Understanding Walkability: Calculating the Index and Investigating Its Connection to Diabetes Prevalence

A. Bose, S. Oberg, A. Cheruvu (Department of Statistics)

I. Abstract

The diabetes epidemic in the United States presents a nuanced public health challenge, shaped by factors such as socioeconomic status and climate. While the influence of these factors on diabetes is well-established, the role of walkability in managing diabetes prevalence remains contested. This study revisits the relationship between walkability and diabetes in the U.S., using walkability indexes calculated from CDC data. Contrary to some studies suggesting that increased walkability reduces diabetes prevalence, our findings, analyzed through Geographically Weighted Regression (GWR), reveal that walkability is not a significant predictor of diabetes prevalence and exhibits notable regional anomalies. Further analysis using Monte Carlo simulations, Global I Moran's Test, and Variance Inflation Ratio (VIR) supports these results. Our study also critiques the current methods of calculating the walkability index, proposing a revised model that incorporates additional relevant variables from the CDC. This nuanced understanding underscores the need for region-specific urban planning and public health strategies that recognize the complex interplay between walkability, environmental, and socioeconomic factors.

II. Introduction

Diabetes, a chronic illness linked to high blood sugar levels, affects approximately 9.7% of U.S. adults, impacting both physical and mental well-being and straining healthcare resources (Xu, et. al). Recent studies suggest that the walkability of a region may impact diabetes prevalence. While some findings indicate that higher walkability index scores are associated with lower diabetes rates, our study explores this relationship, and investigates the process of formulating a walkability index score. Using walkability index score data from the Environmental Protection Agency, we found that the standardized formula for calculating a walkability index score does not taken enough factors into account for enough factors for there to be a statistically significant relationship. It is important to acknowledge, and improve this metric as it is essential for devising effective public health strategies to manage and prevent diabetes.

III. Methods

i. Walkability Index

The walkability index score is a measurement of relative walkability within a given region. According the the Environmental Protection Agency, their official formula for calculating this number for a given group is listed below.

$$\text{Walkability Index} = \frac{w}{3} + \frac{x}{3} + \frac{y}{6} + \frac{z}{6}$$

Figure 1: Formula for calculating walkability index score

In this formula, w represents a block group's block group's intersection density, x represents the distance to the closest transit stop, y represents the employment mix in the block group and z represents the household-employment mix.

ii. Statistical Analysis

To explore the spatial relationships between walkability index score, obesity prevalence, high blood pressure prevalence, low-physical activity prevalence, smoking prevalence, temperature, and median household income, we employed Geographic Weighted Regression (GWR). GWR allows for the examination of spatially varying relationships between variables. Additionally, to assess the robustness of our GWR results, Monte Carlo simulation was utilized. For plotting purposes, gwr.t.adjust() was used to set insignificant coefficients equal to 0 using the Fotheringham-Byrne procedure (Byrne et. al).

iii. Spatial Diagnostics

Global Moran's I test was conducted to examine spatial autocorrelation in the variables, providing insights into the spatial patterns present in the data. Furthermore, Variance Inflation Factor (VIF) analysis was employed to diagnose multicollinearity among the predictor variables.

IV. Results

Our study applied a Geographically Weighted Regression (GWR) model to explore the connection between walkability index score and diabetes prevalence across the U.S. Simulations with artificial data validated the model's effectiveness, showing accurate, reliable predictions with evenly dispersed residuals.
Analysis of real-world data highlighted a notable positive correlation between walkability and diabetes prevalence in the Southern United States, where higher walkability is associated with increased diabetes rates, as depicted in Figure 2. This contrasts with other regions like the West Coast and Pacific Northwest, where higher walkability tends to decrease diabetes prevalence.

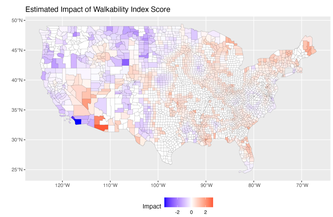

Figure 2: Impact of Walkability Index on Diabetes Prevalence

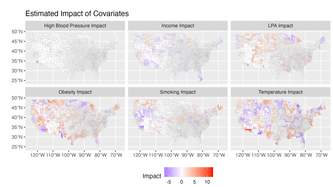

Figure 3: Impact of Covariates on Diabetes Prevalence

Ultimately, we believed that there would be a consistent trend showing that the Walkability Index Score did impact diabetes across the entire U.S. However, from our GWR model's results, that that is not the case. This leads us to believe that the Walkability Index Score calculated by the CDC may not be truly accurate and may fail to account for other additional factors that could contribute to increased diabetes prevalence in the U.S

i. Validation

This is further shown when conducting a Monte-Carlo test. As seen in Figure 4, the National Walkability Index Score is associated with a p-value of 0.45. This value indicates that the National Walkability Index Score is not a statistically significant metric as its p-value is greater than 0.05. Thus, the Walkability Index Score may not have a siginificant impact on predicting diabetes prevalence across the U.S.

Covariates	P_Value
National Walkability Index	0.45
Obesity Prevalence	0.01
High Blood Pressure Prevalence	0.00
Low Physical Activity Prevalence	0.00
Current Smoking Prevalence	0.00
Median Household Income	0.90
Average Temperature	0.00

Figure 4: Results of Monte Carlo Test on GWR Results

In addition, from the Global Moran-Eye Test, our model's Moran 1 Statistic is 0.049, meaning there is slight positive autocorrelation in the covariates. However, we expect this as our covariates tend to resemble similar factors(eg.smoking,obesity,high blood pressure, etc.). However, this is not something to be too worried about as there is not significant autocorrelation or high spatial dependence.
In addition, other validation metrics we looked at include the Variance Inflation Factor(VIF) of each of our covariates. None of our covariates had a VIF score that exceeded 13, indiciating there is not significant multicollinearity in our GWR model that poses an issue. We also looked at the residual plot and we can see that there are fairly evenly dispersed residuals which is a positive sign.

V. Discussion

Our investigation into the relationship between walkability and diabetes prevalence in the U.S. revealed that walkability, as currently indexed by the Environmental Protection Agency, does not significantly predict diabetes outcomes and varies notably across different regions. This challenges the dominant assumption that higher walkability universally lowers diabetes risk and aligns with our analytical results from Geographically Weighted Regression (GWR), Monte Carlo simulations, Global I Moran's Test, and Variance Inflation Ratio (VIR) assessments. Our findings question the efficacy of current walkability indices used by public health and urban planning bodies, suggesting that these indices fail to capture the multifaceted influences on health outcomes effectively. We propose a revised model for calculating walkability that incorporates a broader range of variables provided by the Environmental Protection Agency, which could more accurately reflect the true impact of walkability on health.

i. Relevant Factors

After conducting our analysis, it was important for us to investigate which factors could be added to the formula provided by the EPA in order to allow for walkability to be a significant predictor of diabetes prevalence. Although there are many factors that could play into this, there are a couple that stood out in our reserach. In an article by Creatore et. al, walkability was calculated using population density as one of the factors. In this article, the results found there to be a statistically significant relationship between diabetes and walkability index score. The EPA already has this value as part of their Smart Location Database for each Census Block Group. Another factor that should be considered when coming up with a walkability index score is crime rate. In an article by Nafakh, he found success in generating a specific walkability index using crime rate density as a factor.

VI. References

- G. Xu et al., "Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: population based study,"
- [1] M. I. Creatore et al., "Association of Neighborhood Walkability With Change in Overweight, Obesity, and Diabetes," JAMA, vol. 315, no. 20, p. 2211, May 2016, doi: 10.1001/jama.2016.5898.
- [1] A. J. Nafakh, "Spatial Assessment of Walkability Index - A Case Study in Chicago," Oct. 2018, [Online]. Available: https://indigo.uic.edu/articles/thesis/Spatial_Assessment_of_Walkability_Index_-_A_Case_Study_in_Chicago/10929488