Analysis of Life Expectancy by County

Megan Y. Tan

Stony Brook University

Abstract

Life expectancy is affected by various health and economic factors. This study examines the relationship between life expectancy and smoking prevalence, alcohol use prevalence, obesity prevalence, physical activity prevalence, diabetes prevalence, and Metropolitan Statistical Area (MSA) status on the county level. The data was collected from the Global Health Data Exchanged and a regression was conducted to analyze the relationship between life expectancy and the health metrics. A Oaxaca-Blinder dissociation was also done to decompose the difference in life expectancy between urban and rural counties.

The regression results showed that MSAs had higher life expectancies by 5.17 years for males and 4.67 years for females. Overall, life expectancy is positively correlated with alcohol use prevalence and physical activity prevalence, but negatively correlated with diabetes prevalence and smoking prevalence. The relationship between obesity prevalence and life expectancy is positive for males but negative for females. Contrary to the overall results, the relationship between alcohol use prevalence and life expectancy and the relationship between physical activity prevalence and life expectancy is negative in urban counties.

The Oaxaca-Blinder Dissociation show that much of the difference in the life expectancy between urban and rural counties are not attributed to the difference in the explanatory variables between urban and rural counties, but rather from the differences in the outcomes of those explanatory variables or from other factors not attributed for in this study.

Introduction

Life expectancy is affected by the behavioral health factors of a population. Past studies have suggested that obesity and smoking lead to decreased life expectancy in both males and females. (Peeters, et al., 2003) Besides that, any level of physical activity have also been found to increase life expectancy. (Ferruci, et al., 1999) Previous studies have also shown that life expectancies vary by county, and that life expectancy is inversely related to levels of rurality. (Singh & Siahpush, 2014).

The purpose of this study is to examine the relationship between life expectancy and the prevalence of diabetes, smoking, alcohol use, obesity, and physical activity on the county level and decompose the differences in life expectancies between urban counties, which are counties that are in Metropolitan Statistical Areas (MSAs) and rural counties, which are counties that are not in Metropolitan Statistical Areas (NMSAs). The goal of this study is to determine if there is a significant difference in life expectancy between urban and rural counties and to examine the causes of the difference in life expectancy.

Methods

The county-level adult life expectancy and behavioral health factors data by FIPS code and year was collected from the Global Health Data Exchange. The data combined to form one data set. A subset of years, 1999-2007, was extracted to minimize empty records.

The data was further sorted by the Federal Information Processing Standards (FIPS) codes into urban and rural based on the FIPS code that are in MSAs as provided by the United States Census Bureau. In total, there are 56772 lines of observation, 28386 for each male and female, for 3154 unique FIPS codes, 855 for urban counties and 2299 for rural counties, over the eight years. Table 1 shows the description of the variables used. All variables except alcohol use had data for both males and female, so for alcohol use, the same data is used for both males and females.

Table 1: Description of variables in the data set

|  |  |
| --- | --- |
| **Variable** | **Description** |
| FIPS | The county FIPS code of the observation. |
| Year | The year of the observation. |
| MSA | =1 if the FIPS code is in an MSA. |
| Female | =1 if the observation is for female. |
| Life Expectancy | The average number of years an adult is expected to live. |
| Alcohol Use | The percentage of adults aged 21 and older who consumed at least four drinks if they are women or five drinks if they are men on a single occasion at least once in the past 30 days. |
| Diabetes | The percentage of adults aged 20 and older who reported a previous diabetes diagnosis and/or have high FPG/AIC\*. |
| Obesity | The percentage of adults aged 20 and older who have a BMI of 30 or higher. |
| Activity | The percentage of adults aged 20 and older who reported any leisure-time physical activity in the past 30 days. |
| Smoking | The percentage of adults aged 20 and older who smoke cigarettes daily. |

To observe the relationship between life expectancy and the explanatory variables, a linear model in the form

was estimated on the pooled data without accounting for time and location using an ordinary least squares regression. The Eicker-Huber-White standard errors were used in the regression to account for heteroscedasticity in the data. (White, Halbert, 1980) To observe the effect being in an MSA had on life expectancy, an unrestricted model was tested with MSA and its interaction with other variables as explanatory variables.

Then, the data was analyzed in two groups, MSAs and NMSAs. A Welch Two Sample t-test was run on each of the variables between MSAs and NMSAs. A twofold Oaxaca-Blinder Dissociation was conducted on a randomly selected subset of 5064 urban counties and 13620 rural counties for a total of 18684 counties to decompose the gap in the mean life expectancies between MSAs and NMSAs into explained differences accounted for by the differences in magnitudes of each group’s explanatory variables, and unexplained differences accounted for by the differences in the effects of those explanatory variables, or by factors not in the data. The unexplained differences were further decomposed into the unexplained differences in favor of MSA and the unexplained differences against NMSAs. The group weights chosen for the analysis is -1, which shows results for the coefficients from a pooled regression that does not include the group indicator variable. (Neumark, 1988)

Finally, a fixed-effects and random effects regression was run on the unbalanced panel data on the model

.

To choose the right model, a Breusch-Pagan test (Breusch and Pagan, 1979) was ran on the model to test if there are county-specific effects. Then, the hypothesis of random effects is tested using the standard Hausman test (Hausman, 1978) to test if the life expectancy and country effects are related.

Results

The means of each variable for both urban and rural counties are reported in table 2. There is a significant difference (p-value < 0.001) between the means in urban counties and rural counties for all variables. Urban counties reported higher mean life expectancies by 1.0282 years for males and 0.4247 years for females. Urban counties also reported higher means of alcohol use by 0.3777 percentage points, and physical activity by 4.5299 percentage points for males and 2.7615 percentage points for females. Rural counties reported higher means of diabetes by 0.2943 percentage points for males and 0.513 percentage points for females, obesity by 2.2087 percentage points for males and 2.8608 percentage points for females, and smoking by 1.7399 percentage points for males and 1.7596 percentage points for females.

From the results of the pooled regressions, which is shown in table 3, urban counties showed longer life expectancy compared to rural counties by 4.1604 years for males, and 3.6821 years for females when not accounting for other factors. Like previous literature has suggested, diabetes and smoking are both associated with decreased life expectancy both overall and in urban counties, however the effect of diabetes in urban counties on life expectancy is less severe. Alcohol use and activity shows increased life expectancy overall, but decreased life expectancy in urban counties when compared to rural counties. Strangely, obesity shows increased life expectancy in males but decreased life expectancy in females, and this effect is also seen in urban counties when compared to rural counties.

Table 2: T Test Results for Comparisons between MSAs and non MSAs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Gender | MSAs Mean | NMSAs Mean | *t*-value |
| Life Expectancy Males | Male  Female | 74.4318  79.7312 | 73.4036  79.3065 | 36.551\*\*\*  7.8871\*\*\* |
| Alcohol Use |  | 16.8038 | 16.4261 | 7.8871\*\*\* |
| Diabetes | Male  Female | 13.0647  10.4380 | 13.3590  10.9510 | -11.597\*\*\*  -16.702\*\*\* |
| Obesity | Male  Female | 30.0444  32.0442 | 32.2531  34.9050 | -37.462\*\*\*  -36.556\*\*\* |
| Physical Activity | Male  Female | 58.2406  48.3865 | 53.7107  45.6250 | 38.213\*\*\*  18.91\*\*\* |
| Smoking | Male Female | 21.1643 18.1207 | 22.9042 19.8793 | -32.105\*\*\* -33.431\*\*\* |
| Significant codes | 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ | | | |

The twofold Oaxaca-Blinder Dissociation shows that the difference in the mean life expectancy between urban and rural counties is 1.0395 years for males and 0.4791 years for females with urban counties having the higher life expectancies. These differences are decomposed into 0.9105 years of explained difference and 0.1290 years of unexplained difference, of which 0.0940 years is in favor of MSA and 0.0350 years is against NMSA, for males, and 0.5079 years of explained difference and -0.02880 years of unexplained difference, of which -0.0210 is in favor of MSA and -0.0078 is against NMSA, for females.

Table 3: Coefficients from the regressions on all data, MSA data, and NMSA data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model not accounting for MSA | | Model accounting for MSA | |
| Coefficient | Males | Females | Males | Females |
| Intercept | 77.9996\*\*\* (0.3278) | 81.7385\*\*\* (0.2501) | 77.0506 \*\*\* (0.3528) | 80.5075 \*\*\* (0.2769) |
| MSA |  |  | 4.1604 \*\*\* (0.7142) | 3.6821 \*\*\* (0.5918) |
| Diabetes | -0.4928\*\*\* (0.0143) | -0.2252\*\*\* (0.0130) | -0.5124\*\*\* (0.0111) | -0.2324 \*\*\* (0.0148) |
| Smoking | -0.1764\*\*\* (0.0051) | -0.0763\*\*\* (0.0036) | -0.1420\*\*\* (0.0047) | -0.0532\*\*\* (0.0040) |
| Alcohol Use | 0.0735\*\*\* (0.0034) | 0.0894\*\*\* (0.0032) | 0.0881\*\*\* (0.0036) | 0.0954\*\*\* (0.0036) |
| Obesity | 0.0414\*\*\* (0.0059) | -0.0408\*\*\* (0.0050) | 0.0521\*\*\* (0.0059) | -0.0287\*\*\* (0.0058) |
| Activity | 0.0743\*\*\* (0.0033) | 0.0319 \*\*\* (0.0025) | 0.0703\*\*\* (0.0035) | 0.0383\*\*\* (0.0027) |
| MSA \* Diabetes |  |  | -0.0263  (0.0239) | -0.0159 (0.0292) |
| MSA \* Smoking |  |  | -0.1178\*\*\* (0.0090) | -0.0780\*\*\* (0.0091) |
| MSA \* Alcohol Use |  |  | -0.0351\*\*\* (0.0080) | -0.0202\*\* (0.0075) |
| MSA \* Obesity |  |  | 0.0244 . (0.0104) | -0.0181 . (0.0108) |
| MSA \* Activity |  |  | -0.0199\* (0.0077) | -0.0204\*\* (0.0075) |
| Adjusted R-Squared | 0.733 | 0.695 | 0.7427 | 0.7014 |
| Significant codes | 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ | | | |

The coefficients of the Oaxaca-Blinder Decompositions, shown in figure 1, shows that the higher prevalence of alcohol use and physical activity, associated with increased life expectancy, in urban counties and the higher prevalence of diabetes, smoking, and obesity, associated with decreased life expectancy other than obesity prevalence in males which is associated with increased life expectancy, in rural counties explains only a small degree of the difference in the mean life expectancy between urban and rural counties.

By contrast, a large amount of the difference in life expectancy between urban and rural counties stem from the different effects that the explanatory variables have on those places, or from other factors not in the data. Diabetes decreases life expectancy overall, but the decrease is more severe in urban counties compared to the overall decrease, and higher diabetes prevalence is associated with increased life expectancy in rural counties. Smoking prevalence and alcohol use prevalence decreases life expectancy overall and in both urban and rural counties, but the decrease in life expectancy in urban counties is more pronounced than the decrease in life expectancy in rural counties. Obesity prevalence is positively correlated with life expectancy overall and in urban counties, but it is negatively correlated with life expectancy in rural counties for males; obesity prevalence is negatively correlated with life expectancy overall and in both urban and rural county for females, but the decreased life expectancy is more pronounced in rural counties. Increase physical activity prevalence showed decreased life expectancies overall and in urban counties. However, in rural counties, it showed increased life expectancy for males but decreased life expectancy for females. For females, the decreased life expectancy is more pronounce in urban counties.

**Figure 1**. The coefficients of the explained and unexplained differences of each explanatory variable.

The χ2 from the Breuch-Pagan test (Breusch & Pagan, 1979) is 1542.3 with a p-value of 0.00. Thus, the null hypothesis that there are no individual specific effects may be rejected. The χ2 statistic for the Durbin-Wu-Hausman test (Wu, 1973) is 34.9 with a p-value of 0.00. Thus, the hypothesis that life expectancy and county effects are uncorrelated is rejected, so the fixed effect model is used. The fixed effects regression results, shown in table 4, shows similar relationships between the explanatory variables and life expectancy as the ordinary least square regression.

Table 4: Coefficients from the Fixed Effects Regression

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model not accounting for MSA | | Model accounting for MSA | |
| Coefficient | Male | Female | Male | Female |
| MSA |  |  | 5.1710\*\*\* (0.7211) | 4.6693\*\*\* (0.4780) |
| Diabetes | -0.4363\*\*\* (0.0072) | -0.1300\*\*\* (0.0058) | -0.4310\*\*\* (0.0081) | -0.1399\*\*\* (0.0066) |
| Smoking | -0.1804\*\*\* (0.0027) | -0.0537\*\*\* (0.0019) | -0.1478\*\*\* (0.0032) | -0.0306\*\*\* (0.0022) |
| Alcohol Use | 0.0798\*\*\* (0.0022) | 0.0740\*\*\* (0.0017) | 0.0932\*\*\* (0.0025) | 0.0804\*\*\* (0.0019) |
| Obesity | 0.0627\*\*\* (0.0034) | -0.0356\*\*\* (0.0026) | 0.0587\*\*\* (0.0041) | -0.0207\*\*\* (0.0031) |
| Physical Activity | 0.1130\*\*\* (0.0026) | 0.0901\*\*\* (0.0020) | 0.1134\*\*\* (0.0029) | 0.0957\*\*\* (0.0031) |
| MSA \* Diabetes |  |  | -0.0601\*\*\* (0.0172) | -0.0286\* (0.0141) |
| MSA \* Smoking |  |  | -0.1097\*\*\* (0.0062) | -0.0767\*\*\* (0.0044) |
| MSA \* Alcohol Use |  |  | -0.0398\*\*\* (0.0054) | -0.0187\*\*\* (0.0041) |
| MSA \* Obesity |  |  | 0.0312\*\*\* (0.0074) | -0.0216\*\*\* (0.0055) |
| MSA \* Activity |  |  | -0.0258\*\*\* (0.0068) | -0.0260\*\*\*  (0.0048) |
| Adjusted R^2:  F- statistic: | 0.7335 10287.6\*\*\* | 0.7116  9219.94\*\*\* | 0.7394 4321.19\*\*\* | 0.7179 4324.59\*\*\* |
| Significant codes | 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ | | | |

Discussion

The results from the Hausman test suggests that the fixed-effects regression should be used instead of the pooled regression. The fixed-effects regression shows that for males, a one percentage point increase in diabetes prevalence and smoking prevalence is associated with a 0.43years and 0.18years decrease in life expectancy respectively in a county, and a one percentage points increase in alcohol use prevalence, obesity prevalence, and physical activity prevalence is associated with a 0.08years, 0.06years, and 0.11years increase in life expectancy respectively in a county. For females, a one percentage point increase in diabetes prevalence, smoking prevalence, and obesity prevalence is associated with a 0.13years, 0.05years, and 0.04years decrease in life expectancy respectively in a county, and a one percentage point increase in alcohol use prevalence and physical activity prevalence is associated with a 0.07years, and 0.09years increase in life expectancy respectively in the county.

In urban counties, life expectancy is longer by 5.17 years for males and 4.67 years for females compared to rural counties. Compared to rural counties with no prevalence, a one percentage point increase in diabetes prevalence results in a 0.06years decrease for males and 0.02year decrease for females in life expectancy, a one percentage point increase in smoking prevalence results in a 0.12year decrease for males and a 0.08year decrease for females in life expectancy, a one percentage point increase in alcohol use prevalence is associated with a 0.04year decrease for males and 0.02year decrease for females in life expectancy, a one percentage point increase in obesity is increased with a 0.03year increase and a 0.02year decrease in life expectancy respectively, and a one percentage point increase in physical activity prevalence is associated with a 0.3years decrease in life expectancy for both males and females.

In urban counties, the relationship between alcohol use prevalence and life expectancy, and the relationship between physical activity prevalence and life expectancy are negatively correlated, whereas those relationships are positively correlated overall. This effect may be because there is an omitted variable which is correlated with alcohol use prevalence and physical activity prevalence only in urban counties that is not accounted for in this study.

The relationships between diabetes, physical activity, and smoking for both sexes, and obesity for females, and life expectancy are consistent with previous literature, but the results for the relationship between obesity in males and life expectancy suggests otherwise. (Moore, et al., 2012) This could be because the model used has some omitted variables that are not time-invariant that are correlated with the explanatory variables obesity and physical activity, and thus those omitted variables could not be accounted for in the fixed effects regression.

The Oaxaca-Blinder dissociation showed that a large portion of the difference in life expectancy between urban and rural counties is not caused by the difference in the behaviors between people in urban counties and people in rural counties, but rather by the differences in the effects of those behaviors. This also suggests that our model has some omitted variables which effects life expectancy.

Conclusion

The regression shows that life expectancy, on a county level, is positively correlated with alcohol use prevalence and physical activity prevalence, but negatively correlated with diabetes prevalence and smoking prevalence. The relationship between obesity prevalence and life expectancy is positive for males but negative for females. Contrary to the overall results, the relationship between alcohol use prevalence and life expectancy and the relationship between physical activity prevalence and life expectancy is negative in urban counties.

The Oaxaca-Blinder Dissociation show that much of the difference in the life expectancy between urban and rural counties are from factors not accounted for in this study. Further study on the factors of life expectancy should include more explanatory variables to achieve a deeper understanding of life expectancy.

References

Achim Zeileis, Torsten Hothorn (2002). Diagnostic Checking in Regression Relationships. R News 2(3), 7-10. URL https://CRAN.R-project.org/doc/Rnews/

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica,47(5), 1287. doi:10.2307/1911963

Croissant Y, Millo G (2008). “Panel Data Econometrics in R: The plm Package.” Journal of Statistical Software, 27(2), 1–43. doi: 10.18637/jss.v027.i02

Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, et al. Inequalities in Life Expectancy Among US Counties, 1980 to 2014: Temporal Trends and Key Drivers. JAMA Intern Med. 2017;177(7):1003–1011. doi:10.1001/jamainternmed.2017.0918

Hlavac, Marek (2018). oaxaca: Blinder-Oaxaca Decomposition in R. R package version 0.1.4. https://CRAN.R-project.org/package=oaxaca

Institute for Health Metrics and Evaluation (IHME). Diagnosed and Undiagnosed Diabetes Prevalence by County in the U.S., 1999-2012. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2016.

Institute for Health Metrics and Evaluation (IHME). United States Adult Life Expectancy by State and County 1987-2009. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2012.

Institute for Health Metrics and Evaluation (IHME). United States Alcohol Use Prevalence by County 2002-2012. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2015.

Institute for Health Metrics and Evaluation (IHME). United States Physical Activity and Obesity Prevalence by County 2001-2011. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2013.

Institute for Health Metrics and Evaluation (IHME). United States Smoking Prevalence by County 1996-2012. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2014.

Luigi Ferrucci, Grant Izmirlian, Suzanne Leveille, Caroline L. Phillips, Maria-Chiara Corti, Dwight B. Brock, Jack M. Guralnik, Smoking, Physical Activity, and Active Life Expectancy, American Journal of Epidemiology, Volume 149, Issue 7, 1 April 1999, Pages 645–653, https://doi.org/10.1093/oxfordjournals.aje.a009865

Moore, S. C., Patel, A. V., Matthews, C. E., Gonzalez, A. B., Park, Y., Katki, H. A., . . . Lee, I. (2012). Leisure Time Physical Activity of Moderate to Vigorous Intensity and Mortality: A Large Pooled Cohort Analysis. PLoS Medicine,9(11). doi:10.1371/journal.pmed.1001335

Peeters, A, Barendregt, J.J.M, Willekens, F, Mackenbach, J.P, Al Mamun, A, & Bonneux, L.G.A. (2003). Obesity in adulthood and its consequences for life expectancy: a life-table analysis. Annals of Internal Medicine. Retrieved from http://hdl.handle.net/1765/10043

Singh, G. K., & Siahpush, M. (2014). Information for CME Credit — Widening Rural–Urban Disparities in Life Expectancy, U.S., 1969–2009. American Journal of Preventive Medicine,46(2). doi:10.1016/s0749-3797(13)00672-7

Warner, K. E. (1987). Health and Economic Implications of a Tobacco-Free Society. JAMA: The Journal of the American Medical Association,258(15), 2080. doi:10.1001/jama.1987.03400150072032

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica,48(4), 817. doi:10.2307/1912934

Wooldridge, J. M. (2020). Introductory econometrics: A modern approach. Boston, MA: Cengage.

Wu, D. (1974). Alternative Tests of Independence between Stochastic Regressors and Disturbances: Finite Sample Results. Econometrica,42(3), 529. doi:10.2307/1911789

Zeileis A (2004). “Econometric Computing with HC and HAC Covariance Matrix Estimators.” Journal of Statistical Software, 11(10), 1–17. doi: 10.18637/jss.v011.i10