Investigating the probability of a major outbreak with varying index case age, susceptibility and infectiousness.

1. Introduction:

In this report, I demonstrate that the demographic structure of a population impacts the likelihood of a major outbreak through three age group-dependent factors: intra- and inter-age group contact rates, individual susceptibility and individual infectiousness.

It is evident from previous literature that communicable diseases often have varying impacts depending on the age of the infected individual. In this project I focused on the coronavirus disease 2019 (Covid-19). Pertaining to Covid-19, an example of age affecting the epidemiological outcomes of the pandemic is the case fatality ratio (CFR, the number of deaths per infections). In China, the CFR for infected individuals aged 40-49 is 0.4%, compared to the 14.8% CFR for 80+ year olds. [1] Such differences will affect disease-control policies, and knowing how heterogeneities in the population influence the spread of disease and the probability of a major outbreak (PMO) will allow for better policy decisions.

Interestingly, factors affecting the transmission of disease for distinct age groups may impact the outbreak in the population as a whole. Looking at data provided by the UN [2], as well as data provided by the national health ministries of Italy [3] and South Korea [4], we can see a situation in which the age-dependent contact structure of a population affects the outcome of an epidemic. Fig.1A depicts the similarity between demographic pyramids of Italy and South Korea, whilst Fig.1B shows the difference in total number of infected cases by age group. While such differences can be attributed to differences in severity and timescales of government interventions, another factor that has been postulated to have influenced the result

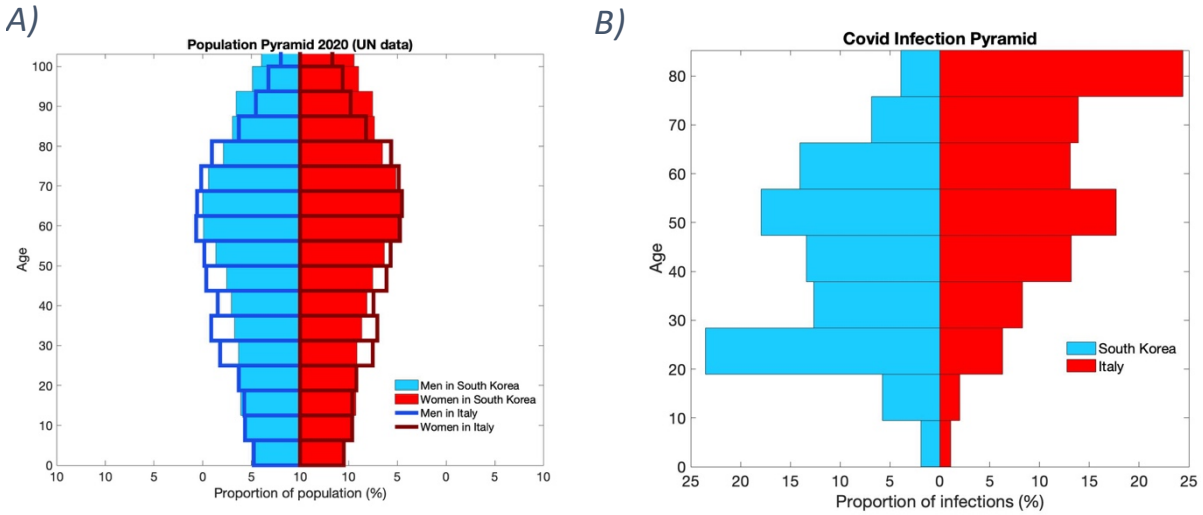

Figure 1: A) Demographic structure of South Korea and Italy, B) the proportion of infections in each age group.

of the pandemic was the differing contact pattern between age-groups.

Dowd et al [1] explain the relatively low infection rate in South Korea through the age structure of the initial outbreak – the first cases were amongst the relatively young Shincheonji religious group, which may have led to the relatively low CFR in South Korea, and most infectious concentrated in the 20-30 age group. Italy has also had proportionately more infections, with approx.. 0.0044% of the total population infected, compared to South Korea's 0.00035% of the population having been infected. As with the low CFR and concentration of cases in younger age groups, Dowd et al [1] attribute the proportionately large size of the outbreak in Italy to the

contact structure of Italy – cross-generational contacts provide greater transmission between older age groups and younger age groups.

Contact structure is by far not the only age-group-dependent factor that affects the probability and size of an outbreak. Recent literature has shown an age-dependent difference in susceptibility to disease: Zhang et al. analysed the contact records and infection status of 7375 symptomatically infected individuals and found that individuals aged 0-14 were a third as infectious as individuals aged 15-64. Individuals over the age of 65 were 1.5 times as infectious as individuals aged 15-64.

The effect of such differences in susceptibility (as well as differences in infectiousness) on the PMO has not been extensively examined. In this report, I develop a simple model based on birth (infection) and death (recovery) processes with incorporated age-group-dependent susceptibility and infectiousness to calculate the PMO.

2. The original model

Recall that the original model (developed last time) was a system of simultaneous non-linear equations that provide the solution to the sub-population-specific PMO (i.e.: the PMO in the whole population if an infected individual is introduced from a specific sub-population). Let r_i be the probability of extinction of the disease conditional on the index case being in population i, so the probability of a major outbreak given that the index case is in i is $1 - r_i$. Considering only the contact rates (number of contacts per day) C_{ij}, we multiply the transmission term by a β-value, which is the probability of an infection given a contact. We assume this is constant across all populations. μ is the recovery rate, and n is the total number of subpopulations. Conditioning on the next event once an infected is introduced, the following system of equations is developed:

$$r_k = r_k \left(\sum_{j=1}^{n} \frac{\beta C_{kj}}{\mu + \sum_{l=1}^{n} \beta C_{kl}} r_j \right) + \frac{\mu}{\mu + \sum_{l=1}^{n} \beta C_{kl}}, for\ k = 1, 2, \dots, n \quad (1)$$

3. The new model with incorporated susceptibility/infectiousness

I incorporate susceptibility and infectiousness by splitting up the probability of infection upon contact. So let $\beta_{ij} = b_i \sigma_j$, where b_i is the 'infectiousness' of individuals from sub-population i, and σ_j is the susceptibility of individuals from sub-population j. Susceptibility has been previously defined in literature as the 'rate' of becoming infected given exposure to a case of infection. Zhang et al [5] define it as a ratio where one sub-population is set as rate 1, and the susceptibility of all other sub-populations are specified in terms of this. I define infectiousness as the rate at which infected individuals will expose a susceptible individual to infection. This can be influences via face-mask wearing, keeping distance etc.

The model with the newly incorporated parameters is as follows:

$$r_k = r_k \left(\sum_{j=1}^{n} \frac{\beta_k C_{kj} \sigma_j}{\mu + \sum_{l=1}^{n} \beta_k C_{kl} \sigma_l} r_j \right) + \frac{\mu}{\mu + \sum_{l=1}^{n} \beta C_{kl}}, for\ k = 1, 2, \dots, n \quad (2)$$

4. Different ways of calculating R0

The models (1) and (2) above give us a way of calculating the probability of a major outbreak given that the index case is in sub-population k. Given that N_i is the proportion of susceptibles in sub-population i, we get that:

$$\mathbb{P}(major\ outbreak) = \sum_{i=1}^{n} \mathbb{P}(major\ outbreak \mid index\ case\ is\ in\ i)\mathbb{P}(index\ case\ is\ in\ i)$$
$$= \sum_{i=1}^{n} (1 - r_i)N_i$$

The standard method of calculating the PMO is $1 - \frac{1}{R_0}$, where R_0 is the reproductive constant of the disease. Calculating R_0 is a non-trivial problem for a metapopulation model – in the standard SIR model, $R_0 = \frac{\beta}{\mu}$, where β is the transmission rate. With varying transmission rates, the standard way to calculate R_0 is the next-generation matrix [6-8]. Keeling and Rohani [6] describe the next-generation matrix for two sub-population. I extend this concept to n sub-populations. The transmission of disease from an individual in sub-population i to an individual in sub-population j occurs at a rate $b_i C_{ij} \sigma_j$, so an individual from j gets infected by someone from i as exactly this rate. This means the infection classes of model (2) satisfied the following system of ODEs:

$$\frac{dI_i}{dt} = -\mu I_i + \sum_{k=1}^{n} b_k C_{ki} \sigma_i I_k S_i\,, for\ i = 1, \dots, n$$

The next-generation matrix is found by linearising about the disease-free steady state, $(I_1, I_2, \dots, I_n) = (N_1, \dots, N_n)$. So R_0 is the dominant eigenvalue of the following matrix:

$$\frac{1}{\mu}\begin{pmatrix} b_1 C_{11}\sigma_1 S_1 & \dots & b_n C_{n1}\sigma_1 S_1 \\ \dots & \dots & \dots \\ b_1 C_{1n}\sigma_n S_n & \dots & b_n C_{nn}\sigma_n S_n \end{pmatrix}$$

Another way one could calculate R_0 is by examining the meaning of R_0 in a probabilistic sense (I call this the 'heuristic method'). When I_s is the number of new infecteds in the second wave,

$$R_0 = \mathbb{E}(I_s) = \sum_{k=1}^{n} \mathbb{E}(I_s \mid\ primary\ infecteds\ in\ N_k)\mathbb{P}(primary\ infective\ in\ N_k)$$

$$Hence\ R_0 = \sum_{k=1}^{n} \left(\frac{1}{\mu} \sum_{l=1}^{n} b_k C_{kl} \sigma_l \right) N_k$$

5. Testing the models: comparing the 2-population case to stochastic simulations

I fix the value of R_0, and use either the heuristic or next-generation method to calculate β values that yield such a value. I use contact matrix data from Prem et. al. [9] Fig.2 depicts the results of calculating the PMOs for a two-population model with equal susceptibilities and infectiousness, corresponding to model (1). The continuous line shows the result of stochastic simulations, which were ran in the form of a 2-population Gillespie algorithm. Calling $b_i \sigma_j = \beta$ for all $i, j = 1, \dots, n$, it can be seen that Fig.2A depicts the results of the models with β-values calculated using the next-generation matrix calculations for R_0. Fig.2B depicts the same result, but with β-values calculated using the heuristic method. Note that in the following figures (Fig 2-5), the dotted lines that represent PMO child and PMO adult are the sub-population-specific

$(1 - r_i)$ values. So 'PMO child' means the PMO in the population as a whole conditional on the index case being a child.

Model 1 matches up with the stochastic simulations, so there is reason to believe that it is a realistic model. I compare Model 2 (keeping either susceptibility or infectiousness constant) to the stochastic simulations. Fig. 3 shows the same comparison, but for a model that incorporates susceptibility whilst keeping infectiousness constant. I used data from Zhang et al on relative susceptibility of children compared to adults. It appears that the model matches up

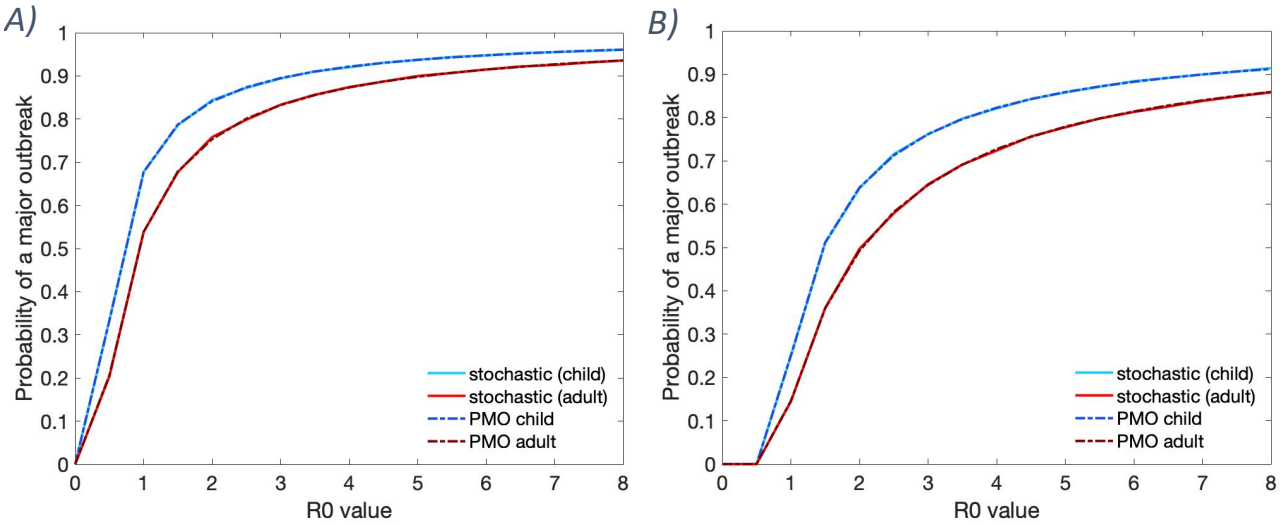

Figure 2: PMOs calculated using model (1) (dotted lines) and stochastic simulations (continuous lines). The figures depict the model with β-values calculated using A) the next-generation matrix and B) heuristic formula for R_0.

very well for the heuristic calculations of R_0 (Fig.3B), but less so for next-generation matrix

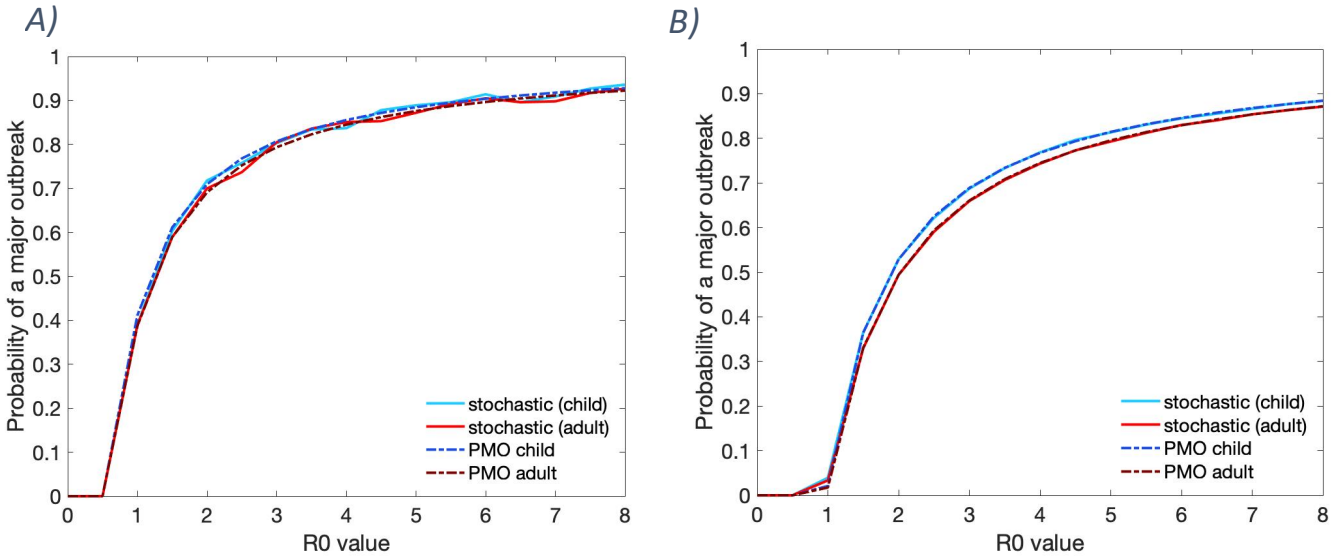

Figure 3: PMOs calculated using model (2) while varying susceptibility and keeping infectiousness constant (dotted lines) and stochastic simulations (continuous lines). The figures depict the model with β-values calculated using A) the next-generation matrix and B) heuristic formula for R_0.

calculations of R_0 (Fig.3A), though it is unclear to me as to why this would be the case.

Lastly, I varied the infectiousness rates, while keeping susceptibility constant, for both heuristic and next-generation calculations of R_0. The results are seen in Fig.4, where I assumed adults were 3 times as infectious as children. I varied infectiousness in children and adults to the same extent as I varied susceptibility in children and adults. Comparing Fig. 4 with Fig. 3, it is clear that varying susceptibility lowers the difference in PMO values for index cases child and adult, whereas varying infectiousness does not (or at least not to the same extent). This implies that if children have varying susceptibility to illness to adults (in this case, less than adults),

then the PMO is equal regardless of whether the index case is a child or an adult. However, if we take into account varying infectiousness of children and adults (with children being less infectious than adults), then there is a significant difference in between the PMO if the index case is a child, and the PMO if the index case is an adult.

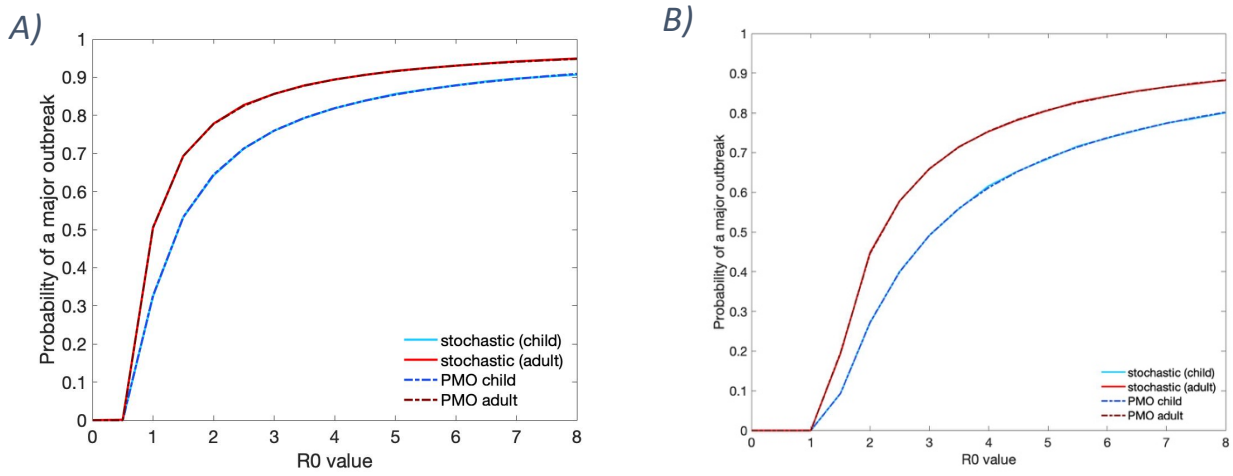

Figure 5: PMOs calculated using model (2) while varying infectiousness and keeping susceptibility constant (dotted lines) and stochastic simulations (continuous lines). The figures depict the model with β-values calculated using A) the next-generation matrix and B) heuristic formula for R_0.

Comparing Fig. 2 and Fig. 4, it is interesting to note that under the assumptions of equal infectiousness and equal susceptibility, children pose a higher risk than adults. However, if children are less infectious than adults, adults pose a higher risk than children. If children are less susceptible to disease than adults, then adults and children pose an equivalent health risk in terms of the likelihood of a major outbreak.

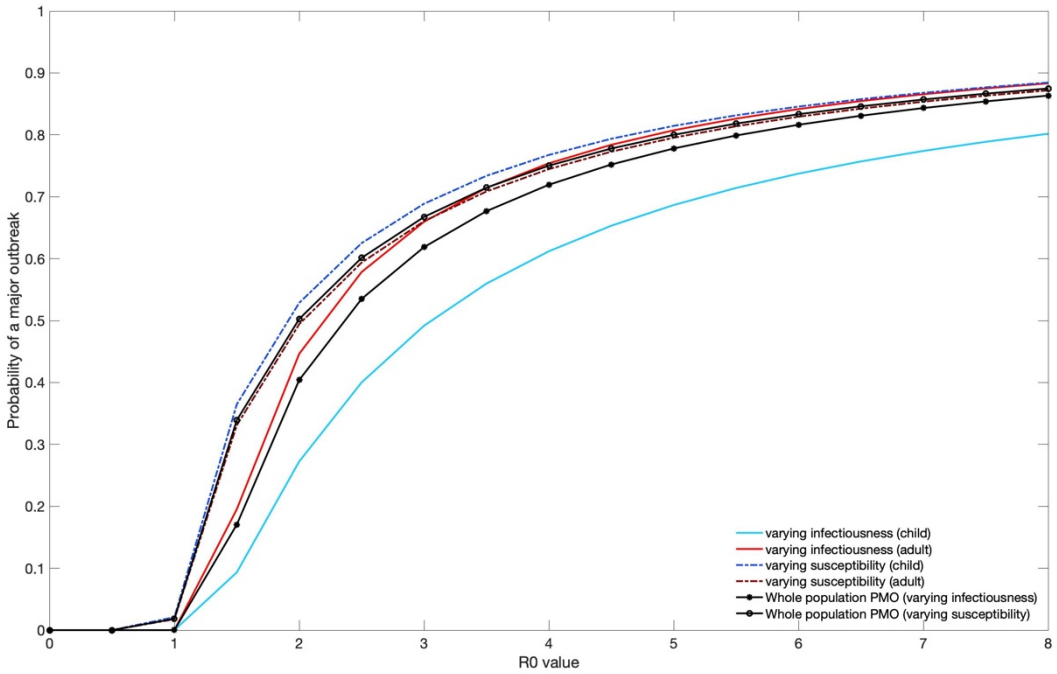

Figure 4: Comparison of sub-population-specific PMO values with total-population PMO values for varying susceptibility and varying infectiousness.

Fig. 5 gives a comparison of the sub-population-specific PMO values with the total-population PMO values for cases of varying susceptibility and varying infectiousness. The coloured lines depict sub-population specific PMO values. Interestingly, the total population PMO values for the two models converge as R_0 is large. However, varying infectiousness significantly changes

the PMO value when the index case is a child, and increases the differences between PMO values for index cases adult and child.

In order to understand this, we can compare the analytic calculations of R_0 when susceptibility is varied and when infectiousness is varied. Varying susceptibility gives us a heuristic formula as follows:

$$R_0 = \sum_{k=1}^{2} \left(\frac{b}{\mu} \sum_{l=1}^{2} C_{kl}\sigma_l \right) N_k$$

where b is the constant infectiousness. If we fix R_0, we need to vary b (as C_{ij}, σ_i, μ and N_i are fixed for $i, j = 1, 2$). Varying infectiousness gives us the following heuristic formula:

$$R_0 = \sum_{k=1}^{2} \left(\frac{\sigma}{\mu} \sum_{l=1}^{2} b_k C_{kl} \right) N_k$$

Again, fixing R_0 means that we must vary σ, the constant susceptibility to disease across the population. The numerical results of keeping R_0 constant and varying either the susceptibility or the infectiousness are shown in Fig 6. From the figure, we can see that the infectiousness parameter (i.e. the value of the constant infectiousness across populations) is larger than the susceptibility parameter (the value of the constant susceptibility across populations). As both of these parameters are directly proportional to the transmission rates, this means that varying susceptibility will lead to a higher transmission rate, and thus a higher PMO.

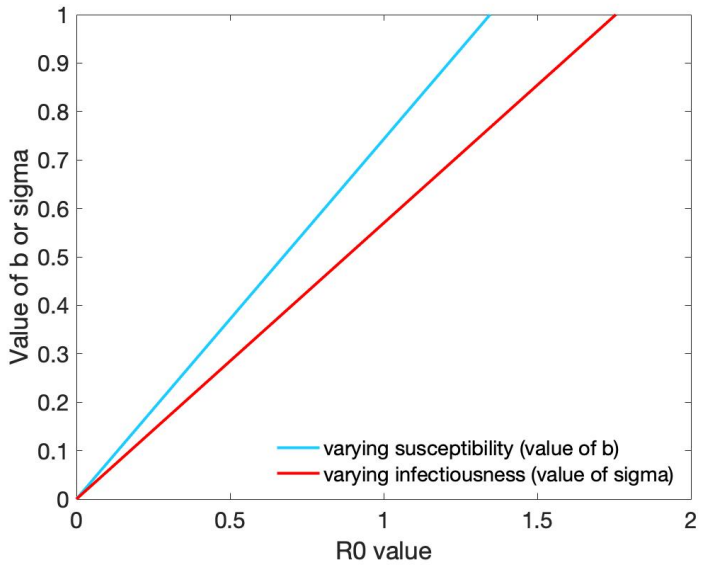

Figure 6: Keeping R_0 constant, we vary either the susceptibility or the infectiousness, and calculate the constant infectiousness parameter (blue) or susceptibility parameter (red), respectively.

Furthermore, from the analytic calculations of R_0 we can see that if all parameters are kept constant and equal, the total number of contacts every individual from one population has with every individual from the other population will vary the value of R_0. Let R_{0k} be the R_0 value calculated by keeping the infectiousness constant and varying the susceptibility. Let R_{0b} be the value calculated by keeping the susceptibility constant and varying infectiousness. Let N_i be the number of people in sub-population i (not the proportion!). Now letting $b = \sigma$ (the value of the constant infectiousness or constant susceptibility) and $b_i = \sigma_i$, we get:

$$R_{0\sigma} = \frac{b}{\mu}(N_1 C_{11}\sigma_1 + N_1 C_{12}\sigma_2 + N_2 C_{21}\sigma_1 + N_2 C_{21}\sigma_2)$$

$$R_{0b} = \frac{\sigma}{\mu}(N_1C_{11}b_1 + N_1C_{12}b_1 + N_2C_{21}b_2 + N_2C_{21}b_2)$$

Cancelling the equal terms, we see that $R_{0\sigma} - R_{0b} = (N_1C_{12} - N_2C_{21})(\sigma_2 - \sigma_1)$. With the numerical values for population sizes, contact rates and infectiousness/susceptibility taken from [2], [9] and [5] respectively, $(N_1C_{12} - N_2C_{21}) > 0$ and $(\sigma_2 - \sigma_1) < 0$. This means that $R_{0k} - R_{0b} < 0$, so $R_{0\sigma} < R_{0b}$. Hence the average number of new infections arising from one primary infected individual in a model where susceptibility is varied and infectiousness is kept the same is higher than the average number of new infections in a model where infectiousness is varies and susceptibility is kept the same (given that infectiousness and susceptibility are varied to the same extent in the same sub-populations). This explains why the PMO values are higher for varying susceptibilities than varying infectiousness.

6. PMO values at small R_0 for the 2-population case

Another interesting phenomenon that arose from conducting simulations and running the model over various R_0 values was the PMO values between $0 < R_0 < 1$. From the definition of R_0 as the 'average number of secondary infections arising from a primary infection', if $R_0 < 1$, then there cannot be a major outbreak. However, from the figures above, the PMO values for both the probabilistic model and the stochastic model are, for some calculations of R_0 and for some transmission rates, greater than 0. Fig. 6 shows the PMO values for small values of R_0 – A) depicts the sub-population-specific PMO values for $0 < R_0 < 1$, and B) depicts the sub-population-specific R_0 values. These values are defined as "the expected number of infections from one infected individual from sub-population i", and calculated as:

$$R_i = \sum_{j=1}^{n} \frac{\beta_j C_{ji} \sigma_i}{\mu}$$

From Fig.7B, it can be seen that the sub-population R_0 values can be greater than 1 even when the total population R_0 value is less than 0. Even if one of the sub-populations have an R_0 value greater than 1, the PMO for both sub-populations can be non-zero.

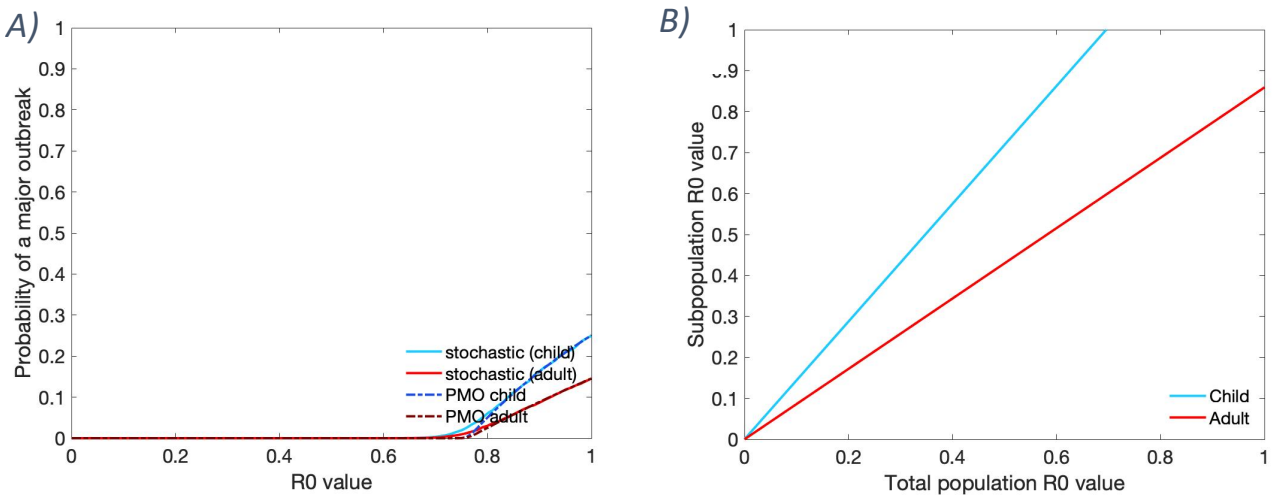

Figure 7: Using heuristic calculations for R_0, A) depicts the sub-population-specific PMO values for $0 < R_0 < 1$. B) depicts the sub-population-specific R_0 values, which were also calculated heuristically.

Similarly, Fig.8 shows that the sub-population-specific R_0 values determine the PMO for model 2 as well, where I vary susceptibility and keep infectiousness constant.

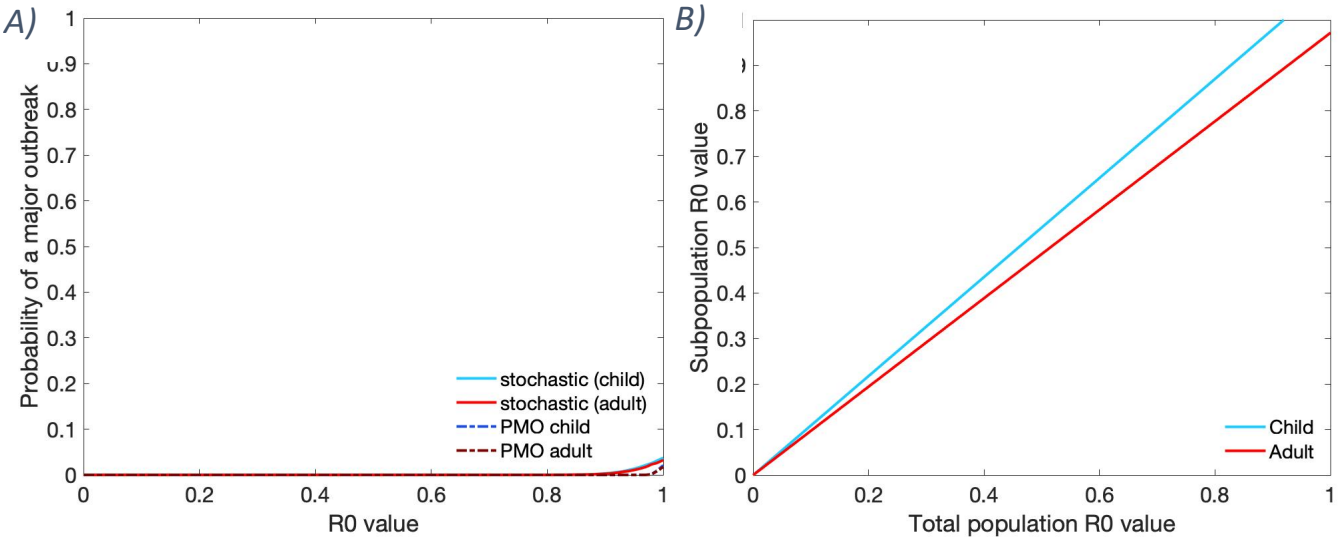

Figure 8: Using heuristic calculations (with incorporated susceptibility) for R_0, A) depicts the sub-population-specific PMO values for $0 < R_0 < 1$. B) depicts the sub-population-specific R_0 values, which were also calculated heuristically.

As can be seen from Fig. 7 and Fig. 8, the PMO is equal to or close to 0 for $0 < R_0 < 1$ when the transmission terms are calculated using the heuristic formula for R_0. However, when the transmission terms are calculated using the next-generation matrix method, the PMO values can be very high for small values of R_0. For example, Fig. 9 depicts the PMO values for $0 < R_0 < 1$ when the transmission parameters are calculated using the next-generation matrix, and the PMO value is not close to 0 for $R_0 > \sim 0.4$.

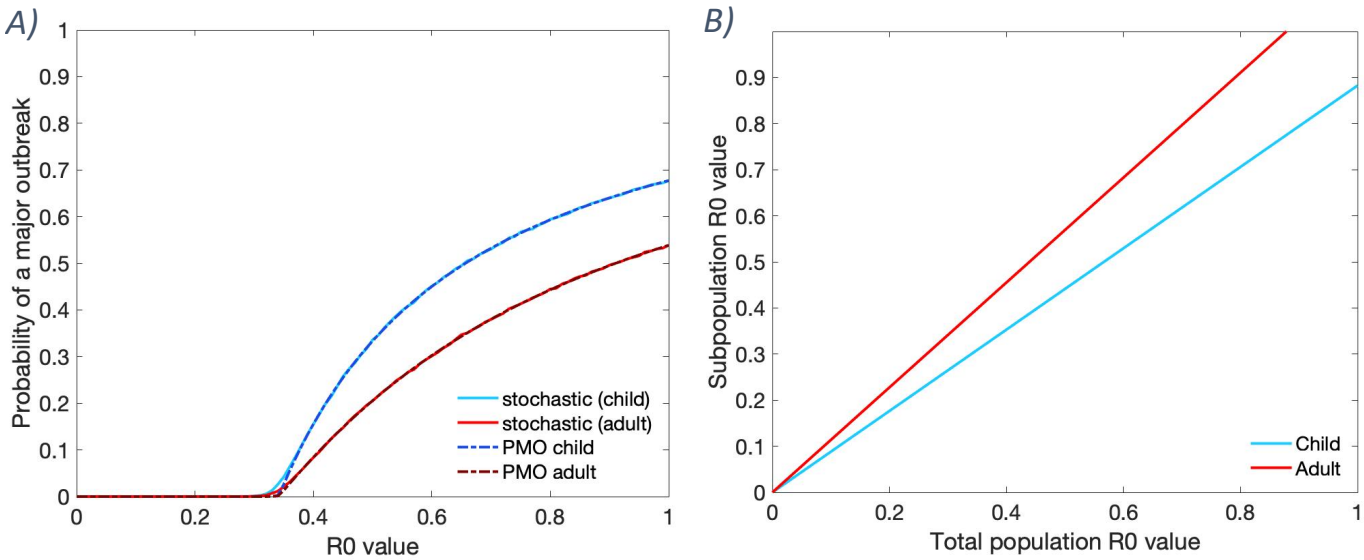

Figure 9: Using next-generation matrix calculations, the PMO values for the model with no varying susceptibility or infectiousness are calculated in A), and the subpopulation R_0 values in B).

In order to examine this phenomenon further, I look at the time series of a major outbreak in an agent-based stochastic simulation for low R_0 values in a model with no varying susceptibility or infectiousness, with R_0 values calculated using the next-generation matrix method. Fig. 9 depicts four simulation runs of the model with $R_0 = 0.8$ where the index case is in the 'child' subpopulation. (I ran the model four times and took the first four runs.) After running the model multiple times, it appears that the adult population ('population y') rarely experiences a major outbreak when the index case is a child. Fig. 10 depicts four runs of the stochastic simulation with the index case in the adult subpopulation. As can be seen, there are cases in which both the adult and child population develop outbreaks of the disease. This is because at $R_0 = 0.8$, the subpopulation-specific R_0 values are close to 1 (although neither of them are 1).

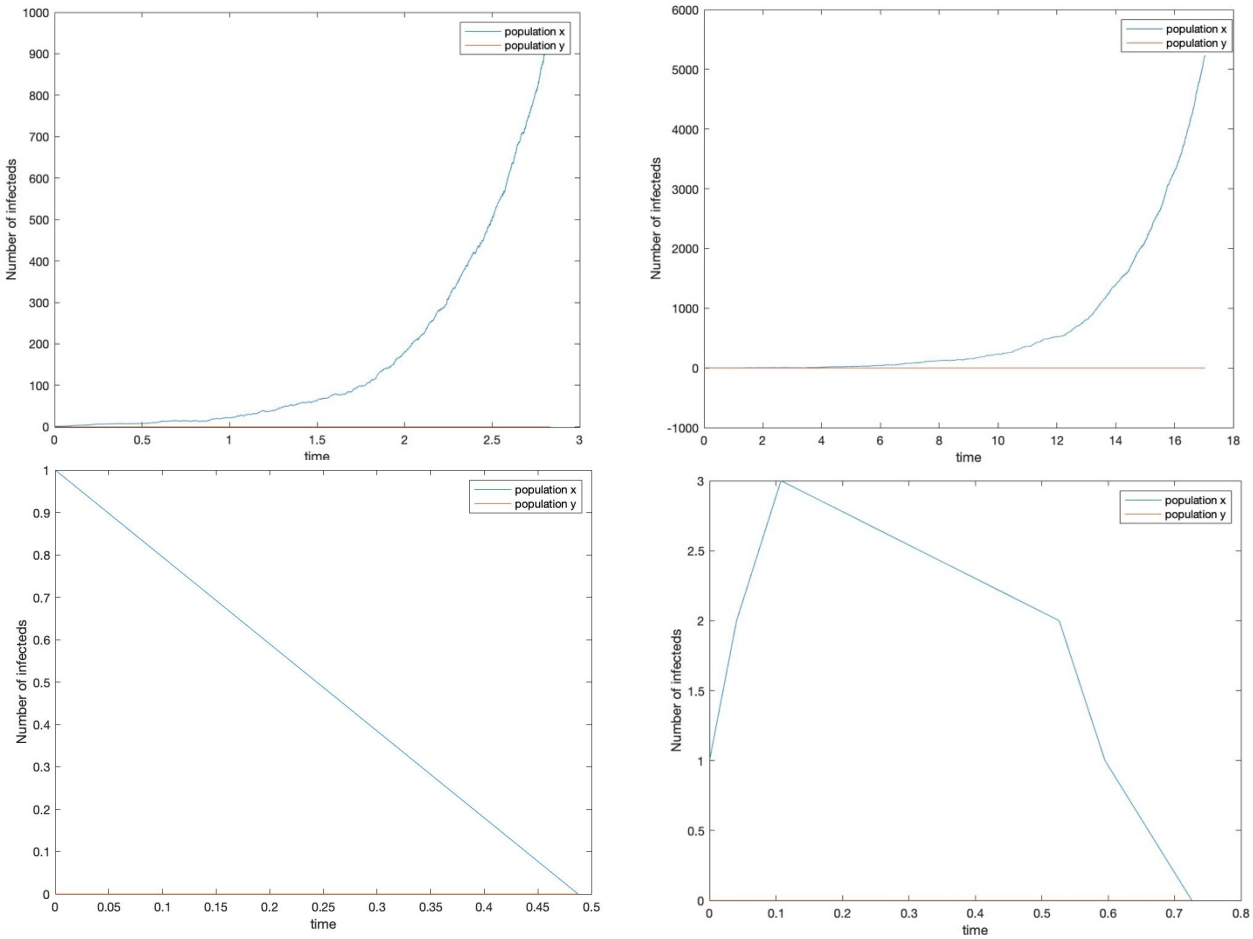

Figure 10: Four runs of the stochastic model with one initial case in the 'child' subpopulation with R_0= 0.8, where 'population x' is the child population, and 'population y' is the adult population.

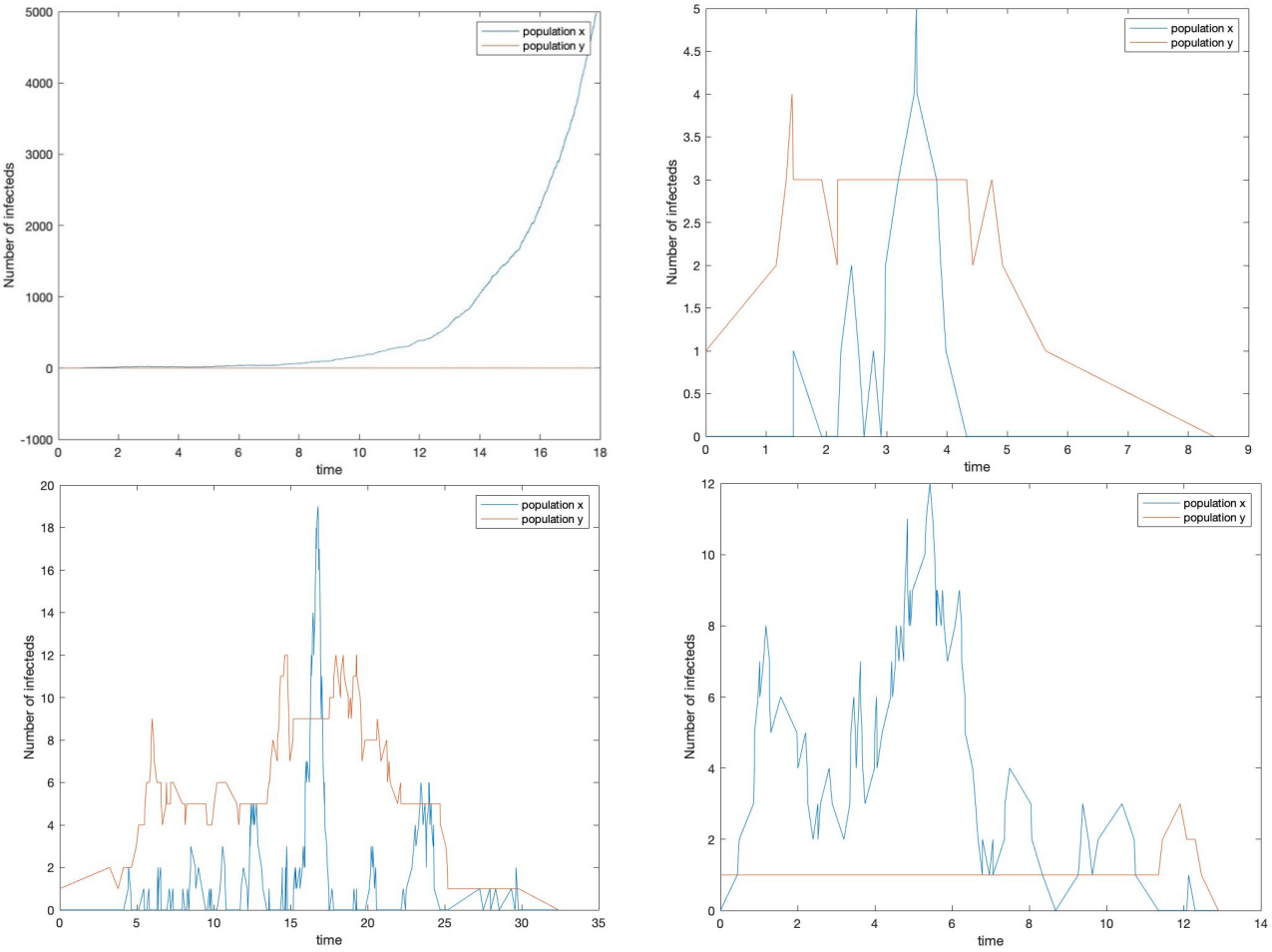

Figure 11: Four runs of the stochastic model with one initial case in the 'child' subpopulation with R_0= 0.8, where 'population x' is the child population, and 'population y' is the adult population.

7. Bibliography

[1] Dowd, Jennifer Beam, et al. “Demographic Science Aids in Understanding the Spread and Fatality Rates of COVID-19.” Proceedings of the National Academy of Sciences of the United States of America, vol. 117, no. 18, 5 May 2020, doi:10.1101/2020.03.15.20036293.

[2] United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Population Prospects 2019, Online Edition. Rev. 1*.

[3] Istituto Superiore di Sanità. Epidemia COVID-19 - Aggiornamento nazionale, 11 agosto 2020, page 11. ID: 1103023.

[4] MOHW (South Korea). August 2020, ncov.mohw.go.kr. ID: 1102730.

[5] Zhang, Juanjuan, et al. “Changes in Contact Patterns Shape the Dynamics of the COVID-19 Outbreak in China.” *Science*, vol. 368, no. 6498, 2020, pp. 1481–1486., doi:10.1126/science.abb8001.

[6] Keeling, Matthew James., and Pejman Rohani. Modeling Infectious Diseases: in Humans and Animals. University Press, 2008.

[7] Jones, James. "Notes On R0" . Stanford University.

[8] Linda J. S. Allen & Glenn E. Lahodny Jr (2012) Extinction thresholds in deterministic and stochastic epidemic models, Journal of Biological Dynamics, 6:2, 590-611, DOI: 10.1080/17513758.2012.665502

[9] Prem, Kiesha, et al. “Projecting Social Contact Matrices in 152 Countries Using Contact Surveys and Demographic Data.” PLOS Computational Biology, vol. 13, no. 9, 2017, doi:10.1371/journal.pcbi.1005697.