SPATIAL ANALYSIS OF DIABETES PREVALENCE IN ASSOCIATION WITH HEALTHY LIFESTYLE BEHAVIOR

IGNM Jaya[1*], N Sunengsih[1]
Department of Statistics, Padjadjaran University, Bandung, Indonesia

Abstract
Diabetes is one of the most prevalent diseases on a global scale. It is the primary cause of global epidemics and has grown to represent a significant component of the global disease burden. Indonesia is one of the countries that has a major diabetes problem, particularly in Bandung. Healthy lifestyle practices are a critical component that may have a significant impact on diabetes at the individual level. Through econometric modeling, we examine the influence of the healthy lifestyle behaviors index on the regional variation of diabetes prevalence rates in Bandung. We discovered that the spatial variation of diabetes is spatially independent using Moran's I. The conventional regression analysis revealed a negative effect of healthy lifestyle habits on the variation in the prevalence rate of diabetes. However, due to the limited sample size, the impact is not statistically independent.

Keywords: Diabetes, Spatial econometrics, Regression, Bandung

---

Introduction
Diabetes is a huge global public health problem (1-3). Diabetes, which is largely a glucose metabolic disorder, is inextricably linked to contemporary lifestyles (4). Diabetes is well recognized as a complex illness resulting from the dynamic interaction of numerous tissues and variables (genetic, epigenetic, and environmental) (3). Diabetes is a substantial public health problem, associated with significant morbidity and mortality. Type 1 diabetes, Type 2 diabetes, gestational diabetes, and prediabetes are the four subtypes of diabetes (5). Excessive thirst and hunger, frequent urination, tiredness or weariness, dry, itchy skin, blurred eyesight, and wounds that heal slowly are all symptoms of diabetes (6). Containing diabetes case growth is crucial, as diabetes can have negative health and economic repercussions at the household and regional level. Direct and indirect medical costs associated with diabetes were predicted to surpass $98 billion in 1997 (7). The community's unhealthy lifestyle is often viewed as the primary cause of the high diabetes rate (8).

Objective of the study
This study aim to evaluate the effect of healthy lifestyle behavior on spatial variation of Diabetes. This study is being conducted in Bandung, Indonesia. Bandung is the capital city of West Java, Indonesia's largest province.

Hypothesis
We hypothesize that given a spatial autocorrelation, healthy lifestyle behaviors have a considerable effect on spatial variance in diabetes prevalence.

Material and Method
Moran's I

The Moran's I measure of spatial autocorrelation is frequently used to determine the presence of spatial autocorrelation in data (9-11). Moran's I is available in both global and local sizes. Global autocorrelation is quantified using global measurements. Moran's I locality is used to determine the degree of autocorrelation within a spatial unit. Calculation of Global Moran's I (12):

$$I=\frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}\,(y_i-\bar{y})(y_j-\bar{y})}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{1}$$

where *n* represents the number of spatial units, y_i represents the observation variable at the i-th location, w_{ij} declares the element of the spatial weight matrix W. We use queen contiguity.
I denotes, Moran's Global Coefficient. Moran's I values, both global and local, range from – 1 < I < 1. Moran's I parameters can also be tested to determine whether autocorrelation is significant or not. The formula for the hypothesis is as follows:
H_0: I= E(I) no spatial dependence (E(I) value is close to zero)
H_0: I ≠ E(I) there is evidence of spatial dependence
To test this hypothesis, the following test statistics were used:

$$T_I = \frac{I - E(I)}{\sqrt{Var(I)}} \tag{2}$$

where $E(I)$ and $Var(I)$ denotes the expectation and variance of I (see [13])

Spatial Econometrics
Spatial econometrics is usually used to model the effect of some covariates $x_{i1}, \dots, x_{iK}$ on the response variable y_i by accounting spatial dependence in response variable or error term.

Spatial econometrics is typically used to model the effect of a set of variables $x_{i1}, \dots, x_{iK}$ on the response variable y_i by taking into consideration the response variable's or error term's spatial dependence. There are two common models are: spatial autoregressive model (SAR) and spatial error model (SEM) (13).
The following definitions apply to the spatial autoregressive (SAR) model:

$$y_i = \rho \sum_{j=1}^{n} w_{ij} y_j + \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i, \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \tag{3}$$

where y_i represents the continuous dependent variable, represents the autoregressive coefficient, w_{ij} is the (i,j) element of the spatial weight matrix, β_0 is the intercept, β_k represents the regression coefficient of the predictor variable x_{ik} and ε_i the error component. The spatial error model is defined as:

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i$$

$$\varepsilon_i = \lambda \sum_{j=1}^{n} w_{ij} \varepsilon_j + \nu_i;\ \nu_i \sim N(0, \sigma_\nu^2) \tag{4}$$

where λ denotes the spatial error autoregressive coefficient. To estimate the model parameters we use maximum likelihood approach and to select the best model we can apply Lagrange Multiplier Test (see (13) for detail)

Result and Discussion
Bandung city has 30 subdistrict with total population over 2.5 million in 2020. In 2020, the total number of cases of diabetes 43,906 with prevalence rate 1.79%. The spatial distribution of prevalence rate is shown in Figure 1.

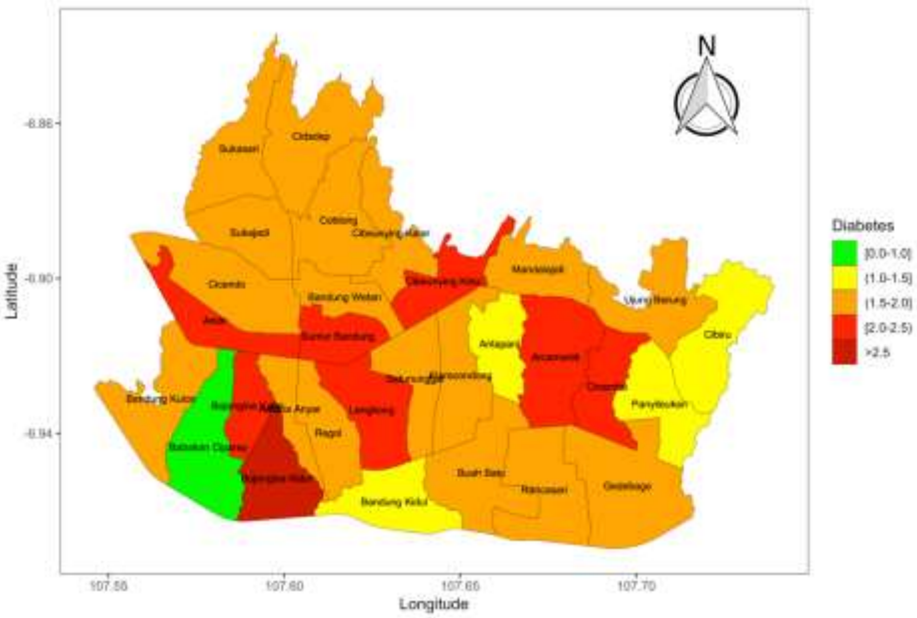

Figure 1. Spatial distribution of prevalence rate in Bandung city, Indonesia

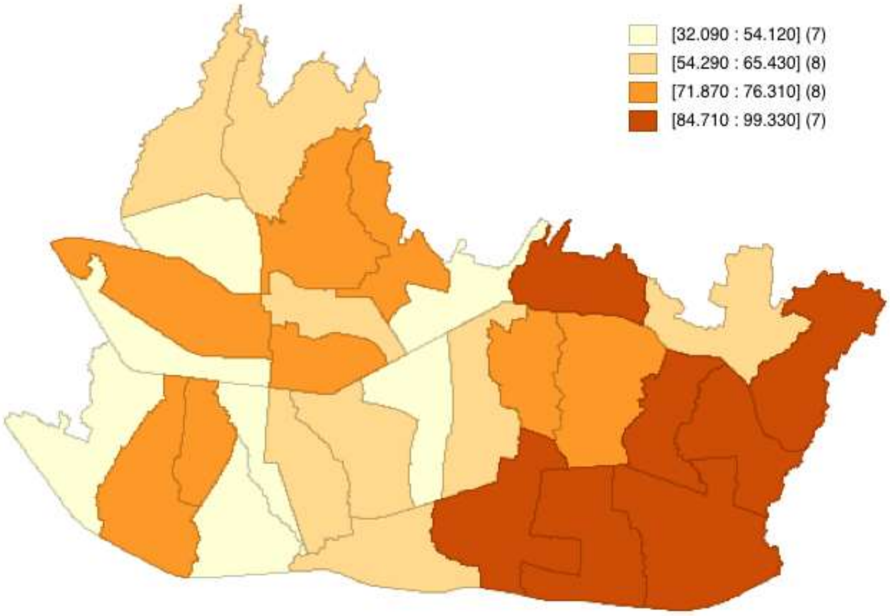

Figure 2. Spatial distribution of healthy lifestyle behavior in Bandung city, Indonesia

The spatial distribution of the healthy lifestyle behavior index is depicted in Figure 2. Batununggal (32.09%), Astana Anyar (41.53%), Bojongloa Kidul (42.63%), Sukajadi (44.76%), and Andir all have indexes below 50%. The highest indexes are found in Gedebage (94.98%) and Panyileukan (94.98%). The areas in southeast Bandung with the highest index of healthy lifestyle behaviors are clustered together.

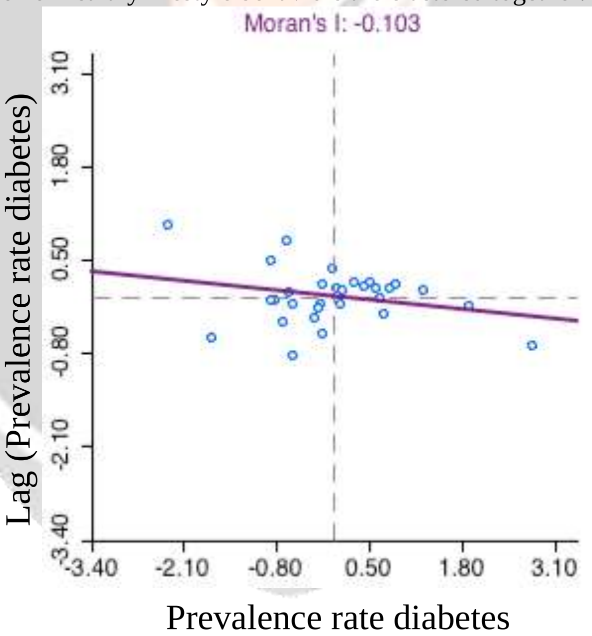

Figure 3. Moran's I: Spatial autocorrelation

Moran's I for prevalence rate is -0.103 (p-value =0.7306), indicating that there is no spatial autocorrelation in Bandung's diabetes prevalence rate.

Table 1. Spatial model selection

Statistics	Value	p-value
Spatial Lag Model	0.121	0.834
Spatial Error Model	0.044	0.728

The Lagrange multiplier test for spatial lag and error model have p-values larger than 0.05, suggests that Spatial Autoregressive Model and Spatial Error Model are not the best candidate model. The Best candidate model for modeling diabetes with healthy lifestyle behavior is the ordinary linear regression model. It agrees with Moran's

I statistics. Hence, we model the healthy lifestyle parameter using ordinary regression with the result is presented in Table 2.

Table 2. Regression parameters

Statistics	Value	p-value
Intercept	2.098	0.000
Healthy behavior	-0.005	0.272

We choose to utilize ordinary linear regression based on the results of the spatial model selection. The parameter estimates are reported in Table 2, and the regression model is shown below:

$$\hat{y}_i = 2.098 - 0.005$$

The fact that the regression parameter slop is -0.005 suggests that if healthy behavior increases by 1%, the prevalence of diabetes decreases by 0.005%. It is quite minimal due to the fact that diabetes is highly variable according to a variety of causes.

Table 3. Model validation

Statistics	Value	p-value
Normality	0.922	0.630
Non Heteroskedasticity	0.084	0.772

For model validation, tree assumptions are examined: autocorrelation, normalcy and heteroskedasticity. Moran's I was used to determine the autocorrelation (Figure 3). Jarque bera statistics were used to determine normality, and the Breusch pagan test was used to determine heteroskedasticity (Table 3). Normality and non-heteroscedasticity tests give p-values larger than 0.05 which indicate that the normality and non-heteroscedasticity assumption are met.

Conclusion

Diabetes prevalence has been increasing globally, especially in Indonesia (14). Bandung is the capital city of West Java, one of the largest provinces in Indonesia. Diabetes disease will affect over 43,000 people in 2020, with a prevalence incidence of 1.75 %. Numerous studies have been conducted on the risk factors for diabetes at the individual level (15-16). We focus on the regional level in our study. We hypothesize that the spatial characteristics and healthy behavior index can account for the regional variance in diabetes prevalence rates in Bandung. The Moran's Index demonstrates that the prevalence rate varies independently throughout space. Given the contiguity weight matrix, no spatial cluster is observed. Although healthy behavior has a detrimental influence on the prevalence rate, the effect is not statistically significant due to the small sample size (n=30). The negative consequences of healthy conduct suggest that by improving the index of healthy behavior, the number of diabetes cases in Bandung could be reduced.

References

[1] L. Anselin, Spatial Econometrics: Methods and Models, Netherlands: Springer, 1988.

[2] J. Lia, S. Wanga and X. Han, "Spatiotemporal trends and influence factors of global diabetes prevalence in recent years," Social Science & Medicine, vol. 256 , no. 2020, p. 113062, Social Science & Medicine.

[3] Y. Wu, Y. Ding, Y. Tanaka and W. Zhang, "Risk Factors Contributing to Type 2 Diabetes and Recent Advances in the Treatment and Prevention," Int J Med Sci, vol. 11, no. 11, p. 1185–1200, 2014.

[4] M. A. Al Mansour, "The Prevalence and Risk Factors of Type 2 Diabetes Mellitus (DMT2) in a Semi-Urban Saudi Population," Int. J. Environ. Res. Public Health, vol. 17, no. 7, pp. 1-8, 2020.

[5] K. Fritzen, L. Heinemann and O. Schnell, "Modeling of Diabetes and Its Clinical Impact," J Diabetes Sci Technol, vol. 12 , no. 5, p. 976–984, 2018.

[6] J. Li, S. Wang, X. Han, G. Zhang, M. Zhao and L. Ma, "Spatiotemporal trends and influence factors of global diabetes prevalence in recent years," Soc Sci Med , p. 256:113062, 2020.

[7] S.-Y. Sun, Z.-P. Liu, T. Zeng, Y. Wang and L. Chen, "Spatio-temporal analysis of type 2 diabetes mellitus based on differential expression networks," Scientific Reports, vol. 3, no. 2268, pp. 1-13, 2013.

[8] T. M. Frayling, "Genome-wide association studies provide new insights into type 2 diabetes aetiology," Nat Rev Genet, vol. 8, no. 9, pp. 657-62, 2007.

[9] CDC, "What is Diabetes?," 16 November 2021. [Online]. Available: https://www.cdc.gov/diabetes/basics/diabetes.html. [Accessed 17 January 2022].

[10] A. Pietrangelo, "Healthline," 28 June 2018. [Online]. Available: https://www.healthline.com/health/diabetes/types-of-diabetes#symptoms. [Accessed 17 January 2022].

[11] N. F. Ray, M. Thamer, E. Gardner and J. K. Chan, "Economic Consequences of Diabetes Mellitus in the U.S. in 1997," Diabetes Care, vol. 21, no. 2, pp. 1-14, 1998.

[12] N. M. Asril, K. Tabuchi, M. Tsunematsu, T. Kobayashi and M. Kakehashi, "Predicting Healthy Lifestyle Behaviours Among Patients With Type 2 Diabetes in Rural Bali, Indonesia," Clinical Medicine Insights: Endocrinology and Diabetes, pp. 1-13, 2020.

[13] A. Getis and J. Ord, "The analysis of spatial association by use of distance statistics," Geographical Analysis, vol. 24, p. 189–206, 1992.

[14] J. Ord and A. Getis, "Local spatial autocorrelation statistics: distributional issues and an application," Geographical Analysis, p. 286–306, 1995.

[15] A. Luc, "The local indicators of spatial association - LISA," Geographical Analysis, vol. 27, pp. 93-115, 1995.

[16] J. Lee and S. Li, "Extending Moran's Index for Measuring Spatiotemporal Clustering of Geographic Events," Geographical Analysis, pp. 1-22, 2016.