**Relationships among Neighborhood Safety, Diabetes rates, and Life Expectancy**

Honors Undergraduate Thesis

Presented in Partial Fulfillment of the Requirements for the Business Administration Degree with Honors Research Distinction in the Fisher College of Business at The Ohio State University

By

Zihan Wang

Accounting Specialization in Honors Contract/Immersion Program

The Ohio State University

Fisher College of Business

2019

## Abstract

Diabetes has long been considered one of the primary threats to public health due to its high mortality rate and treatment cost. Many types of research have been done to understand how lifestyles could affect the risk of diabetes. In fact, every detail in daily life, from sleeping to eating, could potentially increase or decrease this risk. And this research pays attention to one of the most discussed factor, physical activity level. Based on the previous research, the physical activity level is negatively related to the risk of diabetes, suggesting that an individual who does more outdoor activities has a lower possibility to be troubled by diabetes. What attributes to a high activity level? Research in 2004 found that a safer neighborhood leads to an increase in physical activity, as residents do more exercise in a safer region. Then what helps make a neighborhood safer? The objective method to identify if a neighborhood is safe or not is by looking at its crime rates, and a low crime rate is an essential symbol of a safe neighborhood. This research is designed to find out the relationship among crime rate, life expectancy, and diabetes prevalence. The whole analysis took a separate look at time series and cross-sectional datasets; the former dataset focuses on Los Angeles County in California in a ten-year time frame and the latter one includes all counties in the United States in 2010. Regression analysis suggested that the crime rate is negatively related to life expectancy, as the residents in a less safe neighborhood tend to live longer; and the crime rate is positively related to the diabetes prevalence since fewer individuals are troubled by diabetes in a safe neighborhood than in a less safe one.

The relationships discovered in this research could be utilized by healthcare companies to make regional marketing strategies based on neighborhood crime rates. The

government could also apply the models in different neighborhoods to plan on the construction of amenities.

**Acknowledgements**

I would like to express my sincere thanks to Dr. Jason Reece who helped me locate the specific research topic and provided critical insights to the research results. I also want to offer my appreciation to Dr. Roger Bailey, the director of the Honors Contract/Immersion Program, for offering this exclusive opportunity and guiding me through the whole program.

**Vita**

June 2015 Nanjing No.1 Middle School

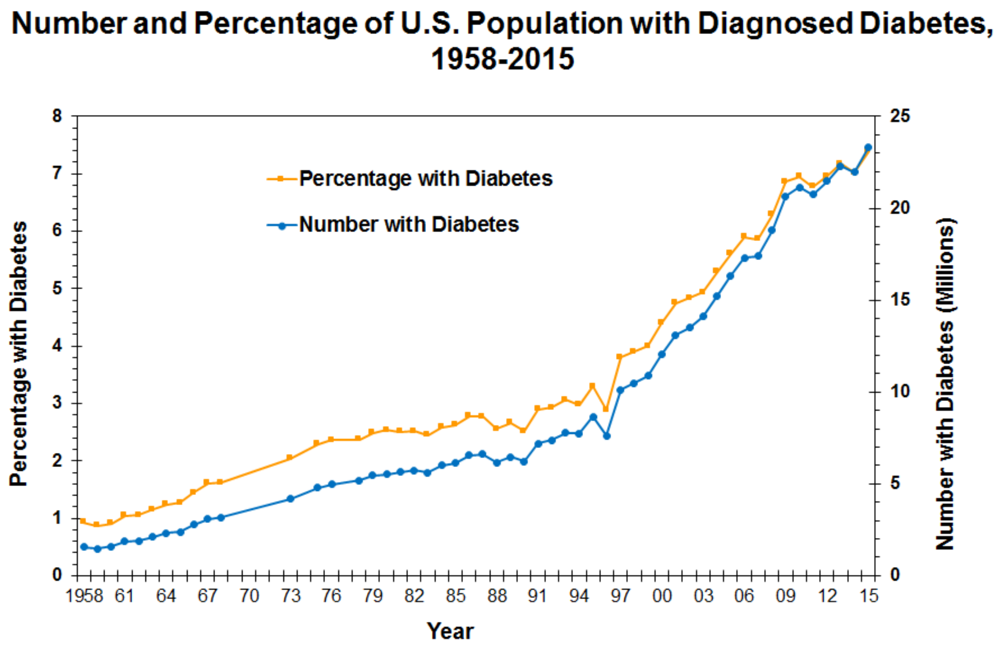
May 2019 B.S.B.A. Accounting, The Ohio State University

**Table of Contents**

## Introduction

Diabetes has long been considered one of the primary health threats that the world is facing in modern days. According to *The National Diabetes Statistics Report (2017)* edited by the Centers for Disease Control and Prevention: in the United States along, “an estimated of 30.3 million people of all ages - or 9.4% of the U.S. population – had diabetes in 2015”, in another word, around one in every ten people is troubled by diagnosed or undiagnosed diabetes, and this number is increasing. What makes diabetes even worse is its high mortality rate and treatment cost. According to the data in 2016 from the Centers for Disease Control and Prevention, around 25 were killed by diabetes in every 100,000 individuals. The government spent over $327 billion on diagnosed diabetes just in the year 2017 based on the statistics of the American Diabetes Association.

**Figure 1:**

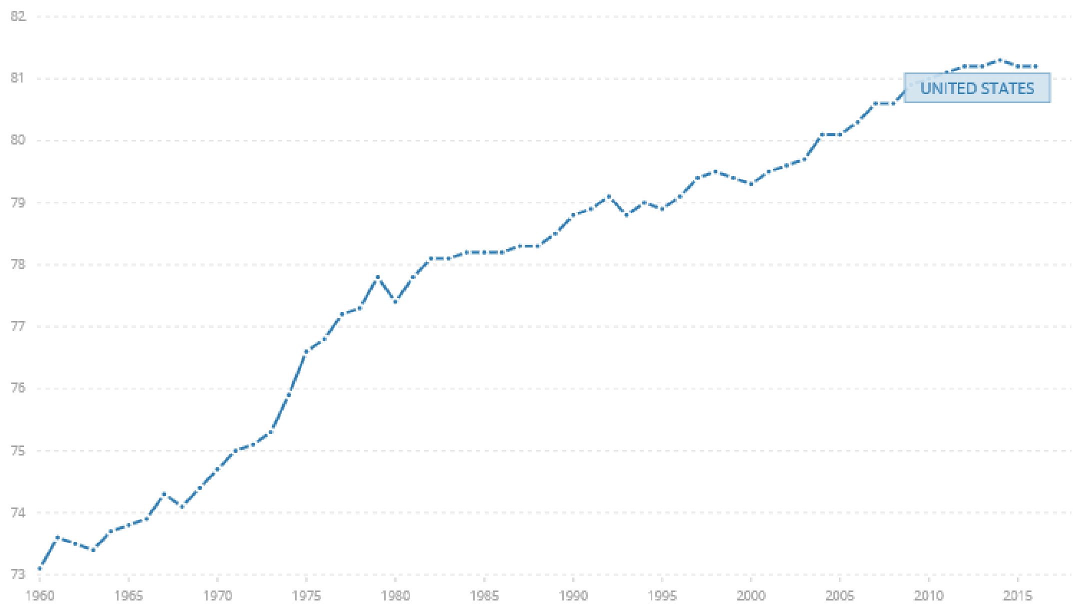


This plot chart suggests how the number and percentage of U.S. population with diagnosed diabetes changed from 1958 to 2015, and both presented increasing trends over the specific periods. There is no doubt that the negative effect of diabetes is expanding on our lives.

There are two main types of diabetes: Type I and Type II, and they could be primarily differentiated based on symptoms and risk factors: Type I patients lose their abilities to produce insulin, which generates glucose to feed the body's cells; while Type II diabetes patients lose their ability respond to insulin and neither produce enough insulin in later periods, in another word, those patients are still producing insulin which, however, could not be effectively utilized by their bodies. Risk factors for Type I diabetes include such congenital dimensions as family history, age, geography, and genetics. Those factors for Type II diabetes, on the other hand, contain more environmental and postnatal elements such as weight, diet, activity levels, and even belly fat (Osborn, C., 2017). This paper focused only on the diagnosed Type II diabetes since there are more environmental factors involved and therefore easier to analyze.

An important element that will be utilized to measure the effects of a safety neighborhood has on the diagnosed Type II diabetes rate is life expectancy. Life expectancy stands for the average years that a person is expected to live. The chart below suggests that the life expectancy of the United States is keeping increasing after World War II, then what are the differences of life expectancy between a healthy individual and diabetes diagnosed patient in a safe neighborhood? What about in an unsafe neighborhood? This research was designed to answer these questions.

**Figure 2: Life Expectancy of the United States, 1960 - 2015**



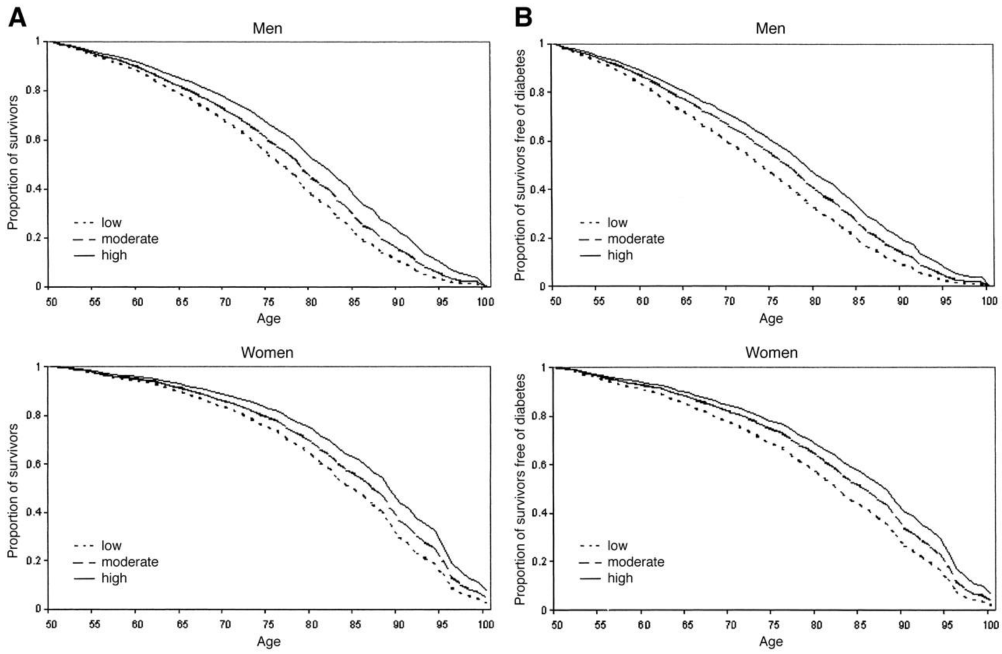

https://data.worldbank.org/indicator/SP.DYN.LE00.FE.IN?end=2016&locations=US&name_desc=false&start=1960&view=chart

The rest of this paper will first review the previous related research projects on diabetes, life expectancy, and neighborhood safety related problems and then identify the research question. Later the methodology part will illustrate the specific variables, tools, and data resources utilized in this research. At last, the results and conclusion will be discussed, follow up with some limitations and next steps.

## Literature Review

In fact, research that focuses on how lifestyles affect health begun decades ago. In 1999, Dr. Hu and his team found that a higher physical activity level contributed to a substantial reduction in risk of Type II diabetes. Four years later, another research team which was also headed by Dr. Hu (2003) proved that sedentary behaviors, such as watching television, elevated risk of obesity and Type II diabetes. Another research conducted by Jacqueline Jonker and her team in 2006 had a more comprehensive description of the relationship between the activity level and diabetes. This research was designed to calculate the differences in life expectancy between individuals with and without Type II diabetes in different levels of physical activity. The team constructed multistate life tables using the data from respondents at their 50 in the Framingham Heart Study. Then they used life expectancy to build connections with different physical activity levels after adjusting for age, sex, and potential confounders.

**Figure 3: Multistate Life Table for the Framingham Heart Study**




http://care.diabetesjournals.org/content/29/1/38.figures-only

The conclusions suggest that individuals with a moderate to a high level of activity have a longer life expectancy and live more years free of diabetes than those with lower activity level. Unfortunately, high activity level does not help reduce the mortality rate for diabetic patients.

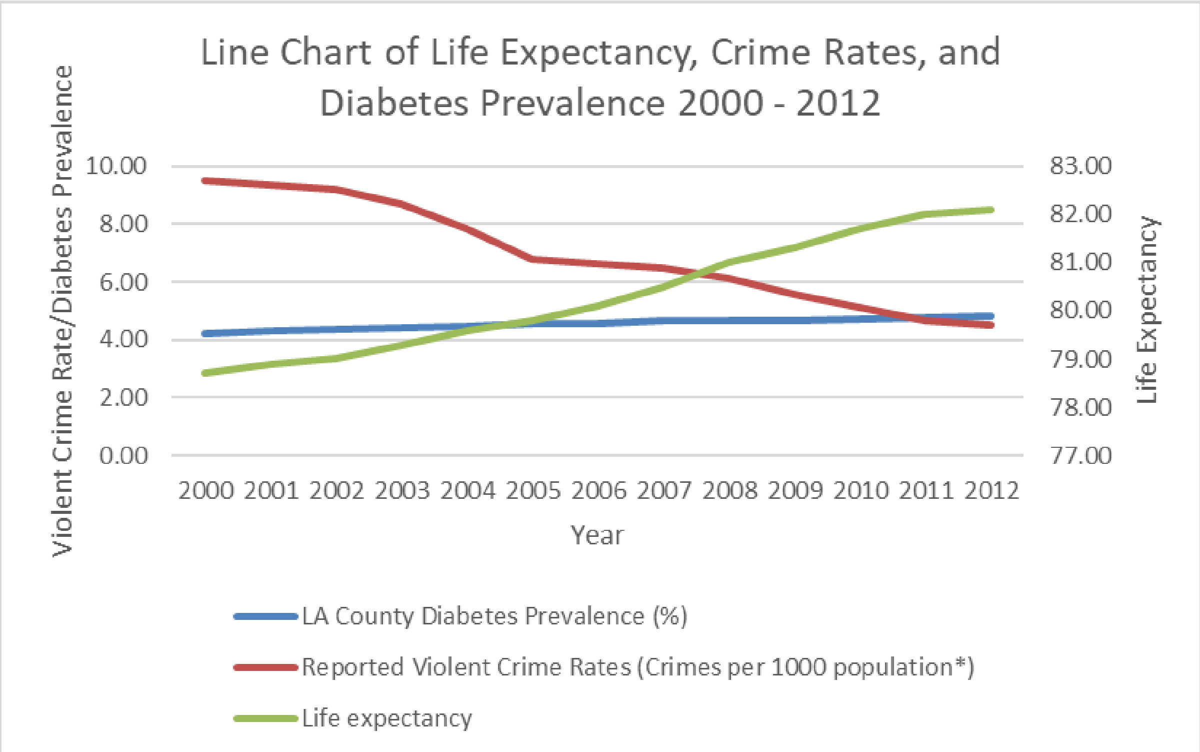
Activity level is such a compounded factor that so many attributes are involved in such as the exercise time length, public transportation density, and the walkability in the nearby region. Among all the related attributes, the neighborhood safety level is one of the most important elements. The research results from Beth and her team (2004) suggest that any methods contributed to a safer neighborhood with fewer disorders help increase physical activities and thereby reduce the risks of overweight. One year later, Hillary Buedette and Robert Whitaker designed a survey to test the hypothesis that if preschool children spend less time on playing outdoors, more time on watching television and therefore have a higher prevalence of obesity when living in neighborhoods that their mothers believe to be unsafe. In this cross-sectional survey on over 20 large U.S. cities, mothers were asked to report how long on average their children spent on outdoor playing and television watching. Children's BMI was also measured as an independent variable in the research. Mothers' perception of neighborhood safety was assessed with the index called "the Neighborhood Environment for Children Rating Scales". The results suggested that mothers' perceptions of neighborhood safety were related to how long their children spend on watching television instead of outdoor playing or to risks for obesity. In another word, neighborhood safety was not one of the reasons for the children's obesity.

However, the concept of neighborhood safety in this survey was based on subjective judgements - the rating scales assessed by mothers themselves. Is there a difference when utilizing more objective variables? This research will discover the relationships among life expectancy, diabetes rates, and neighborhood safety with more objective variables.

## Research Question and Primary Hypothesis

This research was designed to answer one question, that what are the relationships among diabetes prevalence, crime rate, and life expectancy? A line chart of these three variables was created below by utilizing time series data in Los Angeles County from 2000 to 2012. The good news in this chart is that the red line, which represents reported violent crime rates (per capita/10), shows a decreasing trend over the years and that the green line, which is the life expectancy, increases during the period, suggesting that Los Angeles County was becoming safer and people were living longer. However, the blue line, which represents the diabetes prevalence, keeps increasing as well, in an extremely slow rate though.

**Figure 4: Line Chart of Life Expectancy, Crime Rates, and Diabetes Prevalence 2000 - 2012**




It is not hard to explain from the chart that life expectancy may have a negative relationship with violent crime rates. However, the positive relationship between life expectancy and the diabetes prevalence seems abnormal since, based on previous research or common sense, diabetes has a negative effect on individual health and therefore should lead to a lower life expectancy. This irregular relationship might occur due to interactive effects from multiple variables. In this case, it is necessary to introduce an interaction variable in the research.

## Methodology

This research was designed to examine the relationships among diabetes prevalence, crime rates, and life expectancy by running regression analysis separately over time series and cross-sectional data. Time series data analysis focuses on the same neighborhoods in different time periods while cross-sectional data analysis, on the other hand, looks at various neighborhoods at the same time, and for each type of data, an interaction variable will be included, and regression equations with and without the interaction variable will be built and evaluated. The interaction variable was created by simply multiplying the crime rate with the diabetes prevalence, and this makes sense because these two variables are interrelated.

For the time series analysis, the most important step was the selection of the geographic level. It is important to find a proper level of the region because when collecting data at a high level, such as in a city or a state level, there will be tons of records in few categories. However, if the regional data is collected at a low level, the whole dataset will be shattered into too many categories each with little information. The top priority to specify the geographic level is its efficiency and the ability to generate the least biased data in a relatively small sample size. After careful consideration, the county level was believed to be a proper level and Los Angeles County became the research target due to its diversity and richness in data. The time series data analysis focused on Los Angeles County in 2000 – 2012 and the following variables were created and included (Appendix A):

- **Dependent variable Y**: Average Neighborhood Life Expectancy. Retrieved from Department of Public Health in County of Los Angeles;
- **Independent variable X1**: Reported Violent Crime Rates (per capita). Retrieved from

Los Angeles Almanac;

- **Independent variable X2**: Diabetes Prevalence (%). Retrieved from Los Angeles Almanac;
- **Time series variable X3**: consecutive nature number starts at 0
- **Interaction variable X1*X2**: Diabetes*Crime = Independent Variable X1*Independent Variable X2

The regression equation was: Life expectancy $= \beta_0 + \beta_1 Crime\ Rates + \beta_2 Diabetes\ Prevalence + \beta_3 Crime * Diabetes + \beta_4 Time\ Series + \varepsilon$. And It is important to notice that the results of this model summarize the situation in Los Angeles County in the given period instead of providing a universal solution.

In the cross-sectional data analysis, however, a nationwide dataset was collected and analyzed to generate a more comprehensive understanding of the relationships. The cross-sectional data analysis focused on over 2,400 counties in the United States in the year 2010 and the following variables were created and included (Appendix B):

- **Dependent variable Y**: Life Expectancy of Counties. Retrieved from Global Health Data Exchange;
- **Independent variable X1**: Adjusted Reported Violent Crime Index (Crimes*100/population), 2010. Retrieved from Federal Bureau of Investigation;
- **Independent variable X2**: Diabetes Rates (%), 2010. Retrieved from Centers for Disease Control and Prevention, Division of Diabetes;
- **Interaction variable X1*X2**: Crime*Diabetes = Independent Variable X1*Independent Variable X2

The regression equation was: Life expectancy $= \beta_0 + \beta_1 Crime\ Rates + \beta_2 Diabetes\ Prevalence + \beta_3 Crime * Diabetes + \varepsilon$.

There were no existing compiled datasets available, but all data could be obtained from government or healthcare organizations websites. In the datasets collected above, the independent variable diabetes prevalence is the percentages of diagnosed Type II diabetes patients in each region, and this research focuses only on Type II diabetes since there are more environmental factors involved in and therefore easier to analyze than Type I diabetes. Another independent variable crime rates, as an indicator of the neighborhood safety levels, focus only on violent crime (murder and nonnegligent manslaughter, forcible rape, robbery, and aggravated assault). As people are more likely to report a violent crime than any other types, violent crime rates become a more robust indicator of neighborhood safety. At last but not least, the dependent variable life expectancy, which stands for the average year an individual is expected to live, is an overall indicator of living standards in each neighborhood, and the analysis was designed to build an equation of life expectancy with crime rates and diabetes prevalence.

## Results

### Time Series Data Analysis

**Figure 5: Time Series Data Regression Analysis on Life Expectancy, Crime Rates, and Diabetes Prevalence, 2000 – 2012, LA County.**

| Regression Statistics | |
|---|---|
| Multiple R | 0.996808016 |
| R Square | 0.993626221 |
| Adjusted R Square | 0.991501628 |
| Standard Error | 0.11167281 |
| Observations | 13 |

| | Coefficients | Standard Error | t Stat | P-value |
|---|---|---|---|---|
| Intercept | 96.88277816 | 6.056099057 | 15.99755507 | 6.44121E-08 |
| Time | 0.442145621 | 0.058654922 | 7.538082135 | 3.54837E-05 |
| LA County Diabetes Rate | -4.034327422 | 1.236795348 | -3.261919951 | 0.009808582 |
| Crime Incidences (in 10M | -0.123103792 | 0.133110078 | -0.924826983 | 0.379185391 |

$$\text{Life expectancy} = 96.8828 - 0.1231 * \text{Crime Rates} - 4.0343 * \text{Diabetes Prevalence} + 0.4421 * \text{Time} + \text{error}$$

The first regression was built using crime rates, diabetes prevalence, and time series index as independent variables. According to the regression statistics, this model had a low standard error and an extremely high adjusted R square up to over 99%, which means that over 99% of the variation of the dataset could be explained by the model. The coefficients of "Diabetes Prevalence" and "Crime Rates" were negative, indicating that one unit increase in diabetes or crime rates leads to a negative change in life expectancy, such relationship though differed from the hypothesis but makes more sense. However, this model was not good enough to summarize relationships since the p-values of some coefficients were larger than 0.05 (preset alpha), illustrating that such variables were not significant ingredients in the model. To solve this problem, the interaction variable was added.

**Figure 6: Regression Analysis on Life Expectancy, Crime Rates, Diabetes Prevalence,**

**and the interaction term 2000 – 2012, LA County.**

| Regression Statistics | |
|---|---|
| Multiple R | 0.99692351 |
| R Square | 0.993856486 |
| Adjusted R Square | 0.990784728 |
| Standard Error | 0.116287662 |
| Observations | 13 |

| | Coefficients | Standard Error | t Stat | P-value |
|---|---|---|---|---|
| Intercept | 87.1838006 | 18.80151021 | 4.637063704 | 0.001672592 |
| Time | 0.395186398 | 0.105284942 | 3.75349399 | 0.005596175 |
| LA County Diabetes Rates | -1.868901897 | 4.158950341 | -0.449368649 | 0.665090464 |
| Crime Incidences (in 10M) | 0.552833475 | 1.242158355 | 0.445058775 | 0.668073167 |
| Diabetes*Crime | -0.146048356 | 0.266714326 | -0.547583467 | 0.59891829 |

$$\text{Life expectancy} = 87.1838 + 0.5528 * \text{Crime Rates} - 1.8689 * \text{Diabetes Prevalence} + 0.3952 * \text{Time} - 0.1460 * \text{Diabetes} * \text{Crime} + \text{error}$$

The second regression model included the interaction variable apart from the three original ones. However, the edited model did not perform as well as the previous one since the adjusted R square decreased and standard error increased. The positive coefficient of crime rates suggested that each increase in crime rates leads to an increase in life expectancy as well, a relationship that was contrary to common sense. With more variables with p-values larger than 0.05 (alpha), the model had fewer significant ingredients and therefore became less qualified to summarize the relationship.

In conclusion, the model without the interaction variable was a better summary of the relationships among life expectancy, crime rates, and diabetes prevalence in Los Angeles County from 2000 to 2012. And, according to the model, diabetes and crime rates have negative effects on life expectancy. In another word, a safer neighborhood with a lower diabetes rate contributes to a higher life expectancy.

However, this model was not perfect. First of all, the time series effect was so strong that the actual relationships might not be significant; also, one variable in the model had a p-value larger than 0.05 (alpha), indicating that it is not a significant ingredient in the model; at last but not least, the sample size was so small that this model might not be robust when applying to a different region with a larger sample size. Afterall, this model might be enough to summarize the relationships among certain variables in Los Angeles County during the given period, however, it might not be a good option to represent the overall time series relationship.

**Cross-sectional Data Analysis**

**Figure 7: Regression Analysis on Life Expectancy, Crime Rates, and Diabetes Rates 2010, United States.**

| Regression Statistics | |
|---|---|
| Multiple R | 0.772600888 |
| R Square | 0.596912132 |
| Adjusted R Square | 0.596568494 |
| Standard Error | 1.437746797 |
| Observations | 2349 |

| | Coefficients | Standard Error | t Stat | P-value |
|---|---|---|---|---|
| Intercept | 85.28257575 | 0.138804248 | 614.409 | 0 |
| Crime Index | 0.001338375 | 0.008330707 | 0.160656 | 0.872378 |
| Diabetes Rate | -0.717164092 | 0.012167785 | -58.9396 | 0 |

$$\text{Life expectancy} = 85.2826 + 0.0013 * \text{Crime Rates} - 0.7172 * \text{Diabetes Prevalence} + \text{error}$$

The cross-sectional data analysis was first built using cross-sectional data, including crime rates, diabetes prevalence, and life expectancy. Compared to the time series equations generated, this model had a lower, but still relatively strong adjusted R square, which indicates that nearly 60% of the variation of the data could be explained by the model. The standard error, though increased, was low as well. Overall, this model had a relatively good fit to the dataset. Diabetes prevalence had a negative relationship with life expectancy while the crime rate had a positive coefficient, indicating that each increase in crime rate leads to an increase in life expectancy. The p-value of crime rate was high as well. With a high p-value and an abnormal coefficient, the model needed to be improved by adding the interaction variable.

## Discussion

A few points needed to be noticed to the conclusions: first of all, correlation does not mean causality. The regression models focused only on non-directional relationships among variables as the results illustrate the fact that what independent variables have connections with the dependent variable and how the dependent variable would change when the independent variables changes. For example, the coefficient of the crime rate in the cross-sectional data model is around -0.1928, suggesting that one unit increase in crime rates leads to a decrease in life expectancy by 0.1928. However, it is impossible to tell if the increase in crime rates will definitely lead to the decrease of life expectancy as there might be more interrelated variables involved within. This research was specifically designed to discover the correlations among variables, but the causality is absolutely another interesting topic to focus on. Second, the equation might be different when the definition of safety level changes. Actually, the reported crime rate is only part of the story and residence perception on neighborhood safety level could be largely different from the authority's statistic. For example, according to the National Crime Victimization Survey 1996-2013, as the official neighborhood crime rate decreases, the perceptional safety level decreases as well. It is interesting to learn that individuals feel less safe than the government believes they should do and that how the equation would change when using safety perception instead of crime rates. Third, the equation might not be the same when using different geographic scales. The time series data analysis looked only at data from Los Angeles County from 2000 to 2012 and the equation generated might be different when applying data on a larger scale, such as on a state or a national level. In such case, the effect of crime rates on life expectancy might be smaller due to overall lower crime rates. And this rationale applies the cross-sectional model as well when looking at data on a smaller geographic scale, such as on a state or a county level. At last, the equation also needed to be cautiously used when looking at rural versus urban data. In the case of crime rates, urban areas might have relatively higher crime rates with larger standard error than the rural ones, and such fact leads to a higher effect from the crime rates on life expectancy in urban than in rural areas.

## Limitation and Next Step

Even with a reasonable result, the models created from the research above were not flawless. First of all, the time series data model was built using data with a small sample size, which led to not only a high R square but also a high sensitivity to the chosen sample. To be more specific, this model might have a bigger error when applying to larger data. A good way to solve this problem is to include larger sample into the analysis. By including more data, the time series effect is weakened and the standard error could be minimalized, therefore the resulted model would be less biased and applicable to more kinds of situation. As to the cross – sectional data analysis, another improvement could be done is to increase the complexity of the model by including more potential related variables into the analysis. A good example is the walk score. This index was designed to quantify how friendly a region is to walking pedestrians. It is calculated by analyzing hundreds of walking routes to nearby amenities, population density, and intersection density; the region with a walk score closer to 100 is more suitable to walk. Are individuals more likely to do outside activities in neighborhoods with high walk scores? Is a higher walk score correlated to a lower diabetes prevalence? These are all interesting questions that could be answered in further research. At last, the cross – sectional data model was created specifically to summarize the data in the United States in 2010 for Type II diabetes. Does the same model apply to a different year? What about in a different disease? Does it apply to all around the world? There are so many topics that worth taking a look at.

# Appendix

## Appendix A: Compiled Time Series Data in Los Angeles County, 2000 – 2012

| Year | Time | LA County Diabetes Prevalence (%) | Crime Incidences (in 10M) | Diabetes*Crime | Life expectancy |
|---|---|---|---|---|---|
| 2000 | 0 | 4.23 | 9.50 | 40.19 | 78.70 |
| 2001 | 1 | 4.3 | 9.36 | 40.25 | 78.90 |
| 2002 | 2 | 4.36 | 9.20 | 40.11 | 79.00 |
| 2003 | 3 | 4.41 | 8.68 | 38.27 | 79.30 |
| 2004 | 4 | 4.46 | 7.82 | 34.86 | 79.60 |
| 2005 | 5 | 4.56 | 6.76 | 30.82 | 79.80 |
| 2006 | 6 | 4.59 | 6.64 | 30.47 | 80.10 |
| 2007 | 7 | 4.65 | 6.48 | 30.14 | 80.50 |
| 2008 | 8 | 4.66 | 6.11 | 28.47 | 81.00 |
| 2009 | 9 | 4.68 | 5.59 | 26.14 | 81.30 |
| 2010 | 10 | 4.74 | 5.11 | 24.24 | 81.70 |
| 2011 | 11 | 4.76 | 4.68 | 22.29 | 82.00 |
| 2012 | 12 | 4.8 | 4.50 | 21.59 | 82.10 |

## Appendix B: Compiled Cross – sectional Data in 2010

| ID | State | County | Violent crime | Population | Population/100 | Crime Rate per Capita | Crime Index | Diabetes Rate | Crime*Diabetes | Life Expectancy |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Alabama | Autauga | 57 | 54571 | 545.71 | 0.001044511 | 0.104 | 11.8 | 1.233 | 75.74 |
| 2 | Alabama | Bibb | 19 | 22915 | 229.15 | 0.000829151 | 0.083 | 11.1 | 0.920 | 74.13 |
| 3 | Alabama | Calhoun | 38 | 118572 | 1185.72 | 0.00032048 | 0.032 | 13.9 | 0.445 | 74.11 |
| 4 | Alabama | Etowah | 62 | 104430 | 1044.3 | 0.000593699 | 0.059 | 14.1 | 0.837 | 73.86 |
| 5 | Alabama | Geneva | 51 | 26790 | 267.9 | 0.001903695 | 0.190 | 13.6 | 2.589 | 75.65 |
| 6 | Alabama | Hale | 21 | 15760 | 157.6 | 0.001332487 | 0.133 | 16.3 | 2.172 | 73.93 |
| 7 | Alabama | Henry | 31 | 17302 | 173.02 | 0.0017917 | 0.179 | 14.3 | 2.562 | 75.82 |
| 8 | Alabama | Houston | 66 | 101547 | 1015.47 | 0.000649945 | 0.065 | 13 | 0.845 | 76.7 |
| 9 | Alabama | Jefferson | 349 | 658466 | 6584.66 | 0.00053002 | 0.053 | 11.7 | 0.620 | 75.11 |
| 10 | Alabama | Limestone | 37 | 82782 | 827.82 | 0.000446957 | 0.045 | 10.7 | 0.478 | 77.15 |
| 11 | Alabama | Madison | 308 | 334811 | 3348.11 | 0.000919922 | 0.092 | 11.7 | 1.076 | 77.79 |
| 12 | Alabama | Mobile | 250 | 412992 | 4129.92 | 0.000605339 | 0.061 | 12.1 | 0.732 | 74.95 |
| 13 | Alabama | Morgan | 40 | 119490 | 1194.9 | 0.000334756 | 0.033 | 11.9 | 0.398 | 75.92 |
| 14 | Alabama | Russell | 110 | 52947 | 529.47 | 0.002077549 | 0.208 | 13.2 | 2.742 | 73.87 |
| 15 | Alabama | Shelby | 222 | 195085 | 1950.85 | 0.001137966 | 0.114 | 8.9 | 1.013 | 79.18 |
| 16 | Alabama | St. Clair | 13 | 83593 | 835.93 | 0.000155515 | 0.016 | 13.5 | 0.210 | 75.35 |
| 17 | Alabama | Tuscaloosa | 190 | 194656 | 1946.56 | 0.000976081 | 0.098 | 10.5 | 1.025 | 75.86 |
| 18 | Alabama | Choctaw | 4 | 13859 | 138.59 | 0.000288621 | 0.029 | 17.3 | 0.499 | 75.19 |
| 19 | Alabama | Clarke | 63 | 25833 | 258.33 | 0.002438741 | 0.244 | 15.7 | 3.829 | 75.29 |
| 20 | Alabama | Clay | 4 | 13932 | 139.32 | 0.000287109 | 0.029 | 14.7 | 0.422 | 75.25 |
| 21 | Alabama | Cleburne | 24 | 14972 | 149.72 | 0.001602992 | 0.160 | 13.3 | 2.132 | 74.53 |
| 22 | Alabama | Coffee | 21 | 49948 | 499.48 | 0.000420437 | 0.042 | 12.6 | 0.530 | 76.5 |
| 23 | Alabama | Conecuh | 32 | 13228 | 132.28 | 0.002419111 | 0.242 | 17.7 | 4.282 | 74.08 |

## Reference list

Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2017. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services; 2017.

Hu FB, Sigal RJ, Rich-Edwards JW, et al. Walking Compared With Vigorous Physical Activity and Risk of Type 2 Diabetes in WomenA Prospective Study. *JAMA*. 1999;282(15):1433–1439. doi:10.1001/jama.282.15.1433.

Hu FB, Li TY, Colditz GA, Willett WC, Manson JE. Television Watching and Other Sedentary Behaviors in Relation to Risk of Obesity and Type 2 Diabetes Mellitus in Women. *JAMA*. 2003;289(14):1785–1791. doi:10.1001/jama.289.14.1785.

Beth E. Molnar, ScD, Steven L. Gortmaker, PhD, Fiona C. Bull, PhD, and Stephen L. Buka, ScD. Unsafe to Play? Neighborhood Disorder and Lack of Safety Predict Reduced Physical Activity among Urban Children and Adolescents. *American Journal of Health Promotion*. Vol 18, Issue 5, pp. 378 – 386. First Published on May 1, 2004. https://doi.org/10.4278/0890-1171-18.5.378

Burdette, H. L., & Whitaker, R. C. (2005). A national study of neighborhood safety, outdoor play, television viewing, and obesity in preschool children. Pediatrics, 116(3), 657-662.

Jonker, J. T., De Laet, C., Franco, O. H., Peeters, A., Mackenbach, J., & Nusselder, W. J. (2006). Physical activity and life expectancy with and without diabetes: life table analysis of the Framingham Heart Study. Diabetes Care, 29(1), 38-43.

Chetty R, Stepner M, Abraham S, et al. The Association Between Income and Life Expectancy in the United States, 2001-2014. JAMA. 2016;315(16):1750–1766.

doi:10.1001/jama.2016.4226