The connection of COVID-19 infection rate to a healthy diet

Yujun Tong

Georgia State University, Atlanta, Georgia, 30303, United States

ytong3@student.gsu.edu

Abstract. In today's post-epidemic era, small-scale infections occur from time to time. While more research is focused on tackling COVID-19, before finding a way to eliminate its threat, it's even more important that ordinary people take simple steps to reduce their chances of getting infected in their daily lives. This paper focuses on the relationship between the infection rate of COVID-19 and diet and tries to find a nutrient intake way for people to prevent them from being infected by COVID-19. Through the analysis of the database, the Chi-square test, and the linear and logistical relationship between the infection rate and various nutrient intake ways are obtained. By eliminating the effects of obesity and undernourished rate, by analyzing nutrient intake ways, vegetal and animal products are two major parts of nutrient intake ways, and they are considered able to provide better resistance to infection. For more specific components, milk, meats, and cereals are beneficial for the resistance of infection.

Keywords: COVID-19 infection prevention, Nutrient intake, Diet and immune resistance.

1. Introduction

The COVID-19 outbreak had a huge influence on both human health and public health worldwide. The COVID-19 epidemic starts in Wuhan, China's Hubei Province. Subsequently, the global transmission of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) starts [1]. Humans are infected by SARS-CoV-2 through its binding to the spike protein on ACE2 [2]. Until 16 June 2024, the total number of COVID-19 cases has reached 775645882, The United States and China are the two countries that have the most cases [3]. COVID-19 not only infects humans but also brings a severe blow to the world economy, especially bringing challenges to the healthcare systems of many countries [4].

Until this point in time, COVID-19 is still affecting people's lives, both health and the economy, and this coexistence will continue for a long time. In addition, continuous attention should be paid to personal protection measures. Adopting a healthy diet is considered one of the most effective and easy preventive measures in daily life. A healthy diet is considered the best and easiest way most people can do in their normal daily lives. There are already many kinds of research about diet, some of them are about how bad nutrition, like too many fats, sugars, and refined carbohydrates can cause a lower immunity to the virus [5]. Some studies are focused on the efficacy of dietary supplements in combating COVID-19 [6]. Those papers are mostly about how overnutrition can cause a more easily infectious COVID-19. For nutrient intake from the normal healthy diet food, the region where people live can highly influence the nutrient intake, even just inside European countries, nutrient intake can be very different [7]. Despite globalization, eating habits are still very different from country to country and

continent to continent. So, by dividing the countries by their continents, reasonable advice on better nutrient intake ways to prevent different region's populations from infection of COVID-19 can be given.

This article is mainly based on the post-epidemic background, the coexistence of COVID-19 and human beings, also the weak global economy. In today's post-pandemic era, how to make people more resistant to COVID-19 through a more cost-effective way is becoming more and more important, it is also important that the world can return to normal development line faster, avoid as much as possible the continuous impact of COVID-19. Since COVID-19 has hit the economies of many countries, the countries with the higher GDP and the countries that have seen faster economic recovery after COVID-19 are considered to have maintained a consistent domestic standard of living, the influence of COVID-19 on normal nutrition intake of that country is lower [8].

This article hopes to explore a useful way for ordinary people to strengthen their resistance to COVID-19 by simply adjusting their eating habits, and if a beneficial change can be done, what should people do. This research will find a better nutrient composition prevent to being easily infected by COVID-19. It will also attempt to provide a healthy diet based on dietary standards for Chinese [9] and Americans [10].

2. Method

2.1. Dataset introduction

For this research, this paper mainly uses the COVID-19 healthy diet dataset from Kaggle. This dataset is about the basic nutrient intake of the different countries, the sample number is 170. It focuses on the fat, food, and protein supply quantity. This dataset shows the details of the percentage of how the population of different countries intake three kinds of supplies. The supply intake is divided into 23 kinds of different ways. This paper will focus on a few of the main nutrient intake ways that account for the high proportion, containing animal fats, animal products, cereals, meat, milk, vegetables, and vegetal products.

Fish body oil, liver oil, butter, ghee, cream, and raw animal fats are typical examples of animal fats. Eggs, milk, mollusks, cream, crustaceans, cream, cephalopods, meat, butter, ghee, and aquatic animals and plants are all considered animal products. The milk is just milk; butter is not included. The cereals are made up of barley and its products, sorghum and its products, rye and its products, maize and its products, millet and its products, oats, rice, and wheat and its products. Meat from cattle, mutton, goats, pigs, and fowl make up the meat. Vegetable oils, fruits, vegetables, non-foods, and alcohol are samples of vegetable products. Those few kinds of data have proportion in the three kinds of supplies, they contained almost all the basic foods people intake in their daily lives.

This dataset also contains the obesity, undernourished rate, and the data of confirmed, death, recovered, and active cases of COVID-19.

2.2. Data analysis methods of the dataset

2.2.1. Descriptive statistical analysis. To compare how the different supplies and nutrient intake ways affect the not-confirmed case of COVID-19, the arithmetic mean, median, and standard deviation of the main nutrient intake ways of percentage is calculated. An analysis of variance can show whether there is a big difference in the continents' nutrient intake ways. The high proportion of nutrient intake ways data were used. For this research, Select the top three GDP-producing nations in each of the following continents: North America (US, Canada, Mexico), Europe (Germany, UK, and France), South America (Brazil, Argentina, and Colombia), Asia (China, Japan, and India), and Africa (Nigeria, South Africa, Egypt). Because there aren't enough countries in Oceania, don't compare this.

2.2.2. Chi-square test. The chi-square test is mostly used to ascertain if the two categorical variables are independent or associated. In the paper, the Chi-square test is used to see if there are significant differences for countries between cases of COVID-19(rate of confirmed/not confirmed) and

obesity/undernourished rate. As there are no standards for those rates, in this paper, the data with an obesity rate lower than 20% and an undernourished rate lower than 10% are considered to have a low obesity/undernourished rate, this is given by the mean and median of the samples' obesity and undernourished rate. The case of COVID-19 is defined by the not-confirmed rate. As no standards for this also, by the mean and median of the rate of not confirmed cases, the country with a rate more than 99.5% is considered to have a low infection. These statistical analyses were done using the SPSS. By eliminating these additional factors, the result can be more reliable.

2.2.3. Regression analysis. For the following regression analysis, the confirmed case rate of COVID-19 is used. A continuous model describing the connection between a dependent variable and one or more independent variables is constructed using linear regression analysis. This simple linear regression focuses on the relationship between the not-confirmed population of COVID-19 and the high proportion of nutrient intake ways for different supplies. A multiple linear regression is used to determine all the variables related to the not-confirmed case of COVID-19. The Pearson correlation determined the relationship of the variables with the confirmed case, and the P-valve of the variables can determine their significance level to the confirmed case of COVID-19. The linear regression can show the relationship intuitively, but the data may not be linear all the time and may overfit the data.

Then, a binary model is constructed using logistical linear regression to forecast the probability of a dependent variable based on independent variables. The association between the high percentage of nutrient consumption methods and the confirmed COVID-19 case can be demonstrated using logistic regression. For the logistical regression, divide those greater or equal to 1.5% in the confirmed case into one category, and lower than 1.5% into another category. The logistical regression tests the animal products, cereals, and vegetables in the nutrient intake ways. Logistical linear regression can show the probabilities directly, it is a very efficient way of data analysis and can deal with non-linear relationships, but it is limited by only having binary outcomes. By using linear regression and logistical regression, the relationship between the not confirmed case of COVID-19 and the high proportion nutrient intake ways can be shown clearly.

3. Result

3.1. The descriptive statistical analysis result

The descriptive statistical analysis of the data set is trying to display the difference of countries and supplies in the nutrient intake. From the result of descriptive statistical analysis, animal products, and vegetal products are considered two major nutrient intake ways (Table 1). By comparing the food supply to the fat supply and protein supply, the animal products provide more fat and protein with less food supply rate than the vegetal products. Meat and milk are two important components of animal products in all three supplies. For the vegetal products, vegetables are only a small percentage, the major part is cereals. But cereals do not provide much fat supply. From the standard deviation and the result of variance analysis, the data of different nutrient intake ways and different continents all show a very difference.

The Chi-square test of the obesity and undernourished rate is to avoid the two kinds of data interfering with the result of the regression analysis. The Chi-square test result of the Obesity rate shows a significance valve lower than 0.05, which means that it would not interfere with the result of regression analysis, it has not much influence on the COVID-19 case in this dataset. But for the undernourished rate, can influence the result of regression analysis in this dataset. The standard deviation of the nutrient intake ways shows the difference between all the counties, the data of many ways in the three supplies is very large, indicating that food intake varies widely between countries, and the P-valve of the variance analysis also supports this finding. For the countries we choose from the 5 continents, for the animal product percentage of food supply, the final P-valve of the analysis of variance is 0.001, which means that the difference between the continent and the nutrient intake ways is significant at a 95% significance level.

3.2. The Chi-square analysis

Then, the chi-square test of obesity and the case of COVID-19 have a P-valve lower than 0.05, considered no significant difference at 95% significance level (Table 2). For the undernourished and the case of COVID-19, it has a P-valve less than 0.05, so it is a significant difference at a 95% significance level. For the countries with more confirmed cases of COVID-19, the undernourished rates are lower (Table 3).

3.3. The two different regression analysis

For the linear regression analysis, for animal products-related intake ways (animal product, animal fat, meat, milk), it has a P-valve lower than 0.05 at all 3 supplies, considered to have a significant influence on the case of COVID-19 at a 95% significance level. For vegetal products (vegetal products, cereals, and vegetables have a P-valve higher than 0.05 in all 3 supplies and have no significant influence on the case of COVID-19 at a 95% significance level. Between all the nutrient intake ways and the supplies, animal products have the highest personal correlation to fat supplies. For the multiple linear regression, they all have a P-valve lower than 0.05, which means the food, fat, and protein intake can have a huge impact on the case of COVID-19 infection (Table 4). For the logistical linear regression, the correct percentage change from 57.3% to 72.6%, can much butter predict the case of COVID-19 by this model, and for the fat supply, the P-valve of cereals and animal products is lower than 0.05, vegetables are higher than 0.05, shows no significant influence, same result as the linear regression. For the food supply, all three ways have a P-valve lower than 0.05. For the protein supply, all three ways have a P-valve lower than 0.05, but the cereal already has a P-valve close to 0.05 (Table 5).

Table 1. The descriptive statistical analysis of three supplies

Food Supply	Animal Products	Animal fats	Cereals	Meat	Milk	Vegetal Products	Vegetables
Mean	12.1819	0.2211	11.8003	3.3759	6.5198	37.8148	6.0859
Standard Deviation	5.8526	0.2783	5.8249	1.7629	5.0204	5.8529	3.5601
Median	12.0976	0.1169	10.1428	3.4248	5.3369	37.9007	5.0281
Fat Supply							
Mean	20.6957	4.1385	4.3765	9.4943	5.1091	29.3044	0.3095
Standard Deviation	8.0027	3.2875	3.1838	4.6668	3.3317	8.0024	0.2118
Median	20.9431	3.3124	3.3068	9.2590	4.9476	29.0606	0.2475
Protein Supply							
Mean	21.2322	0.1082	19.0249	9.9030	5.9212	28.7682	1.7366
Standard Deviation	7.9218	0.1504	6.4504	4.6891	3.9270	7.9217	1.0029
Median	21.8530	0.0563	17.5689	10.2129	5.6024	28.1526	1.5025

Table 2. The Chi-square test of Obesity rate and Unconfirmed cases of COVID-19

	value	degree of freedom	Progressive significance	Precise significance	Precise significance
Pearson Chi-square	2.761	1	0.097		
Continuity Correction B	2.244	1	0.134		
Likelihood ratio	2.779	1	0.096		
Fisher precision test				0.105	0.067
Number of COVID-19 cases	156				

Table 3. The Chi-square test of Undernourished rate and Unconfirmed cases of COVID-19

	value	degree of freedom	Progressive significance	Precise significance	Precise significance
Pearson Chi-square	30.849	1	0		
Continuity Correction B	28.959	1	0		
Likelihood ratio	31.031	1	0		
Fisher precision test				0	0
Number of COVID-19 cases	155				

Table 4. The Linear regression analysis result of three supplies

Fat Supply	F	P-valve	Food Supply	F	P-valve	Protein Supply	F	P-valve
Animal Product	49.066	0	Animal Product	87.987	0	Animal Product	67.715	0
Animal Fats	57.355	0	Animal Fats	50.365	0	Animal Fats	31.941	0
Cereals	34.518	0	Cereals	39.541	0	Cereals	30.843	0
Meat	4.012	0.047	Meat	15.529	0	Meat	30.386	0
Milk	33.676	0	Milk	83.39	0	Milk	105.077	0
Vegetal Products	49.077	0	Vegetal Products	87.99	0	Vegetal Products	67.701	0
Vegetables	0.005	0.945	Vegetables	2.582	0.11	Vegetables	1.493	0.223
Cereals, Animal Products, Vegetables	20.97	0	Cereals, Animal Products, Vegetables	32.502	0	Cereals, Animal Products, Vegetables	23.767	0

Table 5. The Logistic regression analysis result of three supplies

Fat Supply	Standard error	P-valve	Food Supply	Standard error	P-valve	Protein Supply	Standard error	P-valve
Cereals	0.0923	0.0032	Cereals	0.0440	0.0000	Cereals	0.0719	0.0001
Animal Product	0.0285	0.0002	Animal Product	0.0483	0.0215	Animal Product	0.0776	0.0456
Vegetables	0.9101	0.5289	Vegetables	0.0536	0.0154	Vegetables	0.1926	0.0234

4. Discussion

From the result of descriptive statistical analysis, Chi-square analysis, and the two-regression analysis, animal products, and cereals are two major nutrient intake ways that show more influence on the case of COVID-19.

This result is considered reasonable, maybe due to the following reasons. First, cereals, as the main component of vegetable products, cereals can provide adequate dietary fiber. At the same time, most cereals are processed, and various nutrients will be artificially added in the processing, and the added grains are easy to absorb, and will enhance the absorption of other nutrients [11]. Then, animal products, especially milk, are fairly common and easily available and are an important source of fat and protein intake for people. These fats and proteins provide energy for an organism's immunity, making it better able to defend itself against external risks. Therefore, for most people, it is recommended to eat more cereals and milks, for people who are lactose intolerant, they can eat more meat.

Due to the data of this paper only coming from one dataset, there are some limits to this result, only a few and major part of nutrient intake ways are discussed. Also, the infection rate of COVID-19 can be influenced by many other factors, for example, countries' epidemic prevention policies and medical standards. At the same time, the statistical data of different countries will be biased due to national conditions.

5. Conclusion

The result of descriptive statistical analysis shows that the animal and vegetal products are a major part of supplies among most of the countries. For smaller components, the milk and meats of the animal products are considered major parts, and the cereals of vegetal products are considered the major part.

The Chi-square test result of the Obesity rate shows a significance valve lower than 0.05, meaning that it would not interfere with the result of regression analysis, it has not much influence on the COVID-19 case in this dataset. But for the undernourished rate, can influence the result of regression analysis in this dataset.

The result of the linear regression shows that most of the nutrient intake ways have an influence on it, and the nutrient intake ways with larger F valves have a higher relation to the COVID-19 cases. So, from the three supplies, the animal products show a huge influence, and the milk of the animal products is the one with the biggest impact. The cereals of vegetal products also have a very strong influence on the COVID-19 cases. For the logistical regression analysis, the result is also just like the result of the linear regression, confirming the influence of animal products and cereals again. So, to all those results, the cereals of vegetal products and the milk and meats of animal products can help prevent infection of COVID-19.

This research can help to identify which kinds of nutrients can help resist viral infection, more cereals, milk and meats intake can be added to the daily food intake. It can be used as a reference for the improvement of dietary guidelines and self-daily food intake in the current context. At the same time, it can also help people with mild cases respond faster.

Of course, this study still has limitations due to the small amount of data and the difficulty of excluding more interference, which would be more useful if we could eliminate this interference and explore more detailed nutrient components, such as the role of various vitamins or specific proteins.

References

[1] Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W. C., Wang, C. B., & Bernardini, S. (2020). The COVID-19 pandemic. Critical Reviews in Clinical Laboratory Sciences, 57(6), 365–388.

[2] He, J., Tao, H., Yan, Y., Huang, S.-Y., & Xiao, Y. (2020). Molecular mechanism of evolution and human infection with SARS-CoV-2. Viruses, 12(4), 428.

[3] World Health Organization. (2023). WHO Coronavirus (COVID-19) dashboard > Cases. Retrieved from https://data.who.int/dashboards/covid19/cases

[4] Kaye, A. D., Okeagu, C. N., Pham, A. D., Silva, R. A., Hurley, J. J., Arron, B. L., Sarfraz, N., Lee, H. N., Ghali, G. E., Gamble, J. W., Liu, H., Urman, R. D., & Cornett, E. M. (2021). Economic impact of COVID-19 pandemic on healthcare facilities and systems: International perspectives. Best Practice & Research Clinical Anaesthesiology, 35(3), 293-306.

[5] Wang, S. S., Lay, S., Yu, H. N., & Shen, S. R. (2016). Dietary guidelines for Chinese residents (2016): Comments and comparisons. Journal of Zhejiang University. Science. B, 17(9), 649–656.

[6] Dietary Guidelines Advisory Committee, HHS, Office of Disease Prevention, Health Promotion (US), USDA, & Center for Nutrition Policy Promotion (US). (2015). Dietary guidelines for Americans 2015-2020. Government Printing Office.

[7] Pavlidou, E., Poulios, E., Papadopoulou, S. K., Fasoulas, A., Dakanalis, A., & Giaginis, C. (2024). Clinical evidence on the potential beneficial effects of diet and dietary supplements against COVID-19 infection risk and symptoms' severity. Medical Sciences, 12(1), 11.

[8] Butler, M. J., & Barrientos, R. M. (2020). The impact of nutrition on COVID-19 susceptibility and long-term consequences. Brain, Behavior, and Immunity, 87, 53-54.

[9] Freisling, H., Fahey, M. T., Moskal, A., Ocke, M. C., Ferrari, P., Jenab, M., Norat, T., Naska, A., Welch, A. A., Navarro, C., Schulz, M., Wirfalt, E., Casagrande, C., Amiano, P., Ardanaz, E., Parr, C., Engeset, D., Grioni, S., Sera, F., Bueno-de-Mesquita, B., van der Schouw, Y. T., Touvier, M., Boutron-Ruault, M.-C., Halkjaer, J., Dahm, C. C., Khaw, K.-T., Crowe, F., Linseisen, J., Kroger, J., Huybrechts, I., Deharveng, G., Manjer, J., Agren, A., Trichopoulou,

A., Tsiotas, K., Riboli, E., Bingham, S., & Slimani, N. (2010). Region-specific nutrient intake patterns exhibit a geographical gradient within and between European countries. Journal of Nutrition, 140(7), 1280–1286.

[10] Slow, S., & amid Divergence, R. (2024). World Economic Outlook. World Economic Outlook.

[11] Baniwal, P., Mehra, R., Kumar, N., Sharma, S., & Kumar, S. (2021). Cereals: Functional constituents and its health benefits. Pharma Innovation, 10(2), 343-349.