# Project Proposal

Analysis of the impact of different transmission and recovery rates on the growth rate, peak prevalence, and the cumulative burden of infectious disease using SIR model

## *I. Introduction & background*

This project focuses on developing and examining the effects of different parameters on Susceptible-Infected-Recovered (SIR) models. The COVID-19 pandemic has led to the use of numerous epidemic models based upon SIR. Our goal is to understand the complexities of infectious diseases via 2 models: with and without intervention measures, as well as to underpin such models with real world data to understand their predictive capabilities and limitations.

In the basic SIR model, as susceptible members (S) of the population are exposed to a pathogen via an infected individual (I), they are moved to the infectious compartment by an infection rate, and to the recovered/removed compartment (R) after recovery/death by a recovery rate. Many simplifying assumptions are made in developing this model that could change the spread rate of an infectious disease.

The SIR model is based on differential equations, where the fraction of the population in each compartment changes as a function of time. Thus, the functions can be altered depending on the specific disease mechanism – incubation period of the pathogen, the mode of transmission, etc.

The core formulas to predict each compartment S, I, and R over time are as below:

$$\frac{dS}{dt} = -\beta \bullet S(t) \bullet I(t)$$

*β represents the number of people each infected person interacts with every day on average.*

$$\frac{dI}{dt} = \beta \bullet S(t) \bullet I(t) - \gamma \bullet I(t)$$

*γ represents the fraction of infected people that will recover on a given day.*

$$\frac{dR}{dt} = \gamma \bullet I(t)$$

The parameters used in these differential equations, namely the transmission and recovery rates (β & γ) are estimated based upon several factors, including but not limited to the migration rates of the population within and between countries, and the biological mechanism through which the disease affects individuals (Tang et al. 2020). For diseases like COVID-19, where the disease can be both symptomatic and asymptomatic, the models based on SIR can be extended to involve more compartments, levels to capture the complexity of transmission (Chen et al. 2020).

## *II. Methodology*

A. Basic SIR Model

a. Assumption:

- The recovered (R) compartment will no longer be susceptible
- No treatment is given, disease ends due to herd immunity

b. Parameter - default value:

- Transmission/Infection Rate (β) - 0.01
- Recovery Rate (γ) - 0.7
- Population (n) - 500 (school population)
- Initial number of infected individual - 5
- Percentage of healthy individual with symptoms - 0.2%
- Time (number of days since the day1)

B. Complex SIR Model

*The complex SIR model is built upon the basic model, with more parameters involved in the model and slight changes in assumption*

a. Assumption:

- The recovered (R) compartment will no longer be susceptible
- Intervention actions are taken:
  - ✓ 2 treatments is given to the infected, notated as A and B
  - ✓ Testing (screening) is provided

b. Additional Parameter (with default value, A and B respectively):

- Effectiveness of each treatment - 0.75 and 0.25
- Percentage of infected individuals getting treatment - 1% and 0%
- Cost of each treatment (per person) - $20 and $15
- Cost of testing process (per person) - $1
- Sensitivity of testing process - 0.99
- Specificity of testing process - 0.95

C. Step-by-step procedures:

a. Develop a script for both SIR models (with and without treatment) to calculate the population of each S, I, R compartment over time (number of days)

b. Visualize our models with curves representing the changes of S, I, R compartments

c. Design experiments to examine the effects of adjusting parameters (e.g. changing β and γ) on the growth of each compartment over time, as well as the effect of intervention measures into the development of disease.

**D. Model Evaluation**:

*We plan to test our prediction model on real data of COVID-19 situations in the World. Two potential data sets are listed below with more detailed description. To measure the performance of our models, we propose these following matrices:*

- $R^2$ and Adjusted $R^2$
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2} \qquad MAE = \frac{1}{n}\sum_i |y_i - \hat{y}_i|$$

## *III. Data Sources*

- COVID-19 Data in the World [Covid19clean_complete]
  - Shows day to day data of number of confirmed cases, deaths, active cases, and recovered cases in different regions of the world from January 22, 2020 to July 27, 2020
  - https://www.kaggle.com/imdevskp/corona-virus-report?select=covid_19_clean_complete.csv
- COVID-19 Data in the US [usacountywise]
  - Shows day to day data of numbers of confirmed cases, deaths, active cases, and recovered cases in every county from January 22, 2020 to July 27, 2020.
  - https://www.kaggle.com/imdevskp/corona-virus-report?select=usa_county_wise.csv

## *IV. Bibliography*

Smith, D., and Moore. L. (2004). The SIR Model for Spread of Disease - The Differential Equation Model. MAA Publications.

Chen, R. (2020). A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. Infectious Diseases of Poverty, 9(1), 24–24. https://doi.org/10.1186/s40249-020-00640-3

Tang, L., Zhou, Y., Wang, L., Purkayastha, S., Zhang, L., He, J., Wang, F., and Song, P. X.-K. (2020) A Review of Multi-Compartment Infectious Disease Models. International Statistical Review, 88: 462– 513. https://doi.org/10.1111/insr.12402.