BDP final project - Covid19 data analysis

Moti Marom, Ilan Schaedel

February 18, 2023

1 Abstrct

Background: The Covid-19 global pandemic has risen many questions about the connection between social economical status and the chances of getting infected due the pandemic and/or recover from it. The Spanish influenza(1918) outcomes showed that people of lower socioeconomic status suffered more loses which trigger the same questions about the Covid-19 that we are about to ask and answer.
Objective and methods: Our goal is to analyze data taken from the Israeli CBS and find correlations and statistical significance that might reveal interesting answers about these aspects.

2 Introduction

Due to vast medical progress in the last century which we witness as it reflected through high technology appliances integrated in the level of public hospitals and also through research and developments of medicines short time tables, one may assume that due to a relatively high availability to medical treatment a pandemic will not expose any differences that refer to socioeconomic ranks. As a result, the Null hypothesis is that there are NO socioeconomic differences regarding infection and mortality rates due to Covid-19 pandemic in Israel. Our hypothesis is that there are differences that can be revealed through correlation between socioeconomic status vs infection and mortality rates. On top of that, one can ask if residence area (rural or urban) is also correlated to those rates. In Israel, big health centers and hospitals are usually located in or near Urban areas due to necessity and preferences, so living in vicinity of those centers might influence on the survival rates. Israel, was also a vaccination pioneer. Thanks to its social medical system every citizen had access to 3 vaccination rounds.
We will also examine correlation after every vaccination round.
To verify our hypothesis we grabbed the following data:
1. Aggregated data of infection per town and city from the Israeli CBS.
2. Socioeconomic rank of almost every town and city in Israel.
3. Towns and cities geographical location in lat-long coordinates.

After some data cleaning and filtering we used correlation function and calculated the p-value to see if there is any significance in out findings.

3 Results

3.1 High resolution correlation between infected percentage and socioeconomic sub rank

We got a p-value that points on significance: 5.48e-42.
Correlation value: 0.352.

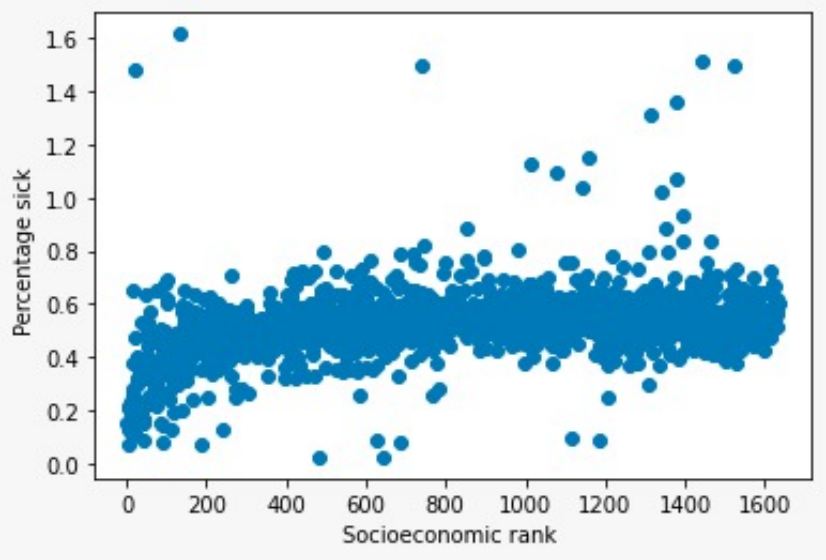

Figure 1: CORR(infected percentage , socioeconomic sub rank)

3.2 Correlation between infected percentage and socioeconomic cluster

In this section we use 10 clusters (as the number of CBS's socioeconomic ranks) to emphasize our findings:
p-value = 5.9e-47, Correlation value: 0.371.

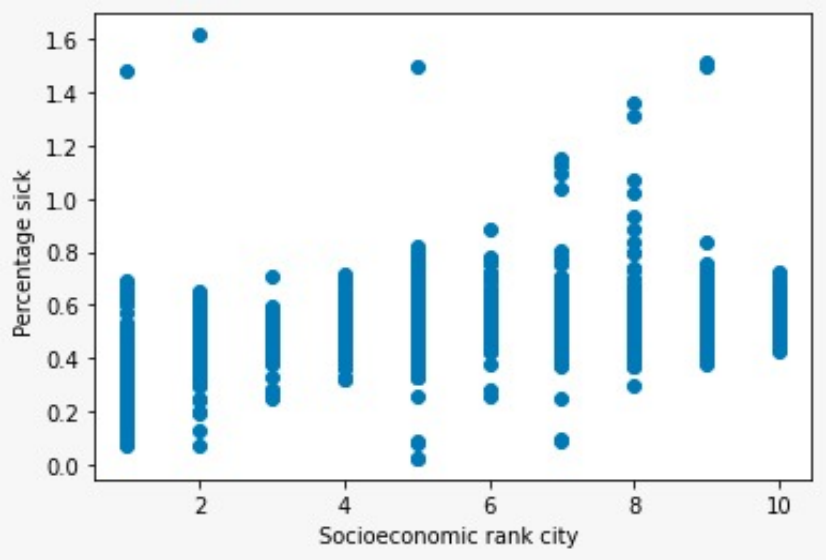

Figure 2: Clusters CORR(infected percentage , socioeconomic sub rank)

3.3 First vaccination correlation between infected percentage and socioeconomic status

We used the same display method to show the statistically significance that we got during our analysis: p-value = 1.2e-148, Correlation value: 0.63.

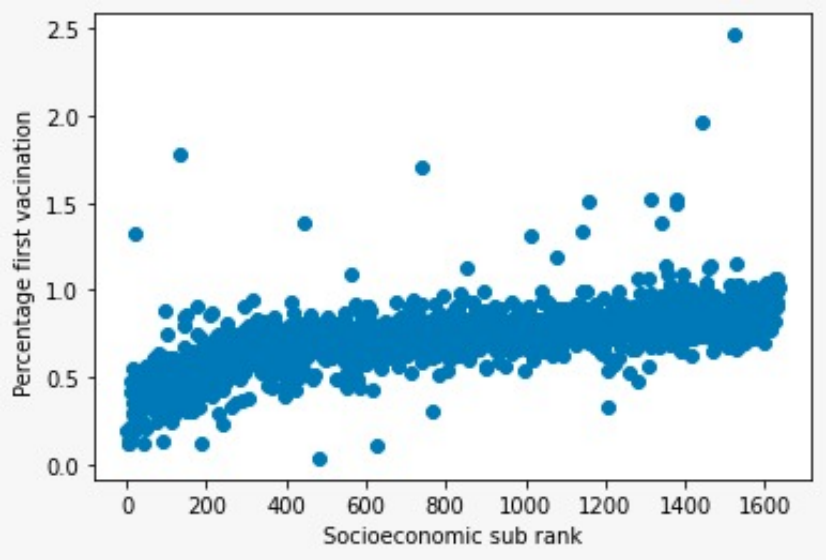

Figure 3: 1st vaccine CORR(infected percentage , socioeconomic sub rank)

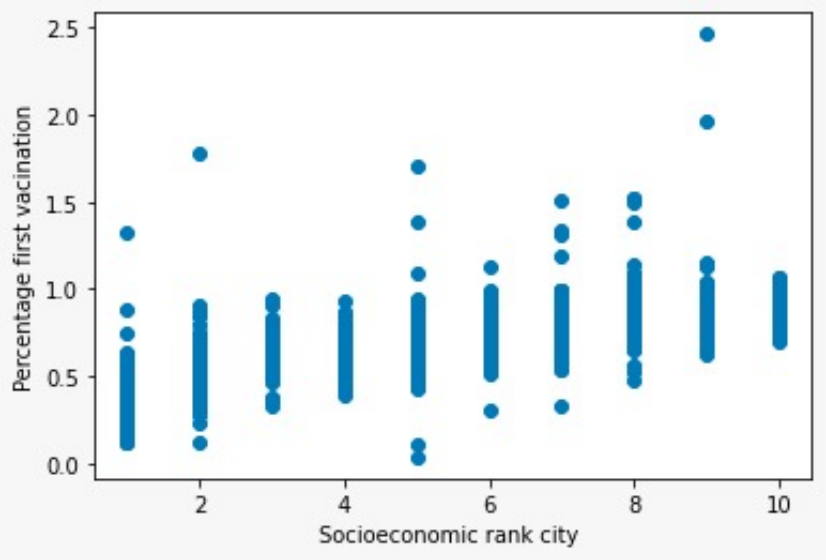

Figure 4: 1st vaccine cluster CORR(infected percentage , socioeconomic sub rank)

3.4 Second vaccination correlation between infected percentage and socioeconomic status

After the 2nd vaccination we got better correlation and statistical significance:
p-value = 2e-196, Correlation value: 0.7.

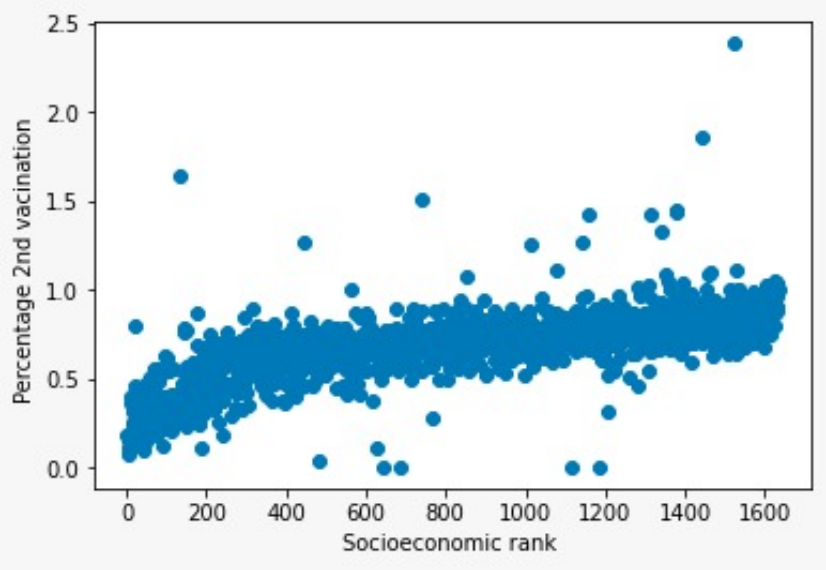

Figure 5: 2nd vaccine CORR(infected percentage , socioeconomic sub rank)

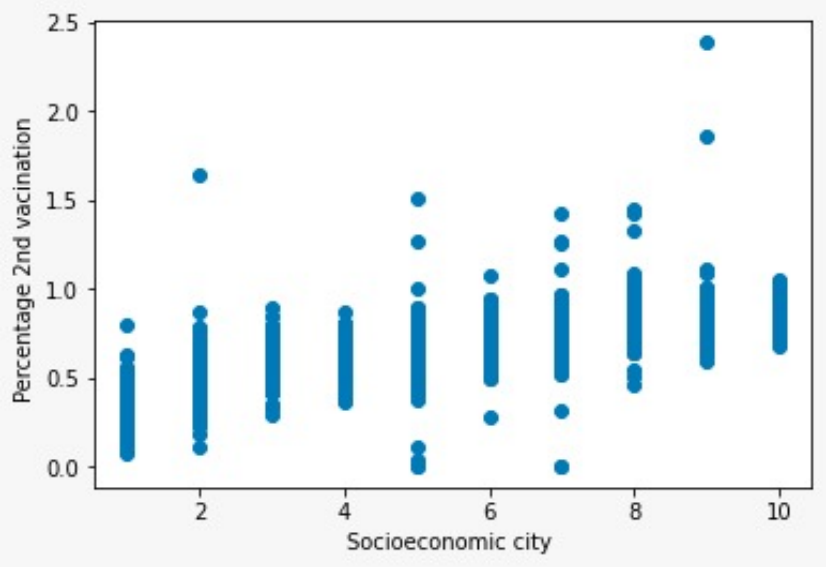

Figure 6: 2nd vaccine cluster CORR(infected percentage , socioeconomic sub rank)

3.5 Third vaccination correlation between infected percentage and socioeconomic status

The correlation and statistical significance improved further after the 3rd vaccination: p-value = 2e-270, Correlation value: 0.78.

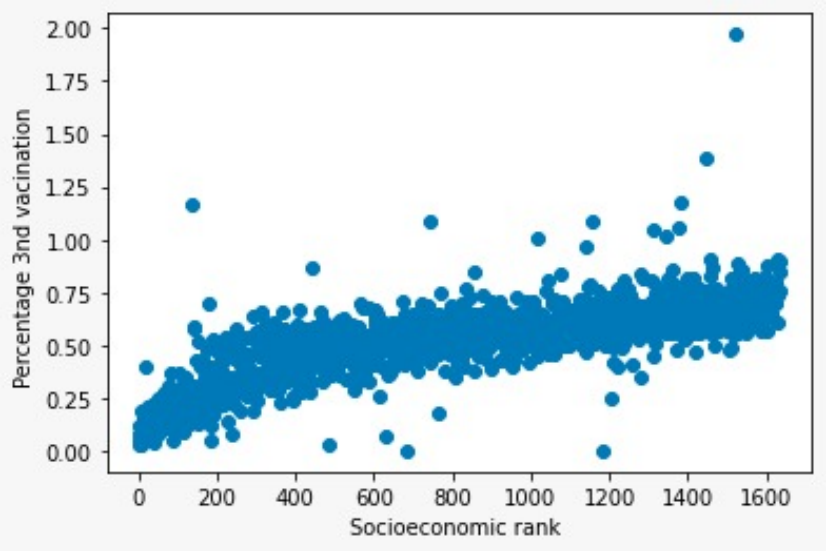

Figure 7: 3rd vaccine CORR(infected percentage , socioeconomic sub rank)

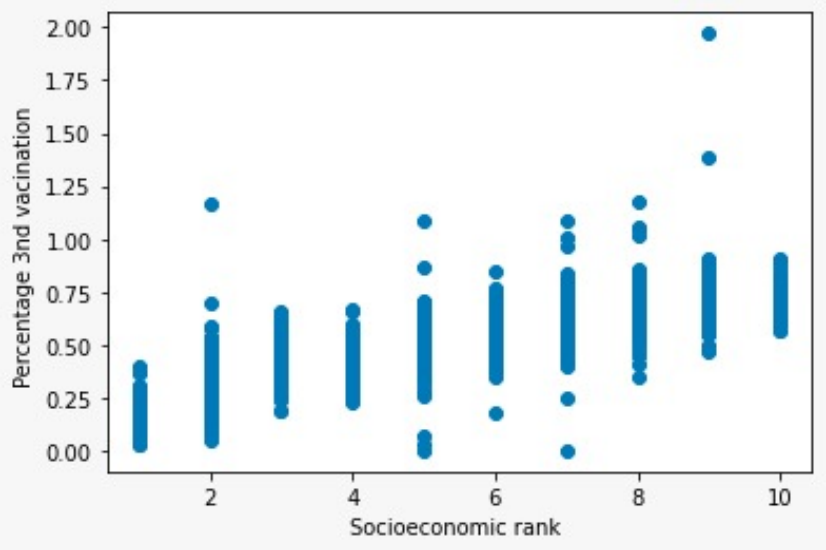

Figure 8: 3rd vaccine cluster CORR(infected percentage , socioeconomic sub rank)

3.6 Correlation between mortality percentage and socioeconomic status

Regarding mortality, the rates are less prominent but still we got statistical significance and positive correlation as you may see in the following graph(No.9):
p-value = 0.0125, Correlation value: 0.308.

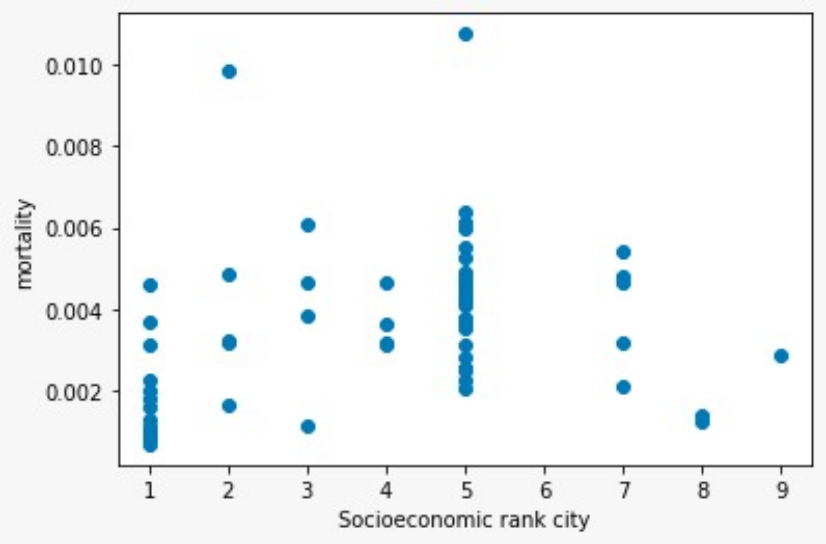

Figure 9: mortality CORR(mortality percentage , socioeconomic sub rank)

3.7 Correlation between lab testing diagnostic percentage and socioeconomic status

Regarding lab testing diagnostics, the rates are still prominent. We got statistical significance and a high positive correlation as you may see in the following graph(No.10):
p-value = 1.76e-127, Correlation value: 0.582.

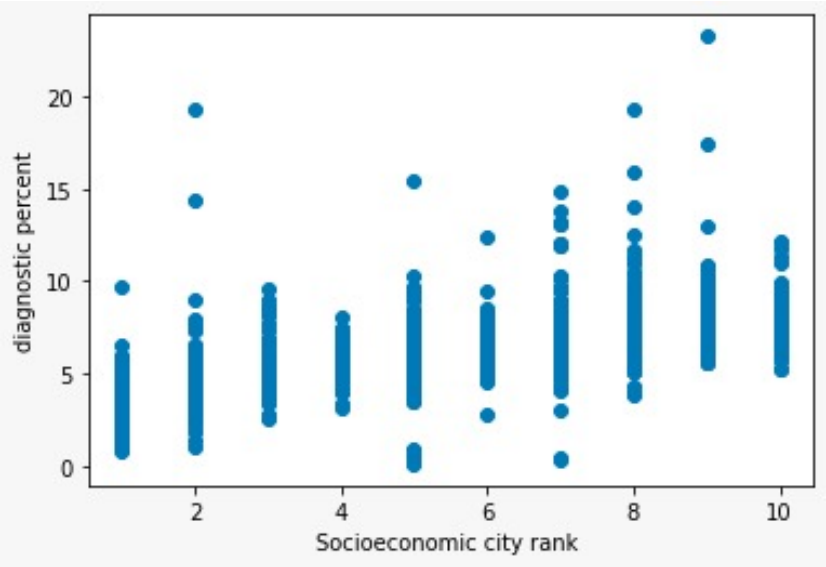

Figure 10: mortality CORR(diagnostic percentage , socioeconomic sub rank)

4 Conclusions

In order to reject the Null hypothesis we set the alpha to 0.05, hence confidence interval was 0.95 throughout all the analysis. We got prominent correlation rates and statistical significance in all aspects which lead to the rejection of the null hypothesis. We can conclude that socioeconomic ranks were reflected in infection and mortality rates during the Covid-19 pandemic in Israel in spite of the the advanced Israeli public health system and 3 vaccination rounds.

References

1. Liao, T. F., 2021. Social and economic inequality in coronavirus disease 2019 vaccination coverage across Illinois counties.
2. Mathieu, E., 2021. A global database of COVID-19 vaccinations.
3. Gustafsson, P. E., 2022. Inequitable impact of infection: social gradients in severe COVID-19 outcomes among all confirmed SARS-CoV-2 cases during the first pandemic wave in Sweden.
4. Samany, N. N., 2021. The most places at risk surrounding the COVID-19 treatment hospitals in an urban environment- case study: Tehran city.
5. Ergo€nul, €., 2020. National case fatality rates of the COVID-19 pandemic.