Capstone Project for CSCI 1820

"The Significance of Regional MMR Vaccination Coverage in COVID-19 Disease Burden"

Beril Erdogdu

Professor: Sorin Istrail

1 Abstract

Many recent pre-print studies regarding the COVID-19 disease focus on investigating potential correlations between childhood vaccination and COVID-19 disease burden. It is noteworthy that the disease has its greatest impact on the elderly. Therefore, it is assumed that childhood vaccination could potentially confer acquired protection over SARS-CoV-2. For the purposes of this project, I have focused on the paper by Young et al., *"Homologous protein domains in SARS-CoV-2 and measles, mumps and rubella viruses: preliminary evidence that MMR vaccine might provide protection against COVID-19"*. The object of my project is to elaborate on the genomic and structural referencing they have conducted on MMR viruses and SARS-CoV-2, and further investigate their methods of epidemiological data analysis by comparing patient data from 20 countries with varying vaccination coverage histories.

2 Introduction

The striking feature of COVID-19 is that it has a disproportionate impact on the elderly population, yet the paediatric populations in many reports escape with mild, if any, symptoms (Ferguson et al., 2020). Recent pre-print article *"Homologous protein domains in SARS-CoV-2 and measles, mumps and*

rubella viruses: preliminary evidence that MMR vaccine might provide protection against COVID-19" by Young et al. suggests a link between MMR vaccination -which was more readily available to the younger generations than the elderly- and the severity of COVID-19 symptoms suffered by the patient.

After separate outbreaks of large pandemics of measles, mumps and rubella in Europe around early 1960s, individual vaccines were developed for each of the viruses. Around 1970s, the individual vaccines were combined into a triple vaccine for infants. There were, however, wide differences between national health systems and their timings for introducing the triple MMR vaccine in their standard childhood vaccination schedules (Hamborsky et al., 2015). If the claims by Young et al. were true, we would expect this to cause the elderly populations to have a better chance of survival in the countries where the MMR vaccine was introduced earlier in comparison to other countries.

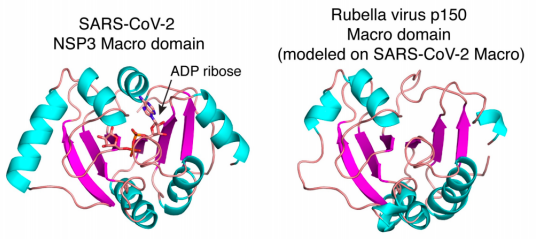
To assess whether or not MMR vaccination coverage is correlated with the outcome patterns in different countires, I have collected COVID-19 patient data from 20 different European countries that were made publicly available by European Center for Disease Prevention and Control. The data consisted of "reporting country", "onset week", "age group", "gender", "hospitalisation", and "outcome" for each tracked COVID-19 patient starting from December 2019 until now.

Furthermore, the records of what percentages of the infant population were vaccinated at each country between the years 1980 - 2018 were collected from World Health Organization's public datasets. The survival risk analysis model that was fit on the patient data combined with regional vaccination coverage history revealed no statistically significant correlation between the MMR vaccine and COVID-19 disease burden amongst the elderly. The details of the model will be presented in section 5, after the discussion of genomic investigation by Young et al. for potential homology between SARS-CoV-2 and measles and rubella in sections 3 and 4.

3 Background: SARS-CoV-2 and Rubella

Young et al. used PSI-BLAST (Altschul et al., 1997) to align sequences of SARS-CoV-2 and the rubella viruses to discover that their Macro domains share 29% amino acid sequence identity. Based on these findings, they generated an atomic model of the rubella virus Macro domain by threading its structure onto the

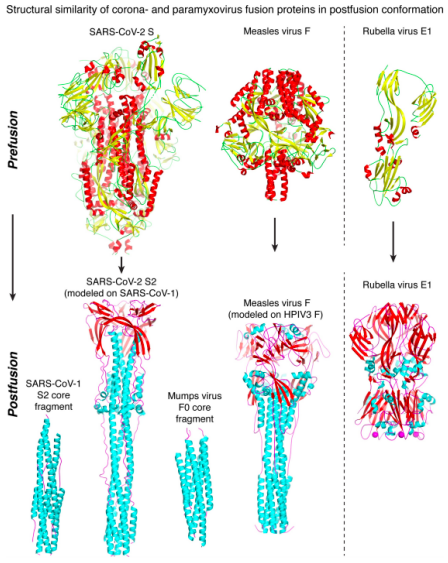
model of SARS-CoV-2 Macro domain. The figure below reveals the similarities between their Macro domain structures.



Note that the marked Macro domain on the right is present n the triple MMR vaccine.

4 Background: SARS-CoV-2 and Measles

Viral envelope glycoproteins play a very important role in cell entry, which is the fusion of the viral and the cellular membrane. It was discovered in previous studies that paramyxoviruses (including measles) share a "a-helical-coiled-coil" structure in their Class I viral membrane fusion proteins. Young et al. disvovered structural homology between the glycoproteins of the measles virus and the glycoproteins of SARS-CoV-2 which are also Class I fusion proteins. The figure below reveals the similarities between the post-fusion structures of the two virues' glycoproteins.



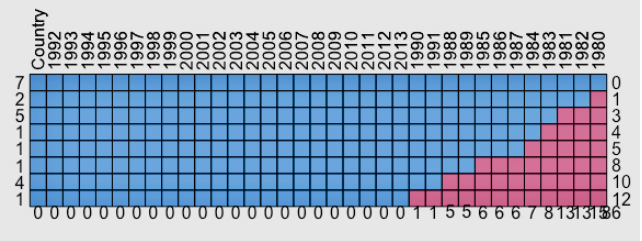
5 Methodology for the Cox Proportional Hazards Survival Analysis

For the epidemiological data investigations, Young et al. chose to gather data from the three European countries with the highest number of reported COVID-19 cases. Their analysis showed that the vaccination coverage percentage in these countries were much lower when the "elderly" age groups were infants. Based on these observations, they concluded that there could be a causal relationship between the higher fatality risks of these age groups and their lack of access to MMR vaccination in the past.

As convincing as the correlation between the two trends sounds, it is difficult to arrive at a conclusion between two completely independent events from a population without running any statistical tests or models on the data sets. Firstly, based on the World Health Organization records, these three countries have quite similar vaccination coverage histories which disables us from comparing any effects of having lower/higher MMR vaccination coverages. Furthermore, the conclusion is arrived without considering any external factors that could be important in disease outcome.

To further investigate the potential correlation between MMR vaccination and disease burden, I have utilized data from 20 different countries with varying vaccination coverage histories. For each patient, the information consisted of "reporting country", "onset week", "age group", "gender", "hospitalisation", "outcome" + **country's vaccination coverage rate during infantry**.

The last covariate for each patient was predicted based on the country's WHO data. As expected, for earlier years the WHO database had many missing values for most countries as represented by the figure below. The missing values for each country were predicted using "time series missing values interpolation" (Moritz and Bartz-Beielstein, 2017).



For each patient, their country's vaccination coverage rate at the exact year they were born was predicted using the same imputation method to create the last covariate listed above. Finally, a Cox Proportional Hazards model was fit on the dataset to discover covariate correlation coefficients with survival probabilities.

6 Results

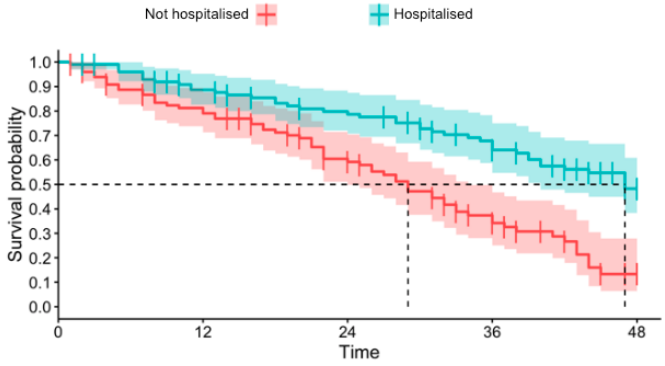
```
> cox.fit <- coxph(Surv(outcome) ~ hospitalisation + cluster(reporting_country) + vacc_cov + strata(age_group),
ties = "breslow", data = coxdata)
> summary(cox.fit)
Call:
coxph(formula = Surv(outcome) ~ hospitalisation + vacc_cov +
    strata(age_group), data = coxdata, ties = "breslow", cluster = reporting_country)

  n= 6481, number of events= 6481

                                    coef  exp(coef)   se(coef)  robust se      z Pr(>|z|)
hospitalisationNot hospitalised -0.2411337  0.7857366  0.0327904  0.0385411 -6.257 3.94e-10
hospitalisationUnknown           0.0193454  1.0195338  0.0297216  0.1055301  0.183    0.855
vacc_cov                        -0.0002152  0.9997848  0.0003988  0.0022195 -0.097    0.923

hospitalisationNot hospitalised ***
hospitalisationUnknown
vacc_cov
```

The model was fit on clustered age groups for each counry, meaning that for each age group the survival probabilities were compared to those of people from the same age group in different countries. The model results revealed that there was not a statistically significant correlation between MMR vaccination coverage and patient survival patterns. Surprisingly, the covariate "hospitalization" was discovered to be highly correlated with the outcome, suggesting that the country's ability to provide sufficient medication and intensive care to its patients impacted the survival rates of the elderly population.
Below is the associated Kaplan-Meier plot with the elderly patients split in to "hospitalized" and "not hospitalized" strata.



7 Conclusion

Based on the comparisons of the elderly population disease outcomes between countries with a wide range of vaccination coverage histories, it appear that there was not a statistically significant correlation between MMR vaccination coverage and COVID-19 disease burden. Based on the "hospitalization" covariate's high correlation coefficient and Kaplan-Meier plots, it could be hypothesized that the countries which are able to provide proper medical care for their elderly population in hospitals with intensive care units are much more likely to attenuate fatality risks. It could also be true that most of these countries are also in the set of countries which were able to introduce MMR vaccine into their standard infant vaccination schedule due to their advanced medicare policies. Since it is quite challenging to control for any such interdependencies and external factors in COVID-19 data analyses, a study using individual based data to compare MMR immunity status in the affected population is needed to conclude whether MMR vaccination can improve the outcomes from COVID-19 infection.

8 Bibliography

- Franklin, Robin, et al. "Homologous Protein Domains in SARS-CoV-2 and Measles, Mumps and Rubella Viruses: Preliminary Evidence That MMR Vaccine Might Provide Protection against COVID-19." 2020, doi:10.1101/2020.04.10.20053207.
- Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. imperial.ac.uk 2020;3–20.
- Hamborsky J, Kroger A, Wolfe S. Epidemiology and Prevention of Vaccine-Preventable Diseases. 13th ed. Washington D.C.: Centers for Disease Control and Prevention; 2015.
- Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res [Internet] 1997;25(17):3389–402.
- Moritz, Steffen, and Thomas Bartz-Beielstein. "ImputeTS: Time Series Missing Value Imputation in R." The R Journal, vol. 9, no. 1, 2017, p. 207., doi:10.32614/rj-2017-009.

- European Centre for Disease Prevention and Control COVID-19, qap.ecdc.europa.eu/public/extensions/COVID-19/COVID-19.html.
- World Health Organisation. The WHO/UNICEF Estimates of National Immunization Coverage (WUENIC) [Internet]. 2019
- "Measles." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 5 Feb. 2018, www.cdc.gov/measles/about/history.html.