Analysis of the Impact of Dietary Elements on National COVID-19 Confirmation Rates

Zidie Zhuang*

Department of Applied Mathematics, University of Washington, Seattle, United States

*Corresponding author: zidiez@uw.edu

Abstract. This work focuses on using the relationship between different diets and the confirmed infection rates of different countries to try and discover other factors that might affect the confirmation rate. Firstly, data on the percentage caloric intake of various countries and their respective COVID-19 confirmation rates were analyzed using BMA variable selections to isolate specific elements of interest. Miscellaneous, stimulants, and obesity were found to have significant correlations with confirmation rate - the first two of which referred to baby food and caffeine, specifically, and the last despite something of interest despite no correlation between it and other food types. Baby food was found to be positively correlated with the percentage child population of a country, and both were found to be associated with lower confirmation rates; as such, baby food is unlikely to be an originating factor. Caffeine consumption may be a factor as it is associated with all-cause mortality in existing literature, but with GDP as a confounding variable, more study is required. Obesity was confirmed to be related to increased COVID-19 confirmation rates, but no relation between obesity and COVID was found between those and any food group. Ultimately some country selection or a study based on a homogeneous group of individuals may be necessary to isolate confounding variables and reach a meaningful conclusion.

Keywords: COVID-19, National diet, Bayesian model, National infection rate, Pandemic.

1. Introduction

Infectious diseases have been a constant and formidable challenge throughout human history, with pandemics such as smallpox and the Black Death causing devastating impacts on societies. These diseases not only led to immense loss of life but also brought significant changes in social structures, economies, and health practices [1]. Smallpox, caused by the variola virus, was a highly contagious and deadly disease that affected communities for thousands of years until its eradication in 1980, a testament to the power of effective public health measures and vaccination campaigns. The Black Death, a pandemic of bubonic plague that struck Europe and Asia in the 14th century, is another stark reminder of the ravages of infectious diseases. It was caused by the bacterium Yersinia pestis and transmitted through fleas on rodents, leading to a massive death toll and profound social transformations [2]. In Duncan's research, the author points out that in the 14th century, health authorities in northern Italy recognized the effectiveness of a 40-day quarantine to combat the plague, understanding its long incubation period and the risks posed by asymptomatic carriers. Despite limited medical knowledge, practices such as isolation and travel restrictions were implemented, based on observations of the disease's infectious nature and transmission methods [3].

Historical analyses of these pandemics have highlighted several critical factors, including the role of human behavior, environmental conditions, and biological characteristics of the pathogens in disease spread [4]. In Fauci's research, he discovered that many chronic diseases, such as peptic ulcers and certain cancers, are caused by microbial infections, highlighting the role of vaccinations in prevention. The hepatitis B vaccine, for example, has successfully reduced hepatic cancer rates in some populations. Approximately 16% of all cancers are linked to infectious agents, emphasizing the significant impact of microbes on cancer development. These studies have shown how societal and technological advancements, such as improvements in sanitation and the development of vaccines, have been pivotal in controlling these diseases. However, these pandemics also revealed the vulnerability of populations with poor nutrition and underlying health conditions, underscoring the importance of overall health and immunity in combating infectious diseases [5].

Fast forward to the present, the world is facing a new challenge with the COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2. First identified in late 2019, COVID-19 rapidly evolved into a global crisis, affecting millions of people worldwide and overwhelming healthcare systems. As with past pandemics, researchers have scrambled to understand the various factors that influence the spread and impact of the disease. In Ciotti's research, they conclude many ways to test the infection of COVID-19 such as PCR and serology and they points out that based on their research, artificial intelligence can be a good tool to use. In Khanna's research, they analyse how different countries' different reactions to the COVID-19 pandemic caused to different outcomes. Among the many aspects being studied, the relationship between underlying health conditions and the severity of COVID-19 has become a critical area of focus. Conditions such as diabetes and heart disease have been identified as significant risk factors for severe COVID-19 outcomes. Intriguingly, these conditions are often linked to dietary habits, suggesting a potential connection between diet and COVID-19 susceptibility [6].

2. Data and Methods

2.1. Data Sources

This study aims to explore this connection by examining the typical diets of different nations and their respective COVID-19 infection rates. By analyzing dietary compositions in terms of the proportions of various food groups consumed, the study seeks to identify dietary patterns that may influence a nation's susceptibility to COVID-19. The initial phase of the research will involve determining which dietary factors are most significantly associated with COVID-19 infection rates. Following this, a more detailed analysis will be conducted on the identified factors to understand their specific roles and the mechanisms through which they may affect susceptibility to the virus. By investigating the link between diet and COVID-19, this study hopes to provide insights that can help individuals and communities make informed decisions about their dietary habits, potentially reducing their risk of contracting the disease. The findings could also offer valuable information for public health strategies aimed at managing and mitigating the impact of the pandemic. Just as historical pandemics have taught us the importance of public health measures and medical advancements, the current crisis presents an opportunity to learn more about the role of diet in infectious diseases, contributing to our overall understanding and preparedness for future health challenges.

This paper uses several comprehensive datasets sourced from Kaggle to investigate the relationship between diet and COVID-19 confirmation rate. The first dataset is titled "COVID-19 Healthy Diet Dataset". This dataset includes the percentage of food intake from various foods such as alcoholic beverages, fruits, vegetables, and aquatic products for 170 countries and the COVID-19 related data such as confirmed rate, death rate, and recovered rate for each country in the year 2020. This dataset is integral to our study as it allows us to explore the potential relationship between dietary fat intake from different food types and the COVID-19 infection rates in various countries [7].

The second dataset used is "Population by Age Group". This dataset includes the population of different age groups of different countries in the year 2021. This dataset is relevant to this paper because some diet intake might have a direct relationship with age. This dataset is important for our paper as dietary habits can vary significantly with age. For example, alcoholic beverages are not usually consumed by children under 10. The relationship between different age groups' populations and diet intake may affect the COVID-19 confirmation rate. This dataset can give us the relationship between the population of different age groups with their diet habits and COVID-19 confirmation rate.

The third dataset used is "COVID-19 Predictors". This paper only use the smoking rate of different countries in this dataset to find out if the smoking rate affects the consumption of stimulants in each country. This can help us do the secondary analysis of stimulants to further discuss the relationship among caffeine consumption, smoking, and COVID-19 confirmation rate.

2.2. Methods

To analyze which factors in the dataset have a possible effect on Covid 19 confirmation rate, one has to first perform a variable selection to see which variables listed in the dataset certainly have an effect on the confirmed rate. Here, the author will use Bayesian Model averaging for its ability to search over various model and provide an approximate parameter of the best performing one. Inputting data from the daily intake proportion dataset, one can try to use various types of food intake and few other factors to predict Covid 19 confirmation rate. As sample result table 1 shows, the p! =0 column shows the possibility that this variable is contained in the model, calculated by searching among various models. Therefore, the author will select variables with this column's value being larger than 90%. Possible models and parameters in the model are shown in latter columns [8].

After deciding variables, the next step is to analyze reasons why these diet or food structures influence Covid 19 confirmation rate. The high relationship does not necessarily indicate causation, for it is likely that they are simply correlated and both resulting from an exterior factor. Discovering the underlying factor that causes such an effect can help identify a more direct cause of change in confirmed rate.

3. Results and Discussions

3.1. Results

Using BMA, the result is shown in Table 1. From the result, it is found that the variables that influence covid confirmed rate. Miscellaneous has a strong negative correlation with Confirmed rate, while Obesity and Stimulant has a positive correlation with the confirmed rate. Among these three factors, Miscellaneous and Stimulants has a strong effect relative to Obesity, therefore, the author can try and discover how these two affects confirmed rates [9].

Table 1. Resulting BMA table

Intercept	p!=0	Intercept	p!=0
Alcoholic Beverages	1.4	Oilcrops	27.7
Animal Products	19.2	Pulses	1.4
Animal Fats	80.9	Spices	1.4
Aquatic Products	2.3	Starchy Roots	4.2
Cereals(Excluding Beer)	6.7	Stimulants	100.0
Eggs	12.4	Sugar/Crops	1.3
Fish/Seafood	1.5	Sugar/Sweeteners	1.3
Fruits(Excluding Wine)	1.3	Treenuts	8.7
Meat	15.0	Vegetal Products	19.2
Miscellaneous	94.0	Vegetable Oils	20.3
Milk(Excluding Butter)	8.4	Vegatables	1.9
Offals	44.7	Obesity	100,0

According to data description, Miscellaneous, the most powerful factor in the data set, refers to baby food. Following that discovery, it is known that a high proportion of consumption of baby food in a country means a high proportion of infant and children population in the country. Therefore, the author doubts the true factor that is affecting the difference in confirmation rate is instead the Children proportion in the country. Using the dataset on population by age group, a scatter plot between confirmed rate and proportion of population of children aged 0-4 shows that there is indeed a significant negative correlation between the proportion of young children and Confirmation rate. This gives us reasons why Miscellaneous, or baby food, influences Covid, while itself does not sound reasonable.

Looking at Fig. 1, one question might arise on whether the high children proportion is a result of low total population. To clear that, the author created a boxplot to show the total population of the countries with high or low proportions of children. From the boxplot it is seen that countries having more or less than 15% as 0-4 years old have no difference in total population, see Fig. 2. This reaffirms the validity of the finding saying that a higher proportion of young children has a correlation with a lower confirmation rate [10].

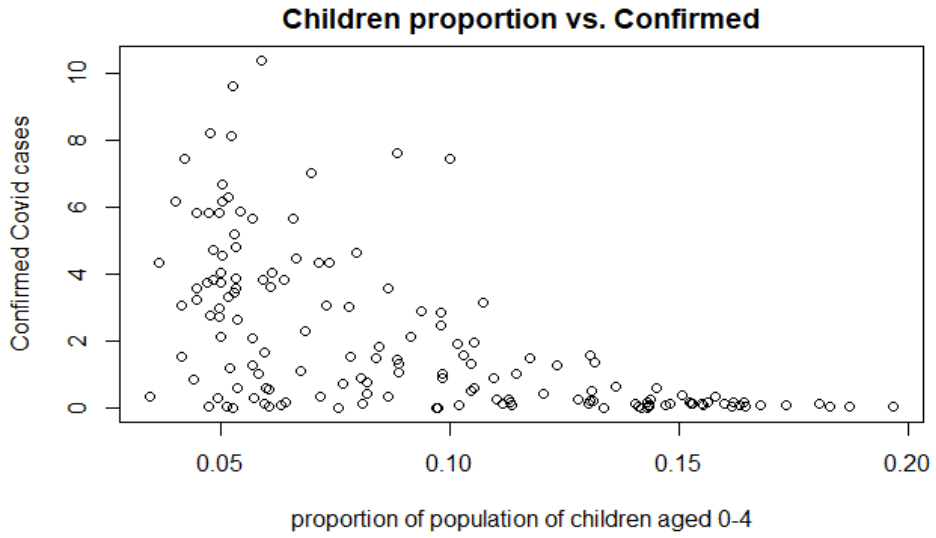

Fig. 1 Children proportion vs Confirmed rate

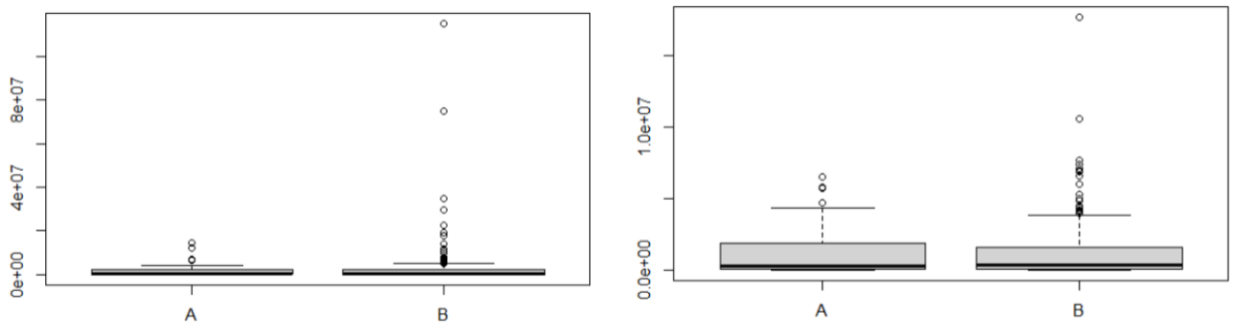

Fig. 2 Total population of countries with more that 15% 0-4 years old (A) vs less than 15%

Now, the author turns to look at Stimulant. The BMA method shows that there is a very high possibility that Stimulant is a covariant that can be used to predict Covid confirmed rate. Data source shows this is a combination of cocoa beans, Coffee, and tea. A research article by Andrew P Smith shows smoking might have a relationship with caffeine intake. However, creating a scatter plot between smoking and Stimulants (as shown in Fig. 3) shows that there is not a clear relationship between the two. Looking over possible factors such as BMI, GDP, and Smoking, the author created another BMA table (see Table 2), finding that none of these factors has a significant effect on stimulant consumption. This makes Stimulant or caffeine consumption a direct factor that affects covid 19 confirmation rate.

Table 2. BMA to predict Stimulants based on GDP, Smoking, and BMI.

Intercept	PPP/GDP per capita	Smoking 2016	Overal mean BMI kg/m2
p!=0	46.3	28.7	16.7

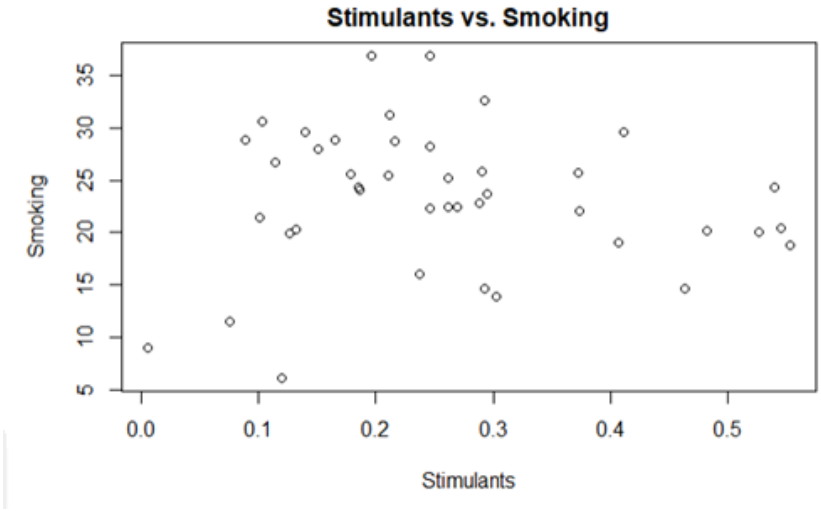

Fig. 3 Stimulants consumption vs. Smoking

3.2. Discussions

Overall, Bayesian Model Averaging found that most dietary elements did not have a significant effect on the COVID-19 infection rate, with only stimulants and “miscellaneous” foods having a significant effect on a nation’s COVID-19 confirmation rate; within the data sources, these were specified to caffeine and baby food. Smoking was considered as a factor that may be associated with increased stimulant use, but the analysis found no clear relation between the two. However, both caffeine consumption and COVID-19 confirmation rates increased with GDP, but a second BMA analysis found a low probability of GDP impacting caffeine consumption. Existing medical literature also indicates that caffeine consumption increases all-cause mortality, which may include COVID-19. Further study is still needed to ascertain the relation between COVID-19 infection rate and caffeine consumption, however.

A negative correlation was found between baby food and COVID-19 infection rate; however, because this food is largely only consumed by infants, the proposition that it would have a genuine impact when applied to a healthy adult was suspect. Indeed, this is more reflective of the infant population within a country, rather than the effect of baby food as evidenced by the fact that the proportion of a population that consists of children was also inversely related to confirmed COVID-19 cases.

Although no actual food group was found to be closely related to COVID-19 infection, the latter was correlated with a diet-related ailment: obesity. Because obesity is associated with overeating in general, this may mean that the impact of any specific food is muted. It is a possibility, for example, that one could eat a great quantity of calories from any combination of fruits, vegetables, fats, or meats and any combination of fat percentage would result in obesity. This is a limitation of this statistical study; it can analyze dietary elements as a proportion of peoples’ diets but does not perhaps capture the absolute number of calories consumed per food group.

4. Conclusion

In conclusion, Bayesian Model Averaging revealed that most dietary elements, except for stimulants and 'miscellaneous' foods (specifically caffeine and baby food), had minimal impact on COVID-19 infection rates. While an increase in caffeine consumption and COVID-19 rates was observed with higher GDP, further analysis suggested a low probability of GDP directly affecting

caffeine consumption. Interestingly, medical literature suggests a link between caffeine consumption and all-cause mortality, potentially including COVID-19, warranting further research into this relationship. Contrastingly, an inverse correlation was noted between baby food consumption and COVID-19 rates, likely reflecting the proportion of infants in a population rather than a direct effect of baby food on adults. Additionally, the study found a correlation between obesity and COVID-19 infection rates. Since obesity is linked to general overeating rather than specific food groups, it suggests that the impact of individual foods may be less significant than overall calorie intake. This study faced limitations, including using entire countries as data points, leading to potential confounding factors such as varying GDP, testing capacities, and pandemic policies. To refine future research, focusing on countries with similar COVID-19 containment measures or analyzing data from individuals within specific regions could provide more homogenous and insulated variables, enhancing the understanding of diet's impact on COVID-19 infection rates.

References

[1] James G., et al. An Introduction to Statistical Learning: With Applications in R, Springer, (2013).

[2] Brice Carolyn, Smith Andrew. Factors associated with caffeine consumption. International journal of food sciences and nutrition. 2002, 53: 55-64.

[3] Michael G Wagner 1, Kapil G Kapoor, Alan L Wagner. Factors Affecting the Association of Coffee Consumption With All-Cause and Cardiovascular Disease Mortality. Mayo Clinic Proceedings, 2013, 88(12): 1491–1492.

[4] C J Duncan, S Scott, What caused the Black Death?, Postgraduate Medical Journal, 2005, 81(955): 315-320.

[5] Anthony S. Fauci. Infectious Diseases: Considerations for the 21st Century, Clinical Infectious Diseases, 2001, 32(5): 675–685.

[6] Marco Ciotti, Massimo Ciccozzi, Alessandro Terrinoni, Wen-Can Jiang, Cheng-Bin Wang, Sergio Bernardini. The COVID-19 pandemic. Critical Reviews in Clinical Laboratory Sciences, 2020, 57(6): 365-388.

[7] Khanna RC, Cicinelli MV, Gilbert SS, Honavar SG, Murthy GSV. COVID-19 pandemic: Lessons learned and future directions. Indian J Ophthalmol. 2020, 68(5):703-710.

[8] COVID-19 Healthy Diet Dataset. https://www.kaggle.com/datasets/mariaren/covid19-healthy-diet-dataset?select=Fat_Supply_Quantity_Data.csv.

[9] Jia Ling. Research on Influencing Factors of Shared Bicycle Demand in Zhangye City Based on Logistic Model. China Journal of Commerce, 2020, (12): 18-20.

[10] Hastie T., et al. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, (2009).