Correlation Analysis of the Global Trend between Nation GDP and Cancer Prevalence Rate

Bojie Li *

Middle School Affiliated to Renmin University, Beijing, 100061, China

* Corresponding Author Email: jdany@email.tjc.edu

Abstract. Through the development of modern society, the prevalence of cancer gradually increased. Many researchers have done a lot of relevant research and experiments from the public health perspective. For example, liver cancer is found to be modifiable through changes in lifestyle. This paper discusses the relationship between cancer prevalence and macro factor, which is the national gross domestic product (GDP). More specifically, this paper discusses the relationship between cancer prevalence and national GDP. This paper applies polynomial regression to study the polynomial between the cancer rate of one country and its cancer rate. Through the collection of public cancer prevalence data and national GDP data, this work found that the cancer prevalence rate has a positive correlation with the national GDP. It is speculated that the reason for this result is that with the increase in national GDP and the development of medical technology, people are less likely to die from less dangerous diseases than before.

Keywords: Cancer; GDP Growth; Polynomial Regressions.

1. Introduction

According to the report released by the International Agency for Research on Cancer, a subsidiary of WHO, in 2020, there were 19 million new cancers and 9.96 million cancer deaths worldwide [1]. The number of new cancers worldwide is expected to reach 30.2 million by 2040. Cancer has become one of the deadliest diseases, seriously threatening human life, health and safety. It has been shown that the crude morbidity and age-standardized incidence rate (ASIR) have a positive correlation with the GDP per capita, while the age-standardized mortality rate (ASMR) was negatively correlated with GDP per capita level [2]. It is thus meaningful to study how gross this index may affect the cancer rate of the whole population and analyze the factors behind it.

Many researchers have done relevant research and calculated relevant formulas [3,4]. But most of the research is from the biological point of view. Excluding the research of biologists, sociologists and economists found that the disease rate of cancer in developed countries is relatively high, and the mortality rate is relatively low. The formula calculated in this way needs to fully understand people's information in the early stage so that it will lead to low practicability. Therefore, it is necessary to find the influencing factors of cancer prevalence from a macro perspective. This paper is not a final answer to this question but an attempt. Therefore, the paper chooses GDP and cancer mortality. This paper mainly discusses the relationship between cancer prevalence and national GDP. Big data analytics is used in this study. The data came from the Global GDP statistics website and the global population and cancer statistics website. The two types of data were fitted, and the data were visualized. Finally, the development trend of national GDP and the trend of national cancer prevalence were compared. Finally, according to the analysis's comparative results, draw the final conclusion and guess the reasons.

2. Method and Data

2.1 Big Data and Data Visualization

In order to ensure the scientific nature, authenticity, and reliability of the data, all the data are found on the official website that publicizes the number of patients, population and GDP of each country. For example, all data related to population sizes are from The World Bank Data on

population growth, as well as the data relating to GDPs is also from [5]. Since the official website showing the GDP of each country only shows some countries, the researchers narrowed down the scope of all data to countries with GDP data. Because the statistical range of different data types is different, the public range is finally selected from 2000 to 2020. Libraries in Python generally refer to third-party libraries and standard libraries. The library has modules to improve the efficiency of writing code. With the import library, this work uses custom modules in the library to achieve code purposes. The library selected for this study is the numpy library, the matplotlib library, and the pandas library in Python, which are used to merge the data into the code, fit and visualize them, and draw conclusions. Through past studies and theories accepted by the public, the researchers speculated that the trend of cancer prevalence was the same as that of the national GDP [6]. Influenced by most of these ideas, this study assumed at the outset that cancer rates would decrease as national GDP grew.

2.2 Polynomial Regression

This work applies polynomial regression to analyze the correlation between the GDP and cancer rate. Polynomial regression is a regression analysis method that studies the polynomial between a dependent variable and one or more independent variables. Polynomial regression is used here because it can find a corresponding relationship in a complex case. This study uses univariate polynomial regression to predict the relation between GDP and cancer rate. Thus, the number of independent variables is one. This study focuses on the relationship between national GDP and cancer incidence. The interpretation of all the graphs created by the polynomial regression is based on the instruction of UCLA Oarc [7].

$$y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \tag{1}$$

3. Results and Discussion

This paper used the data analysis tool from Python. All the data and processes were done using Python codes, including but excluding the conduction of the polynomial regressions and producing all the graphs. The instruction for coding can be found in Python Tutorial [8].

The results show that the trend of cancer prevalence is the same as that of the national GDP. In the following, the fitting results of a few countries are discussed in detail.

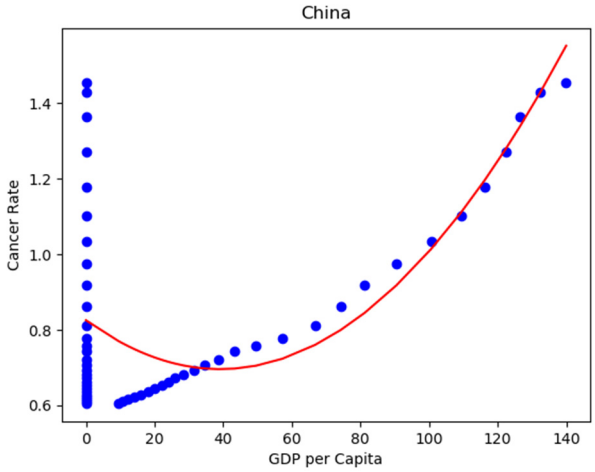

Fig 1. GDP Growth and Cancer Rates in China

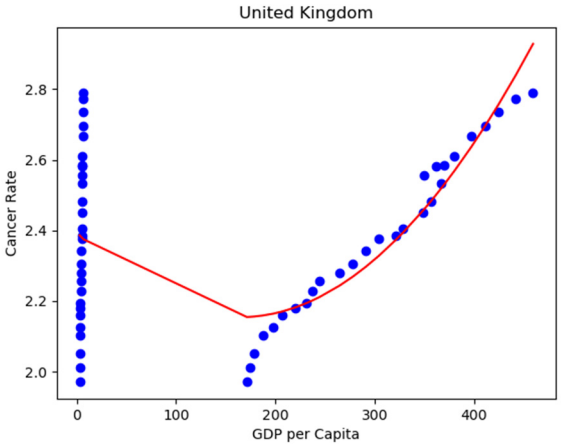

Fig 2. GDP Growth and Cancer Rates in U.K.

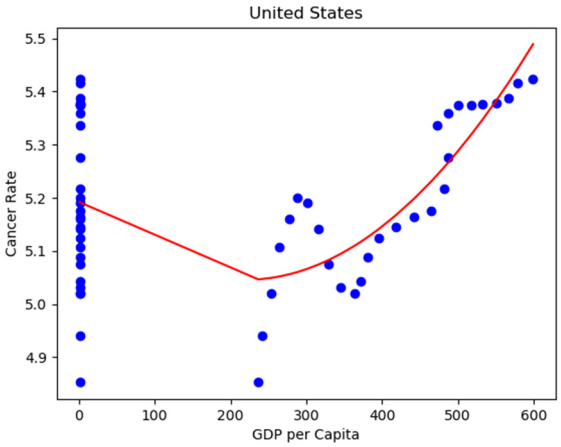

Fig 3. GDP Growth and Cancer Rates in U.S.

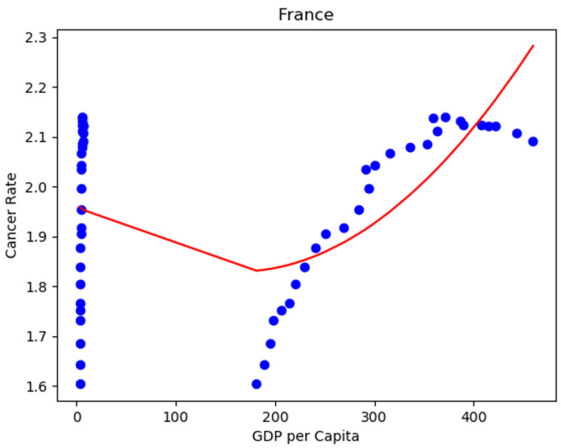

Fig 4. GDP Growth and Cancer Rates in Fiance

The polynomial fitting of GDP and cancer rate in China is shown in Fig. 1. It clearly shows that the incidence of cancer has increased with the development of China's GDP. Fig. 2 shows the fitting result for the United Kingdom. Similar to the case of China, it also shows that with the development of the U.K.'s GDP, the incidence of cancer has increased. Fig. 3 shows the fitting results for the United States. While the overall development of the U.S.'s GDP also has a positive correlation with the cancer rate, unlike in the former two countries, the correlation has fluctuated a lot. Fig. 4 shows the case of France. This figure also shows that the incidence of cancer increases with the development of France's GDP. However, there is a slight decrease in the cancer rate when the GDP per capita is over 400.

The hypothesis is different from the real result because, in the ideal theory, the development of the country has reached its peak and medical technology is extremely developed. The real result is that

the national GDP has indeed been developed, but the development of medical technology has not made great progress in cancer. Therefore, many people avoid dying from diseases that pose little threat but from cancer.

This reminds many countries not to indulge in the ideal theory that with the development of GDP, people's cancer prevalence will decrease. For example, among the causes of cancer, smoking accounted for 30% of total cases [9]. And this may be due to the prevalence of smoking cigarettes when people get wealthier, which is also worth further study. Only by developing GDP, directly developing medical technology, or improving the quality of the living environment can a country truly reduce the incidence of cancer [10].

4. Conclusion

The paper found that cancer incidence increased with the increase in national GDP. This conclusion is different from most people's ideas and predictions. This also reminds the public and other researchers to analyze the cancer prevalence rate from the micro perspective and different macro factors. It is hoped that more studies will be conducted in the future to compare factors such as the average life span of the population, the ratio of men and women, and the cancer prevalence rate. At the same time, it is also hoped that the official website for the statistics of national GDP can expand the scope and grade to mitigate the limitation of cancer incidence being under-detected or under-reported in relatively less developed countries. However, the problem of data initialization has not been solved in this study, resulting in the result of data visualization being inconsistent with the overall result. In the future, this problem needs to be solved in the code. The authors think it would have been more realistic to include more factors, such as the ratio of men to women, life expectancy, and the proportion of older people. At the same time, it is hoped that more macro factors will be considered in the future.

References

[1] Zhixun Yang, Rongshou Zheng, Siwei Zhang, Hongmei Zeng, Changfa Xia, He Li, Li Wang, Yanhong Wang, and Wanqing Chen. "Comparison of cancer incidence and mortality in three GDP per capita levels in China, 2013." Chinese Journal of Cancer Research 29, no. 5 (2017): 385.

[2] International Agency for Research on Cancer, 2021. IARC Biennial Report 2020-2021. Lyon: International Agency for Research on Cancer.

[3] Pritchard, Colin, and T. Hickish. Comparing cancer mortality and GDP health expenditure in England and Wales with other major developed countries from 1979 to 2006. British Journal of Cancer 105, no. 11 (2011): 1788-1794.

[4] Evans, B. T., and C. Pritchard. Cancer survival rates and GDP expenditure on health: a comparison of England and Wales and the USA, Denmark, Netherlands, Finland, France, Germany, Italy, Spain and Switzerland in the 1990s. public health 114, no. 5 (2000): 336-339.

[5] Roser Max. Our World in Data. Economic Growth, 2013. https://ourworldindata.org/economic-growth#citation.

[6] Introduction to SAS. UCLA. FAQ how do I interpret the sign of the quadratic term in a polynomial regression, 2021.

[7] Martin C.S. Wong, Johnny Y. Jiang, William B Goggins, et al. international incidence and mortality trends of liver cancer: a global profile, 2017, 31(03).

[8] Stojiljkovic, Mirko. Linear regression in Python. Real Python. https://realpython. com/linear-regression-in-python/. Accessed 9 (2022).

[9] Ades, F., Christelle Senterre, E. De Azambuja, Richard Sullivan, R. Popescu, Florence Parent, and M. Piccart. Discrepancies in cancer incidence and mortality and its relationship to health expenditure in the 27 European Union member states, 2020, 6(01).

[10] Teasa S. Stewart, Jennifer Moodley, Fiona M. Walter. Population risk factors for late-stage presentation of cervical cancer in sub-Saharan Africa, 2018, 04(02).