

Environmental Effects on Type 2 Diabetes Prevalence in Virginia Counties

SUSAN L. ZYWOKARTE

Department of Geography and Geoinformation Science

George Mason University

The causes of type 2 diabetes are extremely complex. A great deal of research has been conducted regarding genetic and behavioral causes. Newer avenues of research have begun to consider how certain environmental factors explain the prevalence of type 2 diabetes. This study examines two possible environmental explanatory variables. Pollution, specifically particulate matter (PM_{2.5}), and elevation are modeled in a multivariate linear regression analysis along with two known behavioral variables: obesity and inactivity. The case study area was the Commonwealth of Virginia, and the aggregation was at the county level. Although recent research identified a relationship between the prevalence of type 2 diabetes and PM_{2.5} pollution, this study did not show significant results. Elevation also did not have a significant influence on diabetes prevalence in Virginia.

Keywords: Diabetes, pollution, obesity, environment, regression, elevation

Introduction

The prevalence of diagnosed diabetes in the United States continues to increase. In 2010, the number of people diagnosed with diabetes was estimated at 18.8 million, with approximately seven million undiagnosed cases (Geiss et al., 2012). Between the years 1995 and 2010, the number of states with a diabetes prevalence greater than 6% increased from three states in 1995 to all 50 states in 2010 (Geiss et al., 2012). Also in 2010, six states, along with Puerto Rico, had a diabetes prevalence greater than 10% (Geiss et al., 2012). Individuals with type 2 diabetes are

usually able to produce a certain amount of insulin, but not enough to keep sugar levels normal. The acute complications of type 2 diabetes include ketoacidosis, hyperosmolarity, and hypoglycemia.

Saudek, Rubin, and Shump (1997) explain that ketoacidosis occurs when there is a deficit of insulin in the bloodstream, causing the body to break down fat cells for needed energy. This breakdown of fat cells can cause ketones to build up in the bloodstream, resulting in ketoacidosis, which at high levels can lead to coma. Hyperosmolarity occurs when the person's blood sugar level reaches over 1000mg/dl, and causes the blood to thicken, resulting in coma. Hypoglycemia results when blood glucose levels drop below 60 mg/dl on average. Causes can be too much insulin or oral medication, too little food, and/or too much exercise. As the body and brain need a certain amount of sugar to function, untreated cases can result in coma. These acute complications can lead to death if left untreated. Long-term complications can include damage to the eyes (retinopathy), damage to the kidneys (nephropathy), and damage to nerves (neuropathy). Also, arteriosclerosis (the hardening of the arteries) occurs when plaques of cholesterol develop on the inner surface of arteries. Blood flow can be blocked, resulting in stroke or heart attack. Untreated or poorly treated diabetes can be a cause of arteriosclerosis (Saudek et al., 1997).

Ethnicity, race, and heredity are well-known genetic factors that influence type 2 diabetes development. Furthermore, behavioral factors such as overeating and lack of exercise play a significant role in diabetes prevalence. Research on diabetes risk factors suggests that type 2 diabetes diagnoses usually occur in individuals older than 40, who are overweight or obese, have a low activity lifestyle, and have a family history of the illness (Geiss et al., 2012). Certain race and ethnic groups, including African Americans, Hispanics, and Native Americans have a higher prevalence of type 2 diabetes. Obesity, however, has been found to be the most accurate predictor of type 2 diabetes (Saudek et al., 1997). Between 2009 and 2012, the U.S. age-adjusted diagnosed and undiagnosed prevalence of diabetes for individuals 20 years and older was 11.7% (Geiss et al., 2012).

More recently, researchers have attempted to understand the role of environmental factors in diabetes prevalence. Particulate matter (PM_{2.5}) exposure from the environment and elevation are two that have been examined. Particulate air pollution is a mixture of solid and liquid matter differing in origin, composition, and size. PM_{2.5} is considered fine particulate matter and has a particle diameter size of less than 2.5µm (Pope, 2000). The small size of these particles allows

them to penetrate deeper into the lungs than larger particles. Larger particles are often the result of natural processes such as blowing dust or soils. The PM_{2.5} size particles are often industrial in origin, resulting from processes in different industries involving combustion, which produces sulfate and nitrate particles (Pope, 2000). Increased exposure levels of PM_{2.5} has been found to be associated with an increase in the inflammation of insulin responsive organs (Chien, Alamgir, & Yu, 2015). Brook et al. (2013) found that exposure to PM_{2.5} resulted in a decrease of insulin sensitivity. Another avenue of research examined the relationship between elevation and diabetes prevalence. Woolcott et al. (2014) examined the prevalence of diabetes and obesity in high and low elevation counties in the United States and found that counties at high elevations had less diabetes prevalence than counties at low elevations.

The causes of type 2 diabetes are extremely complex. A great deal of research has examined the genetic and behavioral components that lead to type 2 diabetes. However, less work has been dedicated to understanding the role that the environment plays, and thus, more research is warranted. This study examines the association between type 2 diabetes prevalence and environmental factors, while controlling for known behavioral factors. Two models utilizing ordinary least squares multiple regression were developed to explore how PM_{2.5} and elevation might affect the prevalence of type 2 diabetes. This ecological-level case study was conducted using county-level data from Virginia.

Literature Review

Numerous researchers have studied the demographic and behavioral causes of type 2 diabetes. Barker, Kirtland, Gregg, Geiss, and Thompson (2011) examined the geographic distribution of diabetes by U.S. counties in an area the authors describe as a “Diabetes Belt,” located in parts of the Southeast and Appalachian regions. The study evaluated how this region differed from the rest of the United States, finding that counties in the Diabetes Belt had higher obesity and lower activity levels, as well as higher percentages of non-Hispanic African Americans and greater numbers of people 65 years and older. Menke, Rust, Fradkin, Cheng, and Cowie (2014) carried out a series of cross-sectional studies that examined trends in the prevalence of diabetes based on race/ethnicity, age, and body mass index (BMI). They found that between the years of 1976 and 1980, and between 2007 and 2010, diabetes prevalence increased in men from 6.2% to 9.6%, but decreased slightly in women from 7.6% to 7.5%.

Patel, Bhattacharya, and Butte (2010) developed the Environmental Wide Association Study (EWAS). While controlling for age, sex, ethnicity, socioeconomic status (SES) and BMI, the study conducted numerous cross-sectional studies with 266 different environmental factors that could possibly be associated with the prevalence of type 2 diabetes. Logistic regression models were utilized with data obtained from the National Health and Nutrition Examination Survey (NHANES). Researchers found this study method was able to isolate environmental factors with significant effects on type 2 diabetes. Among the 266 environmental factors, there were some that were notably significant. Although already known to increase the type 2 diabetes risk, this study confirmed that factors such as carotenes and polychlorinated biphenyls (PCBs) increase the risk of type 2 diabetes. Also, a component of vitamin E, α -tocopherol, was found to increase diabetes risk. Heptachlor epoxide, a pesticide discontinued for most uses in the 1980s, is still found in the environment, and was previously found to cause type 2 diabetes in pesticide applicators. This study found a broad association of heptachlor levels and type 2 diabetes in the general public using the NHANES data.

More recently, researchers have examined the relationship between $PM_{2.5}$ levels and the prevalence of diabetes. Chien et al. (2015) studied the United States at the county level and found a geographical association between $PM_{2.5}$ levels and diabetes prevalence. Certain regions of the United States were more at risk than others, including the South, Central, and Southwest. Zheutlin, Adar, and Park (2014) performed an ecological analysis on the association between carbon dioxide (CO_2) emissions and the prevalence of diabetes in the United States. However, instead of finding a connection between diabetes prevalence and CO_2 emissions, the results suggested ambient $PM_{2.5}$ was associated positively with diabetes prevalence.

Pearson, Bachireddy, Shyamprasad, Goldfine, and Brownstein (2010), utilizing multivariate regression models, examined county-level data for the United States and found that increased $PM_{2.5}$ levels, while controlling for diabetes risk factors, were associated with an increase in diabetes prevalence. Also, through animal studies, researchers have found that negative response to $PM_{2.5}$ exposure did not occur in lean mice. Mice with diet-induced obesity did respond to increased $PM_{2.5}$ exposure with insulin resistance related to chronic inflammation as a mechanism. The authors utilized these laboratory experiments in mice to gain insight into the association between urbanization and type 2 diabetes in humans (Sun et al., 2009). Brook et

al. (2013) found that sub-acute exposure to $PM_{2.5}$ resulted in a worsening sensitivity to metabolic insulin in humans.

Researchers in the United States have explored associations of both prevalent and incident type 2 diabetes with exposure to $PM_{2.5}$ and NO_2 , as well as proximity to major roadways. After analysis, the researchers found that higher long-term exposure to $PM_{2.5}$ and NO_2 had a significant association with the prevalent cases of type 2 diabetes ascertained at the start of the study (Park et al., 2015).

Other researchers have examined the relationship between elevation and diabetes prevalence in the United States. Woolcott et al. (2014) found that counties at high elevations had less diabetes prevalence than counties at low elevations, and obesity did not explain this inverse relationship. Likewise, Voss, Masuoka, Webber, Scher, and Atkinson (2013) found that the prevalence of obesity, a major risk factor for type 2 diabetes, is inversely associated with elevation and the degree of urbanization while controlling for behavior, demographic factors, and temperature.

Method

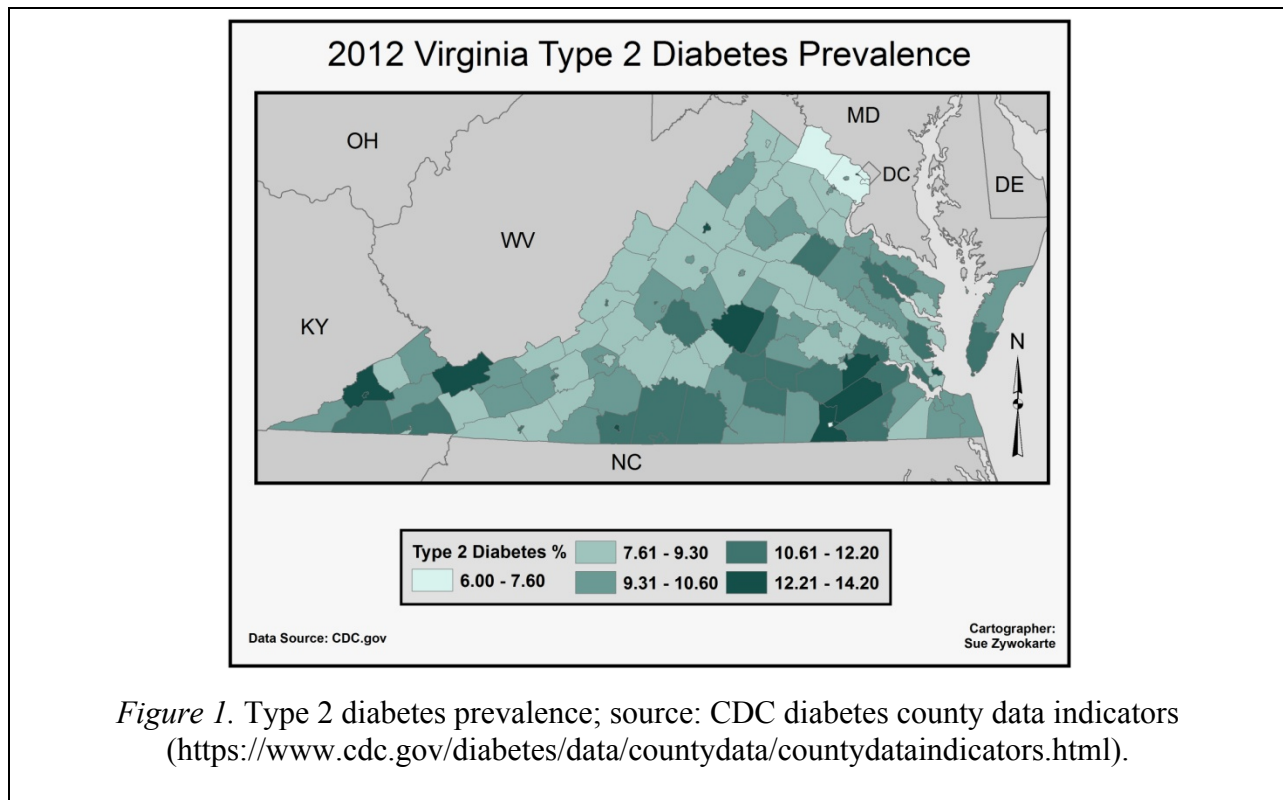
Regression analysis is utilized in research to explore relationships between variables. One variable might influence or affect another, or variables might be tied to each other because of an existing functional relationship. The independent variables are utilized to explain changes in the dependent variable.

This study used an ordinary least squares (OLS) multivariate regression for two separate models. Since research indicated a possible connection between $PM_{2.5}$ and type 2 diabetes, an OLS regression was performed with type 2 diabetes as the dependent variable and $PM_{2.5}$ and obesity as the two independent variables. The second OLS regression model looked at the relationship between type 2 diabetes and elevation, with obesity also as a control factor as an already-known contributor to type 2 diabetes.

The main objective of this analysis was to determine how much variation in type 2 diabetes prevalence can be attributed to obesity and/or $PM_{2.5}$ in the first model, and how much in the second model can be attributed to obesity and elevation. The regression analysis generates standardized regression coefficients, also known as beta values, for each independent variable. These indices show the relative strength of the independent variable contributing to the variance in the dependent variable. Also, for each model, a coefficient of multiple determination (R^2) is

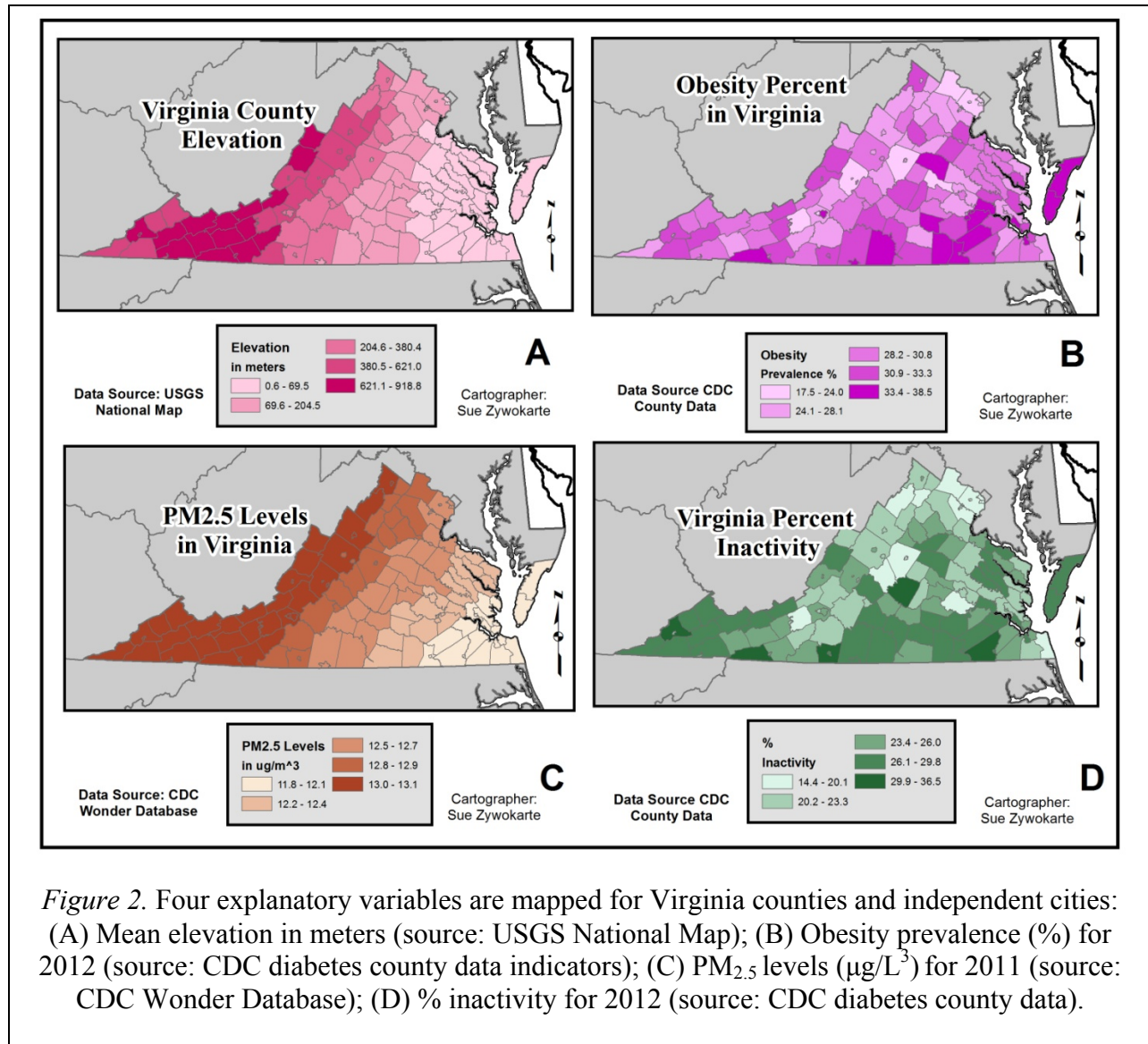
calculated. This measures the ratio of variation in the dependent variable. The beta values undergo a hypothesis test with probabilities calculated to determine whether results are significant.

Age-adjusted type 2 diabetes prevalence county level data for 2012 were downloaded from the U.S. Centers for Disease Control and Prevention (CDC) website, as well as county-level measures for obesity prevalence and percent inactivity (CDC Diabetes County Data, n.d.) . PM_{2.5} data, which are measured in micrograms per liter cubed ($\mu\text{g}/\text{L}^3$) and averaged for the 2011 year, were obtained from the CDC Wonder Database.



A shapefile of Virginia with county and city boundaries was utilized for the analysis. Virginia county and city jurisdictions are different from other states; certain cities are independent with their own court systems and have the same authority as counties. For this study, the city of Bedford was omitted from the analysis because it was missing from the shapefile data. This left 133 county/city observations in the dataset. A raster elevation layer was obtained from the U.S. Geological Survey (USGS) National Map (<http://nationalmap.gov>). The 42 individual 1 x 1 degree data tiles were stitched together to create a single seamless elevation data layer. This layer was re-projected to the Virginia Lambert projection and re-sampled from a

one arc-second cell resolution to a 30-meter cell resolution using a bilinear interpolation. It was processed through ArcGIS utilizing a zonal statistics method, which summarized elevation values within counties and cities of the Virginia shapefile. Average Virginia county elevations were assigned to each county. All five data sources were consolidated in a single file via a table join operation. Figure 1 shows the geographic distribution of type 2 diabetes prevalence in Virginia, and Figure 2 shows the distribution of the four explanatory variables.



ArcMap software was used to calculate the OLS multiple regression for the relationship between environmental factors and the prevalence of type 2 diabetes in Virginia counties and independent cities. The explanatory independent variables were obesity, inactivity, PM_{2.5}, and

elevation. Prior to the regression, the dependent variable (type 2 diabetes) and the explanatory variables were tested for correlation via Pearson's correlation coefficient (r). Obesity and inactivity measures were used as a control for the already-established behavioral variables. (Table 1 contains the correlation matrix.)

Both the elevation/PM_{2.5} pair and the inactivity/obesity pair had a borderline high r value of .796 and .714, respectively. High r values for the environment and behavioral variables suggest multicollinearity, and led to the development of two separate regression models to analyze this dataset. Because of the multicollinearity between the obesity and inactivity variables, it was decided to remove the inactivity variable and use only obesity as the control for the behavior factor. The first model paired obesity and PM_{2.5}, and the second model paired obesity and elevation as explanatory variables. Moran's I was performed on both models to test the residuals for autocorrelation (Moran, 1950). If there is statistically significant clustering, the models could be misspecified, indicating that there is a key explanatory variable missing from the model.

Table 1

Correlation matrix (r)

Variable	Type 2 Diabetes Prev	Obesity	Inactivity	PM_{2.5}	Elevation
Type 2 Diabetes Prev	1	.733	.693	-.38	-.197
Obesity	.733	1	.714	-.275	-.073
Inactivity	.693	.714	1	-.088	.13
PM_{2.5}	-.38	-.275	-.088	1	.796
Elevation	-.197	-.073	.13	.796	1

After performing the OLS regression, the author utilized a Geographically Weighted Regression (GWR) to explore whether the relationships among variables vary geographically across Virginia. Wheeler and Páez (2010) explained that GWR allows a regression model to be

fitted at the geographic location of each observation in the dataset. This model estimates a beta coefficient as well as a R^2 value at each location, whereas in OLS regression, there is only one global, fixed value.

Results

The first model developed consisted of type 2 diabetes as the dependent variable, and $PM_{2.5}$ and obesity as the independent variables. The adjusted R^2 for this model was significant ($R^2 = 0.533, p < 0.001$). Table 2 contains the regression results for this first model.

Table 2

Regression results ($PM_{2.5}/obesity$)

Variable	Coefficient	Robust Std. Error	Robust <i>t</i> Statistic	Robust Probability
Constant	3.0455	3.057	0.996	.321
$PM_{2.5}$	-0.142	0.231	-0.614	.54
Obesity	0.296	0.026	11.441	< .001

For the models to be valid in OLS regression, certain assumptions need to be met. First, the relationships between the independent variables and the dependent variable need to be linear. Second, the models need to be properly specified by including all explanatory variables relevant to the analysis. A misspecified model will be missing a key explanatory, independent variable. To determine whether the model is misspecified, the residuals are tested for spatial autocorrelation utilizing the Moran's I statistic. This tests regression residuals for statistically significant clustering, which could result in invalid statistical inference. The variance of the error residuals is also examined. For the model to be valid, the residual variance should not exhibit heteroscedasticity, which means the variances are unequal. The Koenker (BP) Statistic tests the residuals for heteroscedasticity. A statistically significant result confirms the residuals have unequal variance. A solution for this issue is to consult the robust probabilities, which take this issue into account. The final residual assumption looks at whether or not the standardized residuals are normally distributed. This can be performed visually using a histogram and also by

utilizing the Jarque-Bera test. A statistically significant result ($p < .05$) means the residuals are not normally distributed and that the model predictions are biased.

The probabilities for the PM_{2.5} variable were not significant, whereas the obesity variable was significant. This model was not able to confirm that higher rates of PM_{2.5} were associated with a higher prevalence of diabetes. The Jarque-Bera test probability was not significant; therefore, the residuals were normally distributed. The Koenker (BP) Statistic yielded a significant result, indicating that the residuals exhibit heteroscedasticity. The Moran's I value statistic was .534, indicating that the regression residuals were randomly distributed. The robust probabilities were utilized because of the heteroscedasticity in the residuals. Although this model was valid, it did not confirm previous research findings that higher rates of PM_{2.5} were associated with higher diabetes prevalence. The joint Wald statistic probability indicated that the overall model was significant ($p < .001$). The adjusted R² value was .533, indicating this model explains approximately 53% of the variance in the prevalence of type 2 diabetes.

Table 3

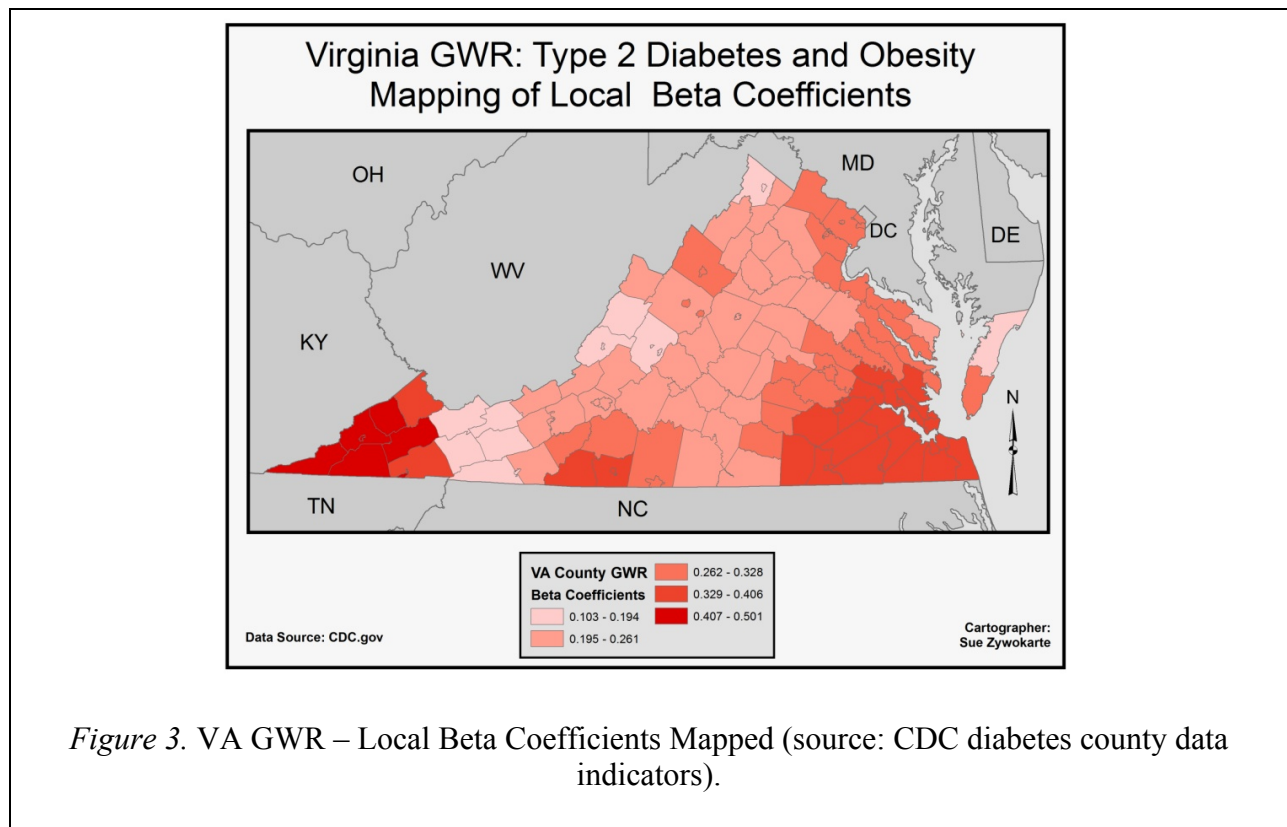
Regression results (elevation/obesity)

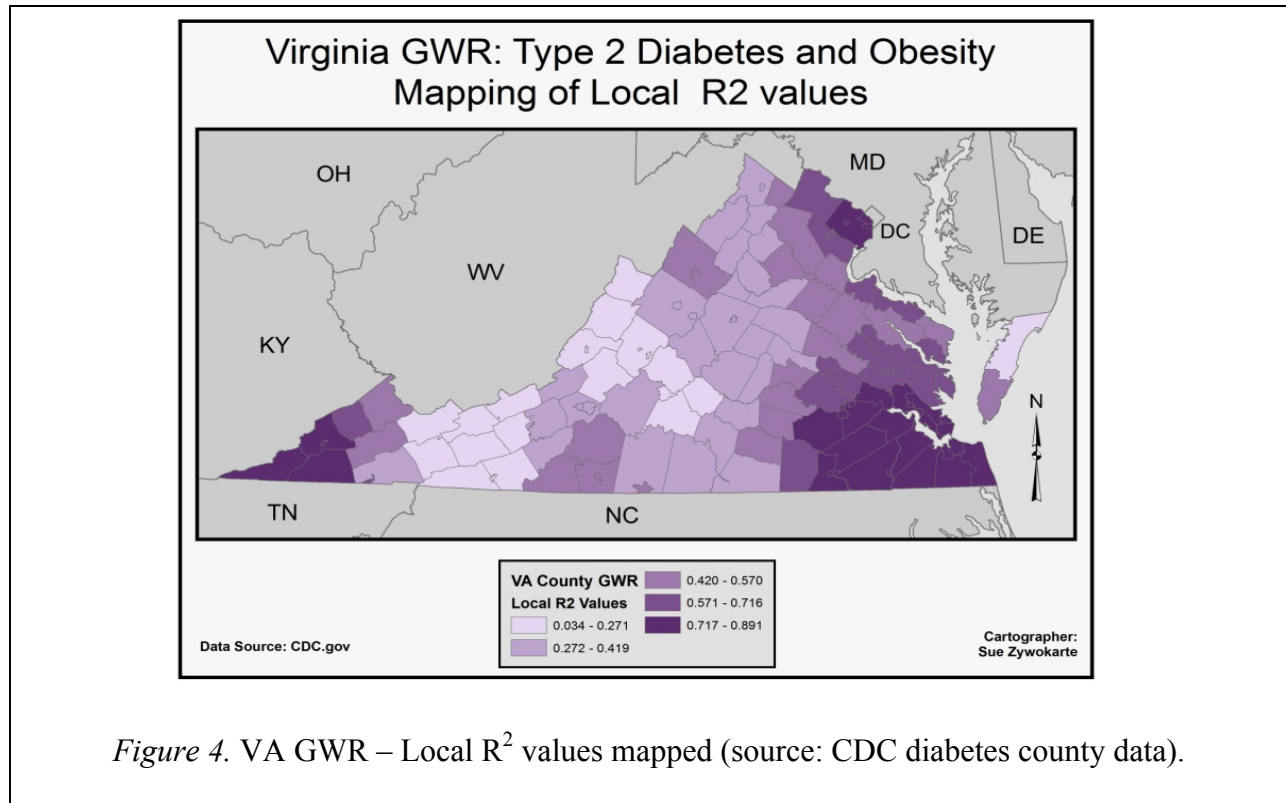
Variable	B-coefficient	Robust Std. Error	Robust <i>t</i> statistic	Robust Probability
Constant	1.251	.714	1.753	.082
Elevation	-0.0002	0.0004	-0.515	.608
Obesity	0.298	0.026	11.531	< .001

The second model developed examined the relationship between type 2 diabetes and elevation while controlling for obesity (Table 3). Again, the probabilities for the obesity variable were significant, but the elevation probability was not. Different model validity tests were employed. Moran's I was equal to .135 ($p = .508$), indicating that the residuals were randomly distributed. The Jarque-Bera probability was not significant; therefore, the residuals were normally distributed. The Koenker (BP) test yielded a significant result, indicating heteroscedasticity. The regression results in Table 3 contain the robust probabilities because of

the heteroscedasticity in the residuals. Also, the joint Wald statistic probability is less than .001, indicating that the overall model is significant. The adjusted R^2 value was .533 for this model, also indicating that the model explained approximately 53% of the variance in the dependent variable.

Figure 3 displays the GWR results, with the beta coefficients for the obesity variable varying across space. Each county observation also receives an R^2 value. It is of interest that there appears to be areas in Virginia where higher and lower beta coefficients cluster. Higher values cluster in the far Southwest and the far Southeast regions of Virginia. These areas show where the independent variable obesity exhibits a stronger influence on the occurrence of type 2 diabetes. Figure 4 displays the local R^2 values for each county/city jurisdiction. This depicts the varying fitness of the model throughout the study area.





Discussion

The two models presented did not confirm that higher rates of PM_{2.5} or higher mean elevations resulted in greater diabetes prevalence. The results from the first model indicated that the PM_{2.5} beta coefficient could not be trusted, as the probability was not significant. Current research findings suggest that PM_{2.5} environmental pollution is related to type 2 diabetes prevalence. The second model's beta coefficient for elevation also could not be trusted because of a non-significant *p* value. Both models showed a strong relationship between type 2 diabetes and obesity, with significant results. After determining both models' validity, obesity accounted for about 30% in the variation in type 2 diabetes for both models. This follows the findings in numerous other studies and is an established relationship. Tests for the validity of both models resulted in model validation with the use of robust probabilities to address the issue of heteroscedasticity in the residuals of both models.

Limitations of the first model were related to the averaging of PM_{2.5} levels over space and time. Different levels of exposure, when averaged, had a tendency to smooth out the

exposure level throughout the 2011 year. Also, high and low exposure regions were smoothed from the aggregation to county level data. This smoothing effect might mask high or low areas, resulting in less apparent variation.

Elevation data, aggregated to county level by interpolation processes, are averages of the high and low values. Woolcott et al. (2014) utilized data aggregated to the state level for the US. The Virginia county data had an elevation range from sea level to 5,729 feet.

Another issue to consider is the “ecological fallacy,” which occurs with data aggregation. For the two models presented in this study, individual data is aggregated to the county level and results in less variance when compared to the original data.

Although this case study of environmental effects on the prevalence of type 2 diabetes in the Commonwealth of Virginia did not show a relationship between the two environmental factors—PM_{2.5} and elevation—it did verify the relationship between obesity and type 2 diabetes. Future research into other types of pollution and other geographic regions is warranted to explore environmental health geography effects on type 2 diabetes. Also, future studies might employ GWR further to explore how variations in geography at the local level could affect data results.

References

- Barker, L. E., Kirtland, K. A., Gregg, E. W., Geiss, L. S., & Thompson, T. J. (2011). Geographic distribution of diagnosed diabetes in the U.S.: A diabetes belt. *American Journal of Preventive Medicine*, *40*, 434-439. <http://doi.org/10.1016/j.amepre.2010.12.019>
- Brook, R. D., Xu, X., Bard, R. L., Dvornch, J. T., Morishita, M., Kaciroti, N., ... Rajagopalan, S. (2013). Reduced metabolic insulin sensitivity following sub-acute exposures to low levels of ambient fine particulate matter air pollution. *Science of The Total Environment*, *448*, 66–71. <http://doi.org/10.1016/j.scitotenv.2012.07.034>
- CDC Diabetes County Data. (n.d.). Retrieved from <https://www.cdc.gov/diabetes/data/countydata/countydataindicators.html> CDC Wonder Database. (n.d.). Retrieved from <http://wonder.cdc.gov/nasa-pm.html>
- Chien, L.-C., Alamgir, H., & Yu, H.-L. (2015). Spatial vulnerability of fine particulate matter relative to the prevalence of diabetes in the United States. *Science of The Total Environment*, *508*, 136–144. <http://doi.org/10.1016/j.scitotenv.2014.11.090>
- Geiss, L. S., Li, Y., Kirtland, K., Barker, L., Burrows, N. R., & Gregg, E. W. (2012, November 16). Increasing prevalence of diagnosed diabetes—United States and Puerto Rico, 1995–2010. *Morbidity and Mortality Weekly Report*, *61*, 918–921. Retrieved from <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6145a4.htm>
- Menke, A., Rust, K. F., Fradkin, J., Cheng, Y. J., & Cowie, C. C. (2014). Associations between trends in race/ethnicity, aging, and body mass index with diabetes prevalence in the United States: A series of cross-sectional studies. *Annals of Internal Medicine*, *161*, 328–335. <http://doi.org/10.7326/M14-0286>

- Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37, 17-23.
<http://doi.org/10.2307/2332142>
- Park, S. K., Adar, S. D., O'Neill, M. S., Auchincloss, A. H., Szpiro, A., Bertoni, A. G., ... Diez Roux, A. V. (2015). Long-term exposure to air pollution and type 2 diabetes mellitus in a multiethnic cohort. *American Journal of Epidemiology*, 181(5), 327–336.
<https://doi.org/10.1093/aje/kwu280>
- Patel, C. J., Bhattacharya, J., & Butte, A. J. (2010). An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS ONE*, 5(5), e10746.
<http://doi.org/10.1371/journal.pone.0010746>
- Pearson, J. F., Bachireddy, C., Shyamprasad, S., Goldfine, A. B., & Brownstein, J. S. (2010). Association between fine particulate matter and diabetes prevalence in the U.S. *Diabetes Care*, 33, 2196–2201. <http://doi.org/10.2337/dc10-0698>
- Pope, C. A. (2000). Epidemiology of fine particulate air pollution and human health: Biologic mechanisms and who's at risk?, *Environmental Health Perspectives*, 108(Suppl 4), 713–723. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1637679/>
- Saudek, C. D., Rubin, R. R., & Shump, C. S. (1997). *The Johns Hopkins guide to diabetes: For today and tomorrow*. Baltimore, MD: Johns Hopkins University Press.
- Sun, Q., Yue, P., Deiuliis, J. A., Lumeng, C. N., Kampfrath, T., Mikolaj, M. B., ... Rajagopalan, S. (2009). Ambient air pollution exaggerates adipose inflammation and insulin resistance in a mouse model of diet-induced obesity. *Circulation*, 119, 538–546.
<http://doi.org/10.1161/CIRCULATIONAHA.108.799015>
- Voss, J. D., Masuoka, P., Webber, B. J., Scher, A. I., & Atkinson, R. L. (2013). Association of elevation, urbanization and ambient temperature with obesity prevalence in the United States. *International Journal of Obesity*, 37, 1407–1412. <http://doi.org/10.1038/ijo.2013.5>
- Wheeler, D. C., & Páez, A. (2010). Geographically Weighted Regression. In M. M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis* (pp. 461–486). Berlin, Germany: Springer Berlin Heidelberg. Retrieved from http://link.springer.com/10.1007/978-3-642-03647-7_22
- Woolcott, O. O., Castillo, O. A., Gutierrez, C., Elashoff, R. M., Stefanovski, D., & Bergman, R. N. (2014). Inverse association between diabetes and altitude: A cross-sectional study in the adult population of the United States. *Obesity*, 22, 2080–2090.
<http://doi.org/10.1002/oby.20800>
- Zheutlin, A. R., Adar, S. D., & Park, S. K. (2014). Carbon dioxide emissions and change in prevalence of obesity and diabetes in the United States: An ecological study. *Environment International*, 73, 111–116. <http://doi.org/10.1016/j.envint.2014.07.012>