December 2021

The Relationship Between Food Environments and Obesity Rates in the United States; Measured by the Modified Retail Food Environment Index

Mary Katherine Miller
Clemson University, markatmiller@gmail.com

THE RELATIONSHIP BETWEEN
FOOD ENVIRONMENTS AND OBESITY RATES IN THE UNITED STATES;
MEASURED BY THE MODIFIED RETAIL FOOD ENVIRONMENT INDEX

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
Economics

---

by
Mary Katherine Miller
December 2021

Accepted by:
Dr. Devon Gorry, Committee Chair
Dr. Jorge Garcia
Dr. Scott Baier

ABSTRACT

The modified Retail Food Environment Index (mRFEI) is a measurement created by the Centers for Disease Control and Prevention (CDC) designed to capture the number of healthy and less healthy food retailers in an area. Out of the total number of food retailers in that area considered either healthy or less healthy, the mRFEI represents the healthy percentage. The mRFEI index is commonly used to measure obesity amongst different levels of food access. Rather than simply relying on broad categorizations of food environments, using the raw index gives a more complete picture of the relationship between food availability and obesity. I use the index on the census tract level and control for demographic variables for each census tract. Through the use of cubic regressions, obesity is regressed on the healthful food index and census-tract level control variables. It was discovered that the mRFEI is, in fact, a good indicator of obesity rates, and each food environment corresponds with a different relationship with obesity rates. Additionally, this study further proves that Food Swamps and Food Deserts are intrinsically different and are entirely separate phenomena. While the direction of the relationship between the index and obesity was revealed, future research should aim to pinpoint more in-depth parameters for each food environment. This study identifies the mRFEI as a more in-depth indicator for obesity instead of using broad parameters that classify geographies into Food Deserts and Food Swamps.

ACKNOWLEDGEMENTS

I am deeply indebted to Devon Gorry for her insightful advice and dedicated support. I would like to thank Devon Gorry, Jorge Garcia, and Scott Baier for serving on my thesis committee. In addition to my committee, I would also like to acknowledge Scott Templeton, Raymond Sauer, and Thomas Hazlett for their support and guidance throughout my time at Clemson University. Lastly, I thank my family, Mom, Dad, Beau, and Harris, without your unwavering support this would not have been possible. Thank you for teaching me the importance of persistence, a passion for education, and the joy and importance of life-long learning. For that, I am forever grateful.

TABLE OF CONTENTS

Page

TITLE PAGE……………………………………………………………………………………i

ABSTRACT……………………………………………………………………………………ii

ACKNOWLEDGEMENTS………………………………………………………………………iii

LIST OF FIGURES……………………………………………………………………………….v

LIST OF TABLES ………………………………………………………………………………vi

LIST OF GRAPHS……………………………………………………………….……..vii

INTRODUCTION…………………………………………………………………………...1

LITERATURE REVIEW………………………………………………………………………4

DATA…………..……………………………………………………………………………..12

METHODOLOGY...………………………………………………………………………….16

RESULTS……….…………………………………………………………………………….22

DISCUSSION…………………………………………………………………………………32

CONCLUSION………………………………………………………………………………..35

SUGGESTIONS FOR FURTHER RESEARCH………………………………………………..37

REFERENCES………………………………………………………………………………..39

LIST OF FIGURES

Page

FIGURE 1……………………………………………………………………………………5

FIGURE 2……………………………………………………………………………………7

FIGURE 3…………………………………………………………………………………..12

LIST OF TABLES

Page

TABLE 1……………………………………………………………………………………15

TABLE 2………………………………………………………………………..…..…22

TABLE 3……………………………………………………………………...………….25

TABLE 4.……………………………………………………………………...……27

TABLE 5……………………………………………………………………...………….29

LIST OF GRAPHS

Page

GRAPH 1……………………………………………………………………………………17

GRAPH 2……………………………………………………………………………………19

INTRODUCTION

The obesity epidemic has plagued the United States for over three decades. According to the CDC (Centers for Disease Control and Prevention), the obesity epidemic rapidly spread across the United States during the 1990s (CDC, 1999). Groups with the highest increases in body weight were those with little college education and those belonging to Hispanic and/or Black origin. To reduce obesity in the United States, the 1999 director of the CDC, Jeffrey Koplan issued a national prevention effort to target obesity rates. Koplan not only called attention to individuals needing to reduce their caloric intake and increase physical activity, but also to the responsibility of health care providers, workplaces, schools, and policymakers. The CDC began promoting healthy options and emphasized schools and workplaces to make it easier for workers and students to obtain healthy levels of physical activity daily. These efforts have certainly brought awareness to the issue, yet countries around the world, with the United States being no exception, continue to climb in obesity rates.

An abundance of research exists on obesity in the United States pertaining to its rapid growth. A surplus of studies also exist on socioeconomic factors contributing to high obesity rates such as access to food; however, less research exists on how food access from location to location causes a direct impact on health. Extreme inequalities in obesity rates exist by geography, driving the research on the relationship between food environments and health. To better capture the severity of food disparity or lack thereof in geographies, the CDC constructed a Retail Food Environment Index (RFEI). The RFEI index is a ratio of describing the relative density of unhealthy food outlets to healthy food outlets on a scale of 1-10. However, the RFEI index has several limitations. The RFEI is undefined when the denominator is zero, this if for environments with a complete lack of healthy food retailers, excluding an entire demographic. To combat this issue, the Centers for Disease Control and Prevention (2011) modified the measure (RFEI x 100) to include environments where small communities may be lacking grocery stores or produce vendors. The new measure was named the Modified Retail Food Environment Index (mRFEI). Currently, the CDC does not offer raw data for mRFEI, because the data on individual retail stores was purchased under contract and cannot be distributed. However, mRFEI scores on the census tract level are available.

Populations who struggle to obtain healthy foods are often categorized into broad parameters of living in a Food Desert or a Food Swamp. A Food Swamp is defined as a geographical location where an abundance of fast-food retailers, junk food outlets, convenience stores, and liquor stores outweigh the available healthy food options (Rose, et al., 2009). A Food Desert is slightly different and is defined as an urban, residential area with limited access to affordable or nutritious foods or where there is a complete lack of food availability. Michelle Obama, the former First Lady, put obesity at the forefront of her mission statement, she specifically emphasized how access to food affects obesity and what could be done to mend the gap. Over the past two decades, obesity has received more attention than other health concerns due to its rapidly increasing rates and the serious health problems that follow. Food Deserts and Food Swamps are becoming a popular focus for policy and government interventions. However, the terms are much too broad to explain individual geographies and are becoming overused.

States could do more when it comes to improving access to food, regulations, and policies to promote healthy eating and fight obesity, according to the CDC. The 2011 Food Environment State Indicator report notes that the communities, childcare facilities, and schools also have roles to play. Thirty-two states and the District of Columbia scored at or below the national average for the Modified Retail Food Environment Index (mRFEI), a measure of the proportion of food retailers that typically sell healthy foods within a state. Scores can range from 0 (no food retailers that typically sell healthy food) to 100 (only food retailers that typically sell healthy food). States with lower mRFEI scores have more unhealthy food retailers, such as fast-food restaurants and convenience stores that are less likely to sell less healthy foods and have fewer food retailers, such as supermarkets, that tend to sell healthy foods, such as fresh fruits and vegetables.

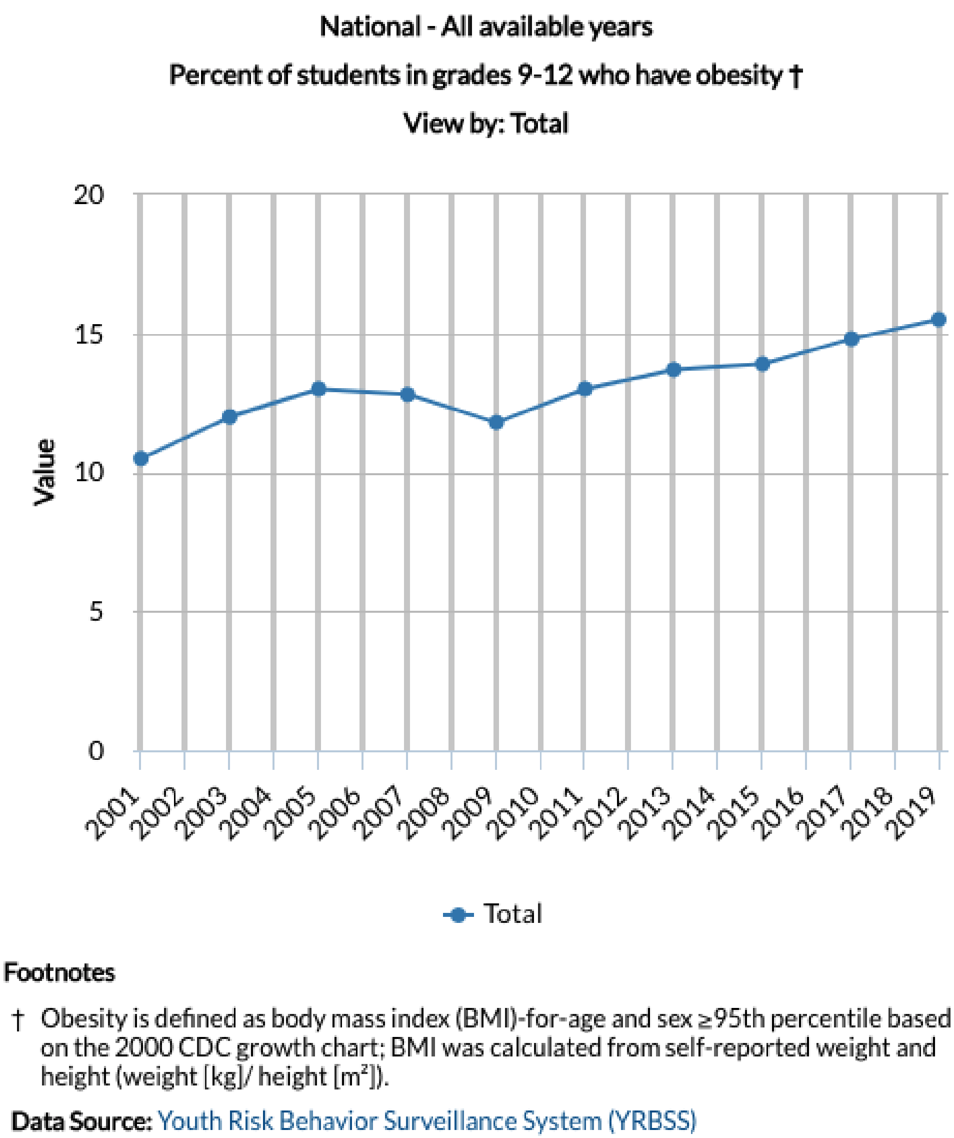
Additionally, obesity requires a unique approach in regards to policy-making given the wide array of causal factors. Because obesity can be explained with various socioeconomic variables, it is often seen as a social justice issue and in turn, receives major attention. Decreased educational attainment often leads to decreased incomes. Lower-income individuals typically live in less developed areas where healthy food is unaffordable, or in less developed areas with an abundance of fast food, junk food, and convenience stores that outweigh the number of healthy options available. Presumably, people that are of low socioeconomic status in Food Deserts are positively correlated with obesity. Similarly, areas deemed as Food Swamps are said to be associated with higher obesity rates. However, I do not believe that it is that simple. By better understanding how food environments affect obesity, further mediation programs can be constructed to decrease obesity rates. While controlling for certain variables, I aim to test the mRFEI index scores including their corresponding food environments' association with obesity. This study was designed to test the efficiency of predicting obesity rates with the food index, as well as the validity of broadly categorizing the index into food environments and its ability to efficiently forecast obesity rates.

LITERATURE REVIEW

Existing literature on Food Swamps is relatively limited, while Food Deserts are a more researched topic. A 2009 study by Rose and colleagues coined the term Food Swamps, on the claim that Food Deserts did not accurately predict obesity and determined there was a need for a more useful metaphor to represent areas where a large relative amount of energy-dense snack foods, inundate healthy food options. Due to the nuance of the term, studies are relatively limited on those particular areas in relationship with obesity rates.

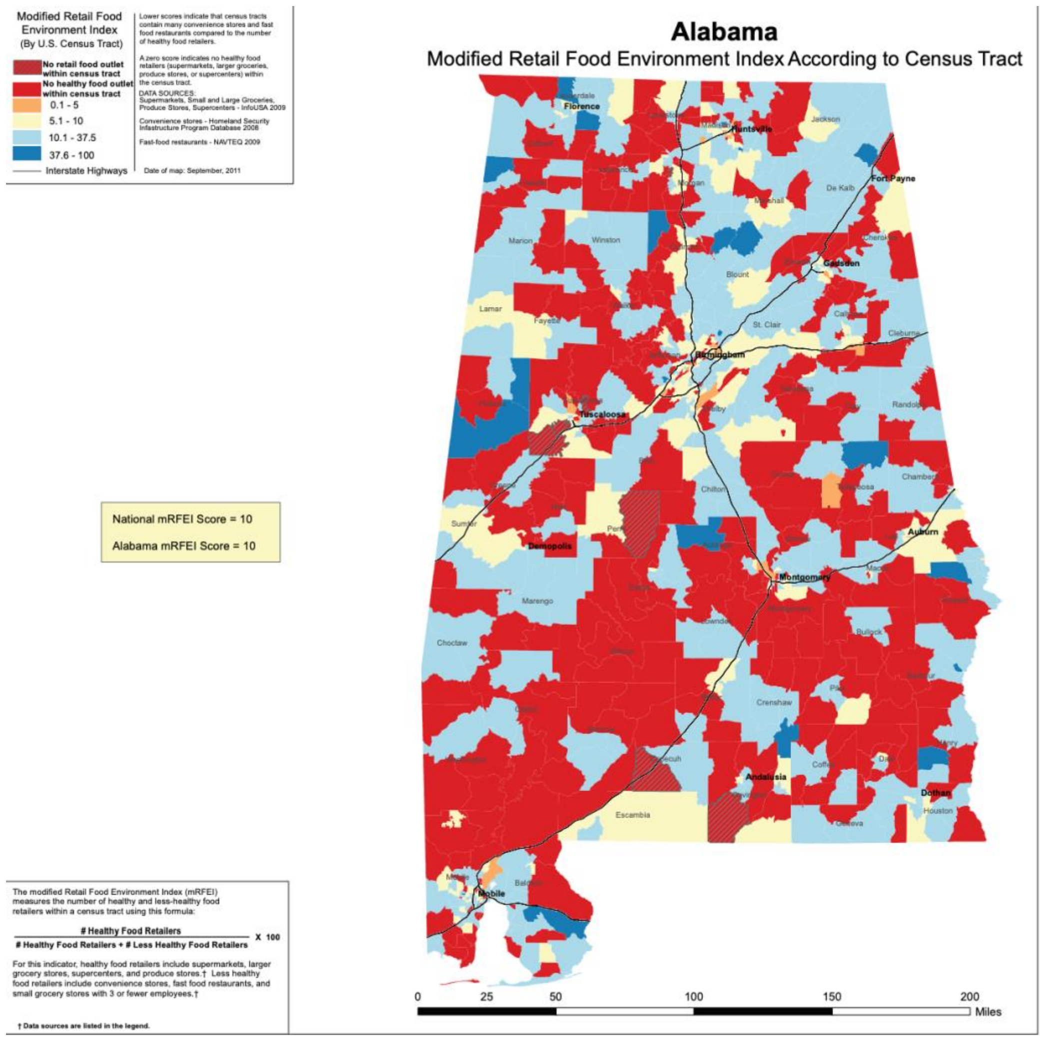
Obesity has plagued not only the United States but developing countries as well. In the 1980s, the United States and Europe experienced rising obesity rates, due to higher-calorie consumption and decreases in physical activity, while developing nations struggled with malnutrition and starvation. However, today, the obesity prevalence is now increasing in developing countries as well. A study performed in a rural Dominican Republic clinic evaluated obesity trends. The study compared obesity rates of 403 Dominican Republic children to that of the United States. The results showed a difference of only 3.6% in overweight children and 4.4% in obese children among the clinic patients (Tay, 2019). The results suggest a concern of increasing rates of obesity in the rural villages in the DR, like the previously observed trend of increasing obesity in the United States. From this study, Tay suggests creating public health interventions that begin with women and children to promote healthy eating and encourage healthy habits at a young age, in pursuit to stop the increasing obesity prevalence. It can be concluded that rural developing countries show similar obesity trends as rural locations in the United States, meaning that locations such as a Food Desert may have extremely malnourished people while also having people with high BMIs (Tay, 2019). In figure 1 below, a time-series portraying the increase of obesity over time in high school age students in the United States is shown.

Figure 1



In April 2011, the CDC's Division of Nutrition, Physical Activity, and Obesity released the *Children's Food Environment State Indicator Report* (CDC, 2011). In the report, the modified retail food environment index (mRFEI) was featured. The index measures the number of healthy and less healthy food retailers within census tracts in the United States as defined in specific types of retail stores (e.g., supermarkets, corner stores, and fast-food establishments). Out of the total number of food retailers, the mRFEI index represents the healthy percentage. With Food Deserts and Food Swamps, the index combines the two concepts into a single measure where scores of zero generally represent a Food Desert, and lower scores greater than zero represent a Food Swamp. Therefore, lower scores on the index represent census tracts that contain many convenience stores and/or fast-food restaurants. An mRFEI score of 10 means that of every 100 food retailers in that census tract there are only 10 retailers, offering healthy foods. The CDC does not provide an exact range of scores that correlate with Food Swamps. However, based on their census tract mRFEI maps, the CDC notes that lower index scores that are greater than zero are associated with Food Swamps. Appearing in Figure 2 below is an example of the state of Alabama color-coded into levels of mRFEI scores (CDC, 2011). Upon observing the CDC map, they categorize "low scores" as between zero and 37.5. However, the CDC fails to define a concrete range of scores for environments characterized as a Food Swamp.

Figure 2



The studies on Food Deserts are primarily found to be positively associated with obesity rates (Ghosh-Dastidar et al., 2014). But recent studies deem Food Swamps as a better predictor of obesity than Food Deserts (Cooksey-Stowers et al., 2017). Previous studies used mRFEI data to measure food accessibility but not as a determinant of Food Swamps (Yaneve et al., 2020, Miyakado et al., 2017). To my knowledge, there are no studies to date that perform a test between obesity and food environments using the mRFEI index as a whole.

In *Distance to Store, Food Prices, and Obesity in Urban Food Deserts,* the authors closely examine the relationship between deserts, swamps, and healthful environments (Ghosh-Dastidar et al., 2014). They conducted the Pittsburgh Hill/Homewood Research on Eating, Shopping, and Health study by asking residents to report where their food shopping was done. The surveyed participants' body weight was recorded and analyzed. Distance to store and prices was found to be positively associated with obesity and when the distance to store and food prices were jointly modeled, only prices remained significant, with higher prices predicting a lower likelihood of obesity. Although varying priced stores did not differ in availability, they significantly differed in the ratio marketing of junk foods to healthy foods. It was concluded that placing supermarkets in Food Deserts to improve access to healthy foods may not be as critical as simultaneously offering better prices for healthy foods relative to unhealthy foods, as well as marketing healthy foods more heavily than junk foods (Ghosh-Dastidar et al., 2014).

A study using mRFEI as a single measure for Food Swamps and Food Deserts further breaks down the severity of Food Swamps (Miyakado et al., 2015). The purpose of this study was to create a pilot test for further research on food environments by census tracts by using Bexar County in Texas. Miyakado et al. found that the mRFEI is a sustainable measure of identifying Food Deserts and Food Swamps in smaller geographic locations rather than the state in its entirety. Additionally, they found that mRFEI is also a user-friendly method to identify Food Deserts and Swamps by visualization through maps.

Predicting Access to Healthful Food Retailers with Machine Learning, a research study driven to better define these parameters for food environments (Dey Amin et al., 2021). Through machine learning Dey Amin and colleagues utilize demographic variables such as race, population density, urbanization, and access to vehicles, predicting Food Deserts and Swamps with approximately 72% accuracy in the independent test data. Similarly, to this research, they use cross-sectional and socioeconomic data. The study notes that Food Deserts and Food Swamps are intrinsically different, in that they are entirely separate phenomena. People residing in Food Deserts are less likely to choose healthful foods because they are simply unavailable, while people in Food Swamps are less likely to choose healthful foods due to targeted marketing of unhealthy foods or the cost of consuming healthy foods (Hager et al., 2016, Dey Amin et al., 2021). Furthermore, the researchers categorized three types of environments using the mRFEI index for Food Deserts, Food Swamps, and healthful, represented as dummy variables where Food Deserts = 0, Food Swamps are between 0.1 and median (mRFEI), and healthful environments are between median (mRFEI) and 100. While this study does not measure obesity, their choice of parameters for each environment is important to note.

Food Swamps Predict Obesity Rates Better Than Food Deserts in the United States investigates the effect of Food Swamps on obesity rates (Cooksey-Stowers et al., 2017). The study merged sociodemographic and obesity data obtained from the United States Department of Agriculture (USDA) Food Environment Atlas, the American Community Survey (ACS), and a commercial street reference dataset. Once the data was merged, geographical locations were split into two sectors: Food Swamps and Food Deserts. The study tested measures of Food Swamps and Food Deserts by utilizing the Retail Food Environment Index (RFEI) as predictors of obesity rates performed on a county level in the United States. A unique variable accounting for the number of highways exits per county was included, to explain certain food environments. I believe that this is somewhat of a limitation, due to the rural environments without highway access and the inconsistent results it derived. Reverse causality was considered because while certain food retailers exist in geographies, it should be noted that people also choose where they live and somewhat demand what type of retailers will be profitable in their area. Their findings result in supporting the position that Food Swamps are an entirely different phenomenon from Food Deserts and play a larger role in obesity rates from county to county than Food Deserts. While controlling for Food Deserts, fitness centers, and natural amenities, Food Swamps showed a statistically positive effect on obesity rates. Cooksey-Stowers et al. (2017), finds evidence that Food Swamps are distinct and separate from Food Deserts as they predict U.S. adult obesity rates more accurately than Food Deserts. Further, the study found that the correlation between Food Deserts and obesity was not statistically significant.

In one paper, *Health-Related Outcomes of New Grocery Store Interventions: A Systematic Review*, a study was reviewed where new grocery stores were placed in low accessible food areas (Abeykoon et al. 2017). Eleven records representing seven grocery store interventions were analyzed. The studies all reported fruit and vegetable consumption, but results were not consistent across all records, some showed a significant increase in consumption, while others showed no change at all. Even in areas with increased fruit and vegetable consumption, BMI and self-rated health did not improve. However, personal perception of food accessibility, the self-reported satisfaction of the neighborhoods and psychological health did show significant improvements. Quasi-experimental and longitudinal studies evaluating the impact of the introduction of new grocery stores in lacking areas surprisingly showed no change in the quality of diet or body mass index (BMI) but only perceived access to healthy food improves (Cummins et al., 2014). The results of these studies propose that by introducing healthier foods into a neighborhood, health results may not change due to continued access to unhealthy foods. The studies found that access to grocery stores does not have a direct causal relationship with obesity and adding in a new grocery store is unlikely to fix obesity rates.

Regarding the existing literature, I aim to further explore the healthful food index (mRFEI) and how the phenomenon of Food Deserts and Food Swamps relate to conditions that directly and/or indirectly affect obesity rates in adults. When only assessing Food Swamps and Food Deserts, information on the individual level of healthful access gets lost by using such broad parameters. This study will use similar regression models as Cooksey-Stowers et al. (2017), by merging the mRFEI variable breakdown used by Miyakado et al. Previous studies explore causal relationships and associations of obesity and the mRFEI index, but most are performed on a state- or county-level, some studies use census tract data, but only for one state or county. While studies in the past have used dummy variables to represent the food environments (Dey Amin et al., 2020), I believe that by relying on dummy variables, important factors of the data become minute and are lost. Therefore, for my research, I will not use dummy variables, and instead, perform tests on the healthful food index itself. This study is unique to using the Modified Food Retail Environment Index (mRFEI) as a food environment indicator without limiting geographies to categories of Food Deserts, Food Swamps, and Healthy Environments and testing obesity on a census tract-level for all census tracts in the United States. I believe through the findings of my research that a deeper understanding of the relationship between food accessibility and obesity rates will be identified, and the impact of the index on health and obesity can be derived, leading to a stronger basis on which policy and reform can be implemented to ease the severity of obesity in the United States.

DATA

The data tested uses census tract-level data by FIPS codes in the United States including the District of Columbia. The data includes demographic variables and the mRFEI to measure food retail environments in proximity to living quarters. This paper uses the Modified Retail Food Environment Index (mRFEI) from the Centers for Disease Control and Prevention (CDC). Lack of access to food retailers in communities to buy healthy foods, such as markets and grocery stores has been associated with lower quality diets that lead to increased rates of obesity (Larson et al., 2009). The mRFEI measures the number of healthy food retailers as well as the number of less healthy retailers in each area. The mRFEI is broken down by census tracts, a geographic region defined for census surveying. The modified retail environment index is calculated as follows:

Figure 3

$$\boxed{mRFEI = 100\ x \frac{\text{\# Healthy Food Retailers}}{\text{\# Healthy Food Retailers} + \text{\# Less Healthy Food Retailers}}}$$

The classification of retailers follows the North American Industry Classification Codes (NAICS), which measures typical food availability in specific types of retail stores. Healthful food retailers include supermarkets and other grocery types of stores (not convenience stores) (NAICS 445110), warehouse clubs (NAICS 452910), and farmer’s markets (NAICS 722211), smaller grocery stores, and convenience stores (NAICS 445120) within census tracts or half a mile from the census tract boundary (CDC, 2011). Small grocers or corner stores are typically those that offer a limited line of products that generally include snacks, milk, bread, and soda (NAICS 2007, CDC 2009, CDC 2011, Grimm et al. 2013).

The modified retail food environment index ranges from 0-100, where locations with a score of zero define areas with no healthy food retailers, otherwise referred to as a Food Desert. A Food Swamp is represented by mRFEI scores ranging from anywhere above zero to lower scores on the index, meaning there is a larger presence of unhealthy food retailers than healthy ones. Rather than using the classification used by Miyakado et al. where mRFEI maps: 0 (no healthy food retailers), 0.1–5 (fewer less healthy food retailers), 5.1–10, 10.1–37.5, and 37.6–100 (more healthy food retailers). To keep all explanatory variables significant, my study does not use dummy variables for each environment but instead observes the raw mRFEI index against obesity prevalence on the census tract level. When using dummy variables in a study, it is common for overlapping and redundancy in the results, by either overstating or understating certain relationships. By not using dummy variables to represent food environments, multicollinearity is less likely. Additionally, without the use of dummies, the relationship between the raw index and obesity can be tested, allowing for a better analysis of the index's ability to predict obesity.

The dependent variable in the study is obesity prevalence, represented by the variable name *obese,* and is in percentage form per census tract. Percentage of the obese population at each census tract was obtained from the Places Project (CDC, 2020). Estimates were provided by the Centers for Disease Control and Prevention (CDC), Division of Population Health, Epidemiology, and Surveillance Branch.

The independent variables are United States census-level statistics broken down by estimates of physical activity participation, vehicle access, gender, race, educational attainment, income, poverty assisted/food stamp recipients, obesity, and the index itself. The variables representing the prevalence of physical activity absence per geography and prevalence of population with no vehicle access were obtained from the American Community Survey (2015-2019). The gender, race, educational attainment, income, access to a vehicle, and cash assisted/food stamp recipients variables were also gathered from the US Census Bureau's American Community Survey (ACS). These variables allow me to control for specific characteristics within the geographical populations and control for any possible endogeneity. Within race, I focus specifically on Black and Hispanic, as the two groups are prevalent minorities in the United States. I look at the male and female percentages per population to compare the obesity rates across gender and account for any existing disparities. Non-Hispanic black women have an obesity rate of 41.8% and Hispanic women with 30.7%, these two groups lead the nation in obesity rates (CDC, 2010).

The American Community Survey reported various levels of educational attainment, the education variable in this particular dataset accounts for the percentage of the population that has attained a high school degree and equivalent (i.e., GED) or higher. Educational attainment is an important variable when looking at health disparities, as people with higher educational attainment tend to live healthier and longer lives when compared to their less-educated peers (Zajacova, Lawrence, 2018). Income was considered, using the median household income per census tract. Next, I included a variable encompassing the percentage of the population who received cash assistance for food through SNAP benefits and/or food stamps. American Community Survey data was obtained through R studio using the packages 'acs' and 'tidy census' (Glenn 2019, Walker 2021). The American Community Survey data were estimates from the years 2015 to 2019.

Lastly, the demographic variables from the American Community Survey (ACS) are merged with mRFEI scores using census tract FIPS (Federal Information Processing Center) codes from the US Census geographical population estimates (United States Census Bureau). Table 1 represents the summary statistics of the dataset, along with the source of each variable and dates of surveying.

Table 1: Summary Statistics

Statistic	Source	N	Mean	St. Dev.	Min	Max
Obese (%)	U.S. Census Bureau (2020)	51,587	32.3	7.1	10.7	59.1
mRFEI	CDC (2011)	50,824	11.3	12.0	0	100
Male (%)	ACS (2015-2019)	51,582	49.1	4.4	3.4	100
Female (%)	ACS (2015-2019)	51,582	50.9	4.4	0	96.6
Hispanic (%)	ACS (2015-2019)	51,582	16.5	21.3	0	100
Black (%)	ACS (2015-2019)	51,582	16.2	23.5	0	100
Education (%)	ACS (2015-2019)	51,582	66.7	13.4	0	160
Food Assistance (%)	ACS (2015-2019)	51,582	5.4	4.8	0	53.7
Income	ACS (2015-2019)	51,417	66,369.4	33,897	2,499	250,001
No Physical Activity (%)	ACS (2015-2019)	51,587	25.7	7.8	8.4	64.4
No Car (%)	ACS (2015-2019)	51,498	10.1	12.9	0	100
Population	ACS (2015-2019)	51,587	4,415.3	2,182.3	0	59,947

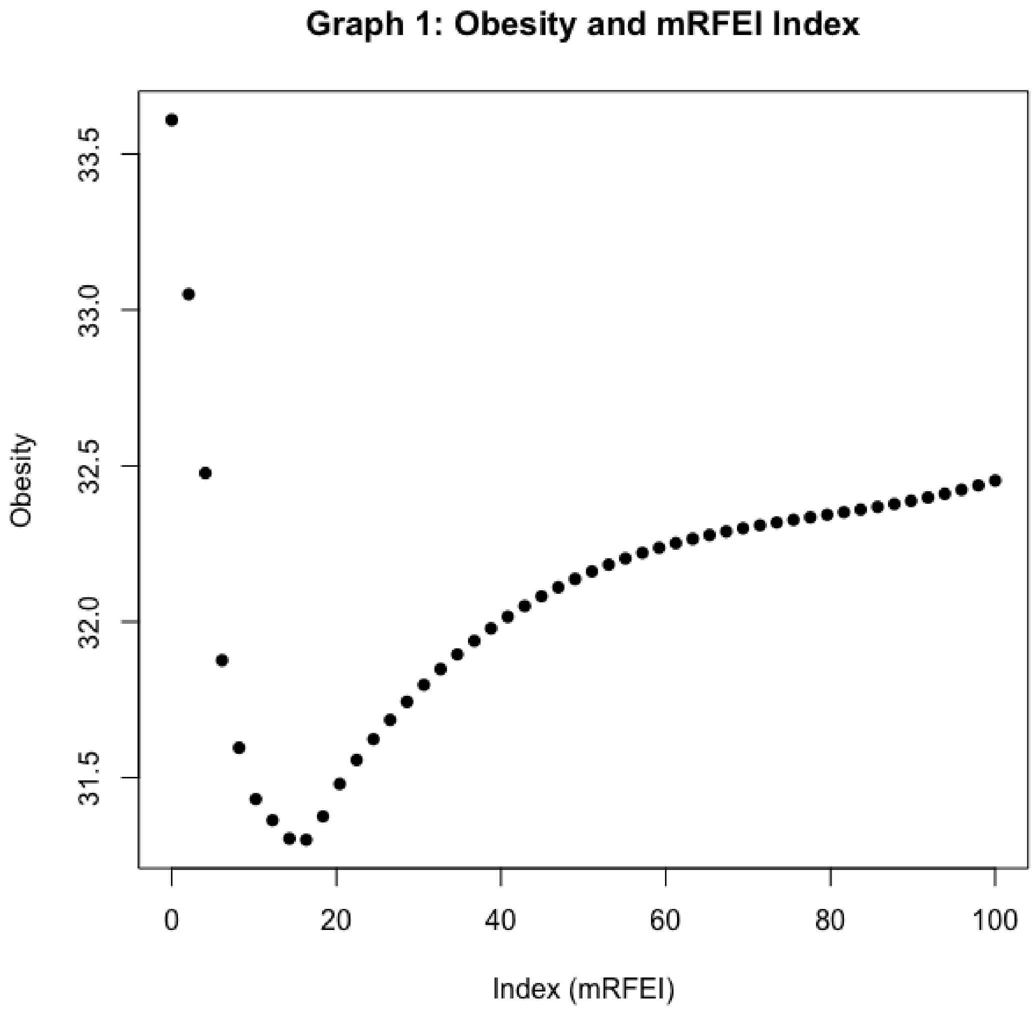
Note: Variables within this study are from the most recent datasets available at the time of research.

METHODOLOGY

Regression models are a test for measuring relationships between a response variable and a predictor variable or variables. The development of the quadratic and cubic regressions are explained in this section. The cubic or polynomial regression analysis was found to be the best fit due to the non-linearity of the relationship between obesity and the explanatory variable *mRFEI*. This helps to examine how much the dependent variable, prevalence of obesity per census tract (*Obese)* changes when change is inflicted upon the independent variable, the food environment index (*mRFEI).* The data is analyzed cross-sectionally, due to the availability of the index offered by the CDC, therefore, all variables are from the most recent data sets available. The index is obtained from the 2011 CDC Food Environment Atlas, obesity data is obtained from the 2020 CDC Places Project, and the explanatory variables are American Community Survey estimates from the years 2015-2019. R Software: a language and environment for statistical computing and graphics, was utilized for all data manipulation, regression tests, and graphing (R Core Team, 2021).

Various regression analyses were performed; quadratic, cubic, and a linear regression testing benchmark values with dummy variables for Food Swamps and Food Deserts. It is important to note that the dataset was standardized by rescaling the distribution values so that the mean observed values are represented as 0 and the standard deviation as 1.

After observing a nonlinear relationship between the index and obesity from the raw data, a locally weighted smoothing technique was implemented through the R program, utilizing the 'loess.smooth' function (Cleveland, 1979). Two scatterplots were constructed using locally weighted smoothing; (1) obesity and the index (2) obesity and the index residuals including all control variables: *pop, male, hispanic, black, educ, inc, nophysicalactivity, food assist,* and *nocar*. The results of these tests are in Graphs 1 and 2 below.



Based on the results of the locally weighted smoothed graphs, distinct inflection points were obvious, and further regressions were then constructed. After viewing the fitted plot of obesity and the mRFEI index, it is apparent that obesity decreases up until an index score between 10 and 20, and beyond that point, obesity begins to increase. With one certain inflection point, this further proved a non-linear relationship, and three quadratic regressions were then tested. The three quadratic regressions were performed with no controls, some controls, and all controls. The first quadratic regression analysis was performed with obesity prevalence being the continuous variable and the food environment index (*mRFEI)* and $mRFEI^2$ being the independent index variables. The no controls quadratic regression equation is as follows:

$$\hat{Y}(obese) = \beta_0 + \beta_1 mRFEI + \beta_2 mRFEI^2 + \epsilon$$

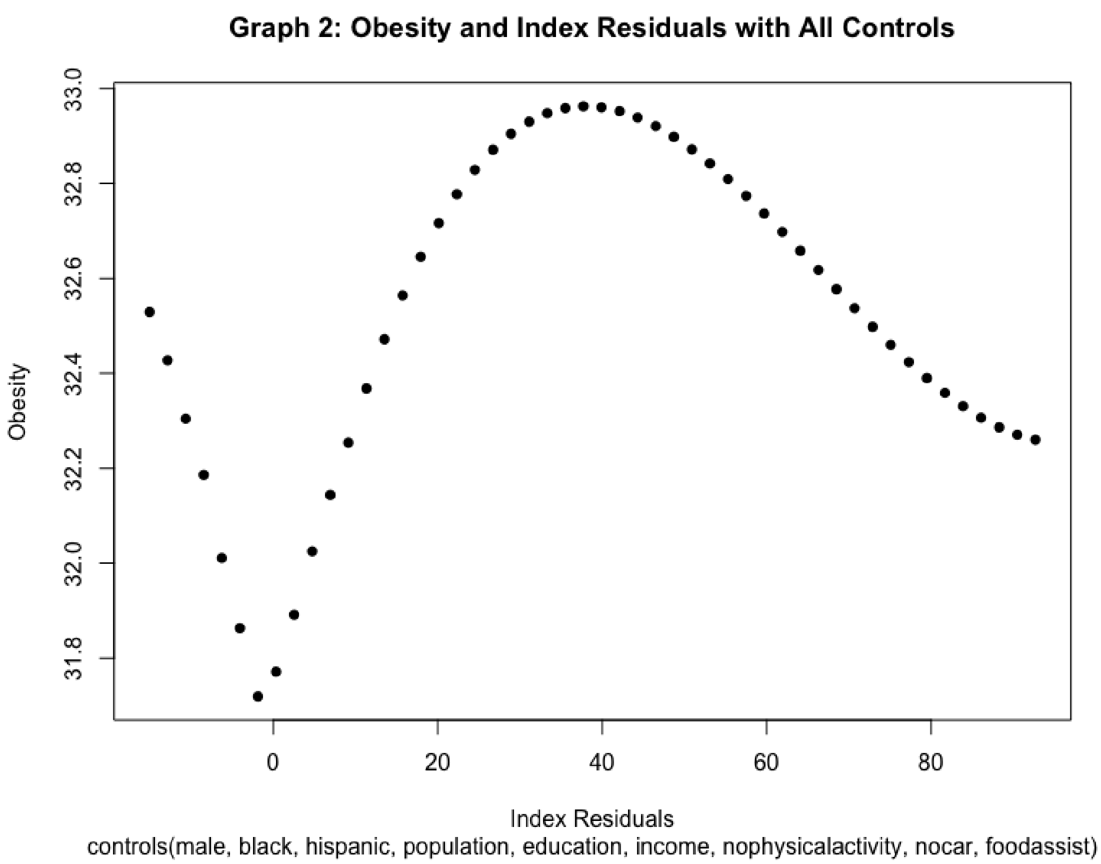
A second quadratic regression analysis was performed with obesity being the continuous variable and *mRFEI,* $mRFEI^2$, and control variables that are not directly causal to obesity; *population, gender, black, and hispanic* being the independent variables. The quadratic regression equation including some control variables is as follows:

$$\hat{Y}(obese) = \beta_0 + \beta_1 mRFEI + \beta_2 mRFEI^2 + \beta_3 pop + \beta_4 male + \beta_5 black + \beta_6 hispanic + \epsilon$$

One final quadratic regression was computed with *obese* being the dependent variable and independent variables being the index *mRFEI,* $mRFEI^2$*,* as well as all of the control variables. The quadratic regression equation including all controls is as follows:

$$\hat{Y}(obese) = \beta_0 + \beta_1 mRFEI + \beta_2 mRFEI^2 + \beta_3 pop + \beta_4 male + \beta_5 black + \beta_6 hispanic + \beta_7 inc + \beta_8 foodassist + \beta_9 nocar + \beta_{10} nophysicalactivity + \beta_{11} educ + \epsilon$$

The quadratic regression model was best fitted when including all controls, as much of the explanatory power is seated in the control variables. The regressions explain a negative relationship between the index and obesity and show a positive relationship between the squared index and obesity. The quadratic model is efficient for the sign change from Food Deserts where the index is 0 and for Food Swamps where the index is greater than zero but fails to explain a second sign change in where healthful environments exist, environments with high index scores and more healthy food retailers. In pursuit of the best fitting model, one where all variables are statistically significant and each food environment within the index is accounted for, an additional locally weighted smoothed scatter plot was created. This scatter plot implements the use of residuals. Graph 2 contains the smoothed regression of obesity on the index residuals containing all variables.



The results from Graph 2 provide support for the idea that the indicators for food environments are each a separate phenomenon and are correlated with different levels of obesity rates. Knowing that there are three separate food environments, Food Desert, Food Swamp, and healthful, the loess graph increases certainty that each environment has a differing effect on obesity rates. Upon analyzing the graph, levels of obesity rates are extremely varied when the residual index score is 0, an increase of obesity rates occur above zero and just below approximately 40, geographies with an mRFEI score larger than 40 experience a decrease in obesity rates. Furthermore, a set of cubic regressions were performed, this time, in an attempt to capture all three sign changes as depicted in the residual locally weighted smoothed plot in Graph 2 above.

Similarly, to the quadratic regressions, the cubic regressions also were performed with no controls, some controls, and all controls. The first cubic regression analysis was performed with obesity being the continuous variable and the food environment index (*mRFEI*), $mRFEI^2$, $mRFEI^3$ being the independent variables. The results of the following cubic regressions are shown in Table 4. The no controls cubic regression equation is as follows:

$$\hat{Y}(obese) = \beta_0 + \beta_1 mRFEI + \beta_2 mRFEI^2 + \beta_3 mRFEI^3 + \epsilon$$

A second cubic regression analysis was performed with *obese* being the continuous variable and *mRFEI*, $mRFEI^2$, $mRFEI^3$, and control variables that are not directly associated with obesity; *population, gender, black, and hispanic* being the independent variables. The cubic regression equation including some control variables is as follows:

$$\hat{Y}(obese) = \beta_0 + \beta_1 mRFEI + \beta_2 mRFEI^2 + \beta_3 mRFEI^3 + \beta_4 pop + \beta_5 male + \beta_6 black + \beta_7 hispanic + \epsilon$$

One final cubic regression was computed with *obese* being the dependent variable and independent variables being the index $mRFEI, mRFEI^2, mRFEI^3$, and all the control variables. The cubic regression equation including all controls is as follows:

$$\hat{Y}(obese) = \beta_0 + \beta_1 mRFEI + \beta_2 mRFEI^2 + \beta_3 mRFEI^3 + \beta_4 pop + \beta_5 male + \beta_6 black + \beta_7 hispanic + \beta_8 inc + \beta_9 foodassist + \beta_{10} nocar + \beta_{10} nophysicalactivity + \beta_{11} educ + \epsilon$$

The cubic regression model succeeded in representing three sign changes in the index explaining obesity. Initially, the index is negatively associated with obesity, at some point, the index becomes positively associated with obesity, and at a third point, the index again becomes negatively associated with obesity. These results closely mirror the relationship depicted in Graph 2. It is important to note that the index variable *mRFEI* was standardized using the z-score in the quadratic and cubic regressions, allowing regression results to be more easily interpreted when observing the relationship between obesity and the modified retail food environment index. Ultimately, the regression model was chosen on goodness-of-fit terms as well as the similarity between regression results and the residual 'LOESS' graph. Many studies have used the food index and tested environments by including dummy variables to represent Food Swamps and Food Deserts (Miyakado et. al., 2015, Dey Amin et al. 2020). I use their results as a benchmark to test if the index itself predicts obesity similarly as food environments do when set into certain parameters i.e. Food Swamps, Deserts.

RESULTS

Studies mentioned above that use strict categorization of food environments limit their explanatory power by using dummy variables and may yield biased and inconsistent results due to not providing a thorough account of the level of the food environment. To test the validity of the index's power to explain obesity, I conducted two separate benchmark tests using dummy variables. The first test uses the same parameters for Food Deserts, Food Swamps, and Healthful Environments as the CDC's original maps do in the Food Environment State Indicator Report in 2011. The categorizations for food environments by the CDC under the mRFEI index are Food Deserts = 0, Food Swamps = 0.1-37.5, and healthy environments = >37.5-100. Other studies such as Miyakado et al. have used this index but did not measure obesity, instead they used the index to test if it was a good method for mapping such environments. Depicted in Table 2 below, are regressions using the CDC's parameters for Food Deserts and Food Swamps, similar to the quadratic and cubic regressions in my study they are run with no, some, and then all controls. The regression with all controls is as follows:

$$Obese = -0.246FoodDesert - 0.421FoodSwamp + 0.00003pop + 0.096male + 0.056black - 0.023hispanic - 0.00003inc + 0.281foodassist - 0.147nocar + 0.564nophysicalactivity + 0.025educ + 13.504$$

Table 2: Dummy Variable Regressions (CDC Parameters)

	Dependent Variable: Obesity Prevalence		
	obese		
	No Controls	Some Controls	All Controls

\`Food Desert\`	1.596***	-0.261	-0.302***
	(0.196)	(0.164)	(0.093)
\`Food Swamp\`	-0.503***	-2.498***	-1.168***
	(0.189)	(0.158)	(0.090)
Population		-0.0003***	0.00003***
		(0.00001)	(0.00001)
Male		0.080***	0.096***
		(0.006)	(0.004)
Black		0.162***	0.056***
		(0.001)	(0.001)
Hispanic		0.048***	-0.023***
		(0.001)	(0.001)
Income			-0.00003***
			(0.00000)
Food Assistance			0.281***
			(0.005)
No Car			-0.147***
			(0.001)
No Physical Activity			0.564***
			(0.003)
Education			0.025***
			(0.001)
Constant	32.240***	28.209***	13.504***
	(0.185)	(0.356)	(0.270)
Observations	50,824	50,819	50,690
R^2	0.015	0.320	0.781
Adjusted R^2	0.015	0.320	0.781

Residual Std. Error	7.073 (df = 50821)	5.876 (df = 50812)	3.331 (df = 50678)
F Statistic	398.349*** (df = 2; 50821)	3,994.373*** (df = 6; 50812)	16,421.930*** (df = 11; 50678)
Note:			*p**p***p<0.01

The results of the linear regression testing Food Deserts and Food Swamps as defined by CDC parameters has a high R^2 score but greatly differs from findings of other studies. Though Cooksey-Stowers (2017) used the RFEI instead of the modified version, they found that both Food Deserts and Food Swamps were positively associated with obesity. Here, a negative association is shown between these food environments and obesity. However, in the regression with no control variables added, Food Deserts is found to be positively associated with obesity. This test fails to accurately represent the effect of environments on obesity, as nearly no study found that Food Deserts or Swamps decrease obesity rates under these parameters. This is evidence of inconsistencies through the use of parameters and proves that categorizing the mRFEI into groups produces inconsistent results.

A second dummy variable regression was conducted using the parameters used in Dey Amin et al. (2021) research on predicting access to food retailers with machine learning. They used the mRFEI to categorize food environments, where Food Deserts (mRFEI = 0), Food Swamp (>0, median(mRFEI)), and healthy environments (>median(mRFEI), 100). Similarly to the CDC parameters dummy variable to test, the regressions below are run with no, some, and then all control variables. The regression with all controls is as follows and results from the dummy variable benchmark test using median parameters are displayed in Table 3 below.

$$Obese = 0.708FoodDesert - 0.360FoodSwamp + 0.00002pop + 0.098male + 0.056black - 0.023hispanic - 0.00003inc + 0.279foodassist - 0.145nocar + 0.564nophysicalactivity + 0.025educ + 12.402$$

Table 3: Dummy Variable Regressions (Median Parameters)

	Dependent Variable: Obesity Prevalence		
		obese	
	No Controls	Some Controls	All Controls
Food Desert	2.233^{***}	1.686^{***}	0.708^{***}
	(0.078)	(0.066)	(0.038)
Food Swamp	0.475^{***}	-1.439^{***}	-0.360^{***}
	(0.077)	(0.065)	(0.038)
Population		-0.0003^{***}	0.00002^{***}
		(0.00001)	(0.00001)
Male		0.084^{***}	0.098^{***}
		(0.006)	(0.004)
Black		0.164^{***}	0.056^{***}
		(0.001)	(0.001)
Hispanic		0.051^{***}	-0.023^{***}
		(0.001)	(0.001)
Income			-0.00003^{***}
			(0.00000)
Food Assistance			0.279^{***}
			(0.005)
No Car			-0.145^{***}
			(0.002)
No Physical Activity			0.564^{***}

			(0.003)
Education			0.025***
			(0.001)
Constant	31.603***	26.033***	12.402***
	(0.044)	(0.319)	(0.253)
Observations	50,824	50,819	50,690
R^2	0.016	0.324	0.781
Adjusted R^2	0.016	0.324	0.781
Residual Std. Error	7.071 (df = 50821)	5.862 (df = 50812)	3.334 (df = 50678)
F Statistic	414.189*** (df = 2; 50821)	4,052.447*** (df = 6; 50812)	16,390.020*** (df = 11; 50678)
Note:			*p**p***p<0.01

The results in Table 3 above show a slight improvement, in that Food Deserts here are positively associated with obesity. The positive Food Desert relationship with obesity is consistent with other studies but varies in the relationship with Food Swamps (Dey Amin et al, 2021, Cooksey-Stowers et al, 2017). The results from the dummy variable tests using two different parameters for food environments are inconsistent with past studies as well as with each other. When tes median parameters, both deserts and swamps are positively correlated with obesity, but change once control variables are included. Logically, in areas with a high density of unhealthy food retailers, obesity would increase, not decrease. Additionally, the direction of the relationship between each food environment and obesity depends on how the parameters are defined, indicating that the use of dummy variables is not an efficient methodology. This further supports the claim that categorizing the index into broad parameters yields inconsistent results for predicting obesity. Therefore a more in-depth approach to test environments and obesity was implemented through the utilization of quadratic and cubic regressions.

In table 4 below are the results from the quadratic regressions with no controls, some controls, and then all control variables. This model closely resembles the loess model in Graph 1, where one distinct inflection point is shown.

$$Obese = -0.110mRFEI + 0.057mRFEI^2 + 0.00001pop + 0.100male + 0.055black - 0.025hispanic - 0.00003inc + 0.281foodassist - 0.150nocar + 0.571nophysicalactivity + 0.026educ + 12.237$$

Table 4: Quadratic Regressions

	Dependent Variable: Obesity Prevalence		
	obese		
	No Controls	Some Controls	All Controls
mRFEI	-0.954***	-0.147***	-0.110***
	(0.044)	(0.037)	(0.021)
mRFEI^2	0.173***	0.104***	0.057***
	(0.011)	(0.009)	(0.005)
Population		-0.0004***	0.00001*
		(0.00001)	(0.00001)
Male		0.092***	0.100***
		(0.006)	(0.004)
Black		0.160***	0.055***
		(0.001)	(0.001)
Hispanic		0.043***	-0.025***
		(0.001)	(0.001)
Income			-0.00003***

			(0.00000)
Food Assistance			0.281***
			(0.005)
No Car			-0.150***
			(0.001)
No Physical Activity			0.571***
			(0.003)
Education			0.026***
			(0.001)
Constant	32.070***	25.929***	12.237***
	(0.033)	(0.324)	(0.254)
Observations	50,824	50,819	50,690
R^2	0.009	0.304	0.779
Adjusted R^2	0.009	0.304	0.779
Residual Std. Error	7.095 (df = 50821)	5.946 (df = 50812)	3.349 (df = 50678)
F Statistic	238.523*** (df = 2; 50821)	3,699.076*** (df = 6; 50812)	16,201.840*** (df = 11; 50678)
Note:			*p**p***p<0.01

The quadratic regression model was best fitted when including all controls, as expected the control variables account for a large portion of the explanatory power in obesity rates. A one standard deviation increase in the index ($mRFEI$) generates a 0.011 percentage point decrease in obesity rates. An increase of one standard deviation of the index squared ($mRFEI^2$) generates a 0.057 percentage point increase in obesity rates. The index demonstrates a small negative correlation with obesity because various underlying factors in environments can lead to individuals being obese or underweight. The index squared shows a positive relationship with obesity and closely mirrors the inflection point from Graph 1. As mentioned before, the second inflection point, where healthful food environments exist, was not accounted for with this model.

A cubic regression was then performed to better match the relationship between obesity and the index, based on Graph 2 including residuals. In table 5 are the results of the cubic regressions, which further explains the validity of the residual loess plot.

$$Obese = -0.246mRFEI + 0.421mRFEI^2 - 0.051mRFEI^3 + 0.00002pop + 0.097male + 0.056black - 0.024hispanic - 0.00003inc + 0.281foodassist - 0.149nocar + 0.566nophysicalactivity + 0.025educ + 12.282$$

Table 5: Cubic Regressions

	Dependent Variable: Obesity Prevalence		
		obese	
	No Controls	Some Controls	All Controls
mRFEI	-1.192***	-0.461***	-0.246***
	(0.046)	(0.039)	(0.022)
mRFEI^2	0.806***	0.942***	0.421***
	(0.037)	(0.031)	(0.018)
mRFEI^3	-0.088***	-0.116***	-0.051***
	(0.005)	(0.004)	(0.002)
Population		-0.0003***	0.00002***
		(0.00001)	(0.00001)
Male		0.084***	0.097***
		(0.006)	(0.004)
Black		0.162***	0.056***
		(0.001)	(0.001)

Hispanic		0.046^{***}	-0.024^{***}
		(0.001)	(0.001)
Income			-0.00003^{***}
			(0.00000)
Food Assistance			0.281^{***}
			(0.005)
No Car			-0.149^{***}
			(0.001)
No Physical Activity			0.566^{***}
			(0.003)
Education			0.025^{***}
			(0.001)
Constant	31.688^{***}	25.630^{***}	12.282^{***}
	(0.039)	(0.321)	(0.253)
Observations	50,824	50,819	50,690
R^2	0.015	0.314	0.781
Adjusted R^2	0.015	0.314	0.780
Residual Std. Error	7.073 (df = 50820)	5.902 (df = 50811)	3.334 (df = 50677)
F Statistic	265.438^{***} (df = 3; 50820)	$3,329.139^{***}$ (df = 7; 50811)	$15,020.800^{***}$ (df = 12; 50677)
Note:			*p**p***p<0.01

The three cubic models, no controls, some controls, and all controls each produced correlation coefficients proving to be statistically significant across all variables. Additionally, the cubic regression model with all controls has the highest R^2 score of all the models, 0.781, further indicating that the model is a better fit than the quadratic regressions. A one standard deviation increase in the index ($mRFEI$) generates a 0.246 percentage point decrease in obesity rates. An increase of one standard deviation of the index squared ($mRFEI^2$) generates a 0.421 percentage point increase in obesity rates. An increase of one standard deviation of the index cubed ($mRFEI^3$) generates a 0.051 percentage point decrease in obesity. These results are representative of the sign changes in the relationship between obesity and the index, visible in the residual loess Graph 2. The sign changes are commensurate with the hypothesis and construct a more meaningful understanding of how obesity rates are closely correlated with food environments and that the CDC's mRFEI index is efficient in predicting them. Without the use of dummy variables, we can learn more from using the raw index as we get a better understanding of the complicated relationship between the index and obesity.

DISCUSSION

I previously mentioned the findings of Cooksey-Stowers et al. and how they found Food Swamps to be a stronger predictor of obesity rates at the county level than Food Deserts are. While my findings are consistent with those results, my study differs in the food environment index used, where I used mRFEI they used the original RFEI index and dummy variables. The researchers in the study found that Food Swamps and Deserts are positively associated with obesity rates, however, my benchmark test results depicted in Tables 2 and 3 show a negative correlation between the environments and obesity rates and only show a positive correlation between Food Deserts and obesity when using median parameters. Furthermore, the researchers fail to create parameters for food swamps within their index, and instead use the entire RFEI range as a Food Swamp measure and create a separate index for Food Deserts. Our studies differ in the fundamental basis of our research questions, Cooksey-Stowers et al. primarily aimed to decipher between the Food Desert and Food Swamp phenomena. My results are consistent with proving that each Food Desert, Food Swamp, and Healthful food environments are individual phenomena, but instead did so by using one scale and proving that the scale can accurately predict obesity in relation to food environments.

The results in Table 5 help support the relationship depicted in Graph 2. The graph shows an initial negative correlation between mRFEI and obesity where the index is approximately zero, then at a certain point after zero, the index becomes positively associated with obesity, after another point in the index, the index becomes negatively associated with obesity again. This is also represented by the correlation in the cubic regression, a one standard deviation increase in the index ($mRFEI$) generates a 0.246 decrease in obesity, when the index squared ($mRFEI^2$) is increased by one standard deviation it generates a 0.421 increase in obesity, lastly, a standard deviation increase in the index cubed ($mRFEI^3$) generates a 0.051 decrease in obesity.

Graph 2 can be used to estimate the parameters of each food environment. In observation, I interpret the initial downward-sloping relationship between the index and obesity to be that of a Food Desert. Logically it is appropriate to assume that when food availability is extremely scarce, obesity will be less prevalent, the negative correlation coefficient supports that claim. However, looking closely at the graph, obesity still exists in low-access food areas. A food environment that has limited access to food can still have obesity, for example people could have no access to food causing malnourishment, while others can only have the financial means to purchase unhealthy items or government-approved items that are unhealthy. It should be noted that a Food Desert is positively associated with high poverty rates (Cooksey-Stowers et al., 2017). The additional criteria for accounting for poverty in food deserts can cause endogeneity in qualitative interpretation.

After the first inflection point in Graph 2, it can be inferred that the positive correlation between the index and obesity at that point is correlated with Food Swamps. At the second inflection point where the sign flips, it can be inferred that the negative correlation between the index and obesity rates is representative of healthy food environments. It would be assumed that with a larger portion of healthy food retailers compared to non-healthy, obesity declines. This means that the parameters used by (Miyakado et al., 2015, CDC, 2011) represent each environment respectively, but that each environment also has a varying, corresponding relationship with obesity. This draws to the conclusion that each food environment is, in fact, a separate phenomenon and that the mRFEI index is a good indicator of such environments. More can be interpreted by using the index as a whole and gives a more thorough analysis of the relationship between food availability and obesity than simply relying on large parameters of deserts and swamps.

CONCLUSION

The CDC's modified retail food environment index that was created in 2011, has been utilized to identify the healthiness of food retail in census tracts in the United States. Understanding the parameters in which a Food Desert, Food Swamp, and healthful food environment affect obesity is difficult to decipher and categorize by one index. Predicting obesity by food environments can be made possible through the utilization of the CDC's mRFEI. However, defining each food environment into levels of the index proves difficult when assessing obesity. It is valid that each Food Desert, Swamp, and a healthy environment are intrinsically different, but should not be confined to distinct parameters using the index.

This study found the mRFEI index to be a good predictor of obesity rates at varying food environment levels. Food Swamps and healthful food environments are better predictors for obesity using the index than Food Deserts. Food Deserts tend to vary greatly due to the lack of available food combined with the high poverty rates of those areas. Moreover, people in Food Deserts could be malnourished due to extreme poverty or obese due to only being able to afford unhealthy, junk foods. Though this seems like a setback for the index, it can prove to be a good indicator for policy in the future. This study further provides evidence for Cooksey-Stowers et al.'s findings that Food Swamps predict obesity rates more efficiently than Food Deserts. Food Deserts lack certainty and with varying environmental aspects, obesity is hard to predict.

By using the index instead of metaphors for defining certain environments, policymakers and interventionists can better assess what can be done to intervene and improve obesity rates respectively. A one-size-fits-all approach has proved to be ineffective with past policies and interventions. By better understanding the problems that occur and the causes of obesity in each given environment, more personalized approaches to decrease obesity rates may prove to be more effective. I believe this study better predicted the food environment's relationship with obesity when using the raw index. Past studies tested primarily county or state-level data that was merged from census-tract level statistics and further combined into dummy variables. When using these dummy variables, the direction of the relationship between each food environment and obesity depends on how the parameters are defined, indicating that the use of dummy variables is not an efficient methodology. Categorizing the index into broad parameters leads to inconsistent results in predicting obesity rates. Here, census-tract level data was used without dummy variables and reduced the possibility for data to become minute or lost as well as allowing the full index to be observed without preset parameters in place. An untestable variable such as personal choice is difficult to account for which is why obesity is hard to predict.

Healthy options may be available but individuals still have access to unhealthy options that are possibly cheaper. To implement policy to decrease obesity, each individual census tracts' mRFEI score should be used rather than a broad approach with parameters. It is apparent that the United States varies from location to location in food access, as do states and even counties, but census tracts allow for interventions to take a more focused approach and to truly understand a specific environment and assess what can be done to ease the severity of obesity respectively.

Overall, this study allows a deeper understanding of the relationship between food environments and obesity. It is apparent that each food environment is different and should be treated as such. The non-linear relationship between the index and obesity further proves that each environment is a separate phenomenon. By using the raw index, more insight is gained into the nuanced relationship between healthy food availability and obesity.

SUGGESTIONS FOR FURTHER RESEARCH

Possible policy implementation can use the index as a guideline to better understand individual geographies based on the severity of food accessibility. More tailored policies respectful to each environment may prove to be more effective in reducing obesity rates, rather than one general policy for the entire United States. Furthermore, the government can offer tax incentives and subsidies to businesses, to attract certain types of healthy retailers to specific areas in need. An example of this would be offering a healthy food business such as a farmer's market a tax exemption to make it more profitable to operate such a business in the area of interest.

Additional policy reform could restructure SNAP benefits and/or WIC, changing them to only cover healthier foods and disallow use at small corner or convenience stores that typically carry energy-dense snack foods. By restricting usage of food assistance to only healthy options, lower-income individuals will be less likely to be persuaded by unhealthy food marketing.

This study could be improved by further narrowing down the index and possibly finding distinct parameters for each food environment. With the certainty of the relationship between obesity with the index, policymakers and interventionists can draw in focus on specific areas and provide particular changes to improve obesity rates respectively. The inflection points in Graph 2 closely mirror the parameters set by the CDC, however, obesity was not a factor in their index parameters. Furthermore, this leads to a question, how did the CDC define the parameters as previously mentioned in the Food Environment State Indicator Report (CDC, 2011)? Since the inflection points in the residual fitted model are similar to the CDC's parameters provided for each environment, how were categorization for terms Food Desert, Food Swamp, and Healthy environment configured?

Additionally, a question of reverse causality arises, while obese individuals may live in an area with a high density of unhealthy foods, the unhealthy retailers could have possibly moved to the areas with high obesity rates, or the unhealthy retailers could have caused the obesity.

REFERENCES

Abeykoon, A., Egnler-Stringer, R., Muhajarine, N. (2017, June 1). *Health-related outcomes of new grocery store interventions: A systematic review*. Public health nutrition. Retrieved September 17, 2021, from https://pubmed.ncbi.nlm.nih.gov/28566095/.

Amin, M. D., Badruddoza, S., & McCluskey, J. J. (2021, February). *Predicting access to healthful food retailers with machine learning*. Food policy. Retrieved September 10, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7564312/.

Behrendt, Stefan. "Add Standardized Regression Coefficients to LM-Objects." *The Comprehensive R Archive Network*, Comprehensive R Archive Network (CRAN), 28 Dec. 2014, https://cran.r-project.org/web/packages/lm.beta/index.html.

Caswell, J. A. (2013, April 23). *Individual, household, and environmental factors affecting food choices and access*. Supplemental Nutrition Assistance Program: Examining the Evidence to Define Benefit Adequacy. Retrieved October 1, 2021, from https://www.ncbi.nlm.nih.gov/books/NBK206912/.

CDC. (2006, October). *CDC Newsroom press release October 26, 2006*. Centers for Disease Control and Prevention. Retrieved September 4, 2021, from https://www.cdc.gov/media/pressrel/r061026.htm.

Centers for Disease Control and Prevention. (2009, August 25). *State indicator report on fruits and vegetables, 2018*. Centers for Disease Control and Prevention. Retrieved August 5, 2021, from https://www.cdc.gov/nutrition/data-statistics/2018-state-indicator-report-fruits-vegetables.html.

Centers for Disease Control and Prevention. (2010, August 3). *Adult obesity*. Centers for Disease Control and Prevention. Retrieved September 14, 2021, from https://www.cdc.gov/vitalsigns/adultobesity/index.html.

Centers for Disease Control and Prevention. (2011). *Access to healthier food retailers - United States, 2011*. Centers for Disease Control and Prevention. Retrieved August 5, 2021, from https://www.cdc.gov/mmwr/preview/mmwrhtml/su6203a4.htm#:~:text=In%202011%2C%2030.3%25%20of%20census,the%202010%20continental%20U.S.%20population.

Centers for Disease Control and Prevention. (2011). *Census tract level state maps of the Modified Food Environment Index (mrfei)*. Centers for Disease Control and Prevention. Retrieved October 17, 2021, from http://stacks.cdc.gov/view/cdc/61367.

CDC. "Places: Local Data for Better Health, Census Tract Data 2020 Release." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 4 Jan. 2021, https://chronicdata.cdc.gov/500-Cities-Places/PLACES-Local-Data-for-Better-Health-Census-Tract-D/cwsq-ngmh/data.

Cleveland, W. S. (1979). *Robust locally weighted regression ... - Iowa State University*. Robust locally weighted regression and smoothing scatterplots. Retrieved 2021, from http://home.eng.iastate.edu/~shermanp/STAT447/Lectures/Cleveland%20paper.pdf.

Cooksey-Stowers, K., Schwartz, M. B., & Brownell, K. D. (2017, November 14). *Food swamps predict obesity rates better than food deserts in the United States*. International journal of environmental research and public health. Retrieved October 17, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5708005/.

Cummins, S., Flint, E., & Matthews, S. A. (2014). New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity. *Health Affairs*, *33*(2), 283–291. https://doi.org/10.1377/hlthaff.2013.0512

Dey Amin, Madhurima; Badruddoza S; McCluskey. (2021) "Predicting Access to Healthful Food Retailers with Machine Learning." *Food Policy*, U.S. National Library of Medicine, https://pubmed.ncbi.nlm.nih.gov/33082618/.

Drewnowski, A., Aggarwal, A., Hurvitz, P. M., Monsivais, P., & Moudon, A. V. (2012, August). *Obesity and supermarket access: Proximity or price?* American journal of public health. Retrieved October 2, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3464835/.

Fonge, Yaneve N., et al. "Examining the Relationship between Food Environment and Gestational Diabetes." *American Journal of Obstetrics & Gynecology MFM*, vol. 2, no. 4, 2020, p. 100204., https://doi.org/10.1016/j.ajogmf.2020.100204.

Ghosh-Dastidar, B., Cohen, D., Hunter, G., Zenk, S. N., Huang, C., Beckman, R., & Dubowitz, T. (2014). Distance to store, food prices, and obesity in urban food deserts. *American Journal of Preventive Medicine*, *47*(5), 587–595. https://doi.org/10.1016/j.amepre.2014.07.005

Glenn, Ezra Haber. “Download, Manipulate, and Present American Community Survey and Decennial Data from the US Census [R Package ACS Version 2.1.4].” *The Comprehensive R Archive Network*, Comprehensive R Archive Network (CRAN), 19 Feb. 2019, https://cran.r-project.org/web/packages/acs/.

Gustafson, A. A., Lewis, S., Wilson, C., & Jilcott-Pitts, S. (2012). Validation of food store environment secondary data source and the role of neighborhood deprivation in Appalachia, Kentucky. *BMC Public Health*, *12*(1). https://doi.org/10.1186/1471-2458-12-688

Haspel, T. (2018, August 27). *Perspective | food deserts don't cause obesity. but that doesn't mean they don't matter*. The Washington Post. Retrieved August 15, 2021, from https://www.washingtonpost.com/lifestyle/food/food-deserts-dont-cause-obesity-but-that-doesnt-mean-they-dont-matter/2018/08/22/df31afc0-a61b-11e8-a656-943eefab5daf_story.html.

Hlavac, Marek. “Package Stargazer.” *CRAN*, Comprehensive R Archive Network (CRAN), 30 May 2018, https://cran.r-project.org/web/packages/stargazer/index.html.

Lawrence, E. M. (2018, April 1). *The relationship between education and health: Reducing disparities through a contextual approach*. Annual review of public health. Retrieved October 2, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5880718/.

Larson, N., Story, M., Nelson, M. (2009, January). *Neighborhood Environments: Disparities in access to healthy foods in the U.S*. American journal of preventive medicine. Retrieved October 2, 2021, from https://pubmed.ncbi.nlm.nih.gov/18977112/.

Miyakado, H. (n.d.). *Modified Retail Food Environment Index (mrfei) map by census tract identifying food deserts and food swamps of Bexar County, Texas, 2015*. cste.confex.com. Retrieved August 1, 2021, from https://cste.confex.com/cste/2016/webprogram/Paper6145.html.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rose, D., Bodor, N., Swalm, C., Rice, J., Farley, T., & Hutchinson, P. (2009). *1 deserts in New Orleans? illustrations of urban food* ... Retrieved September 8, 2021, from https://www.researchgate.net/publication/237579148_1_Deserts_in_New_Orleans_Illustrations_of_Urban_Food_Access_and_Implications_for_Policy.

Salinas, J., Abdelbary, B., Klaas, K., & Tapia, B. (2014, June). *Socioeconomic context and the food landscape in Texas: Results from Hotspot Analysis and* ... Retrieved October 17, 2021, from https://www.researchgate.net/publication/262681982_Socioeconomic_Context_and_the_Food_Landscape_in_Texas_Results_from_Hotspot_Analysis_and_BorderNon-Border_Comparison_of_Unhealthy_Food_Environments/fulltext/03be456c0cf20a342871f073/Socioeconomic-Context-and-the-Food-Landscape-in-Texas-Results-from-Hotspot-Analysis-and-Border-Non-Border-Comparison-of-Unhealthy-Food-Environments.pdf.

Scrucca, Luca. “Model-Based Sliced Inverse Regression [R Package Msir Version 1.3.3].” *The Comprehensive R Archive Network*, Comprehensive R Archive Network (CRAN), 15 Dec. 2020, https://cran.r-project.org/web/packages/msir/index.html.

United States Census Bureau. (2007). *North American Industry Classification System - NAICS*. United States Census Bureau. Retrieved August 10, 2021, from https://www.census.gov/naics/.

U.S. Census Bureau (2019). *Selected housing characteristics, 2015-2019 American Community Survey 5-year estimates*. Retrieved August 10, 2021, from https://www.census.gov/programs-surveys/acs/data.html.plac

Walker, Kyle. “Load US Census Boundary and Attribute Data as 'Tidyverse' and 'Sf'-Ready Data Frames [R Package Tidycensus Version 1.1].” *The Comprehensive R Archive Network*, Comprehensive R Archive Network (CRAN), 23 Sept. 2021, https://cran.r-project.org/web/packages/tidycensus/index.html.

Zajacova, Anna, and Elizabeth M. Lawrence. “The Relationship between Education and Health: Reducing Disparities through a Contextual Approach.” *Annual Review of Public Health*, vol. 39, no. 1, 2018, pp. 273–289., https://doi.org/10.1146/annurev-publhealth-031816-044628.