Proceedings of the Jepson Undergraduate Conference on International Economics

Volume 5 Article 1

7-2023

An Analysis of Socio-Economic, Lifestyle, and Health Care Determinants of National Life Expectancies

Annie Feldman
University of Northern Iowa

An Analysis of Socio-Economic, Lifestyle, and Health Care Determinants of National Life Expectancies

Annie Feldman

Abstract: Life expectancy is an important indicator of a population's health and anticipated life-quality. Life expectancies have continued to rise in recent history, but there is great variance between countries. This paper analyzes potential reasons for country variation in the average human lifespan. Regression analysis is performed on data from 156 countries to infer explanations of observed variance. In this model, I chose factors from past research that have been highly significant. When selecting variables, I considered socio-economic, lifestyle, and health care financing measures. To explain national life expectancies, I tested variables including education, income, urban-to-rural living, alcohol and tobacco consumption, and health care expenditures. My results found education, income, and urban factors were all statistically significant variables within the model. This is consistent with previous research on these factors. However, my model presented contradictions from previous literature on the significance of alcohol and tobacco consumption and health care expenditure factors. Potential reasons for these inconsistencies are briefly discussed.

Acknowledgements: I would like to thank Dr. Imam Alam for his assistance as my professor of Econometrics and faculty advisor. Additionally, I would like to thank Dr. Shahina Amin for organizing the Jepson Undergraduate Research Conference along with challenging me to submit my paper and helping me throughout the process.

1. Introduction

Life expectancy can provide insight into a population's anticipated quality of life. Additionally, life expectancy is an important metric of a population's health and economy to be used for planning and policy. Liou et al. explain that, conceptually, "life expectancy describes the mean additional years of life remaining at a specific age and captures the prevailing patterns of mortality for all the age groups during the period" (2020, p. 1). Throughout human history life expectancies have continued to rise, with experts arguing between values of 80, 90, or 120 years as an inherent upper limit (Lakshmanarao et al., 2022). Yet, even with an upper limit, we have continued to see increasing life expectancies and have therefore not reached this limit. In 1955 the world life expectancy at birth was 46.6 years, but by 2010 the number had risen 21 years to 67.6 years (Novak et al., 2016). This increase in life expectancy around the world is indicative of the improvements seen in global standards of living. Life expectancy reflects social, economic, medical, and technical advancements made over time; therefore, it is a useful health metric (Liou et al., 2020).

However, comparison between countries reveals that there is great variation in the average life expectancy of a nation's population. In 2009, less developed countries had life expectancies 11 years lower than more developed countries (Novak et al., 2016). Thus, Shaw et al. propose an important question, "If societal health can be measured as life expectancy or mortality rates, what are the various socioeconomic factors that increase or decrease it?" (2005, p. 769). Researching lifespan variation between countries is necessary to help answer this question.

This paper will look at several life expectancy determinants among various countries around the world. Using regression analysis, I will test the significance of previously researched factors with updated data in my regression model. In this paper, I will detail previous research on the topic, the data and methodology used in my model, and discuss the results of my analysis.

2. Literature review

A large body of research exists that attempts to determine the factors that influence life expectancy. Within this research, a wide range of variables have been tested for their effect on life expectancy. Shaw et al. (2005) state that although empirical results are mixed, life expectancy is thought to be a function of environmental, lifestyle, and health care measures. Ranabhat et al. (2018) agree with these functions, separating the determinants of life expectancy into corresponding categories. Environmental, or socio-economic, measures include factors like education and economic status. Lifestyle measures include a population's life choices, for example alcohol or tobacco consumption. Finally, health care measures often include access to and consumption of health care goods and services. Previous studies selected various factors from these three measures in their analysis of life expectancy.

A 2005 study found that increasing education among young people positively influences their professional career and standard of living and therefore their life expectancy (Zhang & Zhang). Casper Hansen's study from 2012 found that higher education led to individuals living happier lives as they were better able to finance healthy living and recreational activities they enjoy. This in turn had a positive impact on their life expectancy. In 2016, Novak et al. performed regression analysis on life expectancies with cross-sectional data from 187 countries. Novak et al.'s model included multiple variables but was focused on the impact of education. Through their research, Novak et al. found education to be a highly significant determinant of a country's life expectancy at birth. In their model, an increase in one year of expected schooling resulted in life expectancy at birth rising by 1.087 years. Novak et al. argued that average years of schooling was the strongest determinant of national life expectancy. These studies from 2005, 2012, and 2016 all show that education is a key factor in accounting for variation in life expectancy.

Distribution of a country's population between rural and urban living has also been found to be significant. Singh & Siahpush (2014) found a negative connection between rural populations and life expectancy in a study of the United States. Novak et al. (2016) explain that this finding is due to rural populations having higher rates of heart diseases, injury, lung disease, lung cancer, strokes, suicide, and diabetes. Additionally, rural communities are more likely to be medically underserved and have lower education and income levels that urban communities. Given this research, I will include an urban population distribution factor in my model.

One study from 2006 details that economic activity, or the level of economic standard, is an important variable in determining life expectancy (Marmot). Countries with higher economic prosperity have higher life expectancies according to the same study. Another study from 2018 looked at the impact of wealth on a population's life expectancy. This study found that life expectancy at birth in low-income countries was far lower than that of high-income countries, at 64 and 79 respectively (Ranabhat et al.). Greater wealth has many benefits on an individual's life, including better health care, security, and finances to support a healthy lifestyle. Additionally, residents in countries with greater wealth levels tend to have a higher standard of living. These findings point to the importance of using a variable that measures a country's level of wealth in the regression analysis.

Health care expenditures include a country's spending on health care goods and services. In 2000, a study by Miller & Frech determined that pharmaceutical expenditures had a significant and positive effect of life expectancy. Shaw et al. (2005) used new values to update Miller & Frech's conclusions and again found that pharmaceutical drug consumption had a positive effect on a population's life expectancy.

Lifestyle factors also play a role in the lifespan of a country's population. While there are countless nonmedical lifestyle factors that impact life expectancy, there are specific indicators that are commonly evaluated in regression models. Shaw et al. (2005) found a significant and negative impact of alcohol on life expectancy of males. They also found tobacco consumption had a statistically significant negative effect on life expectancy. Ranabhat et al. (2018) further showed in their study that alcohol consumption negatively impacts life expectancy.

Past literature shows that social, lifestyle, and health care measures are all critical components of determining life expectancy. As a result, I will incorporate these dimensions into my model. I will also use this past research in selecting appropriate variables for each dimension. For my model I will measure environmental factors with education, income, and urban indicators. Alcohol and tobacco consumption data will be used to estimate the lifestyle dimension. Finally, health care expenditures will be used as a health care financing factor. Using regression analysis, I will test if these variables have a statistically significant impact on life expectancy.

3. Data and methodology

3.1 Data

In this section, I will detail the data used and specify sources from which I acquired data values. For my model of life expectancy, I collected cross-sectional data from 156 countries for the year 2018. Due to Covid-19 and its impact on health-related data, I chose the year 2018 to exclude this event from the model. Similar research performed with post-2019 data could look at the impact Covid-19 has had on life expectancy variation between countries.

The first source of data for this research is the United Nations Development Programme - Human Development Reports (2019). I retrieved *The Human Development Indices: A statistical update 2019* report from the UN Data statistical database. The report was a 2019 statistical update of data collected from the completed year prior, 2018. This dataset included values on life expectancy, educational access, and standard of living for 191 countries. Life expectancy is recorded as the years of expected lifespan at birth within a country. Educational data is measured as the population's average years of schooling. Standard of living data in the report is recorded as gross national income per capita.

Data for remaining factors in the model was collected from the World Bank databases (2018). Current health expenditures and urban population data were collected from the World Development Indicators database for 2018 values. Data on alcohol and tobacco consumption was collected from the Health Nutrition and Population Statistics database.

3.2 Variables

In choosing variables to include in my model, I based my selections on past literature and theory. The dependent variable in my model is human life expectancy at birth by country. Following previous research, I have chosen my independent variables based on socio-economic, lifestyle, and health care measures.

Socio-economic factors included in the model are education, income, and urban-to-rural population. Education is measured as the mean years of schooling expected of a country's residents. A country's wealth status is measured as Gross National Income per capita. The country's population composition is measured by including the percentage of the urban population in the model. Education, income, and urban living are important variables that will help capture the social and economic characteristics of a country. Lifestyle measures in my model include alcohol and tobacco consumption levels. The alcohol variable in this model is the total alcohol consumption per capita in a country. This value is measured as the liters of pure alcohol consumed per person (15 years of age or older) over a calendar year. Tobacco consumption is measured as the percentage of adults who currently use tobacco in the country. Alcohol and tobacco consumption variables in the model will address lifestyle influences on life expectancy. Finally, a country's health care financing is considered in the model. To account for this variable, I will use data on the current health expenditures per capita of the country. This variable estimates the current health expenditures on healthcare goods and services consumed within the country in the respective year.

3.3 Model

The dependent variable, life expectancy at birth, and independent variables for the model have now been identified. To determine the significance of my chosen independent variables I performed regression analysis. To do so, I used the following estimated regression model:

$$LEXP = \beta_0 + \beta_1 EDU + \beta_2 GNI + \beta_3 HEXP + \beta_4 ACONS + \beta_5 TCONS + \beta_6 URB + \varepsilon$$

Where:

LEXP = life expectancy at birth of a country's population (measured in years)
EDU = mean education of a country's population (measured in years)
GNI = per capita gross national income (measured in 2011 PPP $)
HEXP = nation's current health expenditures per capita (measured in 2018 PPP $)
ACONS = amount of pure alcohol consumed per capita (measured in liters)
TCONS = tobacco product consumption of the adult population (measured as percentage of the national population)
URB = urban population of the country (measured as percentage of the national population)
ε = a random error term

This theoretical model is used to estimate the population parameters of life expectancy at birth with data gathered from the United Nations Development Programme and the World Bank. Using regression analysis, I have tested the significance of my model's chosen independent variables on the dependent variable.

4. Results and discussion

Table 1. provides summary data on each variables mean, standard deviation, minimum and maximum value, and number of observations.

Table 1.
Summary Statistics of Variables

Variables	Mean	S.D.	Min.	Max.	Observations
LEXP	71.69	7.71	52.53	84.78	156
GNI	20,914.59	20,429.41	731.79	90,918.64	156
EDU	9.14	3.20	2.11	14.09	156
HEXP	1,636.26	1,947.21	36.98	10,515.32	156
ACONS	6.33	4.21	0.003	20.50	156
TCONS	20.49	9.67	3.60	45.10	156
URB	59.23	22.83	13.03	100.00	156

Table 2. details the regression results of this model. An R^2 value of 0.7237 was calculated in the regression which indicates the overall fit of the regression line to the data points. R^2 increases whenever a variable is added regardless of relevance. Due to this, I will also include the $\bar{R}^2$ value as it corrects for this issue. The $\bar{R}^2$ value in this model is 0.7126. This value indicates that approximately 71% of the variation in Y, life expectancy at birth, is explained by the independent variables of the model.

In my model, the coefficient of EDU is positive and explains that one additional average year of schooling would increase a country's life expectancy at birth by roughly 0.705 years. While this coefficient is lower than Novak et al.'s findings in 2016 of 1.087, it is still of the expected sign and the largest coefficient value in the model. The variable is also statistically significant at α levels of 0.10, 0.05, and 0.01. My results support past research findings that education is a critical determinant of life expectancy.

Table 2.
Regression Results

Variables	Coefficient	Robust Standard Error
GNI	0.00011***	0.00037
EDU	0.70471***	0.17661
HEXP	0.00079**	0.00038
ACONS	–0.13336	0.09672
TCONS	0.1357414***	0.03775
URB	0.0542234**	0.02200
F Statistic	65.04***	
R-squared	0.7237	
Adj R-squared	0.7126	

*** Significant at 0.01 level, ** significant at 0.05 level, and * significant at 0.10 level

The coefficient of the income variable in this model, GNI, is also of the expected sign. The model shows that if a country's income per capita increases, it will have a positive impact on life expectancy. This variable is also statistically significant at all α levels. Additionally, the coefficient URB was significant and of the expected sign. This model shows that increasing a country's urban population by 1% could lead to approximately 0.05-year increase in the life expectancy of its residents. The coefficient sign and statistical significance of these findings support past research that stresses the importance of socio-economic variables in explaining the varying human life spans of countries. All three socio-economic variables included in the model were significant and consistent with past research and theory.

Lifestyle factors in this model were not consistent with expected theory. The coefficient of ACONS had a negative impact on life expectancy at birth in the model. This indicates that increasing alcohol consumption negatively affects life expectancy in a country. While this is consistent with past research, the value was not statistically significant within this model. This indicates the ACONS variable was not a determinant of life expectancy in my findings. While TCONS was statistically significant (p-value of 0.000), the coefficient had a positive sign. This implies that an increase in the percentage of a country's population using tobacco products would also increase national life expectancy. This finding is inconsistent with past research and logic that the use of tobacco products negatively impacts health. Further research is necessary to determine the reasoning for this unexpected result between life expectancy and tobacco consumption.

The HEXP variable was not statistically significant at α level of 0.01. However, the variable positively impacted life expectancy at birth which is consistent with past research. Shaw et al. (2005) detail in their analysis of health care financing as a determinant of life expectancy that if age distribution is ignored there is potential for an omitted-variable bias in the elasticity of health care consumption. This is due to the increase in health care consumption as age increases. By ignoring the factor of age distribution, HEXP may be biased. This could explain why the results of the variable in my model are less significant than anticipated.

To test for correlated variables and Near Perfect Multicollinearity, I have run pairwise correlation functions and the Variance Inflation Factor. These results show mean VIF = 2.97 with no variables having values over 10. Additionally, all pairwise correlations are below 0.9. These results indicate the model is not suffering from severe Near Perfect Multicollinearity issues.

Finally, to assess the overall model, I utilized the F-test. This test assumes a null hypothesis where all coefficients are zero; in other words, the estimated model is worthless as no variables impact the dependent variable. If the null hypothesis is rejected, at least one coefficient in the model has a non-zero value and is a determinant of Y. The F-test value of my model was 65.04. With this value we would reject the null hypothesis. Therefore, to some extent, the variables of this model impact the value of life expectancy at birth.

5. Conclusions

Throughout this paper, I have sought to explain why there is variation in the average lifespan of a country's population using previous research as a starting point. While some findings were consistent with existing literature, there were some contradictions. In addition, some variables were not as significant as found to be through past results. A few potential issues with my model will be discussed in response to these findings.

This research affirms the importance of education, income, and urban living on the life expectancy of a country. All three variables were consistent with theory and results of past analyses. These variables are important explanatory factors of varying life expectancies around the world. Due to their significance, countries should consider these factors in their policy and planning. The results of my model indicate that socio-economic measures are the most significant in determining a population's life expectancy.

Lifestyle and health care measures in this model were not entirely consistent with prior research findings. Tobacco usage in my model had the opposite effect than would be expected according to theory. Further, alcohol consumption was not a significant factor. This may indicate that lifestyle dimensions do not explain the variation in national life expectancy as well as socio-economic dimensions do. Research to isolate and identify the reasons for these conflicting results it necessary to derive conclusions from my results. Additionally, the health expenditure variable was not as significant as anticipated. This is potentially due to a lack of an age distribution factor in this model. Previous work indicates that age distribution is an important component of accurately measuring health expenditures. If age is not considered, the effect of the variable may be understated. This is due to an omitted variable, in this case age, creating a bias on the coefficient.

Overall, there is risk that this model suffers from omitted variable bias. To keep a large sample size, potential variables were excluded as many countries did not have recent data available. Life expectancy has many factors and is therefore a nuanced and complicated health estimate. My model only accounted for roughly 71% of the variation observed in life expectancy by country. This indicates there are additional factors that I have not included in the model. This omitted variable bias means my model may have biased results.

This research was an attempt to explain variation in life expectancy across countries using socio-economic, lifestyle, and health care factors. Findings indicate that the socio-economic determinants chosen in the model (education, income, and urban living) are highly important to determining life expectancy at birth. Results also indicate that alcohol and tobacco consumption and health care are less significant. The conclusions drawn from this model highlight the need for continued analysis of this topic to better address variation in national life expectancies.

References

Hansen, C. W. (2012) 'The effect of life expectancy on schooling: some panel data evidence', [online] https://econ.au.dk/fileadmin/Economics_Business/Research/Seminars/Economic_Seminars_Series/2012/ES_12_Hansen_The_effect_of_life.pdf (Accessed 7 December 2022)

Lakshmanarao, A., Srisaila, A., Srinivasa, R. K., Lalitha, G. (2022) 'Life Expectancy Prediction through Analysis of Immunization and HDI Factors using Machine Learning Regression Algorithms', *International Journal of Online and Biomedical Engineering*, Vol. 18, No. 13, pp. 73-74. [online] https://www.researchgate.net/deref/https%3A%2F%2Fdoi.org%2F10.3991%2Fijoe.v18i13.33315 (Accessed 7 December 2022)

Liou, L., Joe, W., Kumar, A., Subramanian, S. V. (2020) 'Inequalities in life expectancy: An analysis of 201 countries, 1950-2015', *Social Science & Medicine*, Vol. 253, No. 112964, pp. 1-3.

Marmot, M. (2006) 'Health in an unequal world', *The Lancet*, Vol. 368, No. 9552, pp. 2081-2094.

Miller, R. Jr., Frech, H. E. (2000) 'Is there a link between pharmaceutical consumption and improved health in OECD countries?', *Pharmacoeconomics,* Vol 18, pp. 33-45.

Novak, A., Čepar, Ž., and Trunk, A. (2016) 'The role of expected years of schooling among life expectancy determinants', *Int J. Innovation and Learning*, Vol. 20, No. 1, pp. 85-99.

Ranabhat, C., Atkinson, J., Park, M. B., Kim, C. B., Jakovljevic, M. (2018) 'The Influence of Universal health Coverage on Life Expectancy at Birth (LEAB) and Healthy Life Expectancy (HALE): A Multi-Country Cross-Sectional Study', *Frontiers in Pharmacology,* Vol 9, Art. 960, pp. 1-3.

Shaw, J., Horrace, W., and Vogel, R. (2005) 'The Determinants of Life Expectancy: An Analysis of OECD Health Data', *Southern Economic Journal*, Vol. 71, No. 4, pp. 768-778

Singh, G. K. and Siahpush, M. (2014) 'Widening rural-urban disparities in life expectancy, US, 1969–2009', *American Journal of Preventive Medicine*, Vol. 46, No. 2, pp.19-29.

United Nations (2019) 'Human Development Report 2019', Complete [online] https://hdr.undp.org/system/files/documents//hdr2019pdf.pdf (Accessed 7 December 2022).

World Bank (2018) Data Catalog, Complete [online] https://datacatalog.worldbank.org/home (Accessed 7 December 2022)

Zhang, J. and Zhang, J. (2005) 'The effect of life expectancy on fertility, saving, schooling and economic growth: theory and evidence', *The Scandinavian Journal of Economics*, Vol. 107, No. 1, pp.45-66.