Supplementary Materials – Under Ascertainment (UE in another file)

Table of Contents

I. Simple Parameter Estimation with SIR: $\begin{cases} \frac{dS}{dt} = -\frac{\beta}{N}SI & (i) \\ \frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I & (ii) \\ \frac{dR}{dt} = \gamma I & (iii) \end{cases}$

A. Introduction:

SIR model is the simplest way to model a pandemic, hence, within a constant rate of infection and the absence of governmental control over the outbreak in the initial stage, this method is useful to see the underascertainment rate.

- N = I + S + R
- Goal: Estimate beta, gamma (infection rate and recovery rate, respectively)
- As I have introduced before, the Bayesian Inference helps us to determine the proportion of population who are highly immune to the current pandemic, hence, Exposed Population ≈ Susceptible Population ≈ N (population excluded those who are immune). At the first day of pandemic (N=S):

$$\Rightarrow \frac{dI}{dt} = \beta I - \gamma I$$

  - *Let* $m = \beta - \gamma$, we will find a non-linear function for computing the m:

$$\Rightarrow \frac{dI}{dt} \sim mI$$

$$\therefore \mathbf{I(t) \sim I_0 e^{mt}} \quad \text{(since I grows exponentially)}$$

#with t is time interval and I_0 is the number of infection in day 1

$$\therefore \ln(I) = mt + \ln(I_0) \qquad I.1$$

From equation [I.1], we can easily estimate the value of m using log plot by substituting the value of I_n from official data then using least square method to fit the line.

Note: Different time intervals might produce different results of m, the higher m means the more uncontrolled pandemic was at the time t interval.

- Then we can estimate the value of recovery rate γ by using 2 ways: using the number of days in infection period certified by health authorities or we can using [iii] equation to reckon the rate γ.
  - R(t) = $\gamma t I_0$
  - $\gamma \approx \frac{R(t+1)-R(t)}{I(t)}$ *#I got a little concern in this part, if we account for the incubation period, it should also be:* $\gamma \approx \frac{R(t)-R(t-\epsilon)}{I(t)}$ *with* ϵ *is the length of incubation period. Let's just stick with R(t+1) in this paper because my final outputs do not show much difference*
  - *After having values of* γ *and m, it is now easier to find the value for* β

B. Application on AUSSIE dataset:

- The scope of this research is the first 60 days of pandemic (because it is the initial stage of pandemic where number of infected = $\sqrt{total\ infected}$ as defined by (Diekmann, 2011))

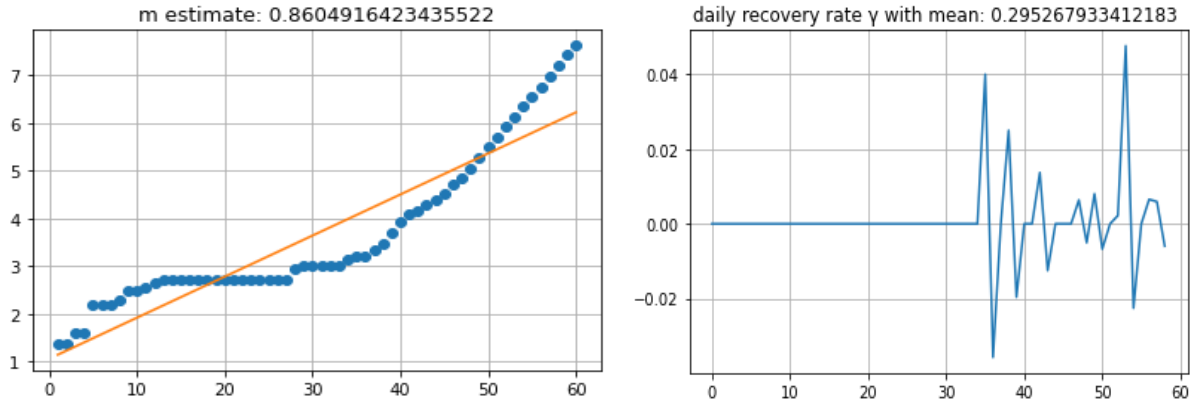


Figure I-1: Estimations of m and γ (from notebook – ADD LINK HERE)

Figure 1 displays that the value of m and gamma is 0.8 and 0.29 respectively. We have final table of parameter estimation for the first method

β=0.565224	**γ=0.295268**	**R_0 = β/ γ = 1.91427**

⇨ Compare with the number of true infection counts, we have:

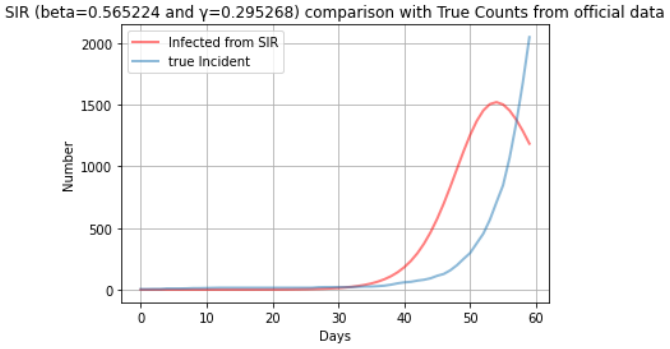


Figure I-2 Compare number of infected

C. Problem of this method?

The fact that SIR is always the one to be chosen as the base model for any infectious disease simulation was due to its simplicity. However, the problem of this method is that it is a heavily data-driven technique that would be bias based on the inputs. This means that SIR only reflects and predicts based on what data we have, ignoring the reporting errors might occur during the process. Let's look back at the equation I.1: **$\ln(I) = mt + \ln(I_0)$.** When we build our model based on this equation, we somehow assume that the number of cases in the first day (I_0) are fully

reported (which definitely is not true). Therefore, a better approach that includes anonymous estimations should be introduced.

II. Model Pandemic using contact network and transmission risk in a structured population:

A. Introduction

The second method is more specific and based on the result of some other peer-reviewed literature upon the rate of contact made by an individual and the risk of infectious transmission. At the effort of attaining the most reliable rate of infection which is calculated by equation (II.1), I have to assume some :

$$\beta = D * c \qquad \text{II.1}$$

$$\begin{cases} \beta \text{ is the rate of infection in the population} \\ D \text{ is the transmission risk that an infected might spread on others} \\ c \text{ is the average contact rate made by an infected individual} \end{cases}$$

The rate of infection is extremely important when analyzing an infectious disease, many models are built on the foundation of this parameter such as ODE's models, Reed-Frost, Branching Process... In this section, I would like to use 3 approaches which are ODE's models for structured population and spatial data, the incident model and branching process from R_0

1. Calculating the average number of contacts c in a structured population:

a) Data Collection Method:

From the survey conducted by European Commission, 7,290 participants recorded characteristics of 97,904 contacts with different individuals during one day, including age, sex, location, duration, frequency, and occurrence of physical contact. Respondents are demanded to fill in Table II-1 and its dataset is attached in the APPENDIX A.

Age (or range)	Gender		Did you touch his/her skin?		How often do you have contact with this person in general?					Where did you have contact? (tick all which apply on your assigned day)						Total time spent with person during whole day				
	F	M	Y	N	Daily or almost daily	About once or twice a week	About once or twice a month	Less than once a month	Never met before	Home	School / College	Work	Transport	Leisure	Other	Under 5 mins	5-15 mins	15 mins – 1 hr	1 – 4 hrs	More than 4 hrs
☐9 (- ☐☐)	☐	☑	☑	☐	☑	☐	☐	☐	☐	☑	☐	☐	☑	☐	☐	☐	☐	☐	☑	☐
25 (- 30)	☑	☐	☐	☑	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☑	☐	☐	☑	☐	☐
25 (- 29)	☑	☐	☐	☑	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐
30 (- 35)	☐	☑	☐	☑	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☑	☐	☐	☑	☐	☐
67 (- ☐☐)	☐	☑	☑	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☑	☐	☐
30 (- 39)	☑	☐	☐	☑	☐	☐	☐	☑	☐	☐	☐	☐	☐	☑	☐	☑	☐	☐	☐	☐
17 (- 22)	☐	☑	☐	☑	☐	☐	☐	☐	☑	☐	☐	☐	☐	☑	☐	☑	☐	☐	☐	☐
35 (- 42)	☑	☐	☐	☑	☐	☐	☐	☑	☐	☐	☐	☐	☐	☑	☐	☑	☐	☐	☐	☐
42 (- ☐☐)	☐	☑	☑	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☑	☑	☐	☐	☐	☐
40 (- 45)	☐	☑	☐	☑	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☑	☑	☐	☐	☐	☐
☐☐ (- ☐☐)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
☐☐ (- ☐☐)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
☐☐ (- ☐☐)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
☐☐ (- ☐☐)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
☐☐ (- ☐☐)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
☐☐ (- ☐☐)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
☐☐ (- ☐☐)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Table II-1: Survey form recording the pattern of contact made by an individual (Source: POSMOD)

Some prominent insights of the dataset are illustrated in Table II-2. It can be instantly acknowledged that the average number of contacts are 19.6 for the whole population (aged from 0-90). However, it would be biased if we ignored assumed that all age ranges have the similar rate of contacts (in reality, infants and senior citizens might not attend public avenues as frequently as young generations since they do not work or travel). Hence, reckoning the true contact rate corresponding to respective age groups is important to identify which groups are the most infective and what measures should be imposed on them.

The implication for this method is that if we can apply the same survey with the form given above to Australian population, we can trace their movements and interactions which can helps authorities to control the pandemic.

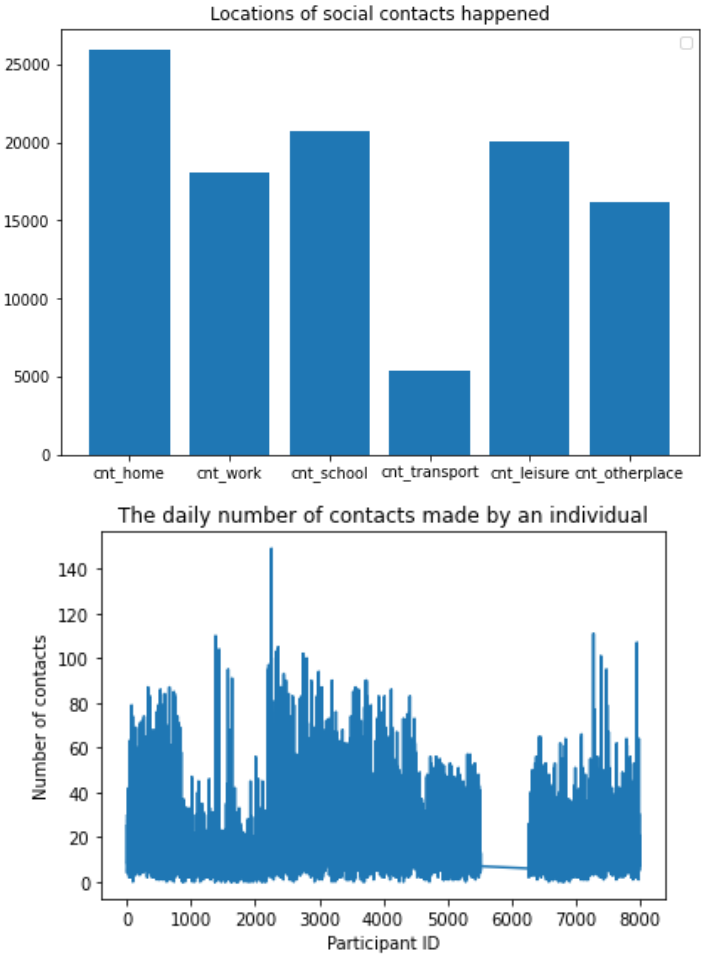


Figure II-1: Contact pattern

b) Method:

A contacts matrix is used to account for the total number of contacts. Let C_{ij} be the number of contacts of people in age group with age group j; P_i be The population of participants reporting their contact incidents; N_i be the total population in group i (including both participants and reported people).

The number of people in group j that an average person from age group i contact with is calculated as:

$$\frac{C_{ij}}{P_i} = m_{ij}$$

The total number of contacts made by an age group to other groups is:

$$total\ count\ T = m_{ij} * N_i$$

Age groups of people that participants contact with	Age groups of participants			
	0-10 (P_1)	10-20 (P_2)	...	80+ (P_n)
0-10	T_{11}	T_{12}	...	.
10-20	T_{21}	T_{22}	...	.
...	...	..	...	.
80+	.	.	.	.

The limitation is that: Logically, T_{21} should be = T_{12}, however, the data from EC is just a sample of the whole population. Therefore, inconsistency in computation can be easily seen where $T_{ij} \neq T_{ji}$. To smoothen the inconsistency, I transpose matrix A and take out the mean of 2 matrices to compute the new matrix B with a more consistent data. Hence:

$$B = \frac{A + A^T}{2}$$

c) Application:

Applying the above method to the dataset from EC (POSMOD), we retrieve the following matrice:

	age group of participant														
age of contact	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70+
00-04	1.92	0.65	0.41	0.24	0.46	0.73	0.67	0.83	0.24	0.22	0.36	0.20	0.20	0.26	0.13
05-09	0.95	6.64	1.09	0.73	0.61	0.75	0.95	1.39	0.90	0.16	0.30	0.22	0.50	0.48	0.20
10-14	0.48	1.31	6.85	1.52	0.27	0.31	0.48	0.76	1.00	0.69	0.32	0.44	0.27	0.41	0.33
15-19	0.33	0.34	1.03	6.71	1.58	0.73	0.42	0.56	0.85	1.16	0.70	0.30	0.20	0.48	0.63
20-24	0.45	0.30	0.22	0.93	2.59	1.49	0.75	0.63	0.77	0.87	0.88	0.61	0.53	0.37	0.33
25-29	0.79	0.66	0.44	0.74	1.29	1.83	0.97	0.71	0.74	0.85	0.88	0.87	0.67	0.74	0.33
30-34	0.97	1.07	0.62	0.50	0.88	1.19	1.67	0.89	1.02	0.91	0.92	0.61	0.76	0.63	0.27
35-39	1.02	0.98	1.26	1.09	0.76	0.95	1.53	1.50	1.32	1.09	0.83	0.69	1.02	0.96	0.20
40-44	0.55	1.00	1.14	0.94	0.73	0.88	0.82	1.23	1.35	1.27	0.89	0.67	0.94	0.81	0.80
45-49	0.29	0.54	0.57	0.77	0.97	0.93	0.57	0.80	1.32	1.87	0.61	0.80	0.61	0.59	0.57
50-54	0.33	0.38	0.40	0.41	0.44	0.85	0.60	0.61	0.71	0.95	0.74	1.06	0.59	0.56	0.57
55-59	0.31	0.21	0.25	0.33	0.39	0.53	0.68	0.53	0.55	0.51	0.82	1.17	0.85	0.85	0.33
60-64	0.26	0.25	0.19	0.24	0.19	0.34	0.40	0.39	0.47	0.55	0.41	0.78	0.65	0.85	0.57
65-69	0.09	0.11	0.12	0.20	0.19	0.22	0.13	0.30	0.23	0.13	0.21	0.28	0.36	0.70	0.60
70+	0.14	0.15	0.21	0.10	0.24	0.17	0.15	0.41	0.50	0.71	0.53	0.76	0.47	0.74	1.47
Total	8.88	14.59	14.80	15.45	11.59	11.90	10.79	11.54	11.97	11.94	9.40	9.46	8.62	9.43	7.33

Table II-2 Matrix of contact numbers

We can inferred from the above table that infants and old people have the lowest number of contacts while children at the primary school age have a higher chance of exposing to an infected individual.

2. Estimating the risk of transmission risk by an infective individual

Transmission risk requires a strong understanding in biomedical domain to compute the desired outputs. There are several methods that are verified to be effective such as the equation of transmission risk proposed by John Hopskin Uni:

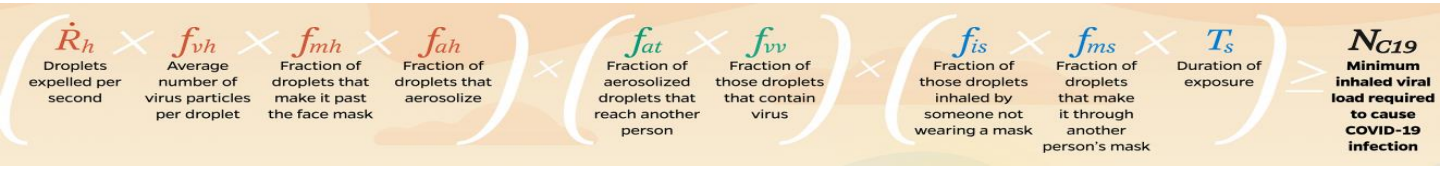


Figure II-2 Transmission Risk Estimation Equation (source: John Hopskin, 2020)

However, this method is hard to apply on a larger scale as only medical specialists are capable of measuring such variables.

A simpler approach of capturing infection risks via the aerosol emission was introduced by a group of chemistry experts from Max Planck Institute that can help individuals to check the risk for themselves (see more: link). ***Despite of its drawbacks that many assumptions were made to perform calculations, but the result from it can be referred on a theoretical level.***

From the article, some inferences can be extracted as the infection risk of a single viral RNA is calculated as :

$$P_{RNA} = 1 - 10^{\frac{\log_{10} x}{Dose_{mean}}}$$

With x = 0.5 represents for the number of susceptible objects to be infected by taking a mean influx of RNA (Dose_mean). The rate of getting infected by a person exposed to an infective in a close proximity is:

$$R(\%) = [1 - (1 - P_{RNA})^{D_{episode}{}^{S}}] \times 100,$$

With s is the people involved in the environment – in our case, s equals to the number of contacts made by an individual, $D_{episode}$ is the number of viral RNA copies inhaled by a person. It would be hard to reckon the value for such $Dose_{mean}$ and $D_{episode}$ since I possess no relevant knowledge to bio-medical field. Therefore, it would be better to take the value from the literation which is 91.9% that a person can be infected by an infective agent if he/she shares a close contact for a duration of 2 hours. That infection rate decreases when the distance of interactions enlarges. However, upon the dataset proposed by POSMOD, the reported contacts are physical ones with varying frequencies, hence, the space of interaction can be equivalent to be in proximity with an infective host. Therefore, the transmission rate is in the range of 60% to 91.9%.

B. Methods

1. Incident Model:

The incident model measures the growth of infective cases in the initial phase of the pandemic. New case at time t result from contacts with individuals that were infected before t and become infectious

at t (we call incubation period τ). The incident i(t) is the number of new cases arising per time t will be executed in the equation:

$$i(t) = \frac{S(t)}{N} D * C * \int_{T1}^{T2} i(t-\tau) d\tau$$

With D = 76% (mean of the probability range in section A.2) and C is the average contact number. We can easily compute the growth rate of incident in the age-structured population, however, I will simplify the work by summing up all the cases to compare with the true counts from official data later on. Set incubation period τ from 4-6, we have the following model:

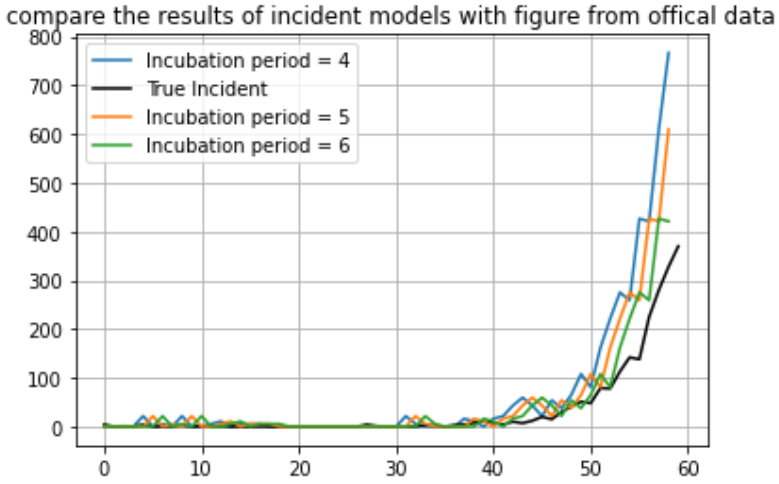


Figure II-3 Comparisons with Incident Model

A quick analysis from the figure II-3 is that it is comprehensible that the lower incubation duration leads to the higher rate of infected cases.

2. ODE SIR application for age-structured population:

The infection rate equation (II.2) will be transformed into:

$$\boldsymbol{\beta_{ij} = D * \sum c_{ij}}$$

c_{ij}: Contact made by an individual from group age i to group age j

The recovery rate $\gamma_1\ \gamma_2$ .... varies upon the age groups (as we may know that older generations have lower immune system than the younger). The new variant SEIR would be as following:

$$\begin{cases} \dfrac{dS}{dt} = -\sum_{i}^{j} \beta_i S_i I_i \\ \dfrac{dI}{dt} = -\sum_{i}^{j} S_i I_i \sum \beta_{ij} - \sum_{i}^{j} \gamma_i I_i \\ \dfrac{dR}{dt} = \sum_{i}^{j} \gamma_i I_i \end{cases}$$

stuck with estimation of gamma for different age group

III. Mixed model retrieve the fully ascertained number of infected:

A. Introduction

All models above can help us to identify the underascertainment to some extent by simply subtracting the new outputs of each model to the original dataset. However, the bias can be clear since some of my assumptions (transmission risks, first day of infection,...) might not be correct. A mixed model of ODE with incidence function can be an option to validate the findings. I came up with a SII model which is illustrated as in the figure III-1 with S is the population of susceptible, I_{rep} is the population of the infected who are reported and I_{unrep} is the population of the infective individuals who are not reported [12].

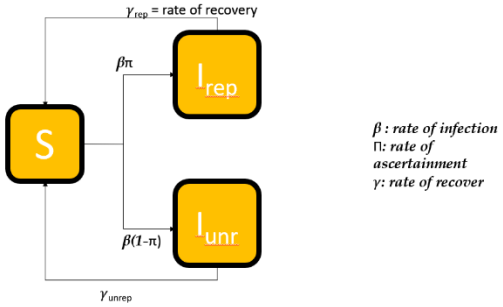


Figure III-1 SII model for the initial phase

It is hard to estimate the value of parameter π within official available data online or some panel reports from surveillance reports. Panel data for estimating the underreporting rate is unreliable since it is a passive method that only attracts the responses from voluntary people. A marginal error of sampling from the survey data will occur and might influence the result of the study substantially. An prominent example of this issue was the presidential debate in USA between Roosevelt and Landon. Among 1.2 million people participating in an opinion poll conducted by some intelligence agencies, 43% of them voted for Roosevelt. However, at the end of the election day, President Roosevelt claimed the victory within 62% of votes. This was because they did not account for the sample error that would occur in such passive surveys (error here is that they only asked for a certain type of people about their voting opinions, many others in population might have different point of view from people they interviewed). Similarly, for infectious disease issues, the number of cases captured by official records might not reflect fully the current situation of the outbreak since:

- There might be some people who are infected but refused to visit testing clinics because they are afraid of being quarantined or the financial cost it might arise.
- There might be some people who are infective but they do not experience any symptoms of the disease (we call this the infective host or patient "0" – a term for a virus carrier who appears with perfect medical condition but can infect other people)

Therefore, parameter π can be understood as the sampling error of an observational study which should be estimated to produce the most exact number of incidents. Since the passive survey methods are ineffective and unreliable, I propose 2 active approach to resolve the issue in the Method section.

After getting the rate of overreporting(?)/underreporting, we will substitute the value of π for the following compartment model below. SII actually is an improved version of SIS (SIS can also be called SI), which is used for the type of infectious disease that possesses no certified cures (meaning that all recovered individuals will have a high chance of second attack from the pandemic

Our ODE now will be as following:

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta}{N}S(I_{rep} + I_{unr}) + \gamma I_{rep} + \gamma I_{unr} \\ \frac{dI_{rep}}{dt} = \frac{\beta\pi}{N}SI_{rep} - \gamma I_{rep} \\ \frac{dI_{unrep}}{dt} = \frac{\beta(1-\pi)}{N}SI_{unrep} - \gamma I_{unrep} \end{cases}$$

According to a peer-reviewed publication by Nishiura, H., [12], he has worked with the Travel & Tourism department of Japan to conduct a survey of under-ascertainment. In his report, 592 of 655 people from the survey showed COVID19-related symptoms and 8 of them are infected but seemed asymptomatic #review later

B. Methods:

1. Method 1: Estimating the sampling error by detecting the disease as soon as possible, then integrating with SII/SIR model to retrieve the number of true counts (or even branching processes model can be used here):

a) Sampling Technique:

Let D(t) denote the set of events that the disease is detected from a sample of size s. The probability of detecting disease, P(D(t) > 0), can be modeled as a binomial random variable. Since we are planning to deploy the SII technique, the proportion of the infecteds at time t is given by I(t)/S(t) (if we use SEIR then the rate would be I(t)/E(t)).

In a sample of size s, P(D(t)=0) is the complement of not detecting any infected individual at time t, and calculated as following:

$$P(D(t) = 0) = 1 - (1 - \frac{I(t)}{S(t)})^s$$

At the initial stage, the ratio of I/S would be relatively minimal (<0), therefore:

$$P(D(t) = 0) \approx 1 - e^{\Sigma_0^n - \frac{sI_t}{S_t}}$$

We also consider the probability of detecting the disease for the first time in the k^{th} sample when sampling occurs regularly at discrete time intervals. A model of detecting the disease in the current sampling period, but not before, is a geometric distribution with time-dependent detection probabilities (geometric distribution is to measure the chance of getting infected cases after having P(D(t)=0) of not detecting any infective individuals). Hence:

$$P(k) = (1-p)^{k-1}p = (1 - e^{\Sigma_0^n - \frac{sI_t}{S_t}})e^{\Sigma_0^n - \frac{sI_t}{S_t}}$$

with π is the probability of positive detection in a single pooled sample at time t, we have:

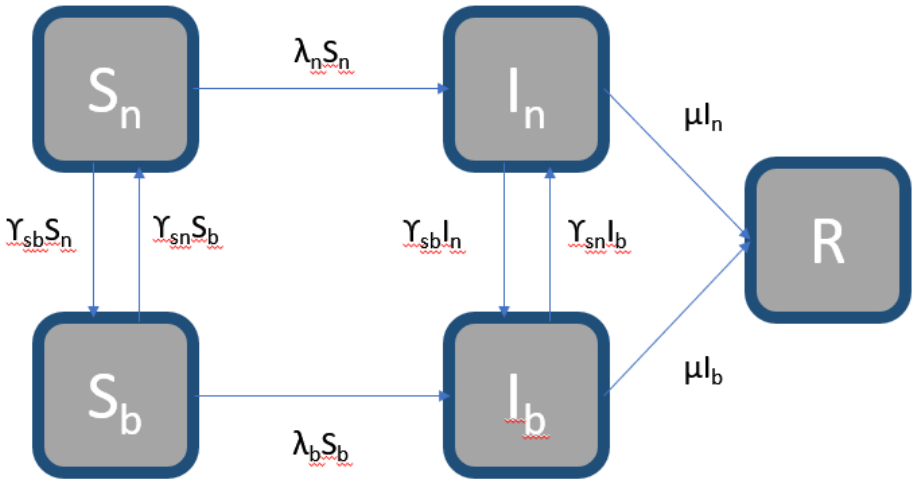
$$P(k) = (1 - e^{\pi \Sigma_0^n - \frac{sI_t}{S_t}}) e^{\pi \Sigma_0^n - \frac{sI_t}{S_t}}$$

- Next step: Simulate data with a pooled sample within a random π variable. Then we would verify the

2. Method 2: Estimating π by using likelihood based methodologies

In detail, we will perform an estimation of R0 then using the equilibrium equation of initial phase of the outbreak to make estimation for transmission rate and recovery rate. Finally, using SII model to estimate the whole under-ascertainment problems.

IV. ODE for next stage of epidemic: Social Distancing – Within results of under-ascertainment rate from (III)



Just plan it out here. Not done yet

V. References

1. Becker, N. (2008). Modeling to Inform Infectious Disease Control. Chapman & Hall.
2. Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S. A., Zhang, T., Gao, W., Cheng, C., Tang, X., Wu, X., Wu, Y., Sun, B., Huang, S., Sun, Y., Zhang, J., Ma, T., Lessler, J., & Feng, T. (2020). Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. medRxiv. https://doi.org/10.1101/2020.03.03.20028423
3. Cantó, B., Coll, C., & Sánchez, E. (2017). Estimation of parameters in a structured SIR model. Advances in Difference Equations, 2017(1). https://doi.org/10.1186/s13662-017-1078-5
4. Cheryl L Gibbons, M.-J. J. M., Dietrich Plass, Arie H Havelaar, Russell John Brooke2, Piotr Kramarz6, Karen L Peterson, Anke L Stuurman, Alessandro Cassini6, Eric M Fèvre, Mirjam EE Kretzschmar. (2014). Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. BMC Public Health 14, 147.
5. Chitnis, S. Y. D. V. J. M. H. N. (2013). MATHEMATICAL MODELS OF CONTACT PATTERNS BETWEEN AGE GROUPS FOR PREDICTING THE SPREAD OF INFECTIOUS DISEASES. Math Biosci Eng, 10.
6. Craft, M. E. (2015). Infectious disease transmission and contact networks in wildlife and livestock. Philos Trans R Soc Lond B Biol Sci, 370(1669). https://doi.org/10.1098/rstb.2014.0107
7. Fernandez-Fontelo, A., Cabana, A., Puig, P., & Morina, D. (2016). Under-reported data analysis with INAR-hidden Markov chains. Stat Med, 35(26), 4875-4890. https://doi.org/10.1002/sim.7026
8. Glasbey, D. J. A. C. A. (2003). A simulation-based method for model evaluation. Statistical Modelling
9. Held, L. (2008). Handbook of Infectious Disease Data Analysis.
10. Isella, L., Romano, M., Barrat, A., Cattuto, C., Colizza, V., Van den Broeck, W., Gesualdo, F., Pandolfi, E., Rava, L., Rizzo, C., & Tozzi, A. E. (2011). Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. PLoS One, 6(2), e17144. https://doi.org/10.1371/journal.pone.0017144
11. Jagan, M., deJonge, M. S., Krylova, O., & Earn, D. J. D. (2020). Fast estimation of time-varying infectious disease transmission rates. PLoS Comput Biol, 16(9), e1008124. https://doi.org/10.1371/journal.pcbi.1008124

12. Kirkeby, C., Halasa, T., Gussmann, M., Toft, N., & Graesboll, K. (2017). Methods for estimating disease transmission rates: Evaluating the precision of Poisson regression and two novel methods. Sci Rep, 7(1), 9496. https://doi.org/10.1038/s41598-017-09209-x
13. Lee, M. P. F. F. C. (2010). Generalized Poisson-Poisson mixture model for misreported counts with an application to smoking data. Journal of data science.
14. Lelieveld, J., Helleis, F., Borrmann, S., Cheng, Y., Drewnick, F., Haug, G., Klimach, T., Sciare, J., Su, H., & Poschl, U. (2020). Model Calculations of Aerosol Transmission and Infection Risk of COVID-19 in Indoor Environments. Int J Environ Res Public Health, 17(21). https://doi.org/10.3390/ijerph17218114
15. Li, C., Pei, Y., Zhu, M., & Deng, Y. (2018). Parameter Estimation on a Stochastic SIR Model with Media Coverage. Discrete Dynamics in Nature and Society, 2018, 1-7. https://doi.org/10.1155/2018/3187807
16. Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science, 368(6490), 489-493. https://doi.org/10.1126/science.abb3221
17. Lyle, D. (2011). Bayesian Inference for Stochastic Processes. Taylor and Francis Group.
18. Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med, 5(3), e74. https://doi.org/10.1371/journal.pmed.0050074
19. Müller, J., & Kretzschmar, M. (2020). Contact tracing – Old models and new challenges. Infectious Disease Modelling. https://doi.org/10.1016/j.idm.2020.12.005
20. Nishiura, H., Kobayashi, T., Yang, Y., Hayashi, K., Miyama, T., Kinoshita, R., Linton, N. M., Jung, S. M., Yuan, B., Suzuki, A., & Akhmetzhanov, A. R. (2020). The Rate of Underascertainment of Novel Coronavirus (2019-nCoV) Infection: Estimation Using Japanese Passengers Data on Evacuation Flights. J Clin Med, 9(2). https://doi.org/10.3390/jcm9020419
21. Omori, R., Mizumoto, K., & Nishiura, H. (2020). Ascertainment rate of novel coronavirus disease (COVID-19) in Japan. Int J Infect Dis, 96, 673-675. https://doi.org/10.1016/j.ijid.2020.04.080
22. Prasanna, A. S. V. K. Data-driven Identification of Number of Unreported Cases for COVID-19: Bounds and Limitations.
23. PRODANOV, D. (2020). ANALYTICAL PARAMETER ESTIMATION OF THE SIR EPIDEMIC MODEL. APPLICATIONS TO THE COVID-19 PANDEMIC. arxiv.

24. Rahimighazikalayeh, G. (2018). Adjusting For Mis-Reporting In Count Data [Theses and Dissertations, University of South Carolina]. University of South Carolina - Columbia.
25. Ruiyun Li, S. P., Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, Jeffrey Shaman. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science Magazine.
26. Salathe, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., & Jones, J. H. (2010). A high-resolution human contact network for infectious disease transmission. Proc Natl Acad Sci U S A, 107(51), 22020-22025. https://doi.org/10.1073/pnas.1009094108
27. Salathe, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., & Jones, J. H. (2010). A high-resolution human contact network for infectious disease transmission. Proc Natl Acad Sci U S A, 107(51), 22020-22025. https://doi.org/10.1073/pnas.1009094108
28. Salles, B. P. B. P. R. (2020). ESTIMATION OF COVID-19 UNDER-REPORTING IN BRAZILIAN STATES THROUGH SARI.
29. Sechidis, K., Sperrin, M., Petherick, E. S., Luján, M., & Brown, G. (2017). Dealing with under-reported variables: An information theoretic solution. International Journal of Approximate Reasoning, 85, 159-177. https://doi.org/10.1016/j.ijar.2017.04.002
30. Seilheimer, R. L. (May 9, 2008). Contact Network Epidemiology: Mathematical Methods of Modeling a Mutating Pathogen on a Two-type Network.
31. Tozzi, A. M. F. G. C. R. A. E. (2013). An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. BMC Infectious Disease, 185.
32. Wood, J. S., Donnell, E. T., & Fariss, C. J. (2016). A method to account for and estimate underreporting in crash frequency research. Accid Anal Prev, 95(Pt A), 57-66. https://doi.org/10.1016/j.aap.2016.06.013
33. Zhang, J., Litvinova, M., Wang, W., Wang, Y., Deng, X., Chen, X., Li, M., Zheng, W., Yi, L., Chen, X., Wu, Q., Liang, Y., Wang, X., Yang, J., Sun, K., Longini, I. M., Halloran, M. E., Wu, P., Cowling, B. J., Merler, S., Viboud, C., Vespignani, A., Ajelli, M., & Yu, H. (2020). Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. The Lancet Infectious Diseases, 20(7), 793-802. https://doi.org/10.1016/s1473-3099(20)30230-9
34. Zitkovic, G. (2014). Lecture 7: Branching processes. In Intro to Stochastic Processes.