Group Member:
Yin-Siou (Hugh) Wang, Chloe (Yangyang) Liu, Daphne Wang, Hazel (I-Chieh) Huang, Isabel Thurston, Yanmin Chen

Executive Summary

This paper explores the relationship between COVID-19 death rates and the number of fast food restaurants on county level. Fast food restaurants are categorized as limited-service restaurants – LSR – which we will be referring to throughout this project. It's the general understanding that LSR provides food that contains high amounts of calories, sugar, salt, and fats. As a result, a chronic, unhealthy diet can lead to health issues such as heart disease, diabetes, obesity, and other chronic diseases. Studies conducted by the CDC have shown that people who have underlying health issues are at a higher risk of dying from COVID-19 (Kompaniyets, 2021). Our hypothesis is that counties with a higher LSR density have higher COVID-19 death rates. Through our regression analysis we will investigate this relationship and see if we could reject the null hypothesis, which is that there is no relationship between the density of LSR in a county and its death rate involving COVID-19. We will use a data-driven approach to obtain insights, examine the data we collect and interpret it in order to determine the relationship and construct recommendation upon it.

For the ideal experiment, we randomly divided all counties into two groups — counties with LSR opening and counties with LSR closed. Therefore, we need to control other variables such as SVI, median household income, population density, vaccination rate, etc. With all control variables remaining the same, we can see how numbers of limited-service restaurants affect the COVID death rate.

Merging eight datasets on a county level from various credible sources including U.S. Census Bureau and CDC, we took the liberty of building a dataset from scratch. Not only did we gather variables that are considered factors associated with death involving COVID-19 after a thorough investigation, but we also created essential variables based on our findings to cater to a more realistic scenario. LSR_per_1000, the main independent variable in our regression analysis, refers to the number of limited-service restaurants per 1000 people; CDR_per_1000, our dependent variable, is the number of deaths involving COVID-19 divided by county population in thousands.

We, however, run into some issues during the process of data merging and data cleaning. A few things worth mentioning include the missing data for the number of LSR and the number of covid deaths. To ensure the integrity and impartiality of the data, we fill in reasonable values. We also find some irregular values in datasets, which we later determine are either an outlier or a rogue data and excluded them in our analysis. The cleaned dataset contains desired information for 3024 counties.

We correlate county-level COVID-19 death rates per 1000 people with the number of limited-service restaurants per 1000 people using multiple linear regression. We include the following control variables in our models: log-transformed median household income, log-transformed population density, the percentage of residents that are aged 65 and older, the number of ICU beds per 1000 people that are aged 65 and older, and Social Vulnerability Index, CDC/ATSDR (2020). The correlation uncovered in our models controls for other variables that we included. For example, we found a positive relationship between limited-service restaurants and COVID-19 death rates, and it is *not* driven by income or population density, since they had been accounted for. As a result, that relationship must come from some systematic factors outside our model, such as higher obesity rates or an unhealthy lifestyle. Even though we cannot control for everything, and correlation does not mean causation, we can help policy makers identify variables that could possibly cause higher COVID-19 death rates and adopt appropriate strategies for the next pandemic.

We use heteroskedasticity robust standard errors in our models and find that, on average, 1 unit increase in the number of limited-service restaurants per 1000 people is associated with a 1.4579 increase in deaths due to COVID-19 per 1000 people, with an Adjusted R-squared of 36.1%. One may speculate that it may be driven by the unhealthy lifestyle that leads to diabetes and obesity that increase COVID-19 death rates. Consequently, policy makers may reduce the number of service restaurants and offer people incentives to consume healthier food in order to reduce the impact of the next pandemic and the costs of healthcare services and government support after investigating further and establishing a causal relationship.

We investigate the relationship further by examining whether modification exists with other variables. A variable modifies a relationship when it depends on that variable. Limited-service restaurants may have different effects on countries depending on their socio-economic status, for example, a wealthier country may have more access to healthcare, even if its residents are regular limited-service restaurant visitors, they are more flexible when the pandemic hits. This informs us that we can focus on specific areas to reduce the death rates more efficiently, since LSR is more relevant in certain counties.

To test whether other variables influence the effect of the main variable LSR density on COVID-19 death rate, we add interaction terms of LSR with income level and population density separately and see whether they are statistically significant. We find that the effect of limited-service restaurants is stronger for low income and high population communities. The reason could be in relatively poor counties, people have higher access to fast food restaurants and unhealthy food, which could be related to causes of diseases. Furthermore, in terms of population density, residents living in crowded households have larger social networks and could be exposed to diseases more in public, which increases the effect of LSR on COVID-19 death rates.

As an important control variable in the regression model, SVI contains four aspects – socioeconomic status, household characteristic, racial & ethnic minority status and housing type & transportation. It is not significant in the original model, and we decompose it into four aspects to see how each one relates to the COVID-19 death rates. We drop socioeconomic status since it is highly correlated with and contains similar information as income in our model. Of the four aspects, housing type & transportation is more relevant to the policy makers. As a result, we interact it with LSR and find that the effect of LSR is significantly stronger for communities that are more vulnerable in terms of housing type & transportation. In addition, racial & ethnic minority status has a significant p-value with negative impact on COVID-19 death rate. A possible explanation could be that after controlling for income, people in culturally-diverse regions may tend to be more receptive to change, such as the mandatory mask rules.

Through the above data analysis, we can conclude that the higher LSR per capita is associated with higher COVID-19 death rates, and the effect is modified and increased by the high population density, low income, and more vulnerable households and transportation. Overall, policy makers could increase the control over limited-service restaurants and promote healthier lifestyles in the areas of low income, populated households and to reduce the impact of the next pandemic. During the future disease outbreak or flu season, the government can take these factors into consideration to mitigate the negative effect of diseases.

For limitations, “the death counts are not tabulated by the decedants’ state of residence” according to CDC (U.S. Department of Health & Human Services, 2022), which affects the calculation of the COVID-19 death rate. A more concrete analysis could be drawn if we have access to time series data, government efficiency, and patient-level data.

Work Cited

Christopher R. Knittel, Bora Ozaltun (Working Paper, June 2020). What Does and Does Not Correlate With Covid-19 Death Rates, from https://www.nber.org/system/files/working_papers/w27391/w27391.pdf

Kompaniyets, L. (2021, July 1). *Underlying Medical Conditions and Severe Illness Among 540,667 Adults Hospitalized With COVID-19*. CDC. Retrieved December 2, 2022, from https://www.cdc.gov/pcd/issues/2021/21_0123.htm

U.S. Department of Health & Human Services (2022, November 16). *CDC/ATSDR Social Vulnerability Index*. ATSDR Agency for Toxic Substances and Disease Registry. Retrieved December 2, 2022, from https://www.atsdr.cdc.gov/placeandhealth/svi/index.html#:~:text=Social%20vulnerability%20refers%20to%20the,human%20suffering%20and%20economic%20loss.