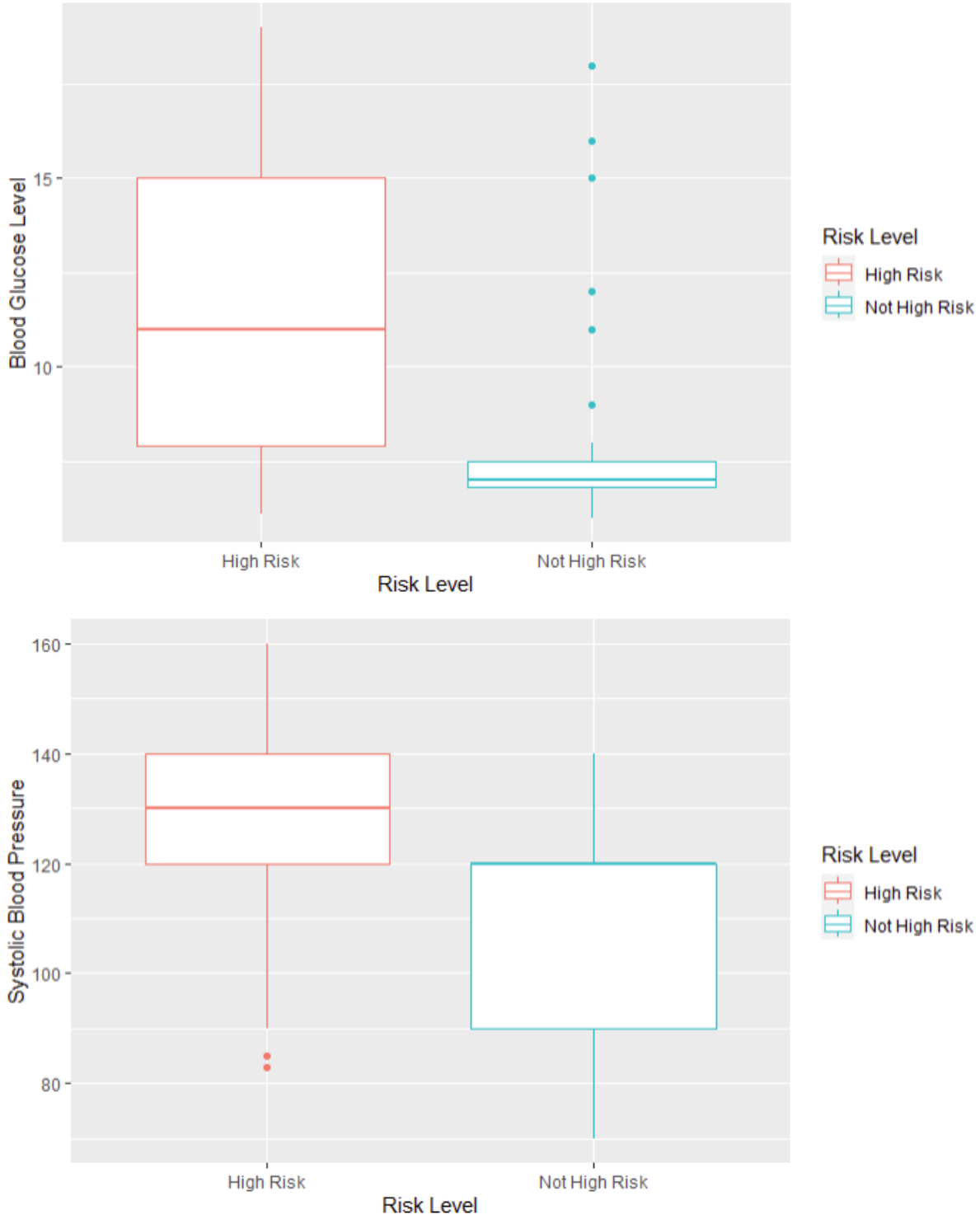
Caption: Histograms of all of the potential predictor and confounder variables included in the dataset to display the spread and variability throughout the patients.

Figure 2:



Caption: Scatterplot exhibiting Blood Glucose Level versus Systolic Blood Pressure and the chance of health risks is displayed in the legend, with red signifying high risk and blue signifying not high risk.

Figure 3:



Caption: Boxplots depicting the variables systolic blood pressure and blood sugar against the risk level outcome variable.

Methods:

Inferential Analysis:

In this part of the analysis, I ran a logistic regression model that included and adjusted for potential confounder, age of the patient measured in years. Adjusting for these potential confounders will help me make more accurate assessments of the association between the variables of systolic blood pressure and blood glucose levels on severity of health risk during pregnancy (outcome variable).

The risk level variable was converted into a binary variable ranging “High Risk” to “Not High Risk”. This allowed a completion of a logistic regression model for both of my statistical questions, which included the age variable as a potential confounder.

The values that were important in these logistic regression models are the coefficients, which estimate the log odds or odds in my model through calculations and a positive coefficient often signifies a positive correlation. The p-values in these models were very important in interpreting results, since they displayed the significance of the variables in the models and the correlation to risk level as well as the confidence intervals that I have generated for each model. If the predictor and the outcome are statistically significantly associated, a p-value of less than 0.05 will be regarded as giving strong evidence against the null hypothesis.

For the first model, I performed a logistic regression in R using the variables blood glucose levels (BS) and the binary variable risk level, adjusting for potential confounder, age.

Null Hypothesis: There is no significant association between changes in blood glucose levels and risk level for health risks during pregnancy. The regression coefficient of the BS variable is equal to 0.

Alternative Hypothesis: There is a significant association between changes in blood glucose levels and risk level. The regression coefficient of the BS variable is not equal to 0.

For the second model, I performed a logistic regression in R using the variables systolic blood pressure and the binary variable risk level, adjusting for potential confounder, age.

Null Hypothesis: There is no significant association between systolic blood pressure changes and risk level for health risks during pregnancy. The regression coefficient of the SystolicBP variable is equal to 0.

Alternative Hypothesis: There is a significant association between systolic blood pressure changes and risk level. The regression coefficient of the SystolicBP variable is not equal to 0.

Results and Interpretation:

Descriptive Analysis:

From the descriptive analysis portion, it is seen that as systolic blood pressure increases in mmHg, blood glucose levels increase in mmol/L as well. The second figure incorporates a positive correlation between the two variables, but in the context of this analysis, is considered a negative outcome since increasing both blood pressure and blood glucose levels tend to lead to a higher risk of parental health risks as explored in Figure 2.

In the third figure, it is seen that as systolic blood pressure increases in mmHg, then the risk level increases as well, which can be seen in the higher concentration of high risk in the increasing region of systolic pressure in the respective box plot. In addition, it is seen that as blood glucose levels increase in mmol/L, then the risk level increases as well, which can be seen in the higher concentration of high risk in the increasing region of blood glucose levels in the respective box plot.

Figure 4: Summary Statistics for Variables included in the MaternalHealthRiskDataset

Descriptive Analysis Table						
Variables:	**Min**	**1st Quart**	**Median**	**Mean**	**3rd Quart**	**Max**
Age (years)	10	19	26	29.87	39	70
SystolicBP (mmHg)	70	100	120	113.2	120	160
DiastolicBP (mmHg)	49	65	80	76.46	90	100
BS (mmol/L)	6	6.9	7.5	8.726	8	19
BodyTemp (F)	98	98	98	98.67	98	103
Heart Rate (bpm)	7	70	76	74.3	80	90

Risk Level:	Count (#)
High Risk	272
Mid Risk	336

Low Risk	406
Not High Risk	742

Inferential Analysis:

Model 1:

The intercept coefficient, 5.510036, signifies the baseline log-odds when both blood glucose level and age are zero. The log-odds of high risk reduce by 0.533044 for every unit higher blood glucose level when age is taken into consideration, according to the blood glucose level coefficient (-0.533044). This coefficient shows a strong association and is very significant ($p < 0.001$).

On the other hand, the coefficient for age (0.009123) indicates that the log-odds of high risk rise by 0.009123 for every one-year increment in age. Although not statistically significant ($p = 0.254$), this coefficient suggests that age may not have a substantial impact on the chance of high-risk during pregnancy in this model.

The area under the receiver operating characteristic (AUC-ROC) curve was computed to be 0.846. This measures the model's capacity to differentiate between high risk and not high-risk individuals. An AUC-ROC value closer to 1 signifies a more effective model in this regard.

In conclusion, the logistic regression model showed a significant correlation between blood glucose level and the likelihood of high-risk during pregnancy after correcting for age. Age was not discovered to be statistically relevant in this situation. The AUC-ROC value of 0.846 indicated a moderate ability of the model in distinguishing high risk from not high-risk individuals.

Model 2:

The intercept coefficient of 7.439318 represents the baseline log-odds when both systolic blood pressure and age are zero. Systolic blood pressure's coefficient (-0.047093) indicates that when age is taken into account, the log-odds of being at high risk fall by 0.047093 for each unit increase in systolic blood pressure. Strong influence is indicated by the coefficient's very significant relationship ($p < 2e\text{-}16$).

Similar to this, when accounting for systolic blood pressure, the coefficient for age (-0.028755) shows that the log-odds of being at high risk reduce by 0.028755 for every year of age increase. This coefficient, which emphasizes how age affects the probability of high risk during pregnancy, is also statistically significant ($p = 1.33e\text{-}06$). The area under the receiver operating characteristic (AUC-ROC) curve, computed as 0.7613.

In conclusion, the second logistic regression model, while controlling for age, reveals strong evidence of a connection between high risk pregnancy and systolic blood pressure with age having a significant influence on the risk level. With an AUC-ROC value of 0.7613, the model demonstrates the ability to classify individuals as either high risk or not high risk.

Assumptions:

For logistic regression analyses, the independence and sample size assumptions are of critical importance. The independence assumption asserts that the observations are mutually independent. The sample size assumption states that the sample size must be sufficient to guarantee accurate parameter estimations and trustworthy inference.

Since patient data is protected under HIPAA laws, the data is known to be independent, which satisfies the assumption of independence.

When looking at the amount of patients in the sample size being 1014, the sample size is greater than 30, which satisfies the assumption of sample size and normality.

Discussion:

This data analysis predominantly aims to comprehend the impacts of blood glucose levels and systolic blood pressure on the probability of detrimental health effects during pregnancy while adjusting for the confounder of age.

Performing a logistic regression analysis for blood glucose level, accounting for age, revealed a reverse correlation between blood glucose level and the probability of heightened risk during pregnancy. This infers that elevated blood glucose levels correlate with a diminished likelihood of high-risk pregnancies. In a surprising twist, the role of age didn't leave a prominent confounding effect in the models, leading us to contemplate that its significance in shaping the prediction of high-risk pregnancies may not be as substantial as I initially conjectured.

When I incorporated systolic blood pressure, adjusting for age, the statistical outcomes of our logistic regression analysis revealed an inverse bond between systolic blood pressure and the likelihood of high-risk pregnancies. This insinuates that the journey towards high-risk pregnancies could be decreased as systolic blood pressure rises. The age variable, in a fascinating turn of events, exhibited a significant inverse correlation with high risk. This suggests that age could contribute to a descending gradient in risk level.

There are many potential constraints that come with the data analysis techniques I used in this research. The influence of potential confounding variables not included in the dataset may sway the observed associations. The sample employed might not fully represent the broader populace, thus constraining the generalizability of the outcomes. Further, potential measurement inaccuracies, underreporting of incidents, or individuals providing incorrect responses could bias the results.

The findings suppose that both blood glucose levels and systolic blood pressure can wield influence over the odds of encountering adverse health scenarios during pregnancy. However, extensive research and consideration of other factors are necessary to fully understand the complexities of high-risk pregnancies and inform clinical protocols.

In conclusion, these discoveries provide essential insights, but they should be interpreted cautiously considering the limitations of the data and the analysis. Future investigations should aim to tackle these limitations and examine additional confounding variables to enhance our understanding of the factors influencing the risk of adverse health outcomes during pregnancy.

Works Cited

Ltd., Arm. "What Are IOT Devices." *Arm*, www.arm.com/glossary/iot-devices#:~:text=IoT%20devices%20are%20pieces%20of,the%20internet%20or%20other%20networks. Accessed 17 May 2023.

"Preeclampsia and High Blood Pressure during Pregnancy." *ACOG*, www.acog.org/womens-health/faqs/preeclampsia-and-high-blood-pressure-during-pregnancy#:~:text=If%20blood%20pressure%20goes%20up,placental%20abruption%2C%20and%20cesarean%20birth. Accessed 17 May 2023.

"Pregnancy If You Have Diabetes - NIDDK." *National Institute of Diabetes and Digestive and Kidney Diseases*, www.niddk.nih.gov/health-information/diabetes/diabetes-pregnancy. Accessed 17 May 2023.