

Problem Statement Formation

How are disease and nutritional factors related to the incidence of COVID-19 and what factors can be mitigated to reduce the viral outbreak?

Context

COVID-19 is a coronavirus outbreak that began in late 2019. The virus has since spread across the entire world and with no effective antiviral nor vaccination, people have relied on social distancing to slow the spread of the virus and hopes of eventually reaching herd immunity. The consequences of social distancing (job and/or income loss, trouble affording rent or other essentials, economic crash) have made this an untenable option for many. Moreover, the hopes for herd immunity fade as new mutations and recurrent outbreaks continue. There is also an increased prevalence of the virus in certain populations (those with preexisting conditions and those with nutrient deficiencies). It follows that the incidence of COVID-19 could both be predicted by the rate of certain diseases and nutritional factors and perhaps mitigated by acting to improve the condition of those with existing disease/nutrition factors.

Criteria for Success

Prediction of the incidence of COVID-19 within a 10% margin of the actual value will qualify as a successful outcome. Deliverables will include a condensed and cleaned dataset created from the data sources, code for each step of the analysis process, and a final slide deck and report.

Scope of Solution Space

The focus of this study is identifying existing conditions and nutritional states that could be mitigated to prevent continued outbreaks of COVID-19.

Constraints

A major constraint of this analysis is the lack of information on social distancing measures. However, these are likely to be highly correlated with information in the nutritional determinants table (sanitation, clean water, etc). Another constraint of this analysis is that, while the considered variables may be successful at predicting COVID incidence, there may be little action that can be taken to mitigate those factors. For example, high rates of cardiovascular disease likely need to be addressed through early and consistent lifestyle changes. Thus, this analysis may prove more useful in providing key insights to address in order to prevent future pandemic events.

Stakeholders

Key insights obtained on the data from the download location. If questions about the dataset exist, search of the site and contact with the site moderator will occur to obtain more information. If clarifying information cannot be obtained, then data points will be removed from the set.

Data Sources

Portions of three sets of data are located and used as follows:

- [Annual number of deaths](#): use last 5 years and average over country for current number of deaths for each disease type relevant to potential vulnerability to COVID-19
- [Nutritional factors dataset](#): Most recent year for all of the tabs “Regional Adult” and “Country determinants” used to collect nutritional metrics and determinants related to nutrition.
- [COVID cases confirmed in each country](#): total cases confirmed per country for training and predicting.