Research Article

Quantitative Modeling of Economic Inequalities and the Prevalence of Food Deserts in Urban and Rural U.S. Census Tracts

Yun Koo

Storm King School, 314 Mountain Rd, Cornwall-On-Hudson, NY 12520, United States

ABSTRACT

This study investigates the relationship between economic inequality and the prevalence of food deserts among 300 U.S. census tracts with quantitative model, focusing on the prediction of food inaccessibility through poverty, unemployment, and lack of transportation access. This study seeks to answer the research question about whether structural socioeconomic factors influence the prevalence of food deserts in the United States, and whether their effects may differ between urban and rural areas. This study hypothesized that each factor of poverty, unemployment, and the lack of transportation access would influence food accessibility, along with compounded effects from multiple disadvantages. Multiple regression model was generated by using data from the USDA Food Access Research Atlas and the American Community Survey, analyzing the stratified modeling and interaction terms between poverty and vehicle access. With the interaction term-included model, 83% of the variance was explained in the prevalence of food deserts (=0.83). All predictors were statistically significant ($p < 0.001$). Unemployment rate proved to be the strongest influence in the prevalence of food deserts, followed by poverty and vehicle access. With stratified models generated for each urban-only (= 0.41) and rural-only (= 0.32) settings, vehicle access was shown to have greater relative importance in rural areas than urban areas. When including the interaction term to generate a unified model, correlation coefficient increased to 0.85, indicating that the impact of poverty was exacerbated when transportation access problem was compounded. These findings provide a coherent quantitative framework to understand food deserts as a structural outcome from intersecting economic inequalities. This study underscores the role of intersectional and geographically-targeted policy interventions to address these structural economic barriers and promote equitable food security.

Keywords: Food deserts; economic inequalities; poverty; unemployment; lack of vehicle; US census tracts; urban area; rural area

INTRODUCTION

There has been a growing public health concern in food deserts in the United States. Food deserts are referred to an area with limited access to affordable and nutritious food. According to the United States Department of Agriculture, more than 19 million American citizens reside in food deserts with low income and low access to affordable food (1). People

Corresponding author: Yun Koo, E-mail: logankoo89@gmail.com.

Accepted September 18, 2025
https://doi.org/10.70251/HYJR2348.35328335

residing in food deserts have limited access to grocery stores to secure nutritious food and tend to rely on fast-food outlets or convenient stores. Because of this, food desert residents are exposed to health problems, including calorie-dense and nutrient poor food, causing obesity, diabetes, and cardiovascular diseases (2).

Food deserts have traditionally been defined by geographic distance where people lived far away from grocery stores and had limited access to secure nutritious food. However, it became more complex to define food deserts only based on geographic distance as a reason why people living far away from grocery stores struggled with food options. There are studies analyzing socioeconomic factors of people, such as poverty, unemployment, and lack of transportation, as critical indicators to understand and predict food access disparities (3). Combined with geographic distance, these socioeconomic factors play a role for people residing in food deserts to suffer from food insecurity. Poverty, one of the main socioeconomic factors, is a reason for reduced purchasing power, while influencing time availability for grocery shopping and also cooking. Similarly, unemployment places financial burden that influences both on their transportation availability or food-assistance programs from the governments. Without vehicles, people in food deserts find it difficult to go to grocery stores. All these issues combined contribute to a food environment that is structurally constrained in food deserts (4).

Prior studies have analyzed individual indicators of food access. A study (5) analyzed spatial accessibility of grocery stores with GIS methods and indicated that it was a main factor of food insecurity. Another study (6) explored a role of economic disinvestment in urban areas for limited access to grocery stores. The characteristics of food deserts in rural areas and urban areas were compared in a study (7). However, they only analyzed them descriptively without using comparative regression framework. Furthermore, it was emphasized how transit system played an important role in the access to food, especially in marginalized urban areas, but without examining how it interacted with economic deprivation (8). Moreover, a conceptual model was proposed (9) by using agent-based simulations to indicate dynamic food access scenarios. However, they did not include regression modeling as a way of quantifying interactions among transportation, poverty, and geography. There was one study who found that increased fast-food consumption was predicted significantly by the presence of fast-food restaurants combined with the lack of access to transportation means (10). However, this study only focused on dietary behavior without analyzing structural access to food. All these studies emphasized a role of individual variable that influenced food access. However, only a few studies provided a quantitative model on how multiple variables interact across geographic classifications.

This literature gap is problematic when making effective public policies. Interventions may face over-generalization or poor-targeting without understanding how food inaccessibility may be exacerbated by the intersection between poverty and transportation, or the difference between rural areas and urban areas in terms of food insecurity. For example, interventions in urban areas may need to focus more on employment access or food retail service, while studies with focus on transportation subsidies may be more effective in rural areas where people suffer more from lack of private transportation.

To address these literature gaps, this study employed a multiple linear regression model to show how poverty, unemployment, and vehicle access jointly influenced the prevalence of food deserts in the United States. This study expanded prior studies in two ways. First, this study used a stratified model through separate regressions for urban and rural food deserts. This allowed comparing how each predictor differs by geographic features directly. Second, this study uses an interaction term between poverty and vehicle access. This enabled to test if these conditions combined had a multiplicative effect on food deserts.

This approach corresponds to the necessity of more intersectional analysis of food access emphasized by prior study (2, 3). With a predictive tool that may be used by policymakers in allocation of resources according to the evidence of disadvantages beyond descriptive statistics and single predictor-based model, this study provides more sophisticated understanding about the structural causes of food inequality, while laying the groundwork for data-driven interventions. Specifically, this study seeks to answer the research question about how poverty, unemployment, and lack of vehicle access predict the prevalence of food deserts, and whether these effects differ between urban and rural areas in the United States. This study hypothesized that the higher the rates of poverty, unemployment, and vehicle inaccessibility were, the more it would be possible to predict greater prevalence of food deserts, especially in rural areas.

METHODS AND MATERIALS

Data Sources

This study used publicly available data from two primary sources: the USDA Food Access Research Atlas to identify the percentage of each tract population classified as having low access to grocery stores in the United States, according to the standard USDA definition (more than 1 mile in urban areas or 10 miles in rural areas) and American Community Survey (ACS) for 5-year estimates of socioeconomic variables. Based on the hypothesis established in this study, the focus was placed on the poverty rate, unemployment rate, and percentage of households without vehicle access. The distribution of poverty rates across the 300 census tracts in the sample was generated (Figure 1A), and the variability of unemployment rates in the dataset was calculated (Figure 1B). In addition, the distribution of households without vehicle access was generated (Figure 1C).

Using the Federal Information Processing Standards (FIPS) code, aforementioned datasets were merged with a unique identifier for census tracts. Datasets were cleaned by removing missing or anomalous values with listwise deletion that all the variables in the final datasets were complete and normalized into decimal form for regression modeling. A binary indicator for urban versus rural tracts was generated to the regression model for urban vs. rural tracts, based on USDA definitions for stratified analysis. A total of 300 census tracts were used in the final dataset, representing a mix of socioeconomic and geographic diversity in the United States.

Simulation Rationale and Procedure

Practical challenges have been posed from directly using tract-level linkage across USDA Food Access Research Atlas and ACS sources due to inconsistent

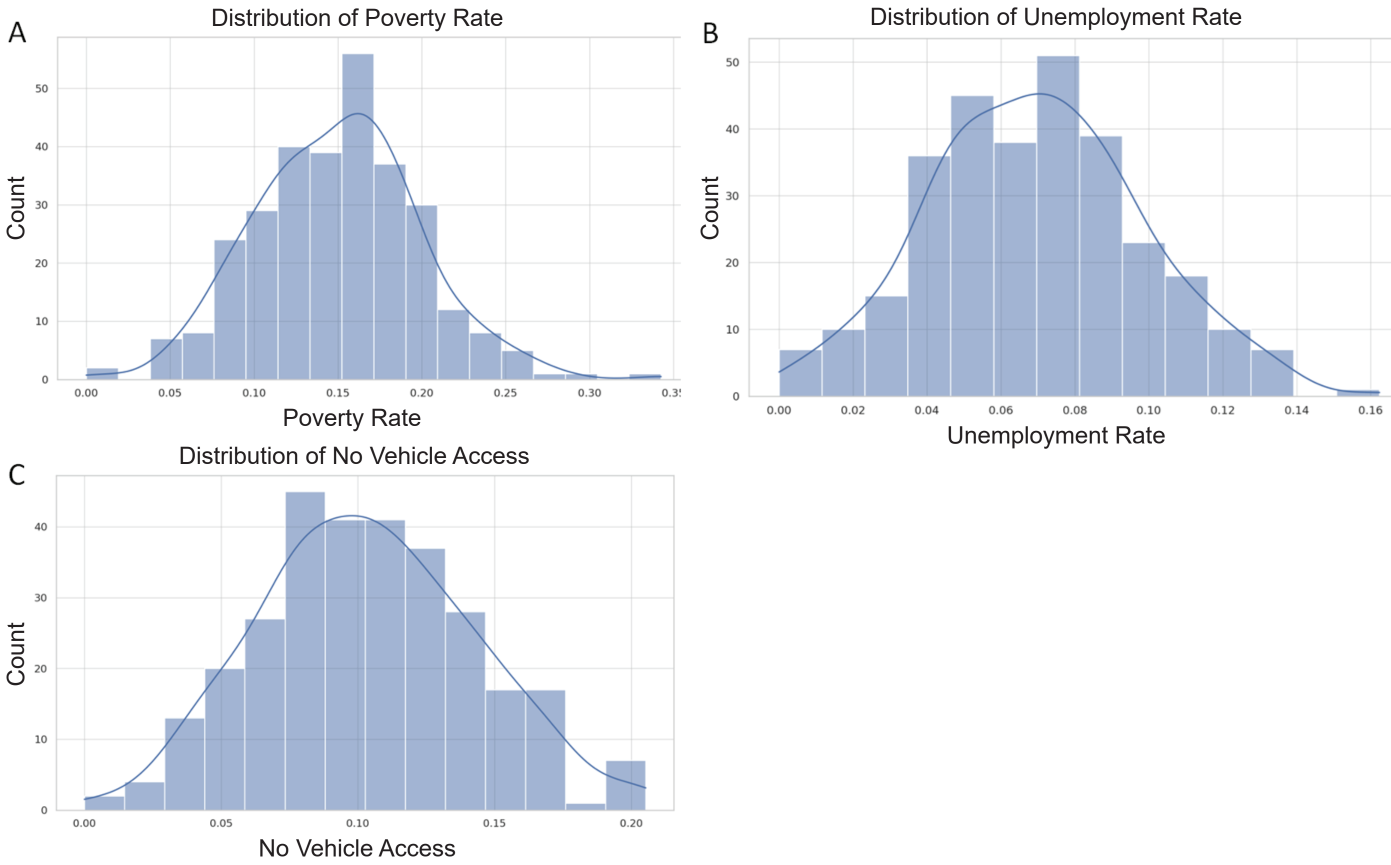

Figure 1. Distribution of Main Predictors Across Simulated Census Tracts. (A) Poverty Rate, (B) Unemployment Rate, (C) Households with No-Vehicle Access. N = 300 traces with variables scaled 0-1. Kernel density overlays were shown for the purpose of visual guidance.

vintages, harmonization of tract identifiers, and missing values. These challenges may have complicated a method-focused study. Therefore, a simulated sample of n = 300 census tracts was generated to mirror real-world distribution to test a simple but reproducible model without suffering from aforementioned issues for transparency in testing the model. Variables, such as poverty, unemployment rate, and no-vehicle rate were sampled from distributions with bound on [0, 1] with means and standard deviations that were matched to published summaries of ACS/USDA tracts. Poverty positively correlated with no-vehicle from pairwise dependencies introduced in the analysis and also moderately related with unemployment. This reflected typical empirical patterns. Using a 1:1 allocation to make balanced comparisons possible in the analysis, tracts were split into urban and rural strata. Low food access percentage was generated from a linear combination of predictors and mean-zero noise. This yielded an overall to be around 0.83, compared to the reported model fit. This study design provided balanced strata with controlled variability, and also transparent assumptions without personal information.

Mathematical Modeling

A multiple linear regression model was used to quantify how economic inequality influenced the prevalence of food deserts. The percentage of the population in each tract with low food access was the dependent variable, and poverty rate, unemployment rate, and the percentage of household without vehicle access were independent variables.

The stratified linear models (without interaction term) were generated for urban-only and rural-only as follows.

$$FDPT_i = \beta_0 + \beta_1 \bullet PovertyRate_i + \beta_2 \bullet UnemploymentRate_i + \beta_3 \bullet NoVehicleAccess_i + \varepsilon_i$$

Where $FDPT_i$ is the food desert prevalence in tract I, PovertyRate, UnemploymentRate, and NoVehicleRate are independent variables, β_0, β_1, β_2, β_3 are model coefficients, and ε_i is the residual error term. Ordinary Least Square (OLS) was used in Python with statsmodels package to estimate model parameters.

For the validation of the applicability of the model, diagnostic tests were performed for linearity, homoscedasticity, multicollinearity, and normality of residuals. For linearity, reasonable linear relationships were shown between each predictor and the response variable. For homoscedasticity, constant variance was confirmed across predicted variables according to the residual plots. For multicollinearity, low-to-moderate correlations were shown among predictors from correlation matrices and variance inflation factor (VIF). In addition, approximately normal residual distribution was shown from histograms and Q-Q plots. All these assumptions were met, confirming the appropriateness of Ordinary Least Square regression in the data analysis.

Moreover, to test compounded disadvantages, an interaction term was added between poverty and vehicle inaccessibility to the math model established and generate the expanded model as follows.

$$FDPT_i = \beta_0 + \beta_1 \bullet PovertyRate_i + \beta_2 \bullet UnemploymentRate_i + \beta_3 \bullet NoVehicleAccess_i + \beta_4 \bullet (PovertyRate_i \bullet NoVehicleAccess_i) + \varepsilon_i$$

Based on this mathematical model, separate regression models were fit for stratification between urban and rural areas to explore whether geographic classification moderated the influence of predictors. In addition, a sensitivity analysis was performed to improve the robustness of the model by removing one predictor at a time and recalculated the correlation coefficient in the model and variables' coefficients. This was to identify the relative importance of each variable in the full model.

RESULTS

Overall Model Results

The original mathematical model established with poverty rate, unemployment rate, and vehicle inaccessibility as predictors yielded R^2 of 0.83. This means that 83% of the variance in the prevalence of food deserts among 300 census tracts used in the model was explained by three variables in the simulation.

All predictors were statistically significant ($p < 0.001$). A 1% increase in poverty rate was correlated with a 0.301% increase in the prevalence of food deserts, while a 1% increase in unemployment rate predicted a 0.401% in food desert prevalence, and a 1% increase in vehicle inaccessibility predicted 0.196% in food desert prevalence. These results indicate that all three variables statistically significantly contribute to predicting food inaccessibility, with unemployment having the strongest correlation with food desert prevalence per unit increase.

Urban-Rural Comparisons

Results in Urban vs. Rural Areas Model

To further assess geographic differences, separate models were estimated for urban and rural tracts. Coefficients of stratified urban-only model and rural-only models were calculated and applied as follows.

Stratified urban-only model:

$$FDPT_i = 0.3751 + 0.1202 \cdot PovertyRate_i - 0.3693 \cdot UnemploymentRate_i - 0.0050 \cdot NoVehicleAccess_i + \varepsilon_i$$

Stratified rural-only model

$$FDPT_i = 0.3242 + 0.1945 \cdot PovertyRate_i + 0.1091 \cdot UnemploymentRate_i - 0.0738 \cdot NoVehicleAccess_i + \varepsilon_i$$

From the results in the urban-only model, the relationship between poverty and food access was positive. However, unemployment coefficient was negative in large magnitude. The coefficient of vehicle access was negligible. This pattern was consistent with urban context that private-vehicle constraints were mitigated by the access to public transportation and retail density. However, this needs to be cautiously interpreted according to inter-correlations among predictors.

From the rural-only model, all coefficients indicated the expected directions. Poverty had a stronger effect in food desert prevalence than in urban areas, and both unemployment and vehicle access also statistically significantly contributed to food desert prevalence. These results suggest that transportation is an issue for people in rural food environment where the infrastructure is limited. Food desert prevalence between urban and rural areas was compared (Figure 2). Poverty rate distributions turned out to vary across geographical categories (Figure 3).

Interaction and Robustness Analyses

Compounded effects were examined, and robustness was tested. Interaction-term combined model is as follows.

$$FDPT_i = 0.3832 + 0.0229 \cdot PovertyRate_i - 0.2052 \cdot UnemploymentRate_i - 0.2210 \cdot NoVehicleAccess_i + 0.9625 \cdot (PovertyRate_i \cdot NoVehicleAccess_i) + \varepsilon_i$$

When the interaction term was included, poverty rate (0.0229) along had a small marginal effect. Unemployment rate showed a negative relationship that was consistent with the previous models, and no vehicle had stronger negative effect in the model. Interaction term was strong and positive, indicating that the combined presence of poverty and vehicle inaccessibility amplifies the prevalence of food deserts.

To improve the robustness of the model in this study, I ran a sensitivity analysis by removing one predictor at a time and recalculating the correlation coefficient score and coefficients. When removing unemployment rate, R^2 dropped from 0.83 to 0.69, showing that even if the coefficient of it in the interaction term-included model was negative, it still explained a significant portion of food desert prevalence. This also suggests that the variable may play a role of suppressor or is related to

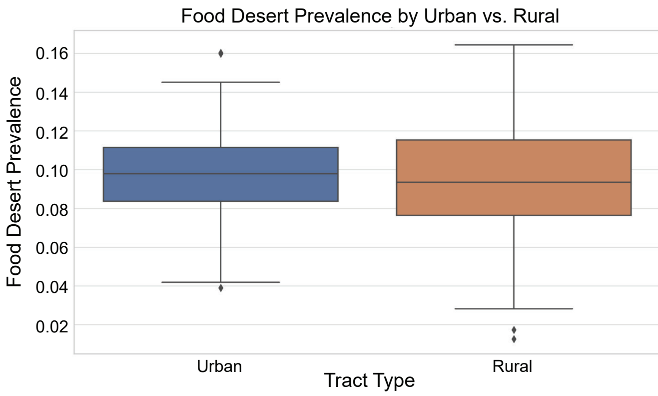

Figure 2. Boxplot of Food Desert Prevalence in Urban and Rural Tracts. Boxplots indicated median, IQR, and 1.5xIQR whiskers. Points behind whiskers showed possible outliers. Outcomes were calculated as a percent of tract population with low food access.

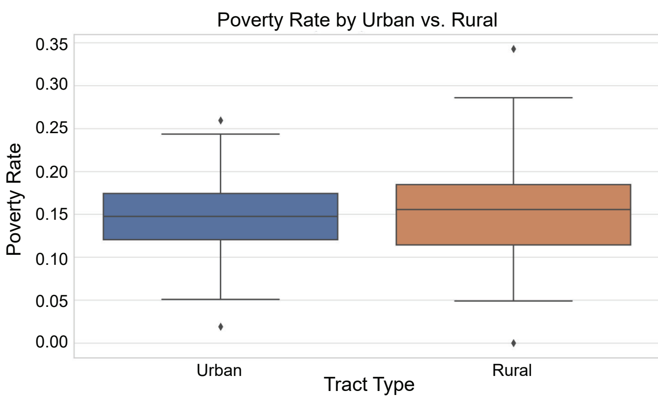

Figure 3. Boxplot of Poverty Rate in Urban and Rural Tracts. Boxplots with rates scaled 0-1. Higher values indicated greater poverty.

other omitted factors. When removing the poverty rate, R^2 dropped to 0.73, confirming that poverty had a central role in predicting the prevalence of food deserts. The coefficient of it also remained positive in the interaction term-included model and became even more powerful in rural-only model. When removing vehicle access, there was a small drop in R^2 from 0.83 to 0.79. This indicates that vehicle access had a relatively smaller independent contribution. Partial R^2 values were calculated by subtracting the R^2 of each reduced model from the interaction term-included model (Table 1).

Even though unemployment showed the strongest partial R^2 contribution (0.14), it had a negative sign in the coefficient in the interaction term-included model and urban-only model. This suggests that further examination of interaction terms may be needed. Poverty rate, with a partial R^2 of 0.1, showed a consistent positive effects across models, and is suggested to be the most theoretically and practically consistent indicator of food inaccessibility. The bar graph of partial R^2 contribution of each predictor was generated (Table 1).

DISCUSSION

This study investigated the extent to which three socioeconomic factors: poverty rate, unemployment rate, and vehicle access predict the prevalence of food desert the most in the United States. The findings in this study confirm that structural economic inequality played an important role in predicting geographic food inaccessibility. In addition, both an interaction term-included model and stratified models provided important evidence that poverty combined with limited transportation access exacerbated the prevalence of food deserts.

The overall multiple regression model explained 83% of the variance in the prevalence of food deserts in this study through a total of 300 simulated census tracts. Among the individual variables, poverty rate and unemployment rate turned out to be the most significant predictors, where the poverty rate remained positive consistently in terms of its effect on food inaccessibility. These findings support the prior studies about how low-income communities tended to suffer limited access to grocery markets and healthy food outlets because of the inability of residents to afford transportation and also the distal location of supermarkets (2, 3).

Interesting finding was that the coefficient of unemployment rate was negative in the interaction-combined model and urban-only model, but was positive in rural-only model. This was an unexpected directional shift, emphasizing the importance of geographical disaggregation in terms of food access. In urban areas, unemployment may not be closely related to food access as initially hypothesized, possibly due to the availability of public transportation, or greater density of grocery stores around (11). However, unemployment turned out to be more strongly correlated with food security in rural areas. There may be many reasons behind this, including the lack of personal vehicle to distant grocery stores with insufficient public transportation. The positive unemployment coefficient aligned well with the findings from prior study suggesting that the food inaccessibility was more of a multidimensional issue in rural areas (12). The negative sign on unemployment in the urban only and interaction models likely indicated an effect of suppression instead of a true protective relationship. Unemployment was moderately correlated with no-vehicle access and poverty in the simulated data. When these covariates were to be applied together, the shared variance may be offset by poverty and transportation, making unique variance of unemployment to indicate the opposite direction. Specifically speaking of urban contexts, unemployment may be decoupled from physical access to food when poverty maintained to be high with greater retail density and public transportation. Putting all of them together with the sensitivity analysis, these results indicated that unemployment was informative for overall variance explained in the analysis when model fit was significantly reduced if removing unemployment. However, the isolated coefficient of unemployment needed to be cautiously interpreted in relation with poverty and transportation access instead

Table 1. Partial Contribution of Poverty, Unemployment Rate, and Vehicle Inaccessibility to Explained Variance in Low Food Access (%)

Predictor	Partial Contribution
Poverty Rate	0.10
Unemployment Rate	0.14
Vehicle Inaccessibility	0.04

Partial R^2 calculated by comparing the full interaction model to the models omitting one predictor. Sample: n = 300 simulated tracts balanced by classifications of urban and rural areas. Variables were scaled 0-1.

of its sign alone.

Vehicle access turned out to have a varied role depending on geographical circumstances. The coefficient of vehicle access was nearly zero in the urban-only model. This suggests that even residents without the cars in urban areas still had access to food through public transportation, walking, or biking. However, it had a strong negative effect in rural-only model, emphasizing that transportation infrastructure was an important component in less densely populated areas for food access. These findings aligned well with prior study about how residents without cars in rural areas were often not able to go to grocery stores, when they were several miles away from where residents lived, due to sparse public transit (13).

Sensitivity analysis also reinforced the robustness of the model. Unemployment was confirmed to have a significant explanatory power in the prevalence of food deserts despite its counterintuitive direction in the interaction term-included model as the R^2 in the model dropped from 0.83 to 0.69 when it was removed. Partial R^2 analysis indicated that unemployment rate had the most unique variance (0.14) followed by poverty ratio (0.10) and vehicle access (0.04). These findings suggest that all three predictors well-contributed to explaining the disparities of food access, and unemployment rate and poverty rate were two critical predictors.

Another notable contribution of this study is that the model included stratified models and interaction term-included model, separately, to bridge the gap from prior studies in regards of the lack of interactions between variables or differences across geographic features. The findings in this study correspond to the necessity of incorporating socioeconomic and spatial factors in food access research (14). Moreover, the findings in this study challenge other public health studies that oversimplify rural and urban areas as homogenous units. This oversimplification may obscure differences between urban and rural areas and come up with policies that will be ineffective or fail to correspond to local needs.

The most effective interventions in the rural areas with high poverty and no-vehicle access rate would be to directly reduce travel frictions to raise food accessibility. Examples may include subsidized grocery delivery, mobile markets, and demand-responsive transit to grocery stores. On the other hand, in urban areas with unemployment as less predictive factor when adjusting for poverty and transportation, effective interventions would include stabilizing household purchasing power and reducing time and costs. Examples may include extended store hours and neighborhood-scale healthy-store programs to target for larger marginal gains.

While this study provides a valuable insight about structural indicators of food deserts in the United States, there are limitations. This study used simulated trace-level data calibrated to public summaries of ACS/USDA instead of raw micro-data linked with tracts. Controlled variability, balanced urban-rural strata, and transparent assumptions were provided by the simulation. However, it may omit place-level dynamics in real communities. With the simulated study design with n = 300 tracts, results may have been viewed as hypothesis-generating and also explanatory, requiring confirmation with fully linked trace-specific datasets. The sample size of n = 300 was chosen for the transparency and balance of the model. However, it may limit precision around estimates of subgroups. Structural biases from the assumed distribution and correlation structure may also be embedded in the simulation process. Therefore, future work is recommended to confirm the results with linked tract data.

CONCLUSION

This study shows that poverty, unemployment, and vehicle access are significant predictors in the prevalence of food deserts in the United States census tracts. With both stratified models and an interaction term-included regression model, it was found that poverty combined with lack of transportation means significantly exacerbated food inaccessibility, especially in rural areas. These findings call for the necessity of context-based policies to address intersecting disadvantaging factors instead of isolated factors. Results in this study also demonstrate that communities suffering from food inaccessibility may require more integrated policy approach according the strong interaction between poverty and the lack of transportation means.

Combined with limitations in this study, I suggest future research to enhance this work by including real-world and longitudinal data and identify more social determinants other than poverty rate, unemployment rate, and vehicle access, including housing instability, health outcomes, or educational attainment. Furthermore, future studies should examine the relationship between the role of food deserts and health disparities to better understand public health relevance of this study and provide more support on comprehensive policy.

CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

REFERENCES

1. USDA. Food Access Research Atlas. United States Department of Agriculture; 2019. Available from: https://www.ers.usda.gov/data-products/food-access-research-atlas/ (accessed on 2025-07-01).
2. Walker RE, Keane CR, Burke JG. Disparities and access to healthy food in the United States: A review of food deserts literature. *Health Place*. 2010; 16 (5): 876-884. https://doi.org/10.1016/j.healthplace.2010.04.013
3. Beaulac J, Kristjansson E, Cummins S. A systematic review of food deserts, 1966–2007. *Prev Chronic Dis*. 2009; 6 (3): A105. https://doi.org/10.1007/s10460-006-9002-8
4. Hendrickson D, Smith C, Eikenberry N. Fruit and vegetable access in four low-income food deserts communities in Minnesota. *Agric Human Values*. 2006;23:371-383. https://doi.org/10.1007/s10460-006-9002-8
5. Apparicio P, Cloutier MS, Shearmur R. The case of Montreal's missing food deserts: Evaluation of accessibility to food supermarkets. *Int J Health Geogr*. 2007; 6: 4. https://doi.org/10.1186/1476-072X-6-4
6. Alwitt LF, Donley TD. Retail stores in poor urban neighborhoods. *J Consum Aff*. 1997; 31 (1): 139-164. https://doi.org/10.1111/j.1745-6606.1997.tb00830.x
7. Ver Ploeg M, Breneman V, Farrigan T, et al. Access to affordable and nutritious food: Updated estimates of distance to supermarkets using 2010 data. USDA Econ Res Serv; 2012.
8. Farber S, Morang MZ, Widener MJ. Temporal variability in transit-based accessibility to supermarkets. *Appl Geogr*. 2014; 53: 149-159. https://doi.org/10.1016/j.apgeog.2014.06.012
9. Widener MJ, Metcalf SS, Bar-Yam Y. Agent-based modeling of policies to improve regional food access for low-income populations. Appl Geogr. 2013;40:1-10. https://doi.org/10.1016/j.apgeog.2013.01.003
10. Morland K, Wing S, Diez Roux A, Poole C. Neighborhood characteristics associated with the location of food stores and food service places. *Am J Prev Med*. 2002; 22 (1): 23-29. https://doi.org/10.1016/S0749-3797(01)00403-2
11. Powell LM, Slater S, Mirtcheva D, Bao Y, Chaloupka FJ. Food store availability and neighborhood characteristics in the United States. *Prev Med*. 2007; 44 (3): 189-195. https://doi.org/10.1016/j.ypmed.2006.08.008
12. Morton LW, Blanchard TC. Starved for access: Life in rural America's food deserts. *Rural Realities*. 2007; 1 (4): 1-10.
13. Sharkey JR, Horel S. Neighborhood socioeconomic deprivation and minority composition are associated with better potential spatial access to fruits and vegetables in nearby stores. *Int J Epidemiol*. 2008; 37 (5): 1178-1186.
14. Liese AD, Weis KE, Pluto D, Smith E, Lawson A. Food store types, availability, and cost of foods in a rural environment. *J Am Diet Assoc*. 2007; 107 (11): 1916-1923. https://doi.org/10.1016/j.jada.2007.08.012