Preventative Maternal Healthcare in Rural Bangladesh

Olivia McCauley, Marlon Fu, Parker Brailow, Kenenth Lam

1 Introduction

While modern medical practices have significantly reduced the risk of maternal mortality, low income women in rural areas face barriers to prenatal care. Transportation and cost can make healthcare near impossible to access. However, portable devices that measure vital signs like blood pressure and glucose have become increasingly accessible to patients, while mobile devices make it easier to communicate that information to medical providers, and could improve healthcare access.

The affordability of blood glucose monitors is promising in identifying high-risk pregnancies in impoverished, isolated communities. Since high blood glucose is causally related to pre-eclampsia, preterm/still births, and other dangerous conditions, treatment can be lifesaving for both mother and child. Treating these conditions often requires healthcare many cannot access, so prevention of these conditions via early detection is the best option. Thus, it is critical to understand how blood glucose levels affect prenatal maternal mortality risk.

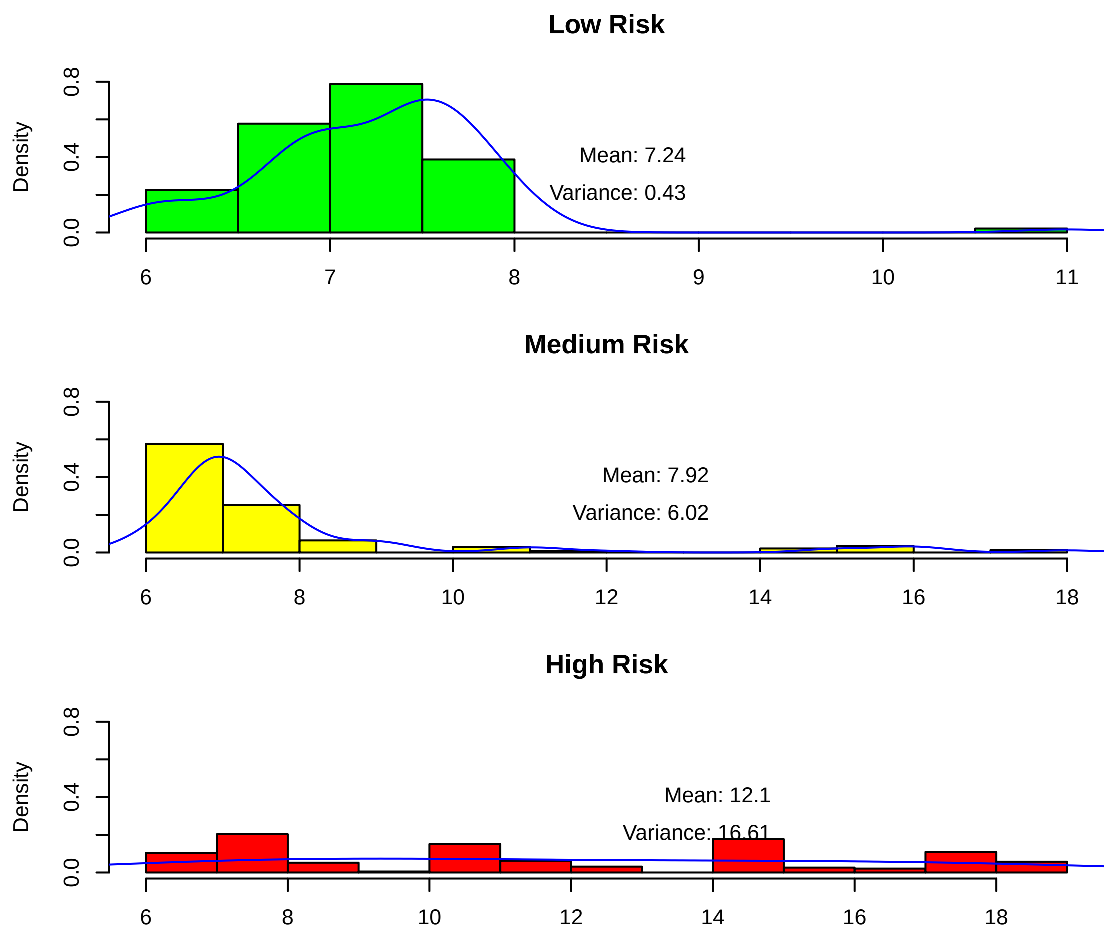
Analyzing health data from approximately one thousand pregnant women in rural Bangladesh, we investigated the effect of blood glucose on maternal mortality risk alongside other vital signs. We seek to measure the degree of reduction in blood glucose necessary to reduce risk — critical information for healthcare providers and pregnant women as they seek to carry healthy pregnancies to term.

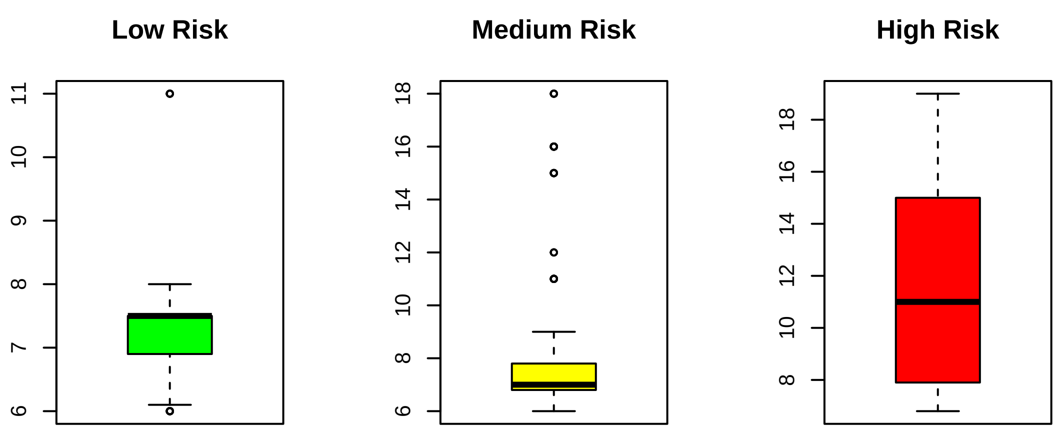
2 Data and Methodology

To operationalize a study of causality between blood sugar and mortality risk of low-income pregnant women in rural areas, we used the Maternal Health Risk dataset from the UCI ML repository. This data was collected from hospitals, clinics, and maternal healthcare centers throughout rural Bangladesh. Each row represents a pregnant patient, and includes their (Age), blood pressure ($SystolicBP$, $DiastolicBP$), blood glucose (BS), body temperature ($BodyTemp$), ($HeartRate$), and maternal mortality risk ($RiskLevel$). Although all variables are strong indicators for risk, risk is determined by a healthcare provider's holistic assessment of the patient and is not based on dataset variables alone.

No major cleaning or reformatting was required. We transformed the target variable $RiskLevel$ from a categorical to ordinal variable. Representing low to high risk on an increasing scale, we converted "low risk" to 1, "mid risk" to 2, and "high risk" to 3. Next, we extracted a 30% subsample of the data for exploratory data analysis, and reserved the remaining 70% of data to generate the report statistics.

The findings of our exploratory data analysis are below:





Using histograms with KDE estimates to analyze the distribution of blood glucose readings by risk level, we found that the expected glucose readings of low and medium risk patients were between 7 and 8 mmol/L, while the average BS readings from high risk patients was ~12 BS mmol/L. Observing that these distributions were right skewed, we applied a log-transform to the BS data, yielding a more linear relationship.

We fit three models to understand the effect of BS changes on risk level.

1. A full model involving all features available in the Maternal Health Risk Dataset.

$$\widehat{RiskLevel} = \beta_0 + \beta_1 \cdot log(BS) + \beta_2 \cdot SystolicBP + \beta_3 \cdot DiastolicBP + \beta_4 \cdot Age + \beta_5 \cdot HeartRate + \beta_6 \cdot BodyTemp$$

2. A model involving all features available in the Maternal Health Risk Dataset excluding `BS`. This model will be used in conducting an F-test to evaluate the significance of the `BS` feature.

$$\widehat{RiskLevel} = \beta_0 + \beta_1 \cdot SystolicBP + \beta_2 \cdot DiastolicBP + \beta_3 \cdot Age + \beta_4 \cdot HeartRate + \beta_5 \cdot BodyTemp$$

3. A minimal model involving only the key variable of `BS` to be used as a benchmark.

$$\widehat{RiskLevel} = \beta_0 + \beta_1 \cdot log(BS)$$

3 Results

The following table reflects results from three regression models predicting maternal mortality.

	Output Variable: Risk Level		
	(1)	(2)	(3)
log(Blood Sugar)	1.56^{***}		1.31^{***}
	(0.06)		(0.08)
Systolic Blood Pressure		0.02^{***}	0.01^{***}
		(0.002)	(0.002)
Diastolic Blood Pressure		0.01^{*}	0.0005
		(0.003)	(0.003)
Age		0.01^{**}	-0.003
		(0.002)	(0.002)
Heart Rate		0.02^{***}	0.01^{***}
		(0.003)	(0.003)
Body Temperature (F)		0.17^{***}	0.14^{***}
		(0.02)	(0.02)
Constant	-1.44^{***}	-18.17^{***}	-17.13^{***}
	(0.14)	(2.37)	(2.02)
Observations	710	710	710
R^2	0.34	0.28	0.45
Residual Std. Error	0.66 (df = 708)	0.69 (df = 704)	0.60 (df = 703)
Note:	$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$		
	HC_1 robust standard errors in parentheses.		

In both models with $log(BS)$, it was statistically significant and the most impactful variable. Without other predictors, $log(BS)$ had a large coefficient of 1.57, which decreased to 1.32 in the model with all predictors.

An F test comparing the full models with and without $log(BS)$ indicate that BS is a significant factor in risk. The coefficient of 1.32 represents the average multiplicative change in the expected value of the risk rating for a one-unit increase in the log-transformed blood sugar. For instance, if the log-transformed blood sugar increases by one unit, the expected risk rating is, on average, 32% higher than it would be with a lower log-transformed blood sugar. The next most important variable was body temperature, and the coefficient of 0.16 indicates that a 6.25 degree increase in body temperature increases risk by a full level. An increase of that size would indicate a severe fever, which would put mother and baby at extreme risk of organ failure and death, explaining the increase in risk level.

4 Limitations

Assumptions: Since our dataset is large (n=1013), we will evaluate the large sample assumptions of linear models.The first large sample assumption is that the observations are IID. As in most real-world situations, this is difficult to guarantee. Geographical clustering is a relevant risk here, as there are cultural trends in diets and lifestyle choices within communities affecting blood sugar and blood pressure. Thus, we cannot say that our data is certainly IID. However, many observational studies have these concerns, and handle this by generalizing the their conclusions only to similarly distributed populations.

The next assumption is that our dataset has a best linear predictor. Because all our features reflect vital signs, they all have finite variance – extremely high or low values would not be conducive to life. Extremely high blood pressure, heart rate, body temperature, and blood glucose would cause organ failure and death. Therefore, there must be a best linear predictor for our data set. Additionally, the lm package in R automatically drops perfectly collinear variables. Since no variables were dropped, we can conclude this assumption holds.

Omitted Variables Biases: One example of a variable omitted from this study that might bias our estimates is accessibility to skilled healthcare. Without the proper care, women might be more susceptible to higher risk levels, so we anticipate a negative correlation between the omitted variable and the key variable of blood sugar. As a result of omitting accessibility to skilled healthcare, blood sugar becomes overly influential in our model. Thus, we expect a positive omitted variable bias, driving the effect away from zero.

Ordinal Target Variable Concerns: A major limitation of our model is the mismatch between predicted and true value types. Our target variable *RiskLevel* is ordinal, so interpreting the residuals of the predicted values is challenging. Thus, when evaluating individual predictions, we can only make conclusions about the model predicting whether they are on the lower or upper end of their risk category.

5 Conclusion

This study estimated the effect of blood sugar on risk levels. Using multiple models, we found that blood sugar had a very strong effect on maternal mortality risk, and that a one-unit increase in log-transformed blood glucose units increased risk by 32%. That means that a BS increase of 8.25 mmol/L from the average low-risk blood sugar of 7.23 mmol/L, which would elevate BS to 15.48 mmol/L, would cause the pregnancy to become medium risk. Furthermore, the average blood sugar for mid-risk mothers was 7.88 mmol/L, which means that a BS increase of 8.92 mmol/L (putting BS at 16.8 mmol/L) would transform a medium risk pregnancy to a high risk one. This is a very strong effect, and was found to be statistically significant both with the p-values from model summaries and an F-test.

In the future, additional data about reproductive health and pre-existing conditions would be useful to identify collinearity between these vital signs and potentially more direct indicators of risk. More direct data about health outcomes (ie, whether the woman experienced complications, or had a successful childbirth) would help construct a stronger argument about the relationship between blood glucose and pregnancy outcomes. The ultimate hope of this line of work is to help healthcare providers and pregnant women use this information to make early lifestyle changes to improve their pregnancy outcomes.