Mental Health Mammoths

Examining Anxiety and Pregnancy During COVID-19

5/12/2025

Abstract

This study examines the impact of COVID-19-related stressors on anxiety levels among pregnant and postpartum individuals in Canada. Titled Mental Health Mammoths: Examining Anxiety and Pregnancy During COVID-19, this project analyzes data from 10,772 participants to assess whether demographic variables, such as maternal age, education, and household income, and perceived COVID-19 threats, concerns for personal and infant health, predict self-reported anxiety during pregnancy. Univariate analyses were conducted to examine the distributions of key variables, including maternal anxiety, measured using the PROMIS Anxiety scale, and perceived threat scores. Bivariate and multivariable analyses were used to identify correlations and build a preliminary multiple linear regression (MLR) model. The final model included household income, perceived threat to the mother's own life, and perceived threat to the baby, accounting for 22% of the variance in anxiety scores. These findings demonstrate that both socioeconomic status and perceived health threats significantly contribute to prenatal anxiety during the COVID-19 pandemic.

Background and Motivation

The COVID-19 pandemic introduced a range of stressors that affected pregnant individuals. Beyond the physical health risks posed by the virus, the pandemic disrupted prenatal care, increased social isolation, and amplified existing mental health challenges. These factors added pressure during a period already marked by emotional and physiological changes.

Psychological distress during pregnancy, especially anxiety, has been linked to adverse outcomes such as preterm birth, low birth weight, and developmental difficulties in children. Understanding what drives this distress is critical for supporting both maternal health and infant well-being.

Research has shown that socioeconomic factors, including maternal age, education, and income, significantly affect mental health during pregnancy. Those from lower socioeconomic backgrounds often experience heightened stress due to financial insecurity and limited access to healthcare. By analyzing how specific stressors related to COVID-19, such as fears for personal health and concerns for the health of their child, interact with socioeconomic factors, we can better understand their impact on maternal anxiety. This insight is essential for developing public health strategies, particularly during global crises, to mitigate adverse mental health outcomes for pregnant individuals.

So, our study aims to answer the following research question: How did socioeconomic factors and COVID-19-related stress impact the mental health of pregnant individuals?

Methods

Data

The data set, mental Health and Pregnancy During COVID-19, was sourced from Kaggle.com and was from the study conducted by Giesbrecht et al., published in 2021. The data includes individuals who were pregnant during the COVID pandemic and were 17 years old or older, at less than 35 weeks of gestation, currently living in Canada, and able to read and write in English or French. The data focuses on various factors related to maternal mental health, particularly anxiety and depression, as well as birth outcomes. Key variables in the data set include maternal age, household income, educational background, and perceived threats of COVID-19 to personal and infant health. In addition, the data set provides raw scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) Anxiety scale, along with birth-related outcomes such as gestational age at birth, birth weight, and NICU admission. This data set is intended to explore the effects of prenatal anxiety and COVID-19-related stressors on mental health and birth outcomes.

Variables

Response Variable

PROMIS Anxiety: The response variable we have chosen is `PROMIS_Anxiety`. This quantitative variable measures the mother's score on the Patient-Reported Outcomes Measurement Information System (PROMIS) Anxiety–Adult Short Form. The scores included in this data set are the raw scores. Scores ranged from 7 to 35 points. The higher scores correspond to elevated symptoms of anxiety.

Explanatory Variables

Maternal Education Level: The variable `Maternal_Education` is a categorical variable that measures the mother's education status. The possible categories are:

- **Less than high school**
- **High school diploma**
- **College/trade school**
- **Undergraduate degree**
- **Master's degree**
- **Doctoral degree**

Household Income: The variable `Household_Income` is a categorical variable measuring the total household income from all sources, before taxes and deductions, for all household members in 2019. The possible categories are:

- **Less than $20,000**
- **$20,000 - $39,999**
- **$40,000 - $69,999**
- **$70,000 - $99,999**
- **$100,000 - $124,999**
- **$125,000 - $149,999**
- **$150,000 - $174,999**
- **$175,000 - $199,999**
- **$200,000+**

Maternal Age: `Maternal_Age` is a quantitative variable measuring the mother's age in years when they initially enrolled in the survey.

Perceived Threat from COVID to Mother's Life: `Threaten_Life` is a quantitative variable measuring the mother's response to the question, "How much do (did) you think your life is (was) in danger during the COVID-19 pandemic?" on a 100-point scale, where 0 represents "Not at all," 50 represents "Somewhat," and 100 represents "Very much so."

Perceived Threat to Baby's Life: `Threaten_Baby_Danger` is a quantitative variable measuring the mother's response to the question, "How much do (did) you think your baby's life is (was) in danger at any time during the COVID-19 pandemic?" on a 100-point scale, where 0 represents "Not at all," 50 represents "Somewhat," and 100 represents "Very much so."

Perceived Risk of Harm to Baby: `Threaten_Baby_Harm` is a quantitative variable measuring the mother's response to the question, "How much are you worried that exposure to the COVID-19 virus will harm your baby?" on a 100-point scale, where 0 represents "Not at all," 50 represents "Somewhat," and 100 represents "Very much so."

Statistical Methods

We began our analysis with univariate exploratory data analysis to understand the distribution and summary statistics of each variable. For the quantitative variables, we examined their distributions using histograms and calculated key summary statistics (mean, median, standard deviation, interquartile range). For categorical variables, we examined their distributions using box plots and calculated proportions for each category.

For bivariate analysis, we used scatterplots to explore relationships between quantitative explanatory variables (Maternal Age, Perceived Threat from COVID to Mother's Life, Perceived Threat to Baby's Life, and Perceived Risk of Harm to Baby) and the response variable (PROMIS Anxiety score). We used boxplots to examine the relationships between categorical explanatory variables (Household Income, Maternal Education) and PROMIS Anxiety score.

Next we fit a kitchen sink model, containing all of our predictors. We performed variable selection using stepwise selection to refine our model and identify the most significant predictors. The final model included Household Income, Perceived Threat from COVID to Mother's Life, and Perceived Risk of Harm to Baby. We then checked the model's conditions, potential issues with multicollinearity and any unusual points. As a result of these checks we removed a subset of unusual points, re-checked our model's conditions using the new data set, and proceeded with interpretations.

Results

Univariate Exploratory Data Analysis

Response Variable

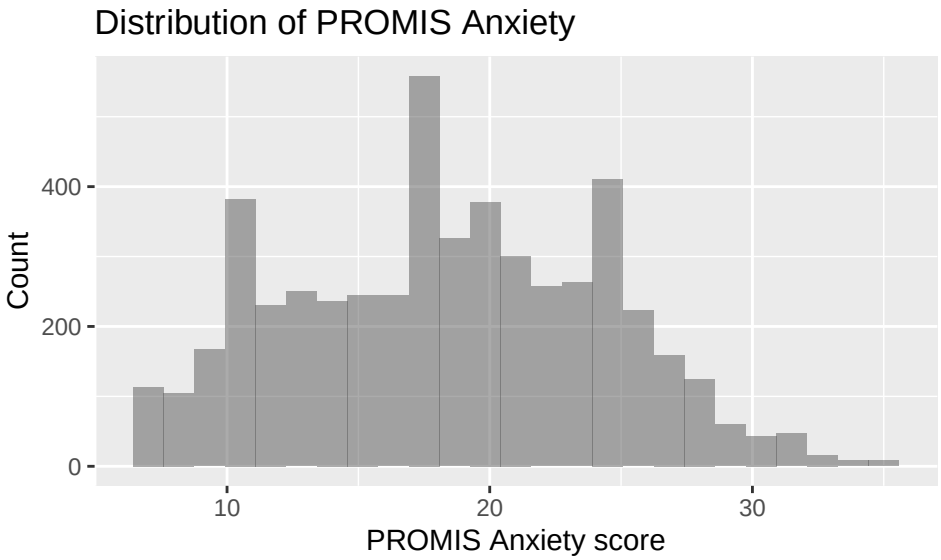

Figure 1. Univariate Response Variable Distribution

The response variable for this study was the PROMIS Anxiety score, recorded on a scale from 7 to 35. Most responses fell between 14 and 23 points, with a median score of 19. The distribution is slightly right-skewed, which is further supported by the median being slightly greater than the mean of approximately 18.7. The distribution is roughly normal, with the center mode around 17-19, though there are two additional spikes at about 10 and 25. The interquartile range (IQR) is 9, and the standard deviation is 5.94, with a mean score of 18.393.

Explanatory Variables - Qualitative

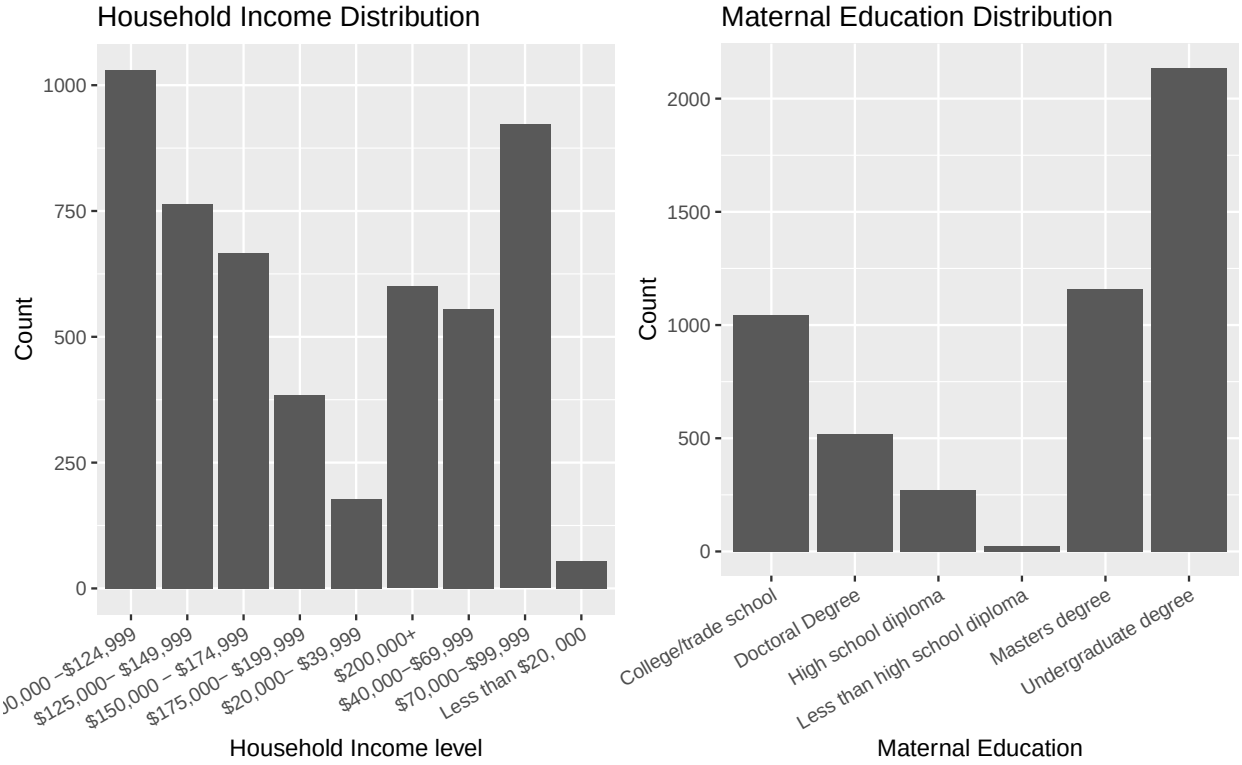

Figure 2. Original Univariate Qualitative Distributions

From the original distributions of Household Income and Maternal Education, the uneven distribution of observations is immediately evident. Household income consists of nine levels, with the distributions ranging from 54 observations in "Less than $20,000" to 1031 observations in "$100,000 -$124,999." Similarly, maternal education consists of six levels, with the distributions ranging from just 23 observations with "Less than high school diploma" to 2136 observations with an "Undergraduate Degree." This uneven distribution could potentially increase error and reduce our ability to detect differences between groups and identify general trends across the income and education levels. Furthermore the sheer number of income and education levels, nine and six respectively, adds unnecessary complexity to the model, increasing the number of predictors and making interpretations confusing. Due to these reasons we simplified household income, and maternal education into three levels.

Household Income was simplified by grouping the categories into three broader categories:

1. **Low**: Includes *Less than $20,000*, *$20,000 - $39,999*, and *$40,000 - $69,999*.
2. **Middle**: Includes *$70,000 - $99,999*, *$100,000 - $124,999*, and *$125,000 - $149,999*.
3. **High**: Includes *$150,000 - $174,999*, *$175,000 - $199,999*, and *$200,000+*.

Similarly we simplified Maternal Education by grouping the categories into three broader categories:

1. **High School or Less**: Includes *Less than high school* and *High school diploma*.
2. **College**: Includes *College/trade school* and *Undergraduate degree*.
3. **Past College**: Includes *Master's degree* and *Doctoral degree*.

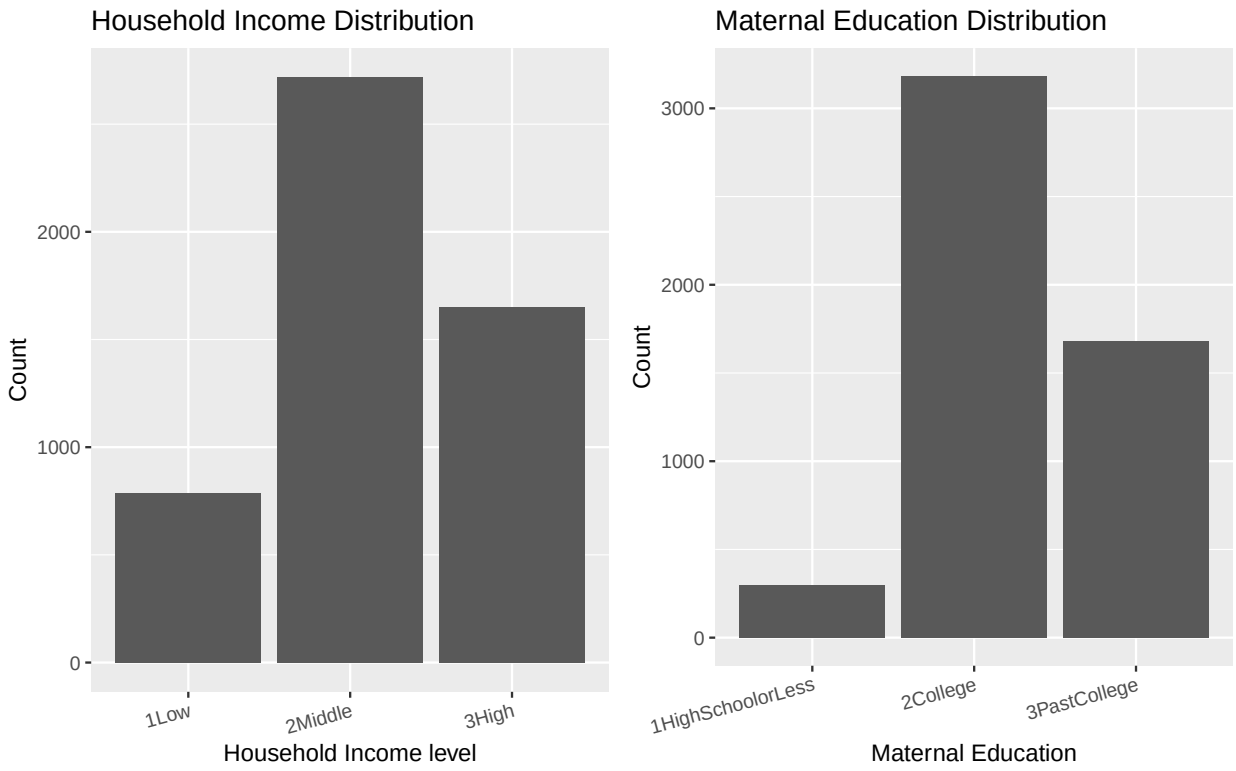

Figure 3. Augmented Univariate Qualitative Distributions

In the case of both Household Income and Maternal Education, the distributions are not perfectly symmetric. The data exhibits some skewness toward the higher levels, with more participants reporting higher income and higher education levels. For instance, there are fewer participants in the lower income and high school or less education categories, and a significant concentration in the middle and high income categories, as well as the College and Past College levels.

For Household Income, the distribution across categories across is as follows:

- Low: 787 (~15.2%) participants
- Middle: 2717 (~52.7%) participants
- High: 1652 (~32.0%) participants

For Maternal Education, the distribution across categories across is as follows:

- High School or Less: 296 (~5.7%) participants
- College: 3182 (~61.7%) participants
- Past College: 1678 (~32.5%) participants

Explanatory Variables - Quantitative

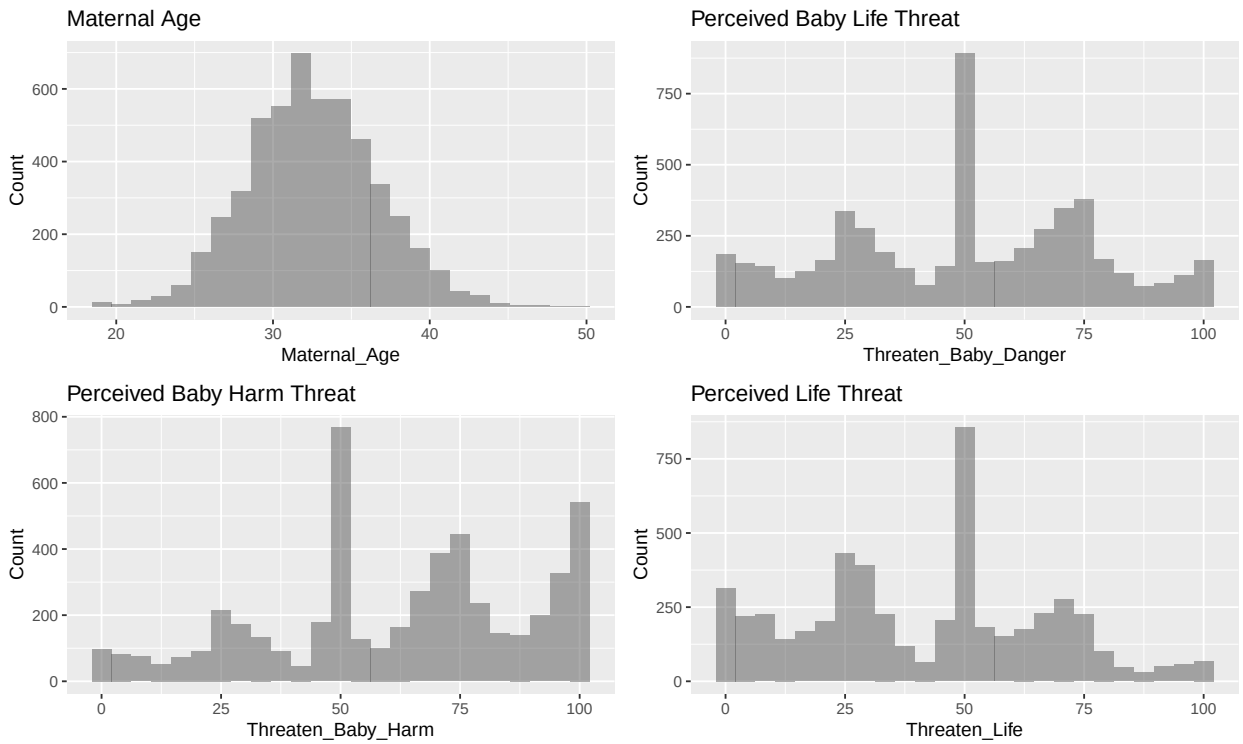

Figure 4. Univariate Quantitative Distributions

Our quantitative explanatory variables were maternal age, perceived threat to one's own life, perceived threat to the baby's life, and perceived risk of harm to the baby. Maternal age appears almost perfectly normally distributed (mean = 32.52 years, SD = 4.14). By contrast, each of the three threat-related scales shows a pronounced spike at the midpoint (50 on a 0–100 scale), suggesting possible central tendency biases Respondents may have gravitated toward the "somewhat" option by default, found the 0–100 range too complex to answer precisely, or clicked the "somewhat" point out of survey fatigue or confusion. Aside from the central spike, the perceived threat to one's own life had a median of 46 (mean = 41.42, IQR = 37, SD = 25.04), threat to the baby's life had a median of 50 (mean = 49.71, IQR = 42, SD = 26.08), and perceived risk of harm to the baby had a median of 66 (mean = 62.40, IQR = 46, SD = 27.05). The wider gap between median and mean in the perceived threat to one's own life and perceived risk of harm to the baby suggests a right skew toward higher worry levels, while the perceived threat to the baby's life central measures suggest no skew. Despite these differences, all three threat variables display very wide distributions and deviate from normality. We explored log, square-root, and power transformations to reduce skew, but these only exaggerated the spikes and failed to normalize the distributions, therefore we retained the original scales for interpretability.

Bivariate Exploratory Data Analysis

Qualitative Variables

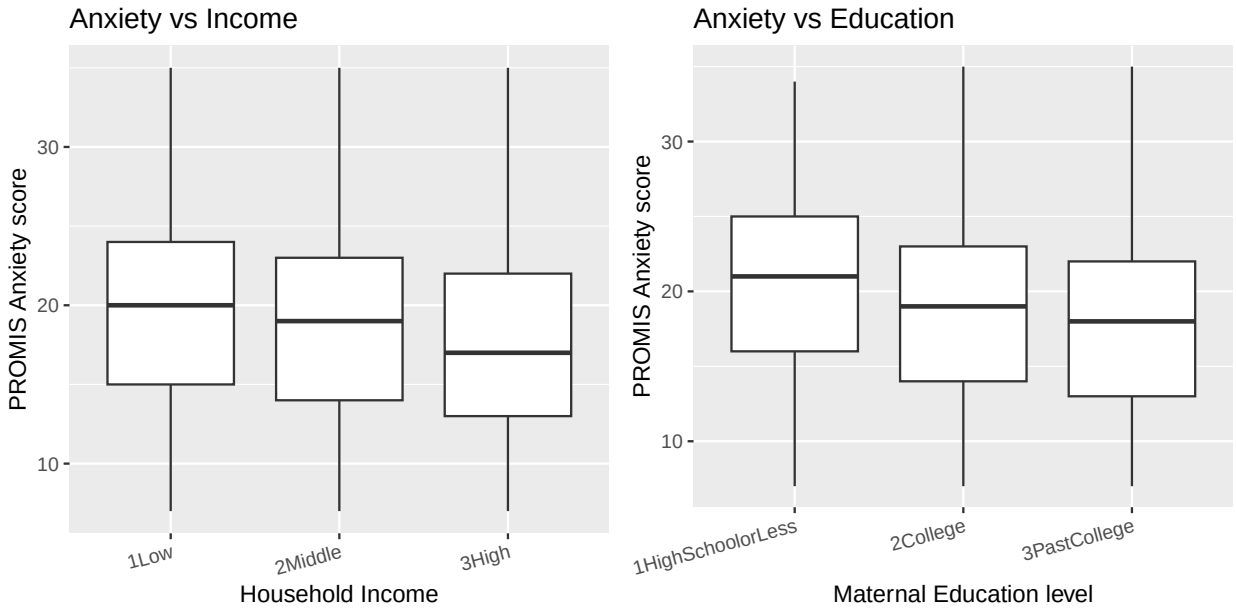

Figure 5. Bivariate Boxplots

For each of the qualitative predictor variables, Maternal Education and Household Income, we created boxplots to show the distribution of PROMIS Anxiety scores by category. The boxplot for Household Income also shows a decreasing trend in median PROMIS Anxiety scores as income increases. Participants in the Low income category (median = 20, IQR = 9) report higher anxiety levels compared to those in the Middle (median = 19, IQR = 9) and High (median = 17, IQR = 9) income categories. The mean anxiety score also reflects this trend, with Low having the highest mean (19.583), followed by Middle (mean = 18.574) and High (mean = 17.528). For Maternal Education, the boxplot reveals a similar trend. As the education level increases, the median PROMIS Anxiety score tends to decrease. Specifically, for the High School or Less group (Median = 21, IQR = 9), anxiety levels are higher compared to the College group (Median = 19, IQR = 9) and the Past College group (Median = 18, IQR = 9). This suggests that individuals with higher education have lower anxiety scores. The mean scores further support this trend, with High School or Less having the highest mean of 20.00, compared to College at 18.618 and Past College at 17.683.

Quantitative Variables

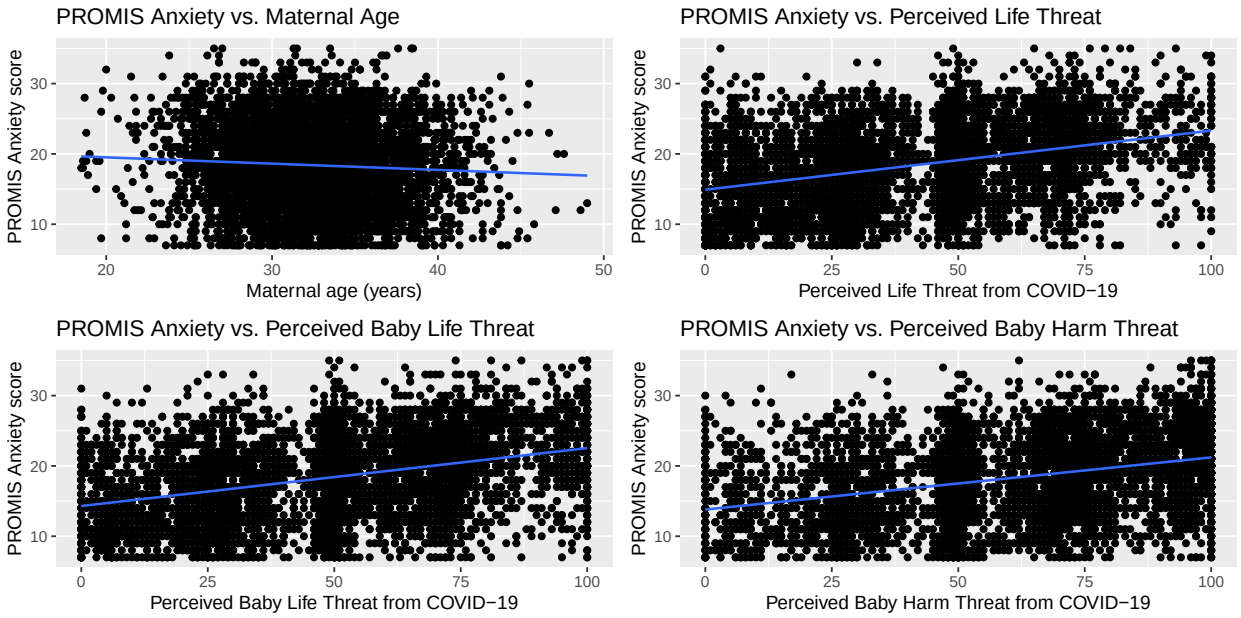

Figure 6. Bivariate Scatterplots

For each quantitative variable, we computed the correlation coefficients and created a scatterplot between PROMIS Anxiety score and the respective variable. The scatterplot between Maternal Age and PROMIS Anxiety score shows a weak negative relationship. As maternal age increases, there is a slight decrease in PROMIS Anxiety scores, as evidenced by the negative correlation coefficient of -0.062212. The very weak relationship suggests that age will not be a useful predictor of PROMIS Anxiety. The scatterplot between Perceived Life Threat from COVID-19 and PROMIS Anxiety score demonstrates a moderate positive relationship, indicated by the correlation coefficient of 0.35402. As the perceived life threat increases, PROMIS Anxiety scores tend to rise as well. This suggests that individuals who perceive a greater threat to their own life due to COVID-19 report higher anxiety levels. The plot shows a fairly linear trend, with some spread around the fitted line, and no significant unusual points.

Similarly, the scatterplot for Perceived Baby Life Threat and PROMIS Anxiety score shows a moderate positive relationship, with a correlation of 0.36166. As the perceived threat to the baby's life increases, the PROMIS Anxiety score also increases, indicating that greater concern for the baby's safety may lead to heightened anxiety in the mother. The relationship is moderately linear, with no major outliers.

The scatterplot for Perceived Baby Harm Threat and PROMIS Anxiety score exhibits a moderate positive relationship, as indicated by the correlation of 0.33783. As the perceived risk of harm to the baby increases, PROMIS Anxiety scores also tend to rise. The points show a moderate linear relationship, with a slight spread and no noticeable outliers.

Together, these patterns suggest that maternal age will add little predictive power to our model and can likely be dropped. Because all threat-related variables exhibit reasonable associations with the mother's anxiety levels, any subset of them would serve as useful predictors; however, the two baby-focused threat measures share almost identical relationships with PROMIS Anxiety, so including both could introduce issues with multicollinearity. Therefore, we will either select one or combine them and formally evaluate any issues using VIF values.

Multiple Linear Regression

```
##                          Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)               16.02208    0.68950   23.24  < 2e-16 ***
## Maternal_Age              -0.02955    0.01911   -1.55  0.12203
## education_cat2College     -1.07940    0.33418   -3.23  0.00125 **
## education_cat3PastCollege -1.33174    0.35478   -3.75  0.00018 ***
## houseinc_cat2Middle       -0.56458    0.22397   -2.52  0.01174 *
## houseinc_cat3High         -1.25227    0.24802   -5.05  4.6e-07 ***
## Threaten_Life              0.04579    0.00414   11.07  < 2e-16 ***
## Threaten_Baby_Danger       0.02709    0.00487    5.56  2.8e-08 ***
## Threaten_Baby_Harm         0.03045    0.00411    7.41  1.4e-13 ***
##
## Residual standard error: 5.41 on 5147 degrees of freedom
## Multiple R-squared:  0.172,  Adjusted R-squared:  0.171
## F-statistic:  134 on 8 and 5147 DF,  p-value: <2e-16
```

Table 1. Kitchen Sink Model Summary

To evaluate all potential predictors of PROMIS Anxiety scores, we ran a kitchen sink model which included Maternal Age, Maternal Education, Household Income, and all three Perceived Threats from COVID. The overall model is statistically significant, $F(8, 5147) = 134$, $p < 2e\text{-}16$. The R^2 value of 0.172 suggests that the model explains approximately 17.2% of the variability in predicting PROMIS Anxiety scores, while the Adjusted R^2 of 0.171 accounts for the number of predictors in the model. Specifically, the kitchen sink model found all of the predictors to be significant except Maternal Age. For Maternal Education, compared to the baseline High School or Less Education group, College (p = 0.00125) and Past College (p = 0.00125) were significant predictors of PROMIS Anxiety scores. For Household Income, compared to the baseline Low Income group, Middle Income (p = 0.01174) and High Income (p = 4.6e-07) were also significant predictors. Perceived Life Threat from COVID-19 ($p < 0.001$), Perceived Threat to the Baby's Life (p = 2.8e-08), and Perceived Harm to the Baby (p = 1.4e-13) were all significant.

Variable Selection - Stepwise

```
##              rsq    adjr2      cp    rss Maternal_Age education_cat2College
## 1  ( 1 ) 0.130797 0.130628 252.765 158138
## 2  ( 1 ) 0.155306 0.154978 102.369 153679
## 3  ( 1 ) 0.161807 0.161319  63.940 152496
## 4  ( 1 ) 0.012153 0.011386 996.499 179723            *                     *
## 5  ( 1 ) 0.169281 0.168474  21.469 151136
## 6  ( 1 ) 0.170701 0.169735  14.636 150878                                  *
## 7  ( 1 ) 0.163415 0.162278  61.943 152203            *                     *
## 8  ( 1 ) 0.172251 0.170965   9.000 150596            *                     *
##          education_cat3PastCollege houseinc_cat2Middle houseinc_cat3High
## 1  ( 1 )
## 2  ( 1 )
## 3  ( 1 )                                                                *
## 4  ( 1 )                         *                  *
## 5  ( 1 )                                            *                  *
## 6  ( 1 )                         *                                     *
## 7  ( 1 )                         *                  *                  *
## 8  ( 1 )                         *                  *                  *
##          Threaten_Life Threaten_Baby_Danger Threaten_Baby_Harm
## 1  ( 1 )                                 *
## 2  ( 1 )             *                                      *
## 3  ( 1 )             *                                      *
## 4  ( 1 )
## 5  ( 1 )             *                   *                  *
```

```
## 6  ( 1 )            *                   *                   *
## 7  ( 1 )            *                   *
## 8  ( 1 )            *                   *                   *
```

Table 2. Stepwise Variable Selection

To select the variables for our model, we used the stepwise variable selection procedure. Based on the results, we determined that the most effective model would include three variables: household income, the perceived threat to the mother's life, and the perceived threat of harm to the baby. This decision was based on the R^2 values, which showed that the model with three predictors explained 16.1% of the variability in PROMIS Anxiety scores. Further models with additional predictors showed only a marginal increase in the R^2 value, with the eight-predictor model explaining only 17.1% of the variance. This indicated that adding more predictors beyond the three selected variables did not substantially improve the model's explanatory power. Therefore, we decided on the three-predictor model as a good balance of accountability of variance and interpretability.

Our Model

```
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         14.22453    0.26669   53.34  < 2e-16 ***
## houseinc_cat2Middle -0.79416    0.22029   -3.60  0.00032 ***
## houseinc_cat3High   -1.64415    0.23579   -6.97  3.5e-12 ***
## Threaten_Life        0.05650    0.00361   15.66  < 2e-16 ***
## Threaten_Baby_Harm   0.04476    0.00334   13.41  < 2e-16 ***
##
## Residual standard error: 5.43 on 5151 degrees of freedom
## Multiple R-squared:  0.164,  Adjusted R-squared:  0.163
## F-statistic:  252 on 4 and 5151 DF,  p-value: <2e-16
```

Table 3. Initial Model Summary

We continued our analysis by fitting a Multiple Linear Regression model using the three-predictor model found via stepwise variable selection in Table 4. The overall model of household income, the perceived threat to the mother's life, and the perceived threat of harm to the baby is significant with an F-statistic of $F(4, 5151) = 252$, $p<0.001$. The R^2 value of 0.164 suggests that the model explains approximately 16.4% of the variability in anxiety scores, while the Adjusted R^2 of 0.163 accounts for the number of predictors in the model. In this model, when compared to the baseline low income group, the middle income ($p = 0.00032$) and high income ($p = 3.5e\text{-}12$) groups were both significant predictors of lower PROMIS Anxiety scores, with the coefficients of -0.79416 and -1.64415, respectively. This indicates that individuals in the middle and high-income categories report lower anxiety levels compared to those in the low-income category. Perceived Threat to Mother's Life also has a significant positive relationship with anxiety ($p < 0.001$, coefficient = 0.05650), suggesting that higher perceived threat to the mother's life is associated with higher anxiety scores. Similarly, Perceived Risk of Harm to Baby is a significant positive predictor of anxiety ($p < 0.001$, coefficient = 0.04476), suggesting that greater concern about harm to the baby leads to increased anxiety levels.

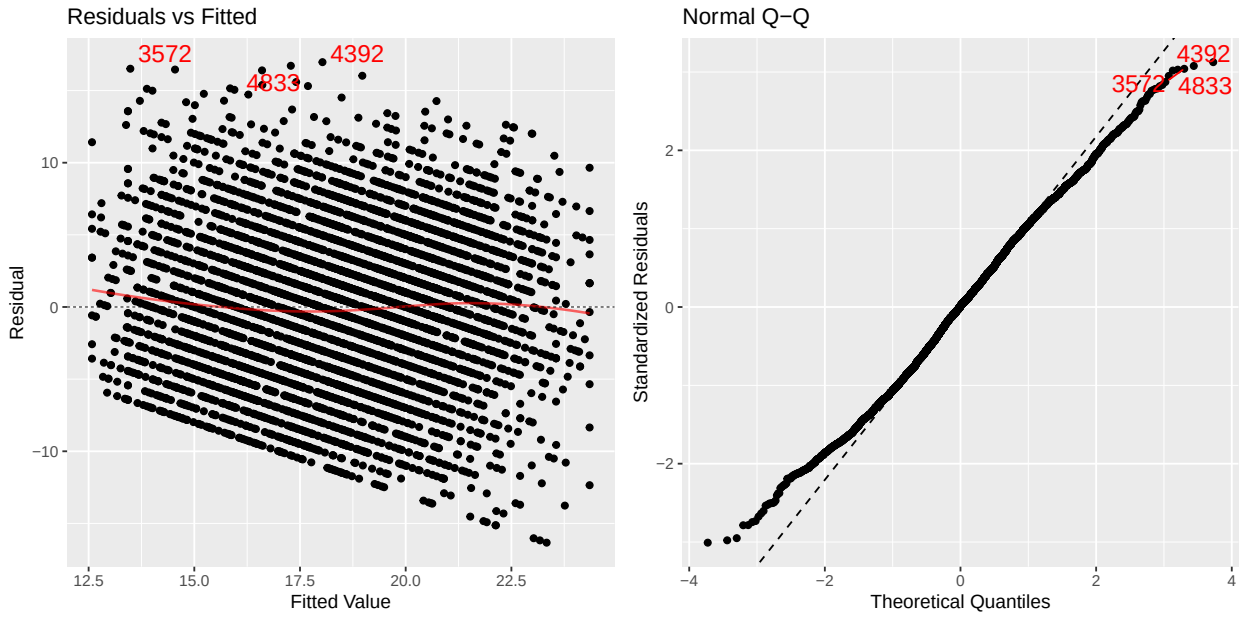

Figure 7. Initial Model Conditions

The residuals vs. fitted plot exhibited no heteroscedastic patterns, as most of the points fall within the same bandwidth around residuals = 0. Since PROMIS Anxiety is measured on an integer-only scale from 7 to 35, there exists banding for each respective score, as well as a floor function effect visible in the plot. This however, does not affect our interpretation of the plot. The model therefore meets the equal variance condition.

The Q-Q plot shows that around the middle, the points are closely clustered around the reference line, but the points fall off towards the left and right ends of the line. Regardless, the normality condition appears to be met. Log, square root, and power transformations of PROMIS Anxiety were attempted to try to make the data meet the normality condition better. However, these transformations only slightly adjust how clustered the data points are to the reference line. Based on this slight change, we decided that it was not worth the additional difficulty of interpretation and did not use any transformations in our final model.

The independence and random conditions are inferred from the source of the data. Since participants were randomly recruited and the data was collected via surveys, we assume both these conditions are met.

Multicollinearity & Unsual Points

	GVIF	Df	GVIF^(1/(2*Df))
houseinc_cat	1.0057	2	1.0014
Threaten_Life	1.4241	1	1.1934
Threaten_Baby_Harm	1.4231	1	1.1929

Table 4. VIF Values

Due to the intuitive similarities between the meaning of the three threat-related variables, as well as the similar results from the bivariate analysis, we tested for any multicollinearity in our model. For the two threat-related variables in our initial multiple linear regression model, perceived threat to the mother's life and perceived threat of harm to the baby, the correlation coefficient was 0.54387. This is a reasonably high correlation and may suggest issues with multicollinearity. Therefore, we produced the VIF values associated with our initial model. Household income had a VIF value of 1.0014, perceived threat to the mother's life was 1.1934, and perceived threat of harm to the baby was 1.929. These VIF values are all well-below the

standard limit of 5, and we concluded that there were no issues with multicollinearity and proceeded with our analysis.

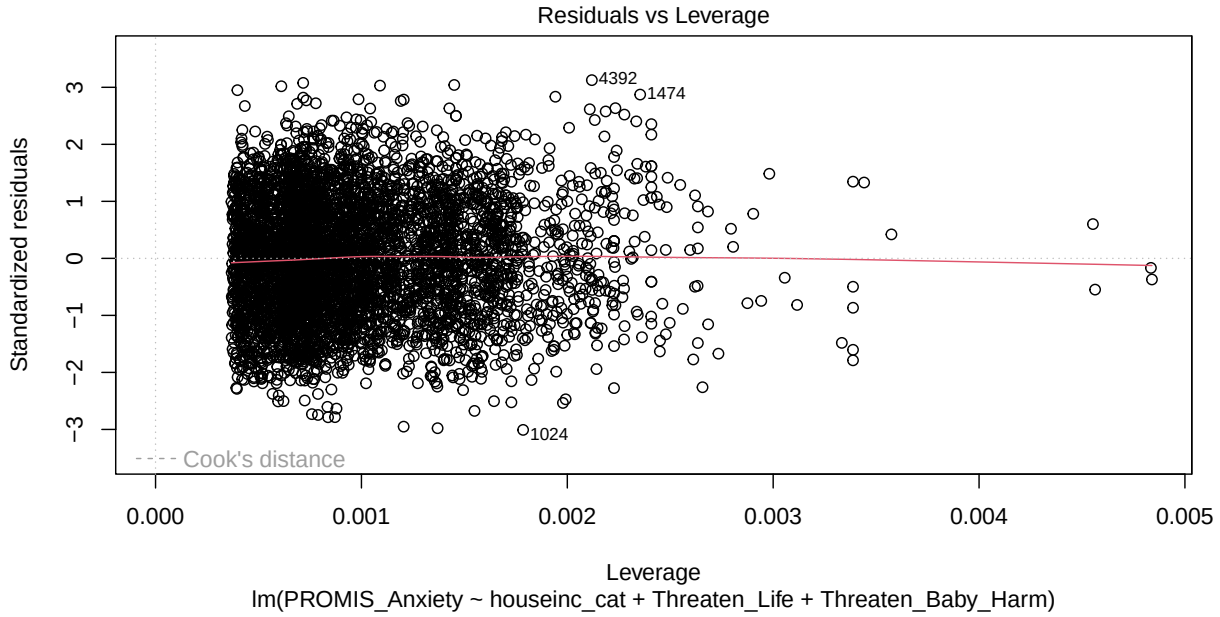

Figure 8. Unusual Points

To continue checking our model's robustness, we checked for any issues with unusual points. We began by calculating the moderate and high leverage cutoffs, which were 0.0015516 and 0.0023274, respectively. From the diagnostic plot, we observed a dense band of points qualifying as moderate leverage, and a sparser collection of high leverage points. Specifically, there were 573 moderate leverage points and 68 high leverage points.

Similarly, for standardized residuals, we observed a dense band of moderate residual values, with values greater than |2|. From the plot, it was difficult to visually identify the number of extreme residual points with values greater than |3|; however, we found 174 potential outliers and 6 extreme residual points.

There were no concerns with Cook's Distance, as the highest value among all observations was 0.00415, well below the moderate cutoff of 0.5.

Due to the relatively high number of points with moderate leverage (573 observations), we chose not to remove them so as not to impact the sample size significantly. However, we did remove all points with high leverage and all potential and extreme residual outliers based on our criteria for residual diagnostics. This resulted in 248 observations being removed.

```
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         13.48942    0.25929   52.02  < 2e-16 ***
## houseinc_cat2Middle -0.75236    0.21123   -3.56  0.00037 ***
## houseinc_cat3High   -1.81017    0.22537   -8.03  1.2e-15 ***
## Threaten_Life        0.06348    0.00348   18.22  < 2e-16 ***
## Threaten_Baby_Harm   0.05125    0.00324   15.81  < 2e-16 ***
##
## Residual standard error: 4.96 on 4908 degrees of freedom
## Multiple R-squared:  0.229,  Adjusted R-squared:  0.229
## F-statistic:  365 on 4 and 4908 DF,  p-value: <2e-16
```

Table 5. Final Model Summary

After checking for multicollinearity and removing unusual points we refit our three-predictor model using household income, the perceived threat to the mother's life, and the perceived threat of harm to the baby. Again we found the overall model to be significant with F(4, 4908), $p < 0.001$. Similarly, the middle income ($p = 0.00037$), and high income groups ($p = 1.2e\text{-}15$) were significant when compared to the low income reference groups. The perceived threat to the mother's life ($p < 0.001$), and perceived threat of harm to the baby ($p < 0.001$) were again significant predictors of anxiety scores.

In the initial model, the residual standard error was 5.43 and the R^2 value was 0.164, meaning the model explains approximately 16.4% of the variability in PROMIS Anxiety scores. In contrast, the final model has a residual standard error of 4.96, which indicates a reduction in error and an improved fit. The R^2 value in the final model improves to 0.229, showing that the final model explains 22.9% of the variability in anxiety scores. This improvement in R^2 reflects a better model, accounting for more of the variability in predicting mothers' PROMIS Anxiety scores.

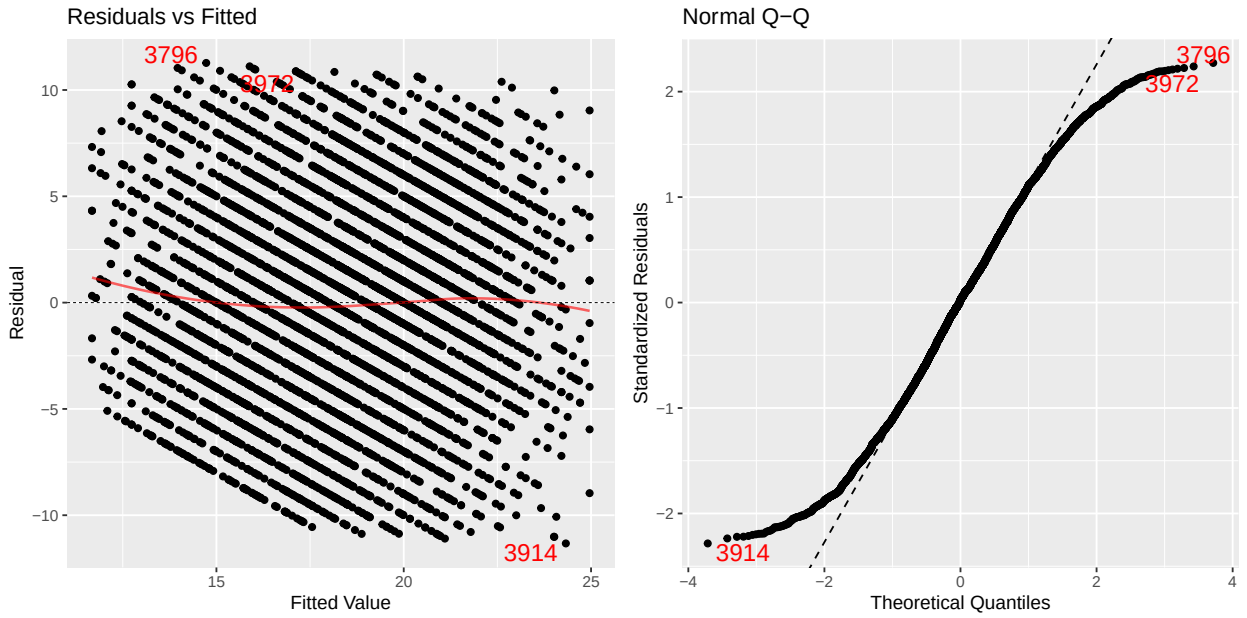

Figure 9. Checking Final Model Conditions

The residuals vs. fitted plot again shows no heteroscedastic patterns, and exhibits the same floor function pattern as before. The model therefore meets the equal variance condition.

The Q-Q plot again shows that around the middle, the points are closely clustered around the reference line, but the points fall off towards the left and right ends of the line. The normality condition appears to be met. We again attempted log, square root, and power transformations of PROMIS Anxiety. However, like in the initial model, these transformations only slightly adjust how clustered the data points are to the reference line. Based on this slight change, we decided that it was not worth the additional difficulty of interpretation and did not use any transformations in our final model.

The independence and random conditions are inferred from the source of the data. Since the data was collected via surveys, we assume both these conditions are met.

Interpreting the Final Model

$$\begin{aligned}\widehat{\text{PROMIS\_Anxiety}} = 13.48942 \\ &- 0.75236\,(\text{houseinc\_cat2Middle}) \\ &- 1.81017\,(\text{houseinc\_cat3High}) \\ &+ 0.06348\,(\text{Threaten\_Life}) \\ &+ 0.05125\,(\text{Threaten\_Baby\_Harm})\end{aligned}$$

Accounting for the other variables, if the mother's household income falls into the middle category ($70,000-$149,999), their average predicted anxiety score on the PROMIS test decreases by 0.75236. Accounting for the other variables, if the mother's household income falls into the high category ($150,000-$200,000+), their average predicted anxiety score on the PROMIS test decreases by 1.81017. Overall, although individuals who had a middle and high income were predicted to have lower anxiety scores, there was a greater decrease in anxiety scores for individuals with a high income.

Accounting for the other variables, for each one point increase in the mother's perceived threat to their life, their average predicted anxiety score on the PROMIS test increases by 0.06348. Accounting for the other variables, for each one point increase in the mother's perceived harm to their baby's life, their average predicted anxiety score on the PROMIS test increases by 0.05125. It appears that the mother's perceived threat to their life had a slightly greater effect on the predicted anxiety score compared to the perceived harm to their baby's life.

Conclusion

This study aimed to examine how socioeconomic status and perceived COVID-19 threats predict prenatal anxiety among pregnant and postpartum individuals. Our final three-predictor model revealed that lower household income, greater perceived threat to one's own life, and greater perceived risk to the baby each independently predicted higher anxiety scores on the PROMIS Anxiety scale. Together, these factors explained 22.9% of the variability in anxiety levels, underscoring the need for targeted mental health resources for lower-income individuals and those with heightened fears about COVID-19.

However, several limitations should be noted. First, our sample of 10,772 Canadian pregnant and postpartum individuals, recruited during the pandemic, may not be representative of populations in other countries or regions. Differences in demographic, healthcare, and cultural factors may limit the generalizability of our findings to populations with different socioeconomic structures, healthcare systems, or public health responses. Future studies should collect data from other countries for a wider picture of the influence of COVID-19 on pregnancy-related stress.

Second, we did not account for other potentially influential factors, such as maternal physical health, pre-existing medical conditions, or pregnancy complications, which could influence anxiety levels or interact with socioeconomic or COVID-19 threat factors. Future work should consider incorporating other factors into predicting anxiety levels.

Third, the distribution of our "threat" variables was heavily concentrated around the midpoint, limiting the ability to meet the normality assumption required for linear regression. Future research should consider alternative methods, such as bootstrapping or randomization, to address this limitation. Lastly, our study focuses exclusively on COVID-19-related stressors. Further research could explore how these stressors compare to other life stressors, examining both pre- and post-pandemic data to better understand the broader impact of such events on maternal mental health.