**Identify Underreporting (UR) issues in covid by using mathematical models**

UR is a composite problem including UA and UE. The rate of data underreported in an chain of events is calculated by equation [1], which is also the goal of the whole research:

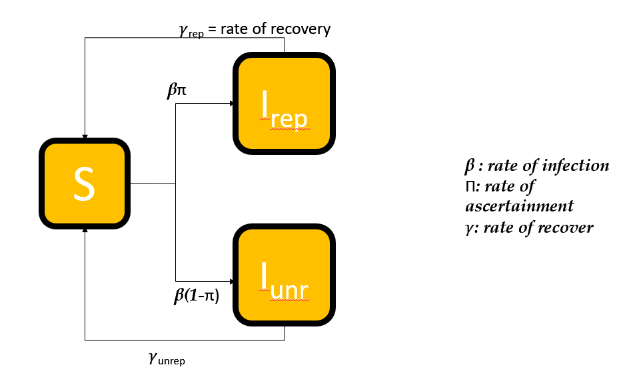
with . Why do we need this parameter? because UA also measure the asymptomatic infectives with true negative testing results, therefore, helps to classify the type of underreported cases in the right categories which eliminates the risk of repetitive counting.

**Heterogeneity of Datasets**

My datasets were inspired and taken from multiple sources to create a synthetic dataset, therefore, I hypothesize that the figures in those data are the real figures of Australian situation.

**UA (Under Ascertainment): this is the underreported number of people who are considered to be infectious but were not captured by official records: (see in the** [**Supplementary Documents**](https://github.com/QUT-Trusted-Networks-Lab/n10648771-stephen/tree/master/doc)**)**

 The aim of identifying of UA issues is to generate an epidemic model that can simulate the initial stage of the pandemic within constant rate of infection and rate of recovery. I use SII model for this work because it can be understood from the epitome of Western countries (like Sweden and England) that the herd immunity strategy was not effective for this particular type of disease, therefore, after a person recovers from the infection, he/she will be set back in the susceptible population



* Before constructing SII, we need to establish some baseline models for final evaluation and comparison. I opted to use Branching Process, SIR, variant SIR with structured population and incident models (see in my file)
* Things will be different in the stage of social distancing where regional areas will behave more normally during the outbreak than people in cities like Sydney. However, cross-regional interactions are certain to occur since the travel ban policy was only imposed on inter-state travelling. A new model should be used for this case - SISIR ( (Manfredi, 2017):

 Figure: SISIR Model used for the social distancing period

* From the infected people we can use the rate of E->I (exposed to infected) from the SEIR model above to measure the potential exposed population => we can find the UA’s number (people who are exposed but refuse to have a medical check-up that can be recorded by official data) just like step iii.
* To classify regions with b or n, I will put the threshold of classification equalling to the mean of the infected number among postcode areas
* Then we came to final stage of covid where vaccination are implemented to preclude the further expansion of the outbreak. There are 3 types of models will be used in this stage:

=> The model with only vaccination => The model with only self-protective action => Hybrid model incorporating 2 types mentioned above

**UE (Under Estimate): Those who attended healthcare testing practices but receive True Negative/False Positive results.**

Since surveillance data are rare and unreliable, we should only use the official dataset from the gov to estimate the false rating of covid testing. Several approaches can be used for UE (which have been used as baseline models):

 Generalized Poisson-Mixed

 Negative-Binomial

 Gaussian Poisson

 Beta-Poisson

The aim is to measure and analyse the specificity and sensitivity of the testing with PCT-RT test kit. We will use probit regression to measure the probability of a false test taking place then use such methods for baseline models to regress and retrieve the real number. According to Hamilton Public Health Institution, the event of false testing occurs when there is a cross-contamination of the specimen (meaning that the specimen can be aerosolized in the testing clinic so the second or third specimen would not show a sign of infection). In addition to that, the chance of false positive testing also depends on the length of the DNA probes, how many and which genes are measured and technical errors.

<<This one has not done yet, due to the data I simulated (inspired by POSMOD) is hard to conduct in real-life experiences>>