Dean&Francis

ISSN 2959-409X

MS003502

Analysing Factors Influencing Human Lifespan Based on Health and Lifestyle Metrics

Wenhao Zhou

School of Mathematical Science, University College Cork, T12 YN60, Cork, Ireland
120111264@umail.ucc.ie

Abstract:

As the global population ages, understanding the factors that influence human lifespan has become increasingly important. This study investigates how health indicators and lifestyle habits impact age. Health data such as blood pressure, cholesterol, BMI, and other lifestyle habits were analyzed. The primary target is to identify which of these key factors most strongly influence human lifespan. Thus, people can better predict their lifespan and give suggestions to avoid some mortality risks. This study takes advantage of both Logistic Regression (LR) and Linear Regression (LinR) analysis to examine the effects of health indicators and lifestyle habits on health over time. Results represent that factors such as blood pressure and bone density are confounders, which are not related to lifespan but impact the actual results of the research. Mental health, sun exposure, and education level are the key factors that significantly promote lifespan. Also, stress levels and pollution exposure are strongly related to the human lifespan as well. Some variables like alcohol consumption or income level are not strongly related to lifespan as public cognition. This paper emphasizes the importance of adopting specific healthy habits that can support longer and healthier lives.

Keywords: Lifespan, Health Factors, Lifestyle.

1. Introduction

Nowadays, with the improvement of healthcare and living conditions, the proportion of elderly people in the world continues to increase. Studies show that the aging process is not just a natural process; it is also influenced by lifestyle, economic income, and environment. Identifying which factors most impact longevity is important. This study will help people understand how lifestyle affects age. It aims to explore this topic through a review of past studies.

A study developed by Stephan A-J and his colleagues explores how the time a person is born, and their lifestyle choices affect their health rate. It shows that older people are more likely to develop chronic diseases [1]. Another study by Vega T found habits like less smoking, less drinking, and more exercise showed little improvement in health levels [2]. Øvrum found out that older adults who were physically active ate well, and didn't smoke had a lower risk of death, even if they had multiple chronic conditions [3]. Some studies also discovered that people with a higher genetic risk of shorter lifespans benefited the most from following a healthy lifestyle [4, 5].

Other social factors rather than health habits might affect longevity as well. People with higher levels of extraversion had a lower risk of death [6]. Social support was also crucial in determining both the quality and length of life [7].

All in all, past research shows that healthy habits can slow down serious illness and lead to a longer age. Regular exercise and a balanced diet can lower the risk of disease. Mental health and personality also play an important role in longevity. Moreover, factors like income and education level are linked to health as well. This paper uses LR and LinR to find the link between lifestyle choices and health. The results can help develop ways to live longer and make a difference for individuals and public health policies. More research on how these factors work together will be important for creating better health strategies and policies for humans.

2. Methodology

2.1 Data Sources and Description

The data used in this study mainly includes two parts: The age of a group of people and health-related habits data. This dataset is called as Human Age Prediction Synthetic Dataset (HAPSD). The data comes from the KAGGLE Database, including factors such as height, weight, BMI, Blood pressure, smoking status, and medication use.

2.2 Indicator Selection and Description

This paper combined all the factors from the dataset and exported the CSV. file as an Excel form as shown in Table 1. Some factors such as stress levels have been transformed into numeric thresholds. This data collection way is inspired by the Lifestyle Score method once mentioned in Jun Wang's paper [8].

Table 1. Data collection form

Gender	Height (cm)	Weight (kg)	Blood Pressure (s/d)	Cholesterol (mg/dL)	Stress Levels	Family History	Sleep Patterns	Age
Male	171	86	151/109	259	2.7	None	Insomnia	89
Female	172	79	134/112	263	9.3	Heart Disease	Normal	77

2.3 Method Introduction

This paper removed some outliers and missing data and analyzed the LinR between different variables as Table 1 shows. The correlations between variables were explained using LinR and LR. LinR is used to identify some confounding factors. LR is used to analyze factors influencing the human lifespan. Other models such as the Cox model and Random Forest model are also being used. Visualizations, like residual plots and scatter plots with colour coding, were important tools for analysis in this paper.

3. Results and Discussion

3.1 Linear Regression (LinR) Analysis

A LinR model among all the variables has been created in this part. The R-squared value of this LinR model is 0.9335, which shows a strong accuracy. By analyzing this model, factors that have a linear relationship with the output variable (Age) should be removed. In lifespan research, variables that are linearly related to age may simply change naturally as people get older and do not impact lifespan. Those variables are known as accompanying variables (or confounding factors). They are more like characteristics of aging rather than factors that influence lifespan.

ISSN 2959-409X

Table 2. R output of the LinR Model.

	Estimate	Std. Error	t value	P value
Blood Pressure	0.0038	0.0003	12.7	< 0.05
Cholesterol Level	0.0335	0.0044	7.6	<0.05
Blood Glucose Level	0.0428	0.0059	7.3	<0.05
Bone Density	-23.8609	0.4577	-52.1	< 0.05
Vision Sharpness	-30.1386	0.9114	-33.1	< 0.05
Hearing Ability	0.1375	0.0094	14.7	< 0.05
Cognitive Function	-0.0719	0.0096	-7.5	<0.05

Table 2 shows those factors with a p-value less than 0.05, which represented a strong linear relationship with age. For example, Blood Pressure has a small positive coefficient (0.0038), suggesting that as age increases, blood pressure tends to increase slightly. Cholesterol Level also has a positive coefficient (0.0335), indicating that cholesterol tends to rise as people get older. The estimated number shows that Bone density, vision, Cognitive Function, and Hearing Ability diminish over time.

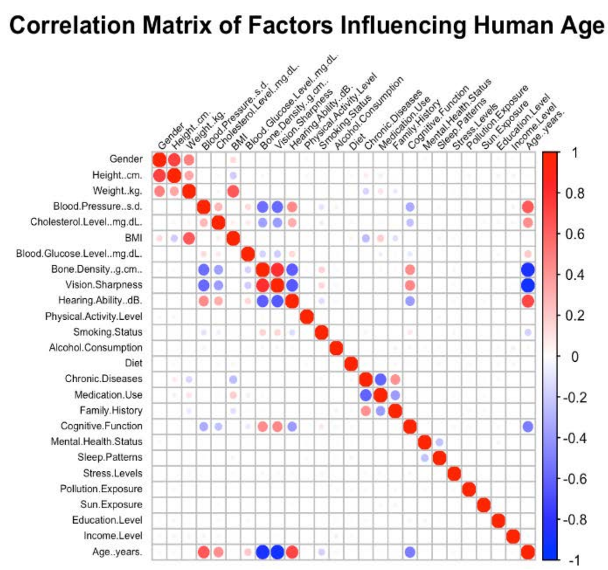

Fig. 1 Corrplot (Photo/Picture credit: Original).

This paper chose to use the Corrplot to find possible interactions. Figure 1 visualizes how significant factors are related to each other, with each circle representing the strength of the relationship between the two factors. Red circles represent positive correlations. Blue circles represent negative correlations. The bigger the circle, the stronger the correlation. Based on the observation of LinR model output and visualization, confounders like blood pressure, cholesterol level, blood glucose level, bone density, vision sharpness, hearing ability, cognitive function, family history, and BMI should not be considered in the LR analysis of lifespan. Since smoking status has a strong correlation with alcohol consumption, only one of them (alcohol consumption) should be kept. The same procedure goes for weight and height, only weight is kept.

3.2 Logistic Regression (LR) analysis

After taking those confounders out of the data, a LR model containing other variables can be created. In the first LR model, those people whose dying age is above 80 are considered as long-living.

Table 3. R output of the LR model (Age>80).

	Estimate	Std. Error	z value	P value
Mental health status 2	0.3553	0.1923	1.848	0.06
Mental health status 3	0.3860	0.1909	2.022	0.04
Sun exposure	0.0338	0.0164	2.064	0.04
Education level	0.2881	0.1600	1.801	0.07

Based on the LR output in Table 3, four variables show varying degrees of association with lifespan. First, Individuals with Mental Health Status level 2 have a positive estimate of 0.3553, with a z-value of 1.848 and a p-value close to significance (0.06). Similarly, level 3 has a slightly higher estimate of 0.3860, a z-value of 2.022, and a p-value of 0.04. This suggests that better mental health status may be positively associated with lifespan. Second, sun exposure shows a positive estimate of 0.0338, with a z-value of 2.064 and a p-value of 0.04. This indicates that higher sun exposure is slightly associated with long-living. Third, the education level has an estimate of 0.2881, a z-value of 1.801, and a p-value of 0.07, a little above the conventional significance level. This implies a potential positive association between higher education and longevity.

Table 4. R output of the LR model (Age>70).

	Estimate	Std. Error	z value	P value
Alcohol Consumption 3	0.1858	0.1102	1.686	0.090
Diet 2	0.2154	0.1101	1.957	0.050
Education level 3	0.3459	0.1188	2.911	0.003

This paper tried different thresholds of long-living. The output of the LR model varies after adjusting thresholds to 50, 60, and 70. For instance, at age 70, as shown in Table 4, alcohol consumption and diet have a positive effect on lifespan.

Table 5. Confidence Interval of the LR model (Age>70).

	OR	2.5%	97.5%
Education level 2	1.2693	1.0039	1.6041
Education level 3	1.4133	1.1194	1.7839

Based on the confidence intervals (CIs) shown in Table 5, there are only two factors' CIs (Education level 2 & 3) lie above 1, which suggests that the association is statistically significant at the 5% level.

Overall, the effectors of the survival rate changed at different ages. Among these factors, the educational level always remains a significant one. This discovery can lead to the conclusion that people are more likely to live longer and avoid early death if they have a higher educational level.

3.3 Cox Survival Analysis

The LR model is suitable for binary outcomes (such as elderly and non-elderly groups) but does not explicitly consider the „time" factor. In comparison, the Cox model can give a hazard ratio (HR) for each variable and displays how each factor affects survival time. This part of the study was inspired by Anne-Julie's study which assessed a unique metabolic signature of a healthy lifestyle. Cox regression models were used to link this metabolomic signature with mortality and longevity outcomes over a 28-year follow-up [9].

Table 6. The Summary Output of the Cox model.

	HR	p-value
Mental Health Status 2	0.7644	0.0077
Mental Health Status 3	0.8005	0.0250

ISSN 2959-409X

In Table 6, Mental Health Status 2 and Mental Health Status 3 both have a significant impact on survival time because their p-values are small (0.00768 and 0.02504 respectively). They might be correlated with lifespan. Thus, a survival curve of mental health status and survival rate can be established.

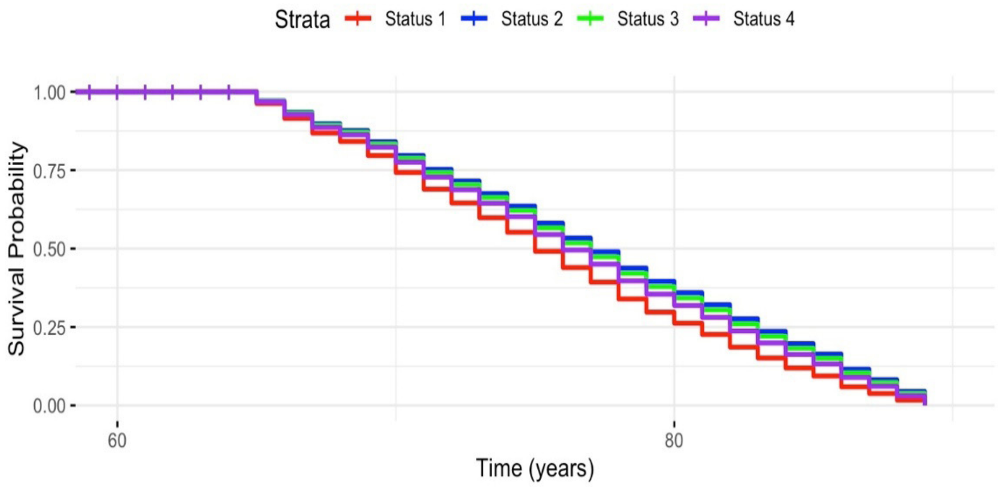

Fig. 2 Survival curves by mental health status (Age 60+) (Photo/Picture credit: Original).

From Figure 2, people with good mental health status (status 2) are more likely to survive as age increases.

3.4 Random Forest Analysis

The paper used a random forest model to capture non-linear relationships, offering a much broader view of lifespan factors compared to previous LinR, LR, or Cox models.

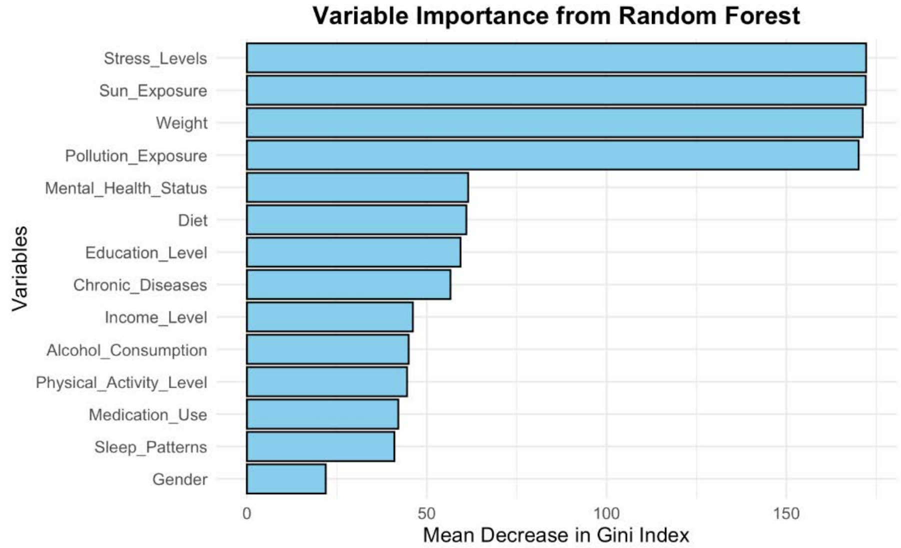

Fig. 3 Random forest plot (Photo/Picture credit: Original).

The variable importance plot from the random forest model highlights key factors influencing lifespan. From Figure 3, Stress Levels and Sun Exposure contribute the most to the model, indicating they have strong associations with lifespan. This result is quite similar to Rizzuto's paper. Some habits that are related to work status might affect health levels. Retirees generally had better health perceptions than those still working because those retirees have regular exercise, and good sleep [10]. Weight and Pollution Exposure also impacts lifespan significantly. Such outcomes are likely due to their effects on health and exposure to environmental risks. Mental Health Status, Diet, Education Level, and Chronic Diseases show moderate importance, they are a bit less influential than the top factors. Alcohol Consumption, Physical Activity Level, Income Level, and Medication Use are lower-importance variables. Gender has the least importance, indicating minimal impact on lifespan within this data set.

4. Conclusion

This study looked at the main factors that influence how long people live. This paper combined health, lifestyle, and environmental data. Using the LinR to extract some confounders in the data to improve the analytic accuracy of LR. Other factors, such as mental health, sun exposure, and education level, showed strong effects on lifespan in multiple models. Stress levels and pollution exposure were also important in the random forest model. Furthermore, the level of significance of effectors changed at different ages. These results suggest that mental health, education level, and environmental conditions are key factors for a longer life. Government Policies that reduce work stress and support mental health could help people live longer and healthier lives. The government should reduce air pollution and give people more opportunities to accept high-quality education. Moreover, this study has some limitations and weaknesses, the data does not come from multiple sources. It might not represent the situation in other groups of people. Some methods like Elastic regression are not used to further analyse the data.

References

[1] Stephan A-J, Strobl R. The times we are born into and our lifestyle choices determine our health trajectories in older age - Results from the KORA-Age study. Preventive Medicine, 2020, 133: 106025.

[2] Vega T, Hilario F, Pérez-Caro M. Genetic, environmental, and lifestyle factors associated with longevity. Protocol and response of the LONGECYL Study. Gac Sanit, 2022, 36(3): 260-264.

[3] Øvrum A, Gustavsen GW, Rickertsen K. Age and socioeconomic inequalities in health: Examining the role of lifestyle choices. Advances in Life Course Research, 2014, 19: 1-13.

[4] Dyer K A. Daily healthy habits to reduce stress and increase longevity. Journal of Interprofessional Education & Practice, 2023, 30: 100593.

[5] Monge-Martín D, Caballero-Martínez F, Forjaz M J, Castillo M J, Rodríguez-Blázquez C. Health state perception of people close to retirement age: Relationship with lifestyle habits and subjects' characteristics. Heliyon, 2023, 9: e17995.

[6] Li Y, Babazono A, Jamal A, Liu N. The impact of lifestyle guidance intervention on health outcomes among Japanese middle-aged population with metabolic syndrome: A regression discontinuity study. Social Science & Medicine, 2022, 314: 115468.

[7] Zhong J, Chen L, Li C, Li J. Association of lifestyles and multimorbidity with mortality among individuals aged 60 years or older: Two prospective cohort studies. SSM - Population Health, 2024, 26: 101673.

[8] Wang J, Chen C, Zhou J. Healthy lifestyle in late-life, longevity genes, and life expectancy among older adults: a 20-year, population-based, prospective cohort study. Lancet Healthy Longev, 2023, 4: e535-e543.

[9] Tessier A-J, Wang F. Plasma metabolites of a healthy lifestyle in relation to mortality and longevity: Four prospective US cohort studies. Med, 2024, 5: 224-238.

[10] Rizzuto D, Mossello E, Fratiglioni L, Santoni G, Wang H-X. Personality and survival in older age: The role of lifestyle behaviors and health status. Am J Geriatr Psychiatry, 2017, 25(12): 1363-1372.