Diabetes, Obesity, Physical Inactivity and COVID-19 Confirmed Cases, Deaths Correlations Across the United States

Wenzhong Liu[1] and Kaiyang Liu[#]

[1]Bellevue High School, Bellevue, WA, USA
[#]Advisor

ABSTRACT

People with certain medical conditions are more likely to get very sick when infected by COVID-19. We examined the correlation between COVID-19 cases, deaths and three common medical conditions including diabetes, obesity, and physical inactivity at county level. We discovered that there are indeed significant, weak to moderate positive correlations between those three common medical conditions and COVID-19 cases and deaths. We strongly recommend maintaining a reasonable physical activity level to reduce the risk of getting very sick when infected by COVID-19.

Introduction

CDC compiled a list of specific medical conditions and informed the public on their website [1] that people with one or more medical conditions on that list are more likely to get very sick when infected by COVID-19, i.e., hospitalized or die. Those medical conditions were not listed in the order of risk. We wanted to understand more on the "likely" part.

Here we studied three common medical conditions in that list: diabetes, obesity, and physical inactivity. We examined how strongly they are associated with COVID-19 mortality across the United States by presenting quantitative data and expanded that to examine the association with infection.

Methods

We extracted five pieces of data from three public data sources. All data is sliced by US county. There are 3221 counties in the US according to US census data.

We downloaded population data for April of 2020 from the US census website [2]. We downloaded COVID-19 confirmed cases and deaths data from the New York Times GitHub site [3]. The COVID-19 data is cumulative data till 2021/2/28 from the end of 2019 when COVID-19 outbreak in the US first started. The CDC website [4] provided us with the diabetes, obesity, and physical inactivity data for 2018. The data is presented as the percentage of people having the specific medical condition over the total population in the corresponding county.

We removed COVID-19 data of counties for New York City and Kansas City which are not aligned with the county definition in both US Census and CDC data sources. We calculated percentages of COVID-19 confirmed cases and confirmed deaths by dividing the affected population with the total population of each county recorded by the 2020 census. Then we joined all the data into a single virtual table by the county FIPS codes [5] included in each data file.

Pearson product-moment correlation coefficient was calculated and used to measure the relationships between medical conditions to confirmed cases and deaths respectively.

In the US, there are many small counties. Out of 3221 counties, there are 1004 counties with more than 50,000 people, but the population numbers of those 1004 counties account for 87.5% of total population. Since there is one data entry for each county, to avoid small counties being overrepresented, we only took counties with population numbers greater than 50,000 into account when calculating the correlation data to make sure we get a more meaningful result.

The data was processed, and correlation coefficients were calculated using Python. Source code on GitHub: https://github.com/jonnyliu2014/Covid-Correlations.

Results

Raw COVID-19 data from New York Times has 3246 counties. After cleaning up, we joined it with population data, diabetes, obesity, and physical inactivity data, removed rows with missing values, and 3131 counties left.

Table 1. Correlations between medical conditions and COVID-19 cases, deaths by US county, included 985 counties with population numbers greater than 50,000.

	Physical Inactivity %s		Diabetes %s		Obesity %s	
	r	*p-value*	*r*	*p-value*	*r*	*p-value*
confirmed cases %s	0.3	**<0.0001**	0.224	**<0.0001**	0.173	**<0.0001**
confirmed deaths %s	0.406	**<0.0001**	0.299	**<0.0001**	0.202	**<0.0001**

r is the Pearson correlation coefficient. It ranges from -1 to 1 with higher absolute value being stronger correlation.

All p-values in boldface are statistically significant at below 0.05 level.

From the results shown in Table 1, we concluded that all three medical conditions have positive correlation with COVID-19 cases percentages and deaths percentages. Since all *p-value*s are less than 0.0001, the result is very conclusive. The correlations here are all in the range roughly between 0.2 to 0.4, which is considered as weak to moderate.

Physical inactivity %s have stronger correlation than diabetes %s and obesity %s. The confirmed deaths %s have stronger correlation than confirmed cases %s.

Discussion

Regarding physical inactivity %s have stronger correlation than diabetes %s and obesity %s, a sensible explanation of this could be that most elders fall into the category of physical inactivity. A research conducted in 2020 [6] shows that "adults over 65 years of age represent 80% of hospitalizations and have a 23-fold greater risk of death than those under 65.".

Overall, the confirmed deaths %s have stronger correlation than confirmed cases %s. So, to avoid severe illness or fatality, people with medical conditions are recommended to take prevention actions, such as taking vaccines.

Limitations

The CDC data about the medical conditions data by county used in this study may not have high accuracy. Also, the CDC medical conditions data were estimates from 2018. This is 2 years prior to the other data; census and COVID-19 data are both from 2020 and 2021.

Conclusion and Future Studies

Results from this study show a moderate positive, but very significant, correlation between physical inactivity and COVID cases and deaths, and a weak positive correlation between diabetes, obesity and COVID-19 cases, deaths. The average U.S. citizen needs to work out more to reduce the risk of severe Covid-19 infection. More opportunities of exercise should be provided during the lockdown.

The COVID-19 social distance policy could cause physical inactivity percentage in the general population to increase. Will this leave people more vulnerable? Workout at home may not be the same as workout outdoors or in gyms. Future studies could be conducted about the trade-off between increasing physical inactivity and decreasing close contact.

Acknowledgments

References

[1] Centers for Disease Control and Prevention. (n.d.). *People with certain medical conditions*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html

[2] US Census Bureau. (2021, October 8). County Population by Characteristics: 2010–2019. Census.Gov. https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html

[3] Nytimes. (n.d.). *Nytimes/COVID-19-data: An ongoing repository of data on coronavirus cases and deaths in the U.S.* GitHub. https://github.com/nytimes/covid-19-data

[4] Centers for Disease Control and Prevention. (n.d.). *Social Determinants of Health - United States diabetes surveillance system*. Centers for Disease Control and Prevention. https://gis.cdc.gov/grasp/diabetes/diabetesatlas-sdoh.html

[5] NRCS. (n.d.). *County FIPS Codes*. Natural Resources Conservation Service. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/nra/nri/results/?cid=nrcs143_013697

[6] Mueller AL, McNamara MS, Sinclair DA. Why does COVID-19 disproportionately affect older people?. Aging (Albany NY). 2020 May 2912:9959-9981. https://doi.org/10.18632/aging.103344