Association between resource inequality and causes of death in 204 countries and regions from 1990 to 2019

Jiaying Hu[1], Tingting Shen[1], Xinying Wang[1],Guangmei Yang[1]

January 2024

1 Introduction

With the progress and improvement of healthcare, the life expectancy of the global population continues to increase [1][2]. For example, in the 1950s, the average life expectancy at birth for the global population was 46 years. But by 2015, it had already reached the age of 71. At the same time, the increase in lifespan goes hand in hand with the compression of the distribution of individual age of death, thereby reducing inequality in lifespan and making the age of death increasingly predictable [3][4][5][6]. The age of individual death has undergone significant changes, with mortality rates generally decreasing across all age groups, and the distribution of causes of death (CoD) has also undergone a significant shift from primarily infectious deaths to most non communicable deaths [7][8][9][10]. Although people have a lot of knowledge about the main causes of death and their evolution over time, they know very little about the diversity of such causes of death.

The diversity of causes of death is an important indicator for evaluating differences in population health, revealing the structure of society and people's behavioral patterns. On the one hand, low diversity of causes of death indicates a more homogeneous population and easier prediction of the main causes of death. On the other hand, the high diversity of causes of death poses a huge challenge to the healthcare system, as efforts to reduce overall mortality become more complex, dispersed, and may have lower outcomes [11]. That is why the composition of causes of death has long been regarded as a key indicator for measuring social well-being and health status [9][11]. Health outcomes are increasingly stratified by socioeconomic status within countries [12][13][14], a phenomenon that could further contribute to diversifying the causes from which individuals die.

We mainly investigated the epidemic trends between different regions from 1990 to 2019. The results of this study will help us better understand how epidemiological transitions have developed in different regions of the world, how

regional differences exist, and the relationship between these transitions and the rapidly aging process happening globally.

The relative details and code are in our Github page:

2 Data and Method

We use publicly available data from the Kaggle database from 1990 to 2019.

2.1 K-means clustering

In the geometric space created in MCA, countries are classified into clusters based on proximity criteria using the k-means algorithm with random initial centers. Each cluster has cluster centers. The optimal number of clusters (k) is evaluated using 100 iterations according to the Calinski Harabaz criterion. The optimal number of clusters is the solution with the highest value of the Calinski-Harabaz index. To assess the internal quality of the clusters, the clustering stability of the best solution is calculated using the Jaccard bootstrap value for 100 runs.

2.2 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical method used for dimensionality reduction in high-dimensional datasets. It transforms correlated variables into uncorrelated ones, called principal components, by capturing the maximum variance in the data. The process involves deriving eigenvectors and eigenvalues from the covariance or correlation matrix.

PCA is widely applied in fields like finance and image processing to simplify complex datasets, aid in visualization, and identify dominant patterns. It serves as a valuable tool for data compression, noise reduction, and extracting essential information from large datasets, contributing to efficient analysis and interpretation.

2.3 Linear Regression Analysis

Linear Regression Analysis is a fundamental statistical technique employed to model the relationship between a dependent variable and one or more independent variables. This method assumes a linear correlation between the variables, where changes in the independent variables are associated with proportional changes in the dependent variable. The model's performance is evaluated using metrics like the R-squared value and Mean Squared Error, which indicate the model's accuracy and how well it captures the variability in the dataset.

In our analysis, linear regression is utilized to explore the trends in global mortality rates across various causes of death. The process involves fitting a linear equation to observed data, where the coefficients represent the relationship's strength and direction. The model's efficacy is assessed using the coefficient of

determination (R^2 value), which quantifies the variance in the dependent variable that can be explained by the independent variables. Additionally, the significance of the regression coefficients is evaluated using p-values, ensuring the robustness of the model. Linear Regression is a widely used tool in epidemiology and public health for trend analysis and forecasting, providing valuable insights for policy development and resource allocation. It facilitates the understanding of complex phenomena by quantifying the impact of various factors on health outcomes, thereby enabling targeted interventions and strategic planning.

3 Results

3.1 Divide 204 regions into three categories based on the cause of death by K-Means and PCA

Perform PCA and clustering based on different causes of death to obtain the two most important principal components (Figure 1-A). Ranking the first principal component coefficients by size, it was found that the top four countries with the lowest coefficients are China, Russia, India, and the United States, all of which are countries with larger land areas and larger populations. The top four countries with the highest coefficients are Tokelau, Niue, Nauru, and Tuvalu, all small island countries located in the Pacific Ocean, with very small areas and relatively small populations. Therefore, the first principal component can be understood as the population component of land area.

Then, ranking the second principal component coefficient by size, it was found that the top four countries with the lowest coefficient are China, the United States, Russia, and Japan. These countries have relatively advanced technology and development levels. The top four countries with the highest coefficient are India, Niger, Ethiopia, and Bangladesh. These countries are all developing countries and face some common socio-economic challenges, such as poverty, inequality Insufficient infrastructure, etc. Therefore, the second principal component can be understood as the economic and technological level component (Figure 1-B).

Finally, based on the results of principal component analysis, clustering was conducted into three categories (Figure 1-C). The first category should be countries with small land area and population, and moderate development level. The second category should be countries with large land area and population, and relatively backward development level, such as India. The third category should be countries with large land area and population, and high development level, such as the United States.

Meanwhile, we have used the K-Means method for our analysis. As indicated by the 'elbow plot', it's preferable to divide the 204 regions into 3 groups (Figure 1-D), and this grouping outcome aligns with the results from PCA (Figure 1-E).

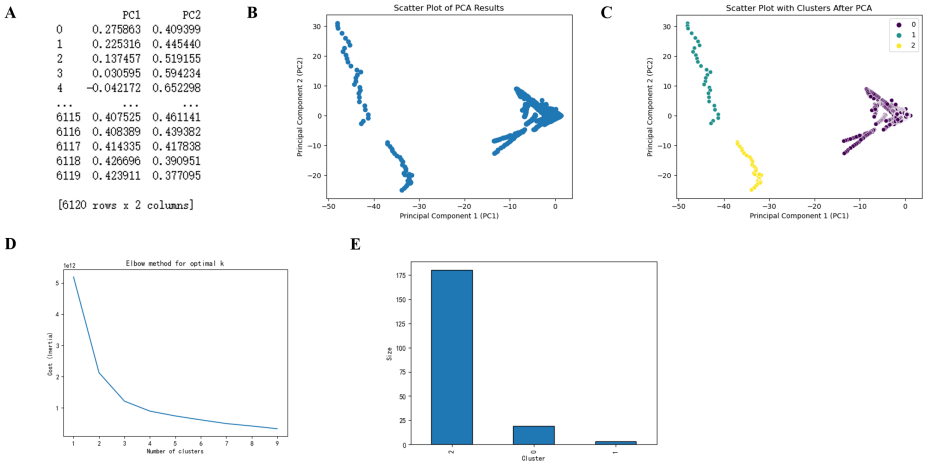

Figure 1: 204 regions are divided into three categories based on the cause of death by using PCA and K-Means

3.2 Study the leading cause of death in economically developed countries by liner regression

Next, we will mainly explore the differences in causes of death among countries with higher GDP. We selected the top five countries in terms of GDP, screened their respective diseases with the highest incidence rate, and drew a regression curve according to the year.

We found that cardiovascular diseases have the highest mortality rate in most countries, except for Japan. China has the highest number of cases and deaths, possibly due to a large population and inadequate medical resources. Cardiovascular diseases are difficult to overcome, and with the advancement of modern medical technology, the number of deaths has not decreased, but fluctuates over time (Figure 2-A). Among developed countries, the United States has the highest number of deaths due to its large population and uneven medical resources. The incidence rate of neoplasms in Japan has gradually expanded with the increase of time, which is inseparable from its aging population and the great change of social structure (Figure 2-B). These data and trends are crucial for public health policy makers as they reveal differences in how different countries face major health challenges.

In our country, the number of deaths from vascular diseases and tumors is showing a significant upward trend, which may reflect the impact of rapid economic development in China, changes in lifestyle (such as dietary habits and reduced exercise), and aging population on public health (Figure 2-C). The mortality rate of tumors (cancer) is also on the rise, which may be related to factors such as environmental pollution, chemical exposure during industrialization, and tobacco use. This requires strengthening cancer screening and early

diagnosis, providing more effective treatment options, and conducting relevant health education.

Overall, these trends indicate that while China is facing rapid urbanization, industrialization, and socio-economic changes, it is also facing a significant burden of non-communicable diseases. This requires systematic public health strategies and health system reforms to address these challenges and reduce the burden of future diseases.

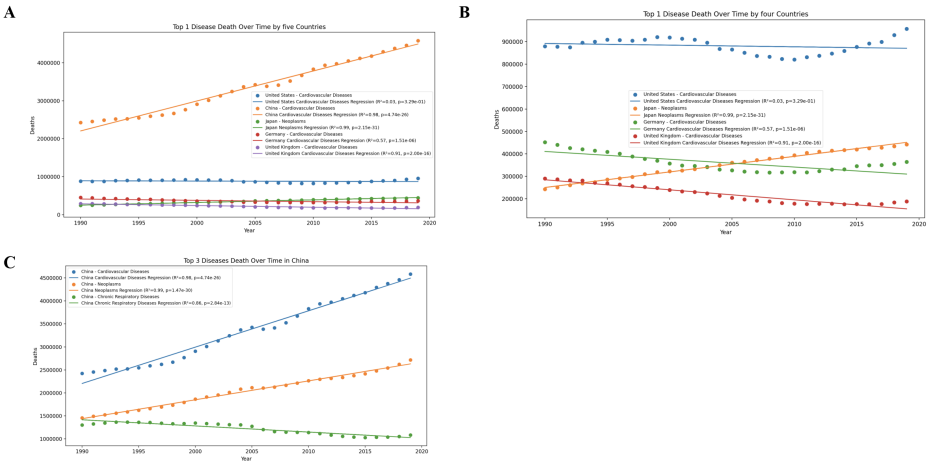

Figure 2: The most common causes of death in the top five GDP countries

3.3 Explore trends of Global Epidemic Diseases by liner regression

Next, we will explore the global trend of cardiovascular disease prevalence. Cardiovascular disease remains a major health challenge on a global scale, especially in Asia, where the growth rate of deaths is of concern (Figure 3-A). In Europe and the Americas, although the growth rate of cardiovascular disease deaths is relatively slow, it does not mean that it can be ignored. Continuous attention and investment of resources are still needed to further reduce the impact of these diseases. The data from Africa reflects the inadequacy of death registration systems or the relatively small impact of cardiovascular diseases compared to other health challenges, such as infectious diseases (Figure 3-B).

We then analyzed the main causes of death trends globally or in large regions (such as Africa, Asia, Europe, etc.) over a time span. Explore which diseases have an increasing or decreasing trend in the number of deaths and evaluate the degree of impact over time, that is, the rate of change in mortality over time. During approximately 30 years from 1990 to 2019, the number of deaths from cardiovascular disease showed a significant upward trend. The number of deaths from tumors is also on the rise. The number of deaths caused by

chronic respiratory diseases, although not as severe as cardiovascular diseases and tumors, still shows an upward trend (Figure 3-C).

In summary, it reflects the increasing trend of deaths from non-communicable diseases worldwide and suggests the need to strengthen public health interventions to prevent and control the development of these diseases. At the same time, this also highlights the importance of global health resource allocation, especially in the fields of prevention and early diagnosis.

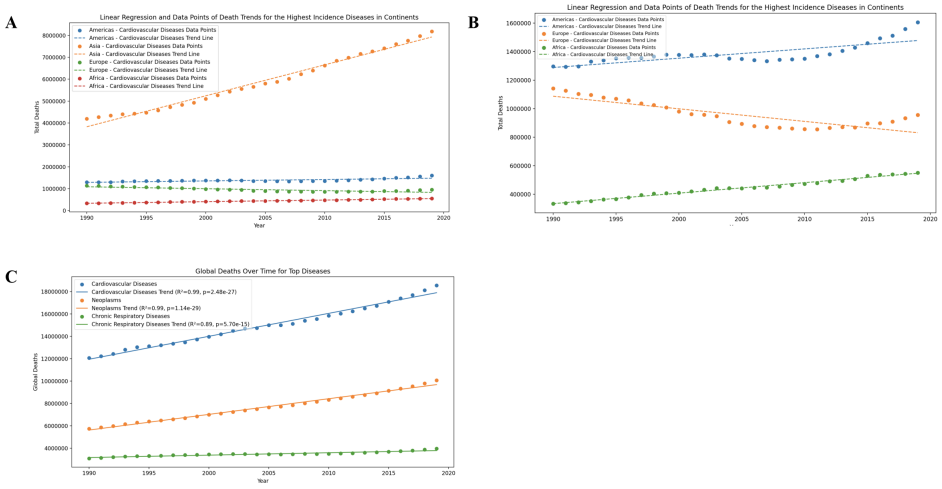

Figure 3: The trend of global causes of death

4 Discussion

Statistics on the annual number of deaths and causes of death are one of the most important means of evaluating the effectiveness of a country's healthcare system, and cause of death statistics help health authorities determine the focus of their public health actions. The increasing trend of deaths from non-communicable diseases worldwide suggests the need to strengthen public health interventions to prevent and control the development of these diseases.

At the same time, this also highlights the importance of global health resource allocation, especially in the fields of prevention and early diagnosis. Our analysis can help us understand the global distribution of diseases. For example, cardiovascular disease remains a major health challenge globally. Meanwhile, studying the causes of death in different countries helps to understand the relationship between mortality rates and socio-economic factors. For example, low-income countries may have higher mortality rates from infectious diseases, while high-income countries may have higher mortality rates from non-communicable diseases such as heart disease and cancer.

Finally, these studies can also reveal specific health issues and challenges, allowing for targeted public health policies and interventions. For example, if a country's suicide rate is exceptionally high, it may indicate the need to invest more resources to improve mental health services and prevent suicide. Our analysis provides important information for global health organizations and governments in developing and implementing prevention and control strategies for cardiovascular disease.

References

[1] J. Oeppen and J.W. Vaupel, "Broken limits to life expectancy," *Science*, vol. 296, no. 5570, pp. 1029–1031, 2002.

[2] J.C. Riley, "Estimates of regional and global life expectancy, 1800–2001," *Population and Development Review*, vol. 31, no. 3, pp. 537–543, 2005.

[3] J. Smits and C. Monden, "Length of life inequality around the globe," *Social Science and Medicine*, vol. 68, no. 6, pp. 1114–1123, 2009.

[4] J.W. Vaupel, Z. Zhang, and A.A. van Raalte, "Life expectancy and disparity: an international comparison of life table data," *BMJ Open*, vol. 1, no. 1, e000128, 2011.

[5] R.D. Edwards, "Changes in world inequality in length of life: 1970–2000," *Population and Development Review*, vol. 37, no. 3, pp. 499–528, 2011.

[6] Permanyer I, Scholl N. Global trends in lifespan inequality: 1950–2015. *PLoS One*. 2019;14(5):e0215742.

[7] Omran AR. The epidemiologic transition: a theory of the epidemiology of population change. 1971: The epidemiologic transition. *Milbank Quarterly* 2005;83(4):731–57.

[8] Omran AR. The epidemiologic transition theory revisited thirty years later. *World Health Statistics Quarterly*. 1998;53(2):99–119.

[9] Wang H, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390(10100):1084–150.

[10] J. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2020;396(10258):1204–22.

[11] Bergeron-Boucher M-P, Aburto JM, van Raalte A. Diversification in causes of death in low-mortality countries: emerging patterns and implications. *BMJ Global Health*. 2020;5(7):e002414.

[12] van Raalte AA, Kunst AE, Lundberg O, Leinsalu M, Martikainen P, Artnik B, et al. The contribution of educational inequalities to lifespan variation. *Population Health Metrics*. 2012;10(1):3.

[13] van Raalte AA, Martikainen P, Myrskylä M. Lifespan variation by occupational class: compression or stagnation over time? *Demography*. 2014;51(1):73–95.

[14] van Raalte AA, Martikainen P. The case for monitoring life-span inequality. *Science*. 2018;362(6418):1002–4.