

# Are Neighborhood Features Associated With Premature Mortality in Toronto Neighborhoods?

Zhehui Zhao<sup>1#</sup> & Jingxin Yuan<sup>1#</sup>

<sup>1</sup> University of Toronto, Toronto, Canada

<sup>#</sup> The two authors contribute equally to the manuscript.

Correspondence: Zhehui Zhao, University of Toronto, Toronto, Ontario, Canada.

Received: June 20, 2022

Accepted: July 21, 2022

Online Published: August 1, 2022

doi:10.20849/iref.v6i3.1252

URL: <https://doi.org/10.20849/iref.v6i3.1252>

## Abstract

**Objective:** The relationship between neighborhood social and economic features and the residents' premature mortality rate is a controversial topic that has brought concerns from many local governments. The purpose of this paper was to determine the impacts of three indicators, including numbers of health providers, numbers of drug arrests, and neighborhood equity scores on premature mortality in the 140 neighborhoods in the City of Toronto.

**Methods:** Conducting regression analysis by using the data from January 2018 to December 2018 obtained from OpenData Toronto. The number of health providers, which shows how many medical service sectors the local community has is generated into a dummy variable ( $<1.5$  or  $\geq 1.5$  health providers per 1000 people), and all datasets are cleaned into the same unit, which is per thousand people. Both single regressions and multiple regression are used to compare the change in premature mortality rate, which means the deaths occurred before 70 years old.

**Results:** Taking all indicators into weighted consideration, the empirical evidence shows that the premature mortality rate increased by 4% on average with every one additional drug arrest incident occurring per thousand people while with every additional health provider per one thousand citizens, the premature mortality rate will decrease by 10% on average; In terms of neighborhood equity score, one point increase is associated with a roughly 1% decrease in premature mortality rate on average.

**Conclusion:** Social and economic factors are closely associated with the local premature mortality rate and actively improving the local living conditions can decrease the premature mortality rate while preventing serious issues before it actually occurs.

**Keywords:** social economics, premature mortality rate, social equity, linear regression

## 1. Introduction

### 1.1 Interest

As one of the most developed cities in Canada, Toronto has excelled in the fields of economics, politics, and education, and many people choose to live here because it is a good choice for their families and personal future development. Toronto has a total of 140 neighborhoods but each neighborhood has a different economic and social performance. This is mainly reflected in the per capita standard of living, the strength of health care, and the community crime index. Some of the rich neighborhoods, such as York Mills-Windfield and Sunnybrook, have seen prices soar in houses and have attracted many expensive schools and private hospitals. People tend to choose these neighborhoods or invest in them because they can expect to see significant returns in the future.

As living standards improve and technology flourishes, life expectancy in Canada is increasing from 81.38 years in 2011 to 82.81 years in 2021. People have more ways to cope with the disease and the likelihood of living longer increases. Premature mortality is defined as death before age 75, which includes both natural and unnatural deaths. Premature mortality is considered preventable and is usually due to health problems such as chronic diseases or poor physical development caused by lifestyle habits, other unpredicted accidents may also lead to premature mortality. At the same time, economic and social inequalities also contribute to the increase in premature mortality. Looking at the 2011 premature mortality data for 140 neighborhoods in Toronto, they are varied among neighborhoods and there are some disproportionately large outliers in several neighborhoods. According to

*Factors Associated With County-Level Variation in Premature Mortality Due to Noncommunicable Chronic Disease in the United States, 1999-2017*, in addition to premature mortality due to disease in the United States, increased socioeconomic inequality also contributes to increased premature mortality (Song et al., 2020).

To explore what factors contribute to premature mortality in 140 neighborhoods in Toronto, the purpose of this paper is to analyze whether the different features of each neighborhood affect premature mortality in each neighborhood. Reducing premature mortality is the goal of every government, by finding links between other factors, the results will provide a better understanding of the causal factors and allow the government to take measures to reduce the number of premature mortality in all possible aspects.

### 1.2 Research Question

Based on interest and effectiveness of the research the research question is are neighborhood features including social factors and economic factors associated with the premature mortality rate in Toronto Neighborhoods?

## 2. Methodology

### 2.1 Data Set

The data sets are provided by the City of Toronto Open Data Catalog website to collect all variables which are needed for analyzing neighborhood's features, objects are 140 neighborhoods in Toronto in 2011, 4 datasets are included to merge a new data called `d_merged_total`.

- `Health2011.csv`: Neighborhood health status and medical conditions.
- `Safety2011.csv`: Neighborhood crime incidents and safety hazards.
- `Equity2011.csv`: Level of social and economic development of the neighborhood.
- `wellbeing-toronto-population-total-2011-2016-and-age-groups-2016.csv`: population 2011.

Neighborhood crime and health are social factors and equity is an economic factor. Some studies have shown that premature mortality is related to some chronic diseases, but they are also closely linked to social inequality. It is more comprehensive to study the premature mortality rate from both economic and social aspects. Three most relevant variables for each data set will be explained in the next part. Mutate functions in R studio convert all three variables whose units are quantities to a number per 1000 people. This is because the number of people in each Canadian neighborhood ranges from a few thousand to tens of thousands and measuring the value per 1000 people makes the data more visible and understandable. Three indicators were selected as independent variables: number of drug arrests, number of health providers, and neighborhood scores, and the dependent variable was the number of premature mortality per 1,000 people. They will participate in the prediction on behalf of their own data set.

### 2.2 Independent Variables

Drug arrest is the first independent variable because it is a prime factor and is most likely to be associated with premature mortality. These drug crimes include abusers of opiates, stimulants, cannabis, hallucinogens, and sedatives/hypnotics. Nyhlén (2011) has a study of the effects of substance abuse and mental disorders on premature mortality. Physical harm from drug abuse and addiction as well as mental disorders are common in drug crimes. The results of the study showed an increased risk of premature mortality before the age of 69 in the group of drug abusers. The cumulative incidence of drug-related mortality expected for opioid and barbiturate abusers was significantly higher during the 37-year observation period. Therefore, drug arrests would be used as a crime factor to estimate premature mortality because they are possibly highly linked.

Health provider is another independent variable. The meaning of this variable is the number of businesses related to the health industry owned by each neighborhood. Examples include hospitals, doctors' offices, pharmacies, and clinics. When there are more health providers in a neighborhood, it means that people have easier access to the health industry. For example, if there are many hospitals or clinics in your neighborhood, then you can go for regular health checkups and monitoring to eliminate or treat chronic diseases. There may also be more health talks and daily health classes in the neighborhood, more education and awareness of chronic diseases caused by daily habits, and people will take their health more seriously and take more action to maintain their health.

For the two independent variables above, it is important to note that their units are the number of drug arrests and health providers in each neighborhood. Although neighborhood context and culture may influence their values, it is important to note that the population base of each neighborhood is also a very large factor. This is because the larger the population, the more hospitals, and drug arrests there are likely to have. The population range for the 140 neighborhoods in Toronto is between 6488 and 53,350, and to eliminate the effect of population on the rest of the

data analysis, both dependent variables were converted to the number of drug arrests and health providers per 1,000 people because the value for premature mortality is also among 1000 people.

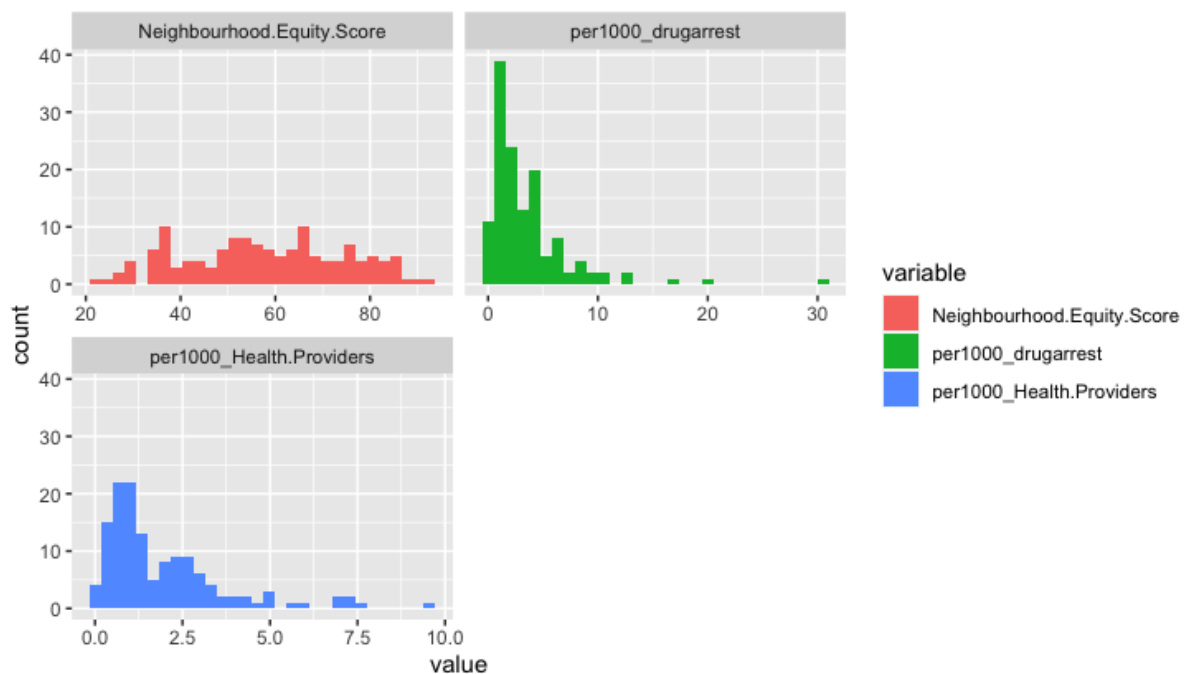


Figure 1. Difference in Tendency of X variables

The last dependent variable is the neighborhood equity score, this is the score of each community's economic development and is provided by the City of Toronto, Social Development Finance & Administration. This score is a good indicator of how developed a community is and the economic level of its people because it scores from a perspective that includes economic opportunities, social development, participation in decision making, natural environment, and healthy living-related outcomes. This variable serves as an economic factor for each community. Initially, income per capita is considered as an economic factor, but since these values are counted every five years from 2006 to 2011, which includes the financial crisis, using per capita income does not remove the impact of the large gap between rich and poor on the average. Therefore, the neighborhood equity score is the most appropriate economic factor because it not only measures the economic level of the year but also covers opportunities and development. Overall, there are three independent variables: per1000\_drugarrest, and per1000\_Health Providers, Neighborhood Equity Score.

### 2.3 Dependent Variable

The independent variable is the number of premature mortality per 1,000 people in each neighborhood. Since life expectancy in Canada had reached 81.38 years in 2011, mortality before the age of 75 was considered premature mortality. The mean value of this variable is 2.1, with some neighborhoods as high as 6.1, and the minimum is 1 premature mortality per 1,000 people. Such a large gap raises our attention. Nowadays, an increasing number of researchers have been noting differences in premature mortality due to inequality, with similar studies in Taiwan in addition to some US cities. "The association between community neighborhood social determinants of health and premature mortality was shown to be region-specific. Race and education explain nearly 84% of the premature mortality disparity. (Weng et al., 2021) also mentioned that these findings provide empirical evidence for the development of site-specific public health programs for geographically prioritized areas."

Since the difference in premature mortality between Toronto neighborhoods is right-skewed according to plot2, log transformation can help those extreme values for better data analysis later.

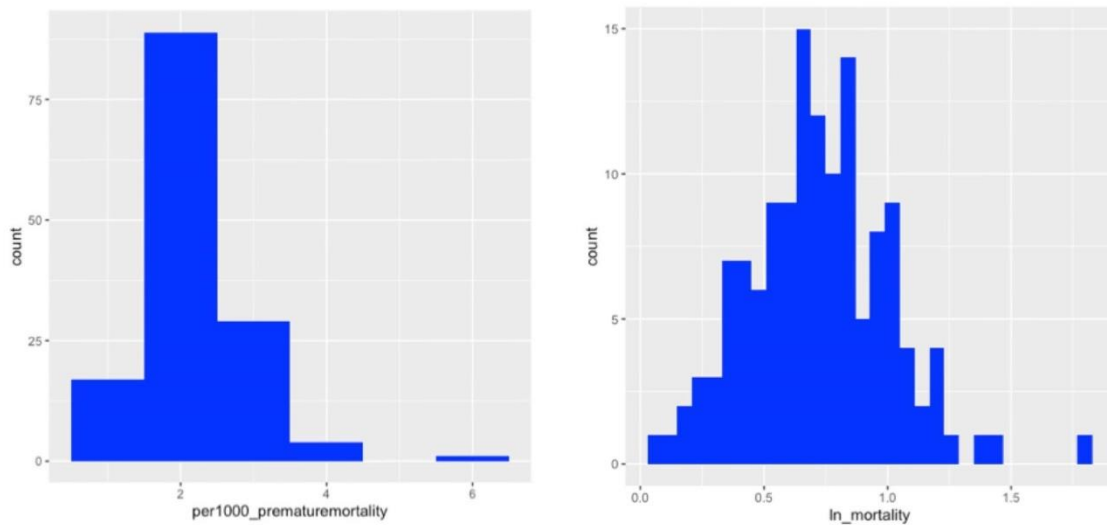


Figure 2. Log transformation on per thousand premature mortality

### 3. Regression Model and Analysis

Starting finding the relationship between our three x variables (drug arrests/per 1000 person, number of health providers/per 1000 person, and neighborhood equity score) and y variable (premature mortality/1000 person) respectively by generating three single regressions. From Figure 3 below, it is clear that both the number of health providers and neighborhood equity scores have negative associations with our y variables, indicating that when people have access to more health providers in their local neighborhoods and live in neighborhoods that have better living conditions and environments, the death occurred before 75 years old would decrease and they would have a longer life expectancy. In terms of a drug arrest, the slope of the regression line is positive, showing that drug arrest is positively related to premature mortality rate. The real-world scenario would be that people have more interactions with drugs, either dealing with them or taking them, which would lead to an increase in premature mortality rate and lower life expectancy.

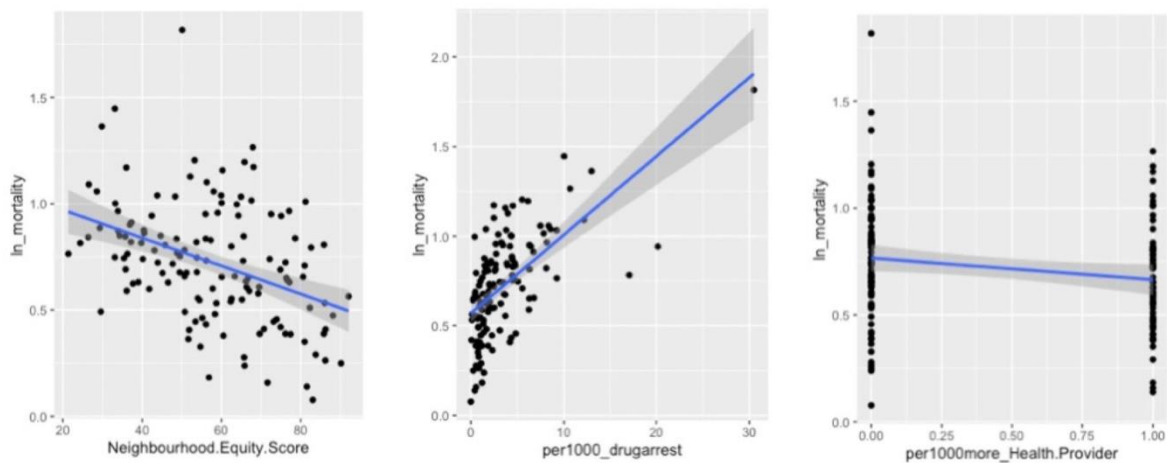


Figure 3. Single Regressions

Considering the fact that statistical significance is very critical in academic research, above three single regressions will examine their significance. The result shows that the all three single regressions are statistically significant. Also, it is worth noticing that the coefficient of the third graph (per1000more\_Health.Provider vs. ln\_mortality) is smaller than the previous two, indicating that the single effect of it on premature mortality is less without taking other factors into consideration.

However, the premature mortality rate is a controversial topic that is influenced by a series of variables. To better analyze how the indicators impact the local premature rate differently and offer reasonable suggestions to the local government, the influences based on weighted significance by creating multiple regressions are combined. One thing worth mentioning is that in the third graph of Figure 1, a new dummy variable called “per1000more\_Health.Provider” generated by replacing all the neighborhoods in the City of Toronto that have less than or equal to 1.5 health providers per thousand people with 0 and replace the neighborhoods that have more than 1.5 health providers per thousand people by 1. The dummy variable mainly serves as a binary variable to make our multiple regression model become a binary variable with continuous, which is easier to interpret compared with continuous-continuous regression models.

Before generating the multiple regression, it is important to ensure that the 4 fundamental assumptions for multiple linear regression (Poole & O’Farrell, 1971). The first assumption is linearity. From Figure 4 below, it is clear that our assumption one is a relatively strong assumption because although the x variables mostly follow the linear relationship, the QQ plots do not perfectly match the 45-degree lines to be completely linear.

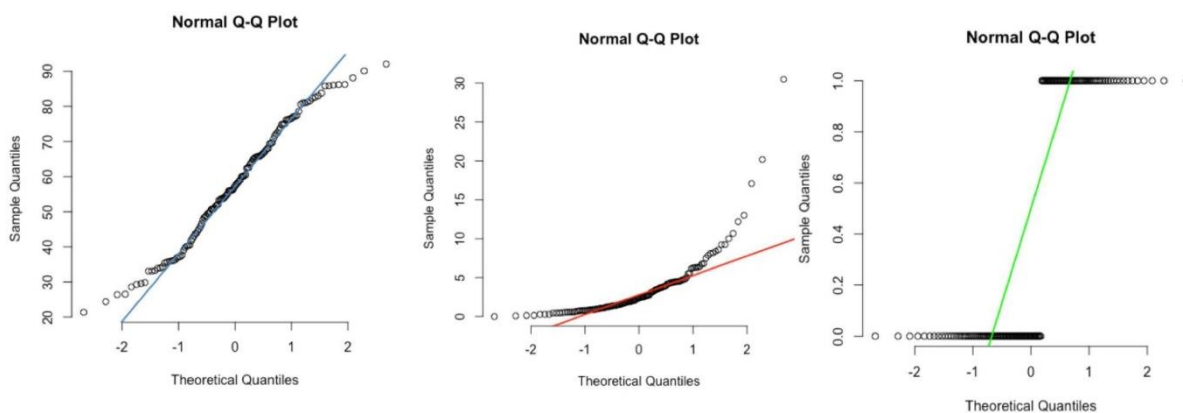


Figure 4. QQ plots examine linearity

In terms of homoscedasticity, which indicate that the datasets should not contain large outliers. Although the initial dataset suggests that there are several large outliers in the Drug.Arrest variable, after we generated the variable into per thousand and eliminated the impact of population, the dataset does not contain outliers anymore and so are the rest two datasets. Therefore, the second assumption is met and would be considered as valid. Thirdly, the assumption of independence which requires that all the observations are independent of each other is met because we would introduce an interaction term to minimize the dependency among observations. Lastly, the normality assumption would be a relatively strong one as well because the data we obtained show mostly normal distribution, especially after we apply the log transformation, they do not show perfect normal distribution and may lead to biased results.

With the four assumptions for multiple linear regression, by using lm\_robust to regress all three x variables on the y variable, the final regression equation is below.

$$Premature.Mortality_i = \beta_0 + \beta_1 * N.E.S_i + \beta_2 * D.A + \beta_3 * H.P + \epsilon_i$$

However, the concern of omitted variables arise as the variable Health Provider may be corrected with another x variable Neighborhood Equity Scores while it also is corrected with the y variable Premature Mortality. The omitted variables would cause bias that will impact the value and even the directions of the three coefficients. In response to possible omitted variables bias there is a single regression to determine the relationship between these two x variables. The result shows that there is a statistically significant relationship between the two variables with

no 0 in the 95% confidence interval (3.01369, 14.449321), and the estimated coefficient is 8.731449, which shows that with one unit increase in the health provider per 1000 people, the Neighborhood Equity Score would experience an increase in 8.731449.

In order to minimize the impact of the omitted variable and better meet the assumption three of independence, an interaction term added into multiple regression model called “Neighborhood Equity Score \* Health providers” and generated a final version of the multiple regression equation that shows the weighted impacts of all the three variables with minimized omitted variable bias:

$$Premature.Mortality_i = \beta_0 + \beta_1 * N.E.S_i + \beta_2 * D.A + \beta_3 * H.P + \beta_4 * N.E.S * H.P + \epsilon_i$$

By tabling the multiple regression in R and generating the coefficients of all four terms, the results show that the coefficients have changed compared with the previous single regressions. From the graph, it is clear that the omitted variables have an impact on the regression model and make the regression coefficients change by various amounts in different directions. From the multiple regression, the main finding is that both the number of health providers per 1000 people and the neighborhood equity score has a positive relationship with the local premature rate (in the unit of per 1000 people) while the drug arrest rate (also calculated by per 1000 people) is negatively associated with the premature mortality variable. Despite the small difference in the magnitudes of coefficients, this finding is most consistent with the finding from the three single regressions we generated previously. Also, the 95% confidence interval shows that all the four variables have statistically significant effects on the y variable. One thing that brought our attention is that the coefficient of the interaction term “Neighborhood Equity Score \* Number of Health Provider” is 0.002968084, which indicated that similar to the variable “drug arrest”, the interaction term has a positive association with the y variable, which is contradicting to the findings from analyzing the coefficients respectively. The possible reasons involve two perspectives. First, after table the regression outcome (Figure 5), the interaction term “Neighbourhood.Equity.Score:per1000more\_Health.Provider” in Model 5, which is our finalized regression model, is not statistically significant, which means the result cannot deny the possibility that the coefficient generated for this interaction term is due to the particularity of the dataset instead of reflecting the real association. Another possible reason is that the units for the two terms are not the same. While standardized all the other variables per thousand to minimize the influence of population, the Neighborhood Equity Score is calculated by the City of Toronto Open Data website with no unit, which would cause the difference in population among neighborhoods.

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	1.10* [ 0.97; 1.24]	0.57* [0.51; 0.62]	0.77* [ 0.70; 0.83]	53.89* [50.17; 57.61]	0.90* [ 0.74; 1.05]
Neighbourhood.Equity.Score	-0.01* [-0.01; -0.00]				-0.00* [-0.01; -0.00]
per1000_drugarrest		0.04* [0.03; 0.06]			0.04* [ 0.03; 0.05]
per1000more_Health.Provider			-0.10* [-0.20; -0.01]	8.73* [ 3.01; 14.45]	-0.25* [-0.50; -0.01]
Neighbourhood.Equity.Score:per1000more_Health.Provider					0.00 [-0.00; 0.01]
R <sup>2</sup>	0.16	0.39	0.03	0.06	0.48
Adj. R <sup>2</sup>	0.15	0.39	0.02	0.06	0.46
Num. obs.	135	135	135	135	135
RMSE	0.26	0.22	0.28	16.59	0.21

\* Null hypothesis value outside the confidence interval.

Statistical models

Figure 5. Multiple Regression Output

#### 4. Discussion

Our research results verify the association between the neighborhood features and the premature mortality rate among 140 neighborhoods in the City of Toronto. The neighborhood features used for analyzing, including the records of drug arrests, the number of health providers, and the neighborhood equity score, demonstrate significant correlations with the early mortality rate. In addition, log transformation is used to process the premature mortality rate prior to running regression, and the log-transformed premature mortality rate is roughly normally distributed with a few large outliers.

According to our regression results, the number of drug arrests has a significant positive correlation with the premature mortality rate in Toronto neighborhoods at a 95% significance level. As a result, the premature mortality rate increased by 4% on average with every one additional drug arrest incident occurring per thousand people. The number of drug arrests could reflect social welfare, the average educational level, and other factors contributing to self-protection awareness and community security. Neighborhoods with fewer drug arrests are perceived as safer and healthier, so the avoidable part of the premature mortality can be prevented. In comparison, neighborhoods with more drug arrests are perceived as having lower security and more drug abuse incidents. Thus there is less control over the premature mortality rate.

Similar to the number of drug arrests, the number of health providers illustrates a negative correlation with the premature mortality rate, and the correlation is significant at the 95% significance level. In order to get a precise result, we use dummy variables and 1.5 health providers per one thousand citizens as the boundary to test the correlation between the number of health providers and the premature mortality rate. The result shows that for every additional health provider per one thousand citizens, the premature mortality rate will decrease by 10% on average. Increasing the number of health providers will be more convenient for citizens to receive immediate treatment and regular healthcare. There will be more medical resources and professionals to help citizens with their concerns, thus reducing the premature mortality rate.

Moreover, the neighborhood equity score represents an area's overall financial and social performance, and the regression result illustrates a strong negative correlation between the neighborhood equity score and the premature mortality rate at 95% significance level. Reaching one point higher in the neighborhood score is correlated with a roughly 1% lower of premature mortality rate on average. As an essential economic indicator, the neighborhood equity score summarizes the average income level, infrastructure development, community services quality, and other contributing factors to the neighborhood performance. Generally, reaching a higher neighborhood equity score shows the specific neighborhood is well-behaved in these areas, affecting the citizens' physical and psychological well-being. Therefore, there will be less early mortality in the neighborhood with a higher equity score.

Our research also used multiple regressions to verify the correlation between the three independent variables and the premature mortality rate. The main findings show a similar pattern as the single regression. Both health providers and neighborhood equity scores negatively correlate with the premature mortality rate. In contrast, the number of drug arrests has a positive correlation with the premature mortality rate. To avoid the overlap of impact, one interaction term combines the neighborhood equity score and the number of health providers. After adding the interaction term, the multiple regression coefficients have moderate changes, but the general trends remain the same. In other words, the community with more health providers has a higher neighborhood equity score.

#### 5. Limitations

While the correlation between neighborhood features and the premature mortality rate is testified, our research has some limitations. First of all, there could be other omitted variables that can mediate the regression result. One example could be the average educational background. Communities with higher neighborhood equity scores might have more budget spending on popularizing education and establishing educational institutes. Thus, it is more convenient and supportive for citizens to receive an education. By acquiring more knowledge and life-essential skills, individuals can increase their health awareness, maintain a healthier lifestyle, and avoid potential disease, so the premature mortality rate will be reduced. However, this research only focuses on the impact of the three most outstanding factors. If doing further research on the omitted variables, the result could be more valid.

Secondly, the outliers from the drug arrest and the neighborhood equity data sets are not removed before running the regression. For example, most communities have less than 10 drug arrest incidents per 1000 individuals, but a few have roughly 30 drug arrest incidents per 1000 individuals. These extremely large outliers could be programming errors or actual records, but it is ambiguous. Therefore, measuring the correlation with the outliers might make the regression less representative in all Toronto neighborhoods.

Thirdly, three major social and economic factors are used to predict premature mortality rate in Toronto, but there could be many more independent variables that can affect the results. The average income level could be one influential contributing factor to explaining premature mortality. Unfortunately, the correlation between income and the premature mortality rate is insignificant. Moreover, data from 140 diverse neighborhoods, and some neighborhoods are relatively more developed and wealthier than the others, and some neighborhoods have their unique community culture. For example, downtown Toronto concentrates on the most influential industries, and Chinatown gathers many early Chinese immigrants. The regression results demonstrate an average correlation between the three independent variables and the premature mortality rate in over one hundred neighborhoods, so the result might not be sufficiently applicable in each identical neighborhood. Therefore, if the diversity and categorize the neighborhoods according to their development level and culture, the result could be more accurate.

## 6. Conclusion

In conclusion, this paper is to study the correlation between three neighborhood features and the premature mortality rate in Toronto. There are three independent variables: the number of drug arrests, the number of health providers, and the neighborhood equity score to run single regressions and multiple regressions. Also, there is an interaction term to avoid the overlap of the number of health providers and the neighborhood equity score. The results present that the number of health providers, the number of drug arrests, and the neighborhood equity score significantly correlate with the premature mortality rate. Based on our research, to effectively reduce the premature mortality rate, Toronto could spend more budget on providing educational opportunities to enhance health awareness and establish more health institutes to monitor and prevent drug abuse. Also, each neighborhood could develop better social facilities and security to protect citizens from drug arrests and avoidable diseases.

The data set we used are records in 2011, but the outbreak of COVID-19 pandemic strongly impact many neighborhood features in the last few years. The number of health providers and average income level will change accordingly, thus the regression results in our research might be less accurate. Therefore, it will be essential to use the recent data to analyze the correlation in the future study. Similarly, the environmental factors and the genetic inheritance are the two possible contributing factors of the premature mortality rate that could be further investigated as the extension of our current research.

## References

- Nyhlén, A., Fridell, M., Bäckström, M., Hesse, M., & Krantz, P. (2011). Substance abuse and psychiatric co-morbidity as predictors of premature mortality in Swedish drug abusers: a prospective longitudinal study 1970-2006. *BMC psychiatry*, *11*, 122. <https://doi.org/10.1186/1471-244X-11-122>
- Poole, M. A., & O'Farrell, P. N. (1971). The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, (52), 145. <https://doi.org/10.2307/621706>
- Song, S., Trisolini, M. G., LaBresh, K. A., Smith, S. C., Jin, Y., & Zheng, Z.-J. (2020). Factors associated with county-level variation in premature mortality due to noncommunicable chronic disease in the United States, 1999-2017. *JAMA Network Open*, *3*(2). <https://doi.org/10.1001/jamanetworkopen.2020.0241>
- Weng, S.-S., Chan, T.-C., Hsu, P.-Y., & Niu, S.-F. (2021). Neighbourhood Social Determinants of health and geographical inequalities in premature mortality in Taiwan: A spatiotemporal approach. *International Journal of Environmental Research and Public Health*, *18*(13), 7091. <https://doi.org/10.3390/ijerph18137091>

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).