Food Insecurity and Child Mortality Rates in the United States

Ella L. Tamayo

BIOS 6344: Introduction to Linear Models

University of Texas Medical Branch, School of Public and Population Health

Dr. Xiaoying Yu

April 26th, 2024

# Introduction

The relationship between food security and child mortality across the United States has yet to be investigated in depth at a county level. Studies show health disparities in how food insecurity impacts individuals living in the United States. In 2020 in the U.S., 18.8% of non-Hispanic black children experienced food insecurity, 15.7% of Hispanic children, and 6.5% of non-Hispanic white children1. Food insecurity is more likely to impact a child living in a large metropolitan area than a smaller metropolitan area, or a completely non-metropolitan area. Individuals living in semi-metropolitan areas are often less likely to experience food insecurity1. A study conducted in North Carolina found that food insecurity, low birth weight and diabetes were positively correlated with infant mortality2. In a study conducted in rural Indonesia, food insecurity was found to be related to child mortality among neonates, infants, and children under 53. However, data is lacking on the relationship between child mortality and food insecurity across the U.S. This study aims to fill the gap in the literature between the relationship of child mortality rates and percentage of food insecurity at county levels, taking into account variables such as race, percentage of rural areas in counties, and rates of disease. The hypothesis being tested in this ecological study is county level child mortality is higher in counties with higher food insecurity.

# Methods

Upon obtainment, the County Level 2023 Rankings data was assessed and cleaned. The dataset included health and community information from all counties across the United States, such as child mortality rates, percentages of race within each county, percentages of food insecurity, percent of uninsured children, and more (n=3143). The dataset originally included totals from each state for each variable, which were removed as they were not needed for this analysis. Missing data from each county were also removed (n=1693). Most of the data missing from the dataset was due to missing cells in child mortality rate (n=1258 missing). This may be due to counties reporting zero child mortality rate as blank, since there are no zeros in the column for child mortality. Another large piece of the data missing is from HIV prevalence rates. However, there are zeros in the column for HIV prevalence rate, so why this data is missing is unclear. Since a large percentage of the data was missing, later the model was recreated with the full dataset to determine whether the model would be similar. The beta coefficients for the two models were mostly different from one another (Table 5). This will be accepted as a limitation of this study. This topic should be investigated further, with more complete data on child mortality rates.

Upon conducting univariate analyses on child mortality rate, it was evident that child mortality rate was highly skewed. To account for this, a log transformation was conducted in order for the variable to have a more linear nature, to ensure the data was suitable for multivariable regression. A log transformation was suitable because the rate had no values of 0. After transformation, the child mortality rate distribution was sufficiently normal.

A multivariable regression was conducted to analyze the association between percentages of food insecurity (primary exposure) and child mortality rate (outcome). A multivariable regression was chosen in order to analyze the relationships and associations between multiple variables (both continuous and categorical) and child mortality rate. Covariates analyzed included percent of individuals with limited access to healthy food, percent of uninsured children, median household income, HIV prevalence rate, percent of individuals who live in a rural portion of county, majority race of county, and region of the U.S. Relationships between child mortality rate and covariates were investigated using bivariate analyses such as scatterplots to determine relationships and correlation matrices (Table 2).

The covariates race and region were created in SAS. Race was created by determining if the majority race in the county was White (percent non-Hispanic White individuals>50), then a new variable called minority was set equal to 0. If the majority race was Non-White (percent Black, Asian, Hispanic, Native Hawaiian, or American individuals>50), then the new variable minority was set equal to 1. Initially the data was separated into each race with multiple factors, but because very few counties were majority non-white, the data was separated into a binary variable. This variable was created because the literature suggested a relationship between race and food security1. The covariate region was created by determining what region of the United States a State and County were located in (West, Midwest, South, and East). North was considered to be the reference for regression purposes. This variable was created to determine whether location posed a significant effect on child mortality rate.

Simple regressions were created analyzing each covariate individually. Each regression produced a significant p-value, indicating significant associations between child mortality rate and covariates (Table 2).

Before model selection, potential interactions and confounding variables were addressed. Looking at all possible interactions with the primary exposure of percent of food insecurity, significant interactions were found between percent food insecurity and percent uninsured children interaction, and percent food insecurity and percent rural interaction (Table 4).

After interactions were assessed, variables were assessed for confounding effects. Significant confounders included region and median household income, because when they were included in models with percent food insecurity, the adding of these variables caused the estimate for food insecurity to change by more than 10% (Table 4).

Covariates were selected to be in the multivariable regression model by using stepwise selection, due to the large number of variables considered. These tests are two-sided. Variables that were forced to be included in the model included the primary exposure (percent of food insecurity), significant interaction terms, significant confounders, and variables that were part of the interaction terms. The model stepwise selection process selected percent food insecurity, percent food insecurity\*percent uninsured children interaction, percent rural, percent food insecurity\*percent rural interaction, median household income, region, percent of individuals with limited access to healthy food, percent of uninsured children, and HIV prevalence rate (Table 3). The beta coefficient for median household income was multiplied by 10,000 so the model would be interpreted as every $10,000 increase in median salary instead of every dollar increase. The R2 of this model is .5376, indicating that 53.76% of the variation in child mortality can be explained by the variables in this model. Model selection ensured that the model had parsimony and was not over inflated with many variables. However, there may be some unaccounted for error because of the remaining high number of variables in the model.

After the creation of this model, collinearity was assessed. High variance inflation factors were found for percent of food insecurity, percent rural, percent of uninsured children, and the interaction terms percent of food insecurity\*percent rural and percent of food insecurity\*percent of uninsured children. This is to be expected because introducing interaction terms often introduces collinearity. This was accounted for by centering the variables in the model, lowering variance inflation factors to less than 10, reducing collinearity. The final transformed, selected, and centered model is shown in Table 3.

Assumptions of multivariable linear regression include homoscedasticity, normality, independence and linearity. These assumptions were assessed using fit diagnostic plots from SAS. The plot of residuals vs. predicted values showed a mostly random scatter of variables with a constant variance, with variables slightly centered around the middle (the upper left hand corner in Figure 1). Because this plot has a mostly random scatter, the assumptions of linearity, homoscedasticity and independence are met. The assumption of independence is met because the health of one county is often not affected by the health of another county, each variable is not influenced by another. The assumption of normality was originally not met since the distribution of the outcome was skewed, this caused the QQ plot to show a curved line across the diagonal axis. This was accounted for by taking the log of child mortality rate. After transformation, the QQ plot showed a straight line on the diagonal, indicating normality (middle left hand corner in Figure 1). These analyses were conducted using SAS.

# Results

This study aimed to discover if there is an association between percentages of food insecurity and rates of child mortality at a county level. Participants included 1693 counties from across regions of the United States with measures of child mortality rates (mean=59.38 (SD=21.39)) and percentages of food insecurity (mean=12.87 (SD=3.28)). Most counties were majority white (n=1570, 92.73%) and were located in the South (n=851, 50.27%). Across the counties included in this dataset, the average median household income was 60,468.17 (SD=15748.61). County percentages of rural areas were a little less than 50% overall (mean=44.32 (SD=26.42) (Table 1).

The final model includes significant p-values for all covariates. The Beta coefficient for percent food insecurity is positive, suggesting that there is a positive correlation between the log of child mortality rates and percentages of food insecurity accounting for other covariates in the model. There are significant interactions between percent of food insecurity and percent rural and percent of food insecurity and percent of uninsured children. Variables that confound the relationship between food insecurity and child mortality rates include the region of the U.S. that a county is in and median household income.

Most covariates had positive associations with the log of child mortality, but the interaction term for percent of food insecurity and percent of uninsured children, median household income, and Eastern region variables had negative relationships. This indicates that the higher median household income a county has, or if a county is located in the Eastern region, the lower the log of their child mortality rates will be.

The relationships between the interaction terms is interesting and needs to be investigated further in future studies. When analyzing the interaction plot between food insecurity and ruralness, it appears that if a county is in a completely urban area (0% rural), and experiences high levels of food insecurity, they are expected to have higher log of child mortality rates than counties in mostly rural area (100% rural) also experiencing food insecurity. When analyzing the interaction plot between food insecurity and child uninsured percentages, the results are unexpected. If a county has high rates of uninsured children (more uninsured children) who are also experiencing food insecurity, the county is likely to have lower log of rates of child mortality. If a county has low rates of uninsured children (more insured children) who are also experiencing high percentages of food insecurity, the log of child mortality is expected to be higher. These relationships are demonstrated in Figure 2. This relationship is unexpected and needs to be investigated further in future studies.

# Discussion

This model shows that the log of county level child mortality is positively associated with higher percentages of food insecurity. Significant covariates included percent food insecurity\*percent uninsured children interaction, percent rural, percent food insecurity\*percent rural interaction, median household income, region, percent of individuals with limited access to healthy food, percent of uninsured children, and HIV prevalence rate (Table 3). These covariates all had significant p-values, indicating that the variables are associated with the log of child mortality rates.

The limitations of this study include that much of the dataset was missing data (removed 1,442 counties). Because of this, the model was used later with the full dataset (not excluding missing data, n=3143). The models had different beta coefficients, indicating that the missing data contributed significantly to the analysis. This is a limitation for this study. Additionally, because there were a high number of covariates involved, this could have potentially inflated R2 of the model. Parsimony was prioritized in model selection, but inflation of the R2 value may still be possible.

The associations discovered in this study can be used to spur further investigation of relationships between covariates and child mortality, as it is a complex subject that is affected by many variables. Future studies could investigate the nature of relationships between interaction terms found between percent food insecurity and percent rural and percent of uninsured children. Future studies could also potentially look into creating a prediction model, looking to see if child mortality rate can be predicted for a county based on these variables.

# References:

1. Ullmann, Heidi et al. (2022). Children living in households that experienced food insecurity : United States, 2019-2020. (432).
2. Cassidy-Vu, L., Way, V., & Spangler, J. (2022). The correlation between food insecurity and infant mortality in North Carolina. Public health nutrition, 25(4), 1038–1044. https://doi.org/10.1017/S136898002200026X
3. Campbell, A. A., de Pee, S., Sun, K., Kraemer, K., Thorne-Lyman, A., Moench-Pfanner, R., Sari, M., Akhter, N., Bloem, M. W., & Semba, R. D. (2009).

Relationship of household food insecurity to neonatal, infant, and under-five child mortality among families in rural Indonesia. Food and nutrition bulletin, 30(2), 112–119. https://doi.org/10.1177/156482650903000202

1. Lee, H. H., Lee, S. A., Lim, J. Y., & Park, C. Y. (2016). Effects of food price inflation on infant and child mortality in developing countries. The European journal of health economics : HEPAC : health economics in prevention and care, 17(5), 535–551. https://doi.org/10.1007/s10198-015-0697-6

# Appendix:

*Tables:*

## Table 1. Descriptive Analyses

| *(n=1693)* |  |
| --- | --- |
| Child Mortality Rate [Mean (SD)] | 59.38 (21.39) |
| Percent Food Insecurity [Mean (SD)] | 12.87 (3.28) |
| Percent People with Limited Access to Healthy Food [Mean (SD)] | 7.68 (5.00) |
| Percent of Uninsured Children [Mean (SD)] | 6.11 (3.22) |
| Median Household Income [Mean (SD)] | 60468.17 (15748.61) |
| HIV Prevalence Rate [Mean (SD)] | 215.05 (219.05) |
| Percent Rural [Mean (SD)] | 44.32 (26.42) |
| Majority White [n (%)] | 1570 (92.73%) |
| Majority Non-White [n (%)] | 123 (7.27%) |
| West [n (%)] | 211 (12.46%) |
| Midwest [n (%)] | 451 (26.64%) |
| South [n (%)] | 851 (50.27%) |
| East [n (%)] | 180 (10.63%) |

## Table 2. Transformed Simple Regression Models and Correlation Matrix

| Variable | Estimate | Standard Error | T-value | Pr>t |
| --- | --- | --- | --- | --- |
| Percent Food Insecurity | 0.0611 | 0.0021 | 28.97 | <.0001 |
| Percent with Limited Access to Healthy Food | 0.0170 | 0.0016 | 10.36 | <.0001 |
| Percent of Uninsured Children | 0.0194 | 0.0026 | 7.5 | <.0001 |
| Median Household Income | -0.000015 | 0.0000 | -36.86 | <.0001 |
| HIV Prevalence Rate | 0.000321 | 0.0000 | 8.23 | <.0001 |
| Percent Rural | 0.0055 | 0.0003 | 18.76 | <.0001 |
| Minority | 0.1800 | 0.0323 | 5.57 | <.0001 |
| Midwest | 0.1424 | 0.0264 | 5.39 | <.0001 |
| South | 0.3090 | 0.0243 | 12.7 | <.0001 |
| East | -0.1005 | 0.0321 | -3.13 | 0.0018 |

*Correlations:*

|  | Pearson Correlation Coefficients with Child Mortality Rate, N=1693 Prob > |r| under H0: Rho=0 |
| --- | --- |
| Percent Food Insecurity | 0.54746  <.0001 |
| Percent with Limited Access to Healthy Food | 0.22745  <.0001 |
| Percent of Uninsured Children | 0.15019  <.0001 |
| Median Household Income | -0.60419  <.0001 |
| HIV Prevalence Rate | 0.19751  <.0001 |
| Percent Rural | 0.41193  <.0001 |
| Minority | 0.15253  <.0001 |
| Midwest | -0.09124  0.0002 |
| South | 0.35203  <.0001 |
| East | -0.25057  <.0001 |

## Table 3. Full Model Regression (Selected, Transformed and centered model)

| Label | Parameter  Estimate | Standard  Error | t Value | Pr > |t| |
| --- | --- | --- | --- | --- |
| Intercept | 4.40841 | 0.04902 | 89.93 | <.0001 |
| Percent Food Insecurity Centered | 0.00606 | 0.00307 | 1.98 | 0.0483 |
| Percent Food Insecurity\*Percent Uninsured Children Interaction Centered | -0.00304 | 0.0005875 | -5.17 | <.0001 |
| Percent Rural Centered | 0.00283 | 0.00028492 | 9.93 | <.0001 |
| Percent Food Insecurity\*Percent Rural Interaction Centered | 0.00024868 | 0.00007242 | 3.43 | 0.0006 |
| Median Household Income (per $10,000) | -0.0945 | 6.63E-07 | -14.26 | <.0001 |
| Midwest | 0.07647 | 0.02054 | 3.72 | 0.0002 |
| South | 0.0562 | 0.01996 | 2.82 | 0.0049 |
| East | -0.09558 | 0.02513 | -3.8 | 0.0001 |
| Percent with Limited Access to Healthy Food | 0.00502 | 0.00135 | 3.72 | 0.0002 |
| Percent of Uninsured Children Centered | 0.00689 | 0.00195 | 3.53 | 0.0004 |
| HIV Prevalence Rate | 0.00029896 | 0.00003057 | 9.78 | <.0001 |

## Table 4. Interaction and Confounding Assessment

*Interactions:*

| Variable | Estimate | Standard Error | t-value | Pr>t |
| --- | --- | --- | --- | --- |
| Percent Food Insecurity | 0.0199 | 0.0149 | 1.34 | 0.181 |
| Percent with Limited Access to Healthy Food | 0.0011 | 0.0060 | 0.19 | 0.8529 |
| Percent of Uninsured Children | 0.0490 | 0.0083 | 5.91 | <.0001 |
| HIV Prevalence Rate | 0.0004 | 0.0001 | 3.19 | 0.0014 |
| Median Household Income | 0.0000 | 0.0000 | -5.29 | <.0001 |
| Percent Rural | 0.0005 | 0.0011 | 0.43 | 0.6697 |
| Minority | 0.0822 | 0.1021 | 0.8 | 0.4211 |
| Midwest | 0.0434 | 0.0888 | 0.49 | 0.6253 |
| South | -0.0322 | 0.0875 | -0.37 | 0.7132 |
| East | -0.0166 | 0.1217 | -0.14 | 0.8918 |
| Food Insecurity\*Access to healthy food interaction | 0.0003 | 0.0004 | 0.7 | 0.4848 |
| **Food Insecurity \* Percent of Uninsured Children** | -0.0032 | 0.0006 | -5.31 | **<.0001** |
| Food Insecurity \* HIV Prevalence Rates | 0.0000 | 0.0000 | -0.96 | 0.339 |
| Food Insecurity \* Median Household Income | 0.0000 | 0.0000 | -0.64 | 0.5219 |
| **Food Insecurity \* Percent Rural** | 0.0002 | 0.0001 | 2.1 | **0.0362** |
| Food Insecurity \* Minority | -0.0076 | 0.0069 | -1.1 | 0.2697 |
| Food Insecurity \* Midwest | 0.0027 | 0.0074 | 0.36 | 0.7181 |
| Food Insecurity \* South | 0.0064 | 0.0069 | 0.93 | 0.3525 |
| Food Insecurity \* East | -0.0080 | 0.0108 | -0.74 | 0.4597 |

*Confounding:*

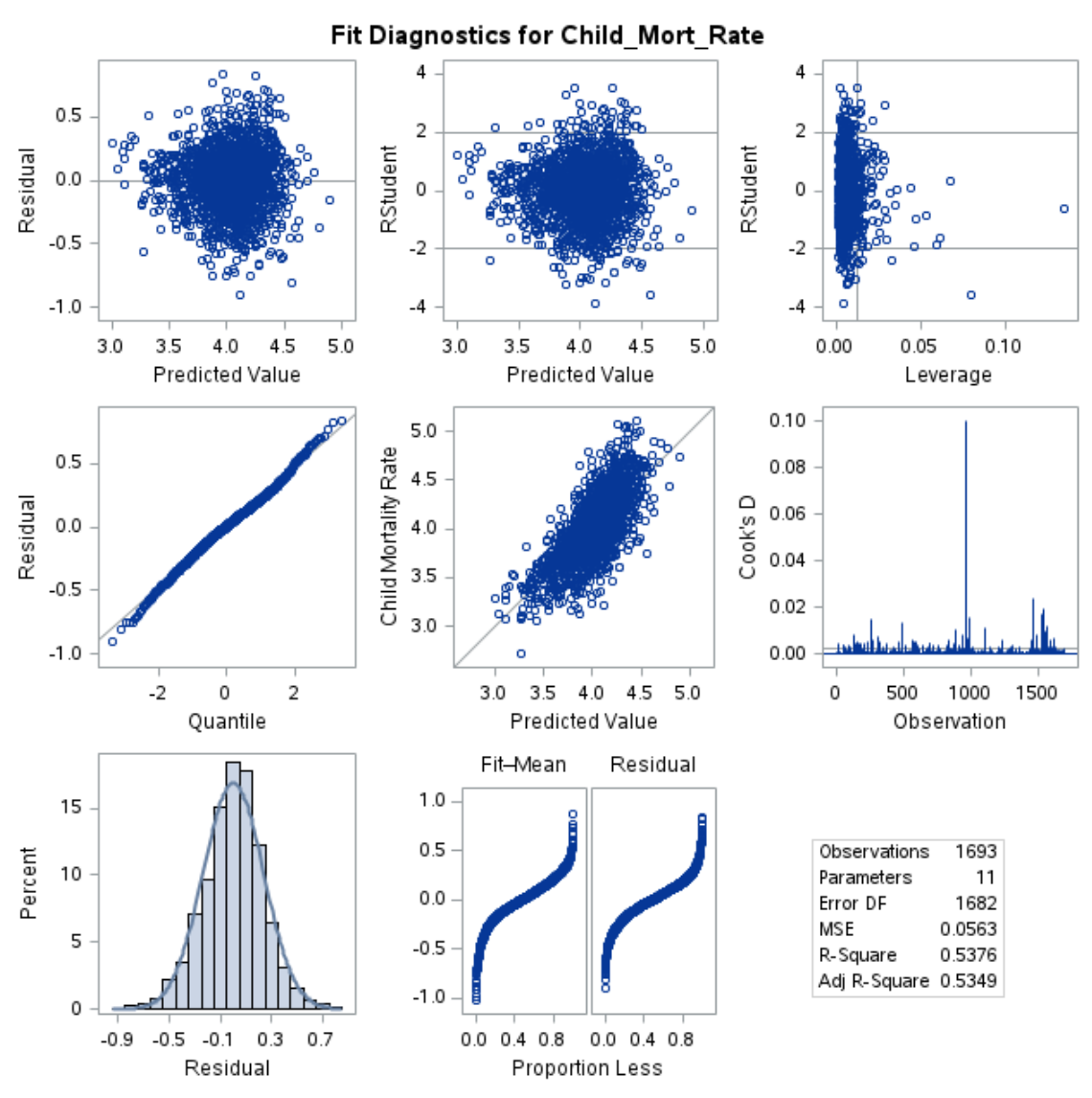
| Food insecurity adjusted by: | Adjusted Parameter | Percent Difference |
| --- | --- | --- |
| Percent with Limited Access to Healthy Food | 0.05912 | 3.35% |
| Percent of Uninsured Children | 0.05959 | NA (Interaction) |
| Median Household Income | 0.01579 | **286.95%** |
| HIV Prevalence Rate | 0.05948 | 2.72% |
| Percent Rural | 0.05212 | NA (Interaction) |
| Minority | 0.06092 | 0.30% |
| Region | 0.05391 | **13.34%** |

## Table 5. Missing Data Analysis

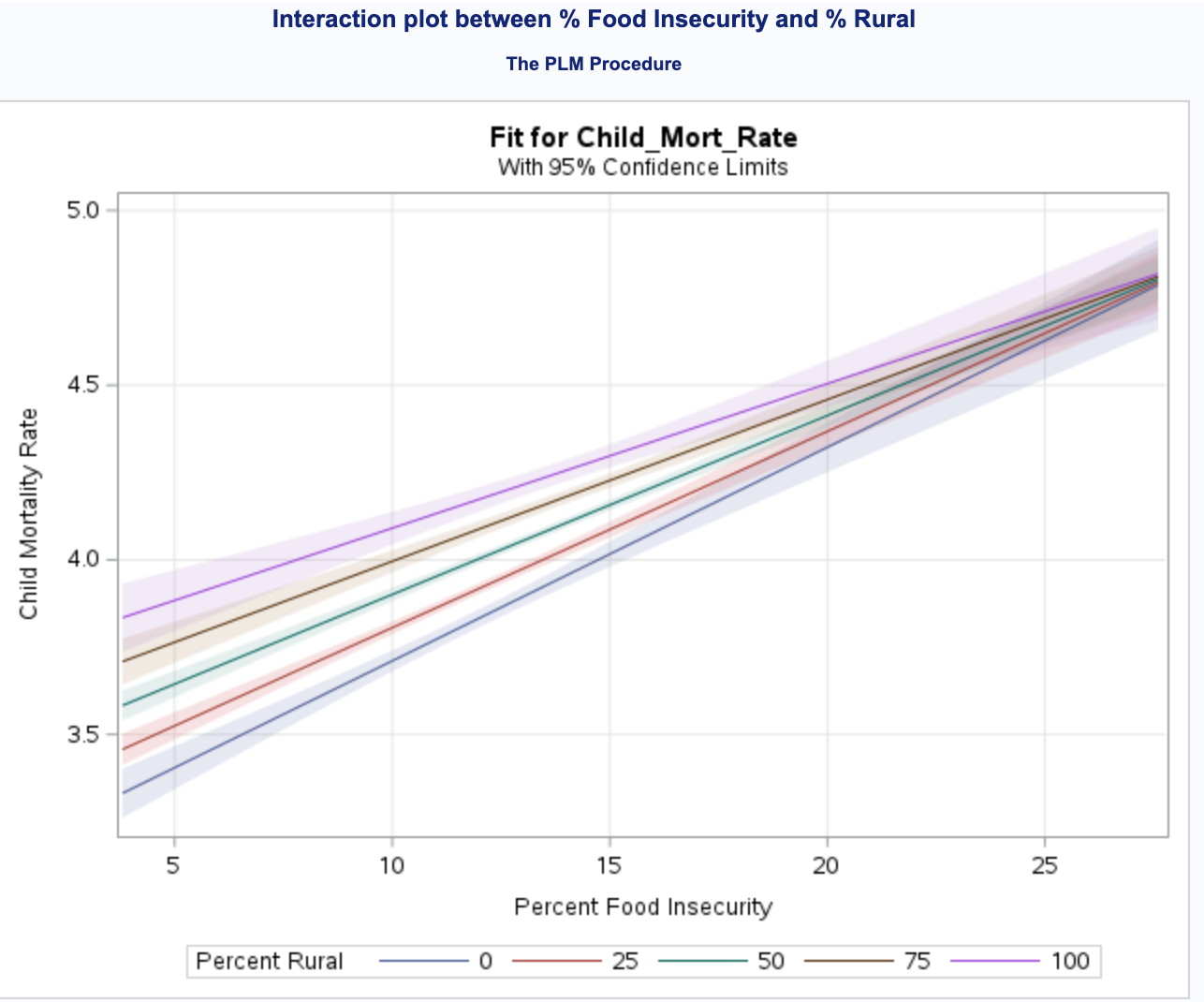
|  | All data (centered): | Missing data removed dataset (centered): |  |
| --- | --- | --- | --- |
| Variable | Parameter  Estimate | Parameter  Estimate | % Difference |
| Intercept | 4.61284 | 4.40841 | -4.64 |
| Percent Food Insecurity Centered | 0.00967 | 0.00606 | -59.57 |
| Percent Food Insecurity\*Percent Uninsured Children Interaction Centered | -0.00336 | -0.00304 | -10.53 |
| Percent Rural Centered | 0.00035962 | 0.00283 | 87.29 |
| Percent Food Insecurity\*Percent Rural Interaction Centered | -0.00001167 | 0.00024868 | 104.69 |
| Median Household Income | 0.09481 | -0.00000945 | 1003380.42 |
| Midwest | 0.0905 | 0.07647 | -18.35 |
| South | -0.06184 | 0.0562 | 210.04 |
| East | 0.00099568 | -0.09558 | 101.04 |
| Percent with Limited Access to Healthy Food | 0.01103 | 0.00502 | -119.72 |
| Percent of Uninsured Children Centered | 0.0002011 | 0.00689 | 97.08 |
| HIV Prevalence Rate | 0.00029896 | 0.00029896 | 0 |

## 

## Figure 1: Diagnosis Plots

****

## Figure 2: Interaction Plots

**

