IEEE Victorian Section

IEEE Student Branch
UNIVERSITY OF MELBOURNE

IDENTIFYING POTENTIAL CLUSTERS FOR COVID-19 VACCINATION IN UNITED STATES USING COUNTY-LEVEL DATA

ISDAIOT 2020

IEEE Computational Intelligence Society

IEEE IoT

Alejandro Rivera, Anish Parajuli, Fitrio Pakana, Kanishka Tamang, Nischay Thapa

RMIT University, Melbourne

1. INTRODUCTION

Introduction

COVID-19 has adversely impacted live of millions of people around the world with economic impact at its core. As of September 2020, the disease has infected around 32.7 million people with close to one million death worldwide where United States currently as the country with highest cumulative number of cases, stands at around 6.7 million cases. Fortunately, many companies have advanced to late stage testing for potential COVID-19 vaccines. However, when a vaccine is finally approved, likely many countries including United States will experience limited supply and therefore an effective, data-driven vaccine deployment strategy is essential.

Literature Review

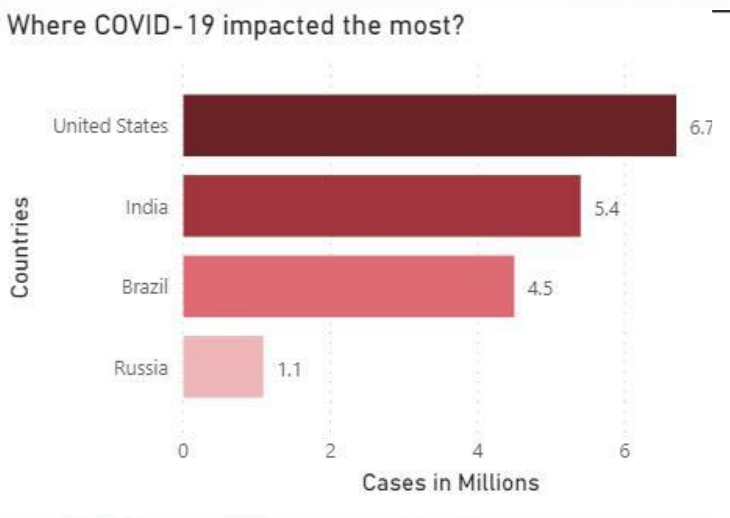
Driven by the limited supply of vaccines, it is imperative for countries to decide on prioritisation of vaccination target groups because such decision will define the deployment and vaccination activities that works best for them [1]. For influenza A (H1N1) 2009 pandemic, SAGE recommended divided population into four vaccination target groups and noted that the order of priority should be determined by each country based on country-specific conditions [7]. COVID-19 is a respiratory illness caused by a large family of viruses called coronavirus which shares similar traits with influenza in terms of the disease presentation and transmission [7] and therefore it is likely that the target groups for vaccination will be similar as well.

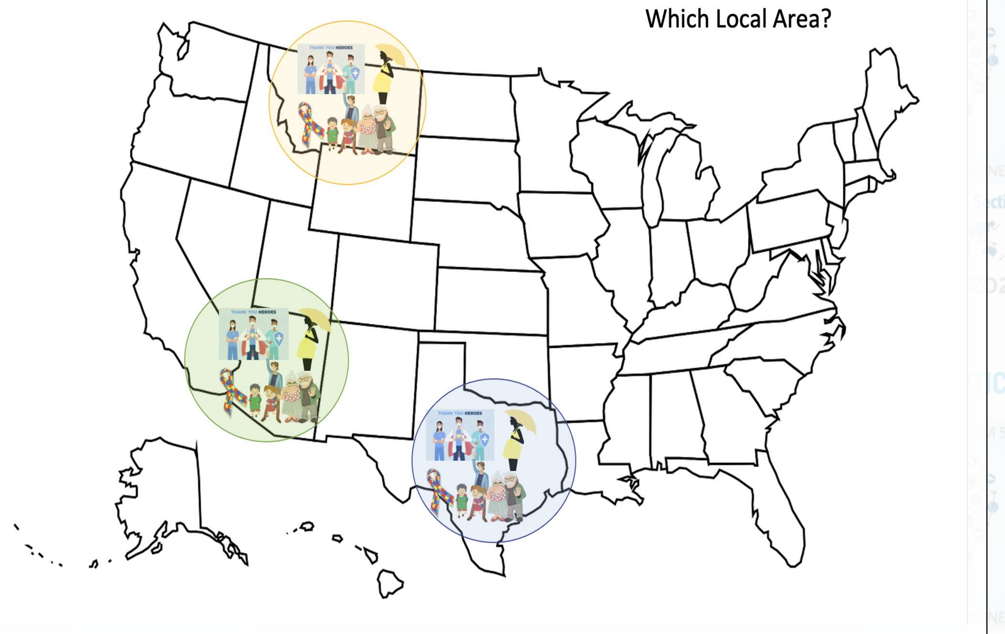
Impact of a disease intervention such as vaccination to the dynamics of cases in communities can be projected using some epidemiological modelling technique, such as SIR model. SIR is acronym for Susceptible, Infectious, and Recovered, represent the number of people in each compartment at a particular time [5]. For the normal cases projection (i.e. with no vaccine intervention), the model takes input of number of cases at starting point (T0) then using mathematical equations, it projects the dynamics of cases given epidemiology parameter estimates (i.e. transmission rate and recovery rate) within specified time continuum [4].
For clustering, the most recent work is by Zarikas et. al. [6] that clusters countries in respect to their active cases, active cases per population and active cases per population and per area.

2. RESEARCH MOTIVATIONS

In reference to the 2009 H1N1 pandemic, WHO issued recommendation on vaccination priority of the target groups. The target groups are health-care workers, pregnant women, person with underlying conditions, and age-specific populations. However, there is still a question on which local area within a country, should be considered having higher priority.

The objective of this study is to devise a novel framework that may offer an alternative guideline for public health officials in making informed decision on which local area within a country should be considered having higher priority for COVID-19 vaccination. Such decision in turn can be used as consideration point in crafting an effective vaccine deployment strategy.





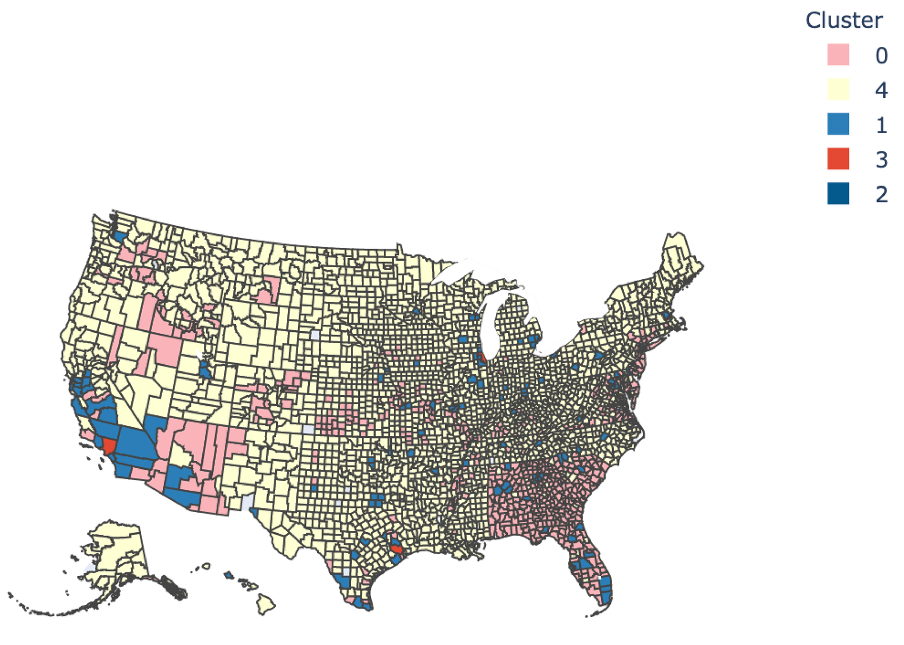
3. METHODOLOGY

Methodology

In the proposed framework, cases dynamics and other relevant demographic data of local area are investigated to get insights whether there are similar patterns that can be used to construct a set of logical groups or "clusters". The process involved investigating some reliable sources from COVID-19 historical data collected by John Hopkins University, and US demographic data (US Census Bureau). Then, an agglomerative hierarchical clustering technique was applied to group counties based on the standardised features: active cases per 1000 people, population density, and 14 days average of new daily cases with Euclidean distance as affinity, and Ward's method as linkage.

Later, the generated clusters were simulated using the SIR model, an acronym for Susceptible, Infectious, and Recovered, which is a standard disease spread modelling technique considered having reasonable predictive power for infectious disease that transmits between humans.

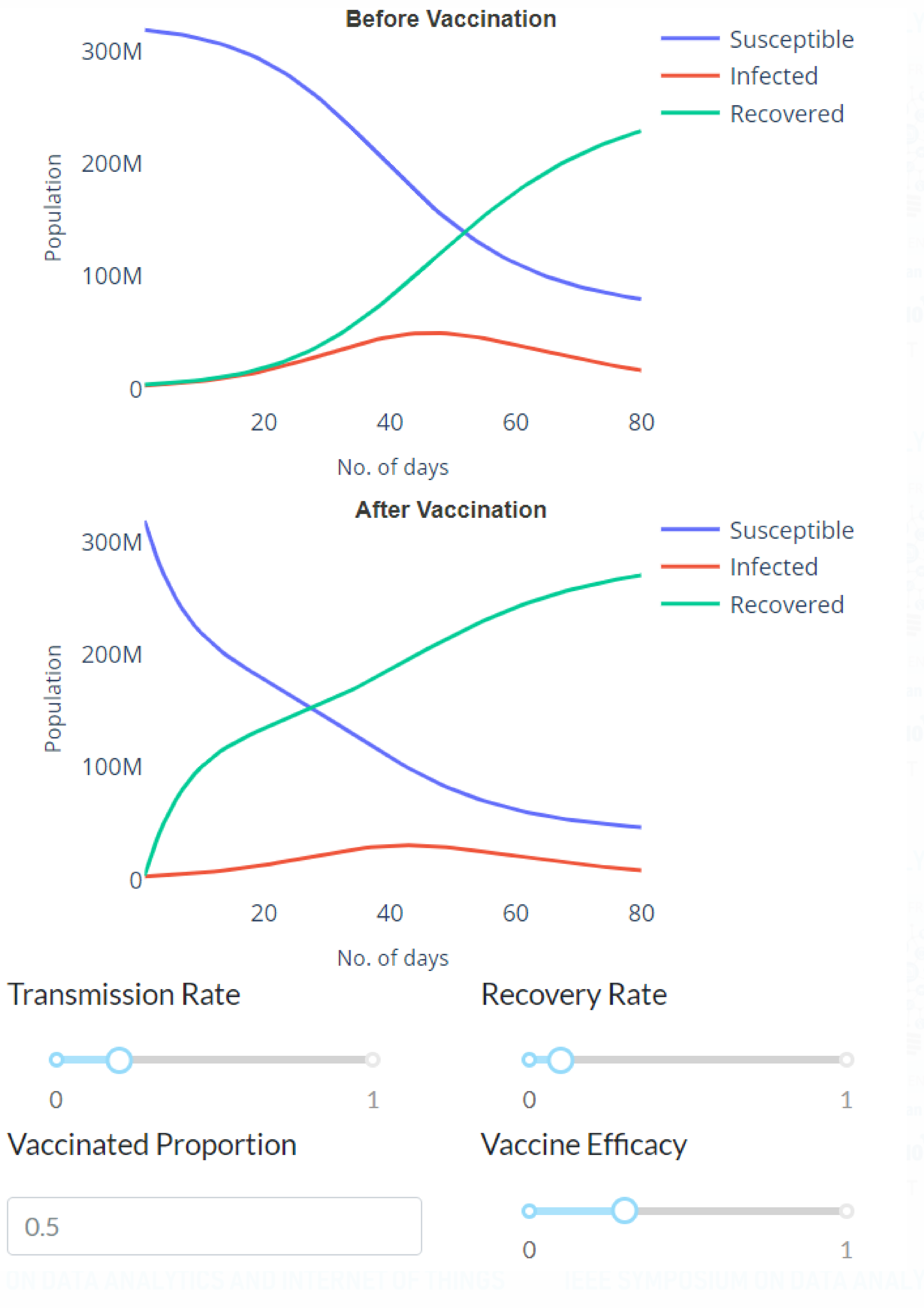
Based on domain expertise, public health officials and/or epidemiologists, then make hypothesis of vaccination prioritisation for these clusters e.g. based on some prevailing vaccination objective, which cluster or set of clusters should have higher priority over the others and observe the dynamics of cases in the population if vaccination is to be administers in the selected cluster or set of clusters.



Discussion

The clustering technique resulted with 5 clusters which are identified with a visual intuition looking at a dendrogram and further validating it through a silhouette score of 0.8 determining each counties being similar to others in the same group.

Based on the experiment, it is evident that Cluster 4 should have the highest priority as it covers 39% of the entire US population followed by Cluster 1 with 35% and Cluster 0 with 18%. Based on the SIR simulation with the assumption of at least 30% vaccine efficacy as recommended by WHO[8], If counties within cluster 4 are vaccinated, the number of infected cases would reach the maximum of 30 million in 44 days. In comparison to the others, Cluster 1 records 30 million in 40 days and Cluster 0 in 36 days.



Discussion points:

1. Would the predefined target groups within the local area affect the clustering result?
2. What is the better way to measure our proposed technique performance in comparison with the currently used technique prior to vaccine deployment

4. CONCLUSION / FUTURE WORK

Conclusion

This proposed framework provides guidelines that uses machine-learning and interactive simulation to support the United States public health officials in developing an effective, data-driven vaccine deployment strategy in anticipation of imminent limited supply of vaccine when it is approved for safety and efficiency.

Future work

There are opportunities for further improvements including but not limited to:

- Dataset enrichment to include features related to social living (social index) and non-vaccine mitigation measures.
- Adoption of epidemiological compartment model that has better performance for COVID-19 pandemic based on latest research, such as the SEIR-QD and SEIR-PO [16].
- Live data feeds to allow most up-to-date data insights and expanding coverage to not only limited to United States but also for many other countries worldwide.

References

[1] X. Chen, M. Li, D. Simchi-Levi and T. Zhao, "Allocation of COVID-19 Vaccines Under Limited Supply", 2020.
[2] "Vaccines for Pandemic Threats | History of Vaccines", Historyofvaccines.org, 2020. [Online]. Available: https://www.historyofvaccines.org/content/articles/vaccines-pandemic-threats. [Accessed: 28 Sep 2020].
[3] GUIDANCE ON Development and Implementation of a National Deployment and Vaccination Plan for Pandemic Influenza Vaccines.WorldHealthOrganization,2012.[Online].Available: https://apps.who.int/iris/bitstream/handle/10665/75246/9789241503990_eng.pdf [Accessed: 28 Sep 2020].
[4] W. Yang, D. Zhang, L. Peng, C. Zhuge and L. Hong, "Rational evaluation of various epidemic models based on the COVID-19 data of China", 2020. Available: 10.1101/2020.03.12.20034595 [Accessed 3 August 2020].
[5] "A Simple Model for Vaccination - Dynamics of Susceptibles | Coursera", Coursera, 2020. [Online]. Available: https://www.coursera.org/lecture/developing-the-sir-model/a-simple-model-for-vaccination-rHPrN. [Accessed: 28 Sep 2020].
[6] C. Shalizi, Distances between Clustering, Hierarchical Clustering. Carnegie Mellon University, 2009. [Online]. Available: http://www.stat.cmu.edu/~cshalizi/350/lectures/08/lecture-08.pdf. [Accessed: 24 Aug 2020].
[7] Weekly epidemiological record, 30th ed. World Health Organization, 2020, pp. 301-308. [Online]. Available: https://www.who.int/immunization/sage/SAGE_July_2009.pdf. [Accessed: 28 Sep 2020].
[8] P. Krause, T. R. Fleming, I. Longini, A. M. Henao-Restrepo, R. Peto, N. Dean, M. Halloran, Y. Huang, T. Fleming, P. Gilbert, V. Degruttola, H. Janes, P. Krause, I. Longini, M. Nason, R. Peto, P. Smith, A. Riveros, P. Gsell, and A. Henao-Restrepo, "COVID-19 vaccine trials should seek worthwhile efficacy," The Lancet, vol. 396, no. 10253, pp. 741–743, 2020.

Acknowledgement: This study was carried out under supervision of Dr. Damiano Spina, RMIT University.