An SEIR Model with Contact Tracing and Age-Structured Social Mixing for COVID-19 outbreak

Ali Teimouri*

*London Institute for Mathematical Sciences,
35A South St, Mayfair, London, United Kingdom.*

July 5, 2020

Abstract

In December 2019 a severe acute respiratory syndrome now known as SARS-CoV-2 began to surge in Wuhan, China. The virus soon spread throughout the world to become a pandemic. Since the outbreak various measures were put in place to contain and control the spread, these interventions were mostly based on compartmental models in epidemiology with the main goal of controlling and monitoring the rate of the basic and effective reproduction number. In this paper, we propose an SEIR model where we incorporate contact tracing and age-structured social mixing. We show the explicit relation between contact tracing and social mixing and other relevant parameters of the proposed model. We derive a formula for the effective reproduction number which is expressed in terms of reported cases, tracing quantities and social mixing. We use this formula to determine the expectation value of the effective reproduction number in London, UK.

1 Introduction

COVID-19 is a fast-spreading infectious disease which is caused by the novel coronavirus SARS-COV-2 [1, 2]. The rapid rate of the spread and lack of preparation soon translated to a pandemic causing major disruption globally [3, 4]. From the early days of the spread, many well-established models were proposed to contain and control the outbreak. These models were mainly the extensions of the simple SIR model [5]-[11] where the population is being divided into susceptible, infected and recovered compartments. Various parameters can be incorporated into the models to study and analyse the effect of intervention methods and also to predict the spread of the infectious disease.

In this paper, we propose an SEIR model integrated with contact tracing and age-structured social mixing. Contact tracing is an important intervention method when it comes to controlling an infectious outbreak [12]. As the name suggests it involves tracing the contacts of a reported infectious individual and remove them into isolation if needed, so further spread would be prevented. There are numerous methods of contact tracing and this intervention strategy proved to be effective in controlling the COVID-19 outbreak in countries such as China, South Korea, Taiwan and Japan [14]-[16]. In this study, we adopt contact tracing appropriate to the characteristics of the COVID-19. It is now established that infected individuals with the COVID-19 are distinguished into two groups of asymptomatic and symptomatic [17]. The asymptomatic individuals are infectious individuals who show no symptoms while symptomatic individuals are those who show symptoms, and hence are reported and might need hospitalisation. To this end, our model traces the contact of those individuals who are symptomatic and reported. The rationale of the model is that if the the traced contacts show symptoms they will be removed into

* Electronic Address: ilia.teimouri@gmail.com

isolation, and if they do not show symptoms upon tracing they will be monitored by healthcare workers or via a mobile application for the recommended time of two weeks by the WHO [12].

Another feature that we integrate into our model is age-structured social mixing matrix. This component of the model is important when it comes to effectively classifying the risk within different age groups in a society. One of the known characteristics of the COVID-19 is that it affects mid-age and elderly people more severely than those who are in the young age group [18]. As a result, to implement interventions such as social-distancing within households or at workplaces, it is important to know the mixing between different age groups. In this work, we show the theoretical scaling effect of the social mixing matrix on the value of the effective reproduction number.

We begin by giving an overview of the SEIR model in section 2.1, we then incorporate contact tracing and social mixing respectively into our model in sections 2.2 and 2.3. We then analyse our results in section 3 and provide a deterministic formula for keeping the effective reproduction number below one. We calculate the change in the effective reproduction number when contact tracing and social mixing are involved. In section 3.1 we provide an example where we determine the value of the expected effective reproduction number for London, United Kingdom (UK).

2 Model and Methods

2.1 Toy SEIR model

We begin by considering a simple and adequate SEIR compartmental model which captures the essence of the spread of the COVID-19 through a homogenous population. In this model, the population N is the the sum of Susceptible (S), Exposed (E), Symptomatic (I_s), Asymptomatic (I_a) and Recovered (R) individuals (i.e. $N = S + E + I_s + I_a + R$). By definition, exposed individuals are those incubating the disease and will be infected by the symptomatic and asymptomatic individuals, symptomatic individuals are those who are infected, show symptoms and will be reported or hospitalised, asymptomatic infected individuals are those who have mild or no symptoms and hence will not be reported and do not require hospitalisation. Both infected groups will eventually move into the recovered compartment, which includes recovered and dead cases. The system of the ordinary differential equations (ODE) which describe the dynamics of this model is,

$$\begin{aligned} S'(t) &= -\frac{\beta I_s S(t)}{N} - \frac{\mu \beta I_a S(t)}{N}, \\ E'(t) &= \frac{\beta I_s S(t)}{N} + \frac{\mu \beta I_a S(t)}{N} - \frac{E(t)}{\tau}, \\ I'_s(t) &= \alpha \frac{E(t)}{\tau} - \frac{I_s(t)}{T_s}, \\ I'_a(t) &= (1 - \alpha) \frac{E(t)}{\tau} - \frac{I_a(t)}{T_a}, \\ R'(t) &= \frac{I_s(t)}{T_s} + \frac{I_a(t)}{T_a}. \end{aligned} \tag{2.1}$$

The ODEs are parameterised by β which is the transmission rate for infectious individuals, μ that is the multiplicative factor reducing the transmission rate of the asymptomatic infected patients, τ , the mean incubation period in days, α , the fraction of the reported/hospitalised cases, T_s , the time from symptom onset until hospitalisation/isolation or reporting in days and T_a , the mean infectious period in days for the asymptomatic individuals. Note that T_s is the same for the cases which are hospitalised, go to isolation or just being reported. The distinction of infected individuals into two groups is essential in this work as in reality only the symptomatic cases are being recorded and hence can trigger contact tracing [31]. Note that there are studies that take T_s and T_a as equal [32].

The basic reproduction number $\mathcal{R}_0$ is defined as average number of secondary infections produced by an infected individual in a population where everyone is susceptible, this number is used to measure the transmission potential of a disease. In a realistic world a population can rarely be wholly susceptible. Some individuals can be immune, for instance due to prior infection. As a result, not all contacts will become infected and the average number of secondary cases per an infectious case will be lower than the basic reproduction number. To this end, the effective reproduction number $\mathcal{R}_e$ is defined as the average number of secondary cases in a population made up of both susceptible and non-susceptible individuals. The effective reproduction number is proportional to the product of the basic reproduction number and the fraction of the host population that is susceptible. Using the next generation matrix method [19] the effective reproduction number, $\mathcal{R}_e$, is calculated by

$$\mathbf{K}_L = -\mathbf{T}\mathbf{\Sigma}^{-1} \quad (2.2)$$

where the matrix $\mathbf{T}$ corresponds to transmissions and the matrix $\mathbf{\Sigma}$ corresponds to transitions. Note that all the epidemiological events that lead to new infections are incorporated in the model via $\mathbf{T}$ and the rest of the events via $\mathbf{\Sigma}$, hence progress to either dead or recovery makes $\mathbf{\Sigma}$ invertible. We obtain $\mathbf{K}_L$ as,

$$\begin{aligned} \mathbf{K}_L &= - \begin{pmatrix} 0 & \beta & \mu\beta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} -\tau & 0 & 0 \\ -\alpha T_s & -T_s & 0 \\ -\tau T_a T_s \left(\frac{1}{\tau T_s} - \frac{\alpha}{\tau T_s} \right) & 0 & -T_a \end{pmatrix} \\ &= \begin{pmatrix} \beta(\alpha T_s + (1-\alpha)\mu T_a) & \beta T_s & \beta\mu T_a \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \end{aligned} \quad (2.3)$$

the dominant eigenvalue of this matrix is equal to the effective reproduction number,

$$\mathcal{R}_e = \beta(\alpha T_s + (1-\alpha)\mu T_a) = \mathcal{R}_0 \left[\alpha \frac{T_s}{T_a} + \mu(1-\alpha) \right] \quad (2.4)$$

where $\mathcal{R}_0 = \beta T_a$ is defined as the basic reproduction number without cases being reported or hospitalised. We assumed that the population size N is constant over the short time of the outbreak and hence we absorb it into parameter β (i.e. we do not consider birth and death rates). Furthermore, we take the total population $N = 1$ so the results would be per-capita. We note that for $T_s = T_a$ and $\mu = 1$ the relation in the above equation would reduce to $\mathcal{R}_e = \mathcal{R}_0$ where the assumption is that the basic reproduction number is only proportional to transmission rate and the mean infectious period. Namely this ratio would be the expected number of new infections from a single infection in a population where all individuals are susceptible.

2.2 SEIR model with contact tracing

To incorporate contact tracing into our model we follow the steps of [5]. Contact tracing in this manner is based on three elements, contact identification, contact listing and follow-up. This process begins once an infected (dead or alive) case is being reported, at this stage one lists the contacts of the reported person, then a team of contact tracers track the listed contacts (this could also be done via a mobile application). If the traced contacts show symptoms at the time of tracking they will be removed by isolation or quarantine. Otherwise, they will be instructed to update their status if they develop any symptoms over 14 days. The follow-up team would then look after the traced contacts during this period and will remove them if symptoms appear.

There are two assumptions to proceed with this model:

- Only the reported or hospitalised cases can trigger contact tracing.
- Contact of a reported patient are traced with probability ϕ . There is also a fixed delay δ between index case reporting and contact tracking. If the traced individuals are symptomatic at the time of contact, they will be removed immediately, otherwise they follow up the monitoring protocol.

Untraced contacts have probability $(1 - \phi)$, these are the cases that are missed and did not trigger tracing from the reported cases. We shall now define the cumulative number, at time t , of traced infected individuals who are in exposed or infectious group when first tracked,

$$C'_e(t + \delta) = \underbrace{\frac{I_s(t)}{T_s}}_{(i)} \int_0^\infty \underbrace{\frac{e^{-z/T_s}}{T_s}}_{(ii)} \underbrace{\int_{t-z}^t}_{(iii)} \underbrace{F_E(t-r+\delta)}_{(iv)} \underbrace{\phi\beta S(r)}_{(v)} dr dz, \quad (2.5)$$

$$C'_i(t + \delta) = \frac{I_s(t)}{T_s} \int_0^\infty \frac{e^{-z/T_s}}{T_s} \int_{t-z}^t (1 - F_E(t-r+\delta)) \phi\beta S(r) dr dz, \quad (2.6)$$

where:

- i is the rate of reporting of infectious individuals,
- ii is the reported case infectious period probability density,
- iii is the integral over infectious period of length z ,
- iv is the probability still incubating $t - r + \delta$ units after infection and
- v is the traced transmissions infected at $r \in [t - z, t]$.

We assumed that the incubation period cumulative distribution F_E is exponentially distributed. Since the average infectious period is short the the change in population size is negligible and one assumes that S is approximately constant over this period. Hence, the total number of the traced transmissions at time $t + \delta$ is,

$$C'(t + \delta) = C'_e(t + \delta) + C'_i(t + \delta) \approx \frac{\beta e^{-\delta} \phi S(t) I_s(t)}{T_s + 1} + \beta \phi S(t) I_s(t) \left(1 - \frac{e^{-\delta}}{T_s + 1}\right) = \beta \phi S(t) I_s(t). \quad (2.7)$$

We now define four additional compartments, E_c^e are the exposed individuals who will be traced and *incubating* when they were tracked, E_c^i are the exposed individuals who will be traced and *infectious* when tracked. I_c^e are the monitored (and under monitoring) infectious individuals who *have shown no symptoms* (while incubating) at the time of first contact, and I_c^i are the infectious individuals who should be removed as they were *found symptomatic* at the time of first contact.

The probability p_e of a traced contact being incubating when first tracked should depend on the mean infectious period T_s , the delay between index case reporting and tracing δ , and the incubation period of the traced case τ , as a result:

$$p_e = \frac{C'_e(t)}{C'_e(t) + C'_i(t)} = \frac{\tau e^{-\frac{\delta}{\tau}}}{T_s + \tau}, \quad (2.8)$$

for $t \rightarrow \tau$. Note that we assumed the incubation period is exponentially distributed i.e. $F_E(x) = e^{-x/\tau}$, and that $S(t)$ is constant over the infectious period.

The average infectious period, prior to being tracked, for the fraction $(1 - p_e)$ individuals is defined by T_i and can be approximated by the expected value of the time after symptom onset when the contact is being tracked, $\hat{T}_i$,

$$\frac{1}{T_i} = \frac{1}{\bar{T}} + \frac{1}{\hat{T}_i}, \quad (2.9)$$

where $\bar{T} = \alpha T_s + (1 - \alpha) \mu T_a$ is the weighted average infectious period independent of contact tracing and

$$\hat{T}_i = \frac{1}{T_s} \int_0^{T_s} \left[T_s + \delta - \left(\int_r^{T_s+\delta} \frac{x e^{-(x-r)/\tau}}{\int_0^{T_s-r+\delta} e^{-s/\tau} ds} dx \right) \right] dr. \quad (2.10)$$

We assume monitored infectious individuals, I_c^e have transmissibility β_m and mean infection period T_m in days. The value of β_m and T_m directly depends on the efficiency of the contact tracing protocol. Note that for the infectious individuals, I_c^i , traced during their infectious period, we assign transmissibility β prior to tracking and mean infectious period T_i .

Contact tracing which is triggered by individuals in I_c^e and I_c^i compartments counts as secondary tracing, tertiary tracing and so on and thus we call them *higher order tracing*. In contrast, the first order tracing are the contacts of the fraction ϕ who will be traced. We shall then define probabilities p_e^e and p_e^i as probability of traced contact, infected by I_c^e individuals, and probability of traced contact, infected by I_c^i individuals, as

$$p_e^e = \frac{\tau e^{-\frac{\delta}{\tau}}}{T_m + \tau}, \quad \text{and} \quad p_e^i = \frac{\tau e^{-\frac{\delta}{\tau}}}{T_i + \tau}, \quad (2.11)$$

where we remind that T_m is the mean infectious period for the monitored infectious individuals, I_c^e , (time between symptom onset and isolation) and T_i is the mean infectious period (prior to being tracked) for the contact traced infectious individuals, I_c^i . The modified SEIR model with contact tracing is,

$$\begin{aligned} S' &= -\beta I_s S - \mu \beta I_a S - \beta I_c^i S - \beta_m I_c^e S, \\ E' &= \mu \beta I_a S + (1 - \phi) [\beta I_s S + \beta I_c^i S + \beta_m I_c^e S] - \frac{E}{\tau}, \\ (E_c^e)' &= \phi [p_e^e \beta I_s S + p_e^i \beta I_c^i S + p_e^e \beta_m I_c^e S] - \frac{E_c^e}{\tau}, \\ (E_c^i)' &= \phi [(1 - p_e^e) \beta I_s S + (1 - p_e^i) \beta I_c^i S + (1 - p_e^e) \beta_m I_c^e S] - \frac{E_c^i}{\tau}, \\ I_s' &= \alpha \frac{E}{\tau} - \frac{I_s}{T_s}, \\ I_a' &= (1 - \alpha) \frac{E}{\tau} - \frac{I_a}{T_a}, \\ (I_c^e)' &= \frac{E_c^e}{\tau} - \frac{I_c^e}{T_m}, \\ (I_c^i)' &= \frac{E_c^i}{\tau} - \frac{I_c^i}{T_i}. \end{aligned} \quad (2.12)$$

In the same manner as in the previous section the effective reproduction number can be obtained by,

$$\mathbf{K}_L = \begin{pmatrix} 0 & 0 & 0 & \beta(1 - \phi) & \beta\mu & (1 - \phi)\beta_m & \beta(1 - \phi) \\ 0 & 0 & 0 & \beta\phi p_e & 0 & \phi p_e^e \beta_m & \beta\phi p_e^i \\ 0 & 0 & 0 & \beta\phi(1 - p_e) & 0 & \phi(1 - p_e^e)\beta_m & \beta\phi(1 - p_e^i) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\tau} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\tau} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\tau} & 0 & 0 & 0 & 0 \\ -\frac{\alpha}{\tau} & 0 & 0 & \frac{1}{T_s} & 0 & 0 & 0 \\ -\frac{1-\alpha}{\tau} & 0 & 0 & 0 & \frac{1}{T_a} & 0 & 0 \\ 0 & -\frac{1}{\tau} & 0 & 0 & 0 & \frac{1}{T_m} & 0 \\ 0 & 0 & -\frac{1}{\tau} & 0 & 0 & 0 & \frac{1}{T_i} \end{pmatrix}^{-1}, \quad (2.13)$$

where the dominant eigenvalue of this matrix is,

$$\mathcal{R}_e = \beta((1 - \alpha)\mu T_a - \alpha(\phi - 1)T_s) = \mathcal{R}_0 \left[\alpha \frac{T_s}{T_a} (1 - \phi) + \mu(1 - \alpha) \right] \quad (2.14)$$

as before $\mathcal{R}_0 = \beta T_a$ is the basic reproduction number independent of T_s . α is the fraction of untraced cases which are reported and ϕ is the fraction of reported cases which are traced. The main difference of the above equation and Eq.(2.4) is the probability ϕ which measures the likelihood of contacts of a reported case being traced. In the limit $\phi \rightarrow 0$ we recover Eq.(2.4) where there is no contact tracing.

2.3 SEIR model with contact tracing and social mixing

To take the social mixing into account we rewrite the transmission coefficient as the product of a transmission β times a constant matrix C whose element $C_{\lambda\sigma}$ measures the average number of physical contacts, on daily basis, among an individual in group age λ with an individual in group age σ . It should be noted that the probability of a susceptible individual, in age class λ , to come to contact with an infected individual in class σ is equivalent to the product of the contact rate $C_{\lambda\sigma}$ times the probability I_σ that individual in class σ is infected. Hence we can denote age groups appropriately via λ and σ and our model modifies to,

$$\begin{aligned}
 S'_\lambda &= -\beta S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)s} - \mu \beta S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)a} - \beta S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)c}^i - \beta_m S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)c}^e, \\
 E'_\lambda &= \mu \beta S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)a} + (1 - \phi) \left[\beta S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)s} + \beta S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)c}^i + \beta_m S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)c}^e \right] - \frac{E_\lambda}{\tau}, \\
 (E_c^e)'_\lambda &= \phi \left[p_e \beta S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)s} + p_e^i \beta S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)c}^i + p_e^e \beta_m S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)c}^e \right] - \frac{E_{(\lambda)c}^e}{\tau}, \\
 (E_c^i)'_\lambda &= \phi \left[(1 - p_e) \beta S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)s} + (1 - p_e^i) \beta S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)c}^i + (1 - p_e^e) \beta_m S_\lambda \sum_{\sigma=1}^n C_{\lambda\sigma} I_{(\sigma)c}^e \right] - \frac{E_{(\lambda)c}^i}{\tau}, \\
 I'_{(\sigma)s} &= \alpha_\lambda \frac{E_\lambda}{\tau} - \frac{I_{(\sigma)s}}{T_s}, \\
 I'_{(\sigma)a} &= (1 - \alpha_\lambda) \frac{E_\lambda}{\tau} - \frac{I_{(\sigma)a}}{T_a}, \\
 (I_{(\sigma)c}^e)' &= \frac{E_{(\lambda)c}^e}{\tau} - \frac{I_{(\sigma)c}^e}{T_m}, \\
 (I_{(\sigma)c}^i)' &= \frac{E_{(\lambda)c}^i}{\tau} - \frac{I_{(\sigma)c}^i}{T_i}.
 \end{aligned} \tag{2.15}$$

We note that $C_{\lambda\sigma}$ denotes the interaction between different age groups and hence it does not matter what groups of infected individuals it is coupled to. It should also be noted that the age structured mixing of individuals in the age group λ change their likelihood of being exposed to the virus given certain number of infectious people in the population. Furthermore, in our model we made a dependency between α and λ (i.e. $\alpha \rightarrow \alpha_\lambda$) due to the assumption that younger individuals are more likely to be asymptomatic and less infectious [9]. The effective reproduction number is now calculated as,

$$\begin{aligned}
 \mathbf{K}_L &= \\
 &\begin{pmatrix} 0 & 0 & 0 & \beta(1 - \phi) \sum_{\sigma=1}^n C_{\lambda\sigma} & \beta \mu \sum_{\sigma=1}^n C_{\lambda\sigma} & (1 - \phi) \beta_m \sum_{\sigma=1}^n C_{\lambda\sigma} & \beta(1 - \phi) \sum_{\sigma=1}^n C_{\lambda\sigma} \\ 0 & 0 & 0 & \beta \phi p_e \sum_{\sigma=1}^n C_{\lambda\sigma} & 0 & \phi p_e^e \beta_m \sum_{\sigma=1}^n C_{\lambda\sigma} & \beta \phi p_e^i \sum_{\sigma=1}^n C_{\lambda\sigma} \\ 0 & 0 & 0 & \beta \phi (1 - p_e) \sum_{\sigma=1}^n C_{\lambda\sigma} & 0 & \phi (1 - p_e^e) \beta_m \sum_{\sigma=1}^n C_{\lambda\sigma} & \beta \phi (1 - p_e^i) \sum_{\sigma=1}^n C_{\lambda\sigma} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\
 &\cdot \begin{pmatrix} \frac{1}{\tau} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\tau} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\tau} & 0 & 0 & 0 & 0 \\ -\frac{\alpha_\lambda}{\tau} & 0 & 0 & \frac{1}{T_s} & 0 & 0 & 0 \\ -\frac{1 - \alpha_\lambda}{\tau} & 0 & 0 & 0 & \frac{1}{T_a} & 0 & 0 \\ 0 & -\frac{1}{\tau} & 0 & 0 & 0 & \frac{1}{T_m} & 0 \\ 0 & 0 & -\frac{1}{\tau} & 0 & 0 & 0 & \frac{1}{T_i} \end{pmatrix}^{-1}.
 \end{aligned} \tag{2.16}$$

The dominant eigenvalue of this matrix is of the form,

$$\mathcal{R}_e = \beta((1 - \alpha_\lambda)\mu T_a - \alpha_\lambda(\phi - 1)T_s) \sum_{\sigma=1}^n C_{\lambda\sigma} = \mathcal{R}_0 \left[\alpha_\lambda \frac{T_s}{T_a} (1 - \phi) + \mu(1 - \alpha_\lambda) \right] \sum_{\sigma=1}^n C_{\lambda\sigma}. \quad (2.17)$$

where $n \geq 1$ is the number of age groups one wishes to consider. It can be seen from the formula above that the social mixing matrix $C_{\lambda\sigma}$ has a scaling effect on increasing or decreasing the effective reproduction number.

3 Results

The impact of contact tracing on the effective reproduction number: In order to keep the spread of the virus under control we need to keep the effective reproduction number below one. If we assume the same mean infectious period for both symptomatic and asymptomatic case (i.e. $T_s \approx T_a$) and also if we take the multiplicative factor $\mu = 1$, we get,

$$\alpha_\lambda \phi < 1 - \frac{1}{\mathcal{R}_0 \sum_{\sigma=1}^n C_{\lambda\sigma}}, \quad (3.1)$$

where $\alpha_\lambda \phi$ is the *critical* proportion of the total cases which need to be traced in different age groups in order for $\mathcal{R}_e < 1$.

Furthermore, the impact of contact tracing can be quantified by considering the relative change in the effective reproduction number. From the simple SEIR model let us recall the effective reproduction number and call it, $\mathcal{R}_e^o = \mathcal{R}_0 \left[\alpha_\lambda \frac{T_s}{T_a} + \mu(1 - \alpha_\lambda) \right] \sum_{\sigma=1}^n C_{\lambda\sigma}$, where we added the social mixing term appropriately. Then the relative change in $\mathcal{R}_e$ with contact tracing would be of the form,

$$\frac{\mathcal{R}_e^o - \mathcal{R}_e}{\mathcal{R}_e^o} = \frac{\alpha_\lambda \phi T_s}{(1 - \alpha_\lambda)\mu T_a + \alpha_\lambda T_s}. \quad (3.2)$$

For $T_s \approx T_a$ and $\mu = 1$ the relative change takes the critical form of $\alpha_\lambda \phi$ which is independent of the social mixing matrix $C_{\lambda\sigma}$ and is only proportional to the probability that contacts of a reported case being traced, ϕ , and fraction of untraced cases that are reported in a particular age group, α_λ .

In literature [3], μ has a unconstrained distribution of $\mu \sim \text{uniform}(0.2, 0.55)$. α_λ has a unconstrained distribution of $\alpha_\lambda \sim \text{uniform}(0.1, 0.99)$. In Fig. 1 we plot the relative change for upper and lower values of μ while keeping $T_s \approx T_a$. Higher value of μ means higher transmission rate for asymptomatic individuals, as a result the relative change in the effective reproduction number with tracing is proportional to $\sim \alpha_\lambda \phi$ for upper bound of μ while proportional to $\sim \phi$ for lower bound of μ .

Imperfect tracing: So far we assumed perfect tracing, that is to consider that traced individuals do not cause secondary infections due to an effective monitoring and tracking policies. In other words, we assumed transmission rate for monitored individuals is zero (i.e. $\beta_m = 0$). Let us now relax this condition and assume that $\beta_m > 0$, then the full dominant eigenvalue of (2.16) is,

$$\mathcal{R}_e = \left(\mathcal{R}_0 \left[\alpha_\lambda \frac{T_s}{T_a} (1 - \phi) + \mu(1 - \alpha_\lambda) \right] + \phi \beta_m T_m p_e^e + \phi \beta T_i (1 - p_e^i) \right) \sum_{\sigma=1}^n C_{\lambda\sigma}. \quad (3.3)$$

To capture the effect of the tracing let us assume $p_e^e = 1$ that is to consider that higher order traced contact is 100% infected by I_c^e individual, let us also assume that $p_e^i = 1$ meaning that higher order traced contact is 100% infected by I_c^i individual. In that case, the effective reproduction number reduces to,

$$\mathcal{R}_e = \left(\mathcal{R}_0 \left[\alpha_\lambda \frac{T_s}{T_a} (1 - \phi) + \mu(1 - \alpha_\lambda) \right] + \phi \beta_m T_m \right) \sum_{\sigma=1}^n C_{\lambda\sigma}. \quad (3.4)$$

Figure 1: Relative change in the effective reproduction number, $\mathcal{R}_e$, given by $\frac{\alpha_\lambda \phi}{(1-\alpha_\lambda)\mu + \alpha_\lambda}$. The plot varies the values of α_λ according to its distribution (i.e. $\alpha_\lambda \sim \text{uniform}(0.1, 0.99)$) and ϕ (where $0 \leq \phi \leq 1$) for two values of $\mu = 0.2$ and $\mu = 0.55$.

We remind that I_c^e are the monitored infected individuals who have shown no symptoms upon contacting them and I_c^i are the symptomatic infectious individuals who were traced and removed into isolation. It is now trivial that the efficiency of contact tracing has a linear contribution in the value of the effective reproduction number given by $\phi\beta_m T_m$. The formula above shows that in order to have an effective monitoring policy in place the mean infectious period for monitored infectious individuals should be minimised. In other words, the time between symptom onset and isolation, T_m should be at its minimum.

The impact of age-structured social mixing on the effective reproduction number: To impose an effective distancing policy one needs to study the social mixing matrices that illustrate the average number of contacts between different age groups. Measuring the social mixing varies for different locations and could depend on the structure of households, schools and workplaces [9, 20]. In this paper, we are using the social mixing matrices constructed by [20]. The study by [20] conducted in the UK by surveying a sample of adults and recording their contact patterns after the British government put social distancing measures in place. In Fig. 2 we reconstructed the social mixing matrix from CoMIX study [20], there are six age groups participated in their study, namely 18-29, 30-39, 40-49, 50-59, 60-69, and 70+. As a result $\sum_{\sigma=1}^6 C_{\lambda\sigma}$. The CoMIX matrices show the average number of contacts between different age groups in different setups from the 24th of March 2020 one day after the lockdown was imposed in the UK. To control the spread of the virus one should monitor the number of contacts between each age group and minimise them when necessary. In the next section we use the mixing matrices in Fig. 2 to illustrate the effect of social mixing on the expected value of $\mathcal{R}_e$.

3.1 The expected value of the effective reproduction number in London

Let us recall Eq. (3.4) and rewrite it as:

$$\mathcal{R}_e = \left(k \left(\frac{1-q}{q} \right) + k_m \right) \sum_{\sigma=1}^n C_{\lambda\sigma}, \quad (3.5)$$

where we define $k = \phi\mathcal{R} = \phi\beta T_s$, in this definition $\mathcal{R}$ is the basic reproduction number in the absence of contact tracing, note that this is different than $\mathcal{R}_0$ which was defined earlier.

Figure 2: Reconstruction of CoMIX matrices as given in [20], each matrix represents a social setup and show the average number of contacts between different age groups after the lockdown was imposed in the UK on the 23rd of March 2020.

Moreover, let

$$q = \frac{\phi}{\phi + (1 - \phi)\alpha_\lambda + \mu(1 - \alpha_\lambda)\frac{T_a}{T_s}}, \quad (3.6)$$

be the fraction of reported cases which are traced contacts. In fact, q can be interpreted as the conditional probability that a transmission is traced given a cases is being reported. Finally, $k_m = \phi\mathcal{R}_m = \phi\beta_m T_m$ where $\mathcal{R}_m$ is the reproduction number of the contact traced / monitored individuals. Since k and k_m are the infected traced contacts identified per untraced and traced reported cases respectively, it is sensible to define $\mathcal{C}_j$ and $\mathcal{T}_j$ as the cumulative reported cases and traced-reported cases in week j respectively. We now define the average number of traced infected individuals identified in week $j + i$ per untraced and traced reported case from the preceding week j as,

$$k^{j,i} = \frac{\mathcal{T}_{j+i}}{\mathcal{C}_j - (1 - \theta)\mathcal{T}_j} \quad \text{and} \quad k_m^{j,i} = \frac{\theta\mathcal{T}_{j+i}}{\mathcal{C}_j - (1 - \theta)\mathcal{T}_j}, \quad (3.7)$$

where $0 \leq \theta \leq 1$ is tracing/monitoring effectiveness, in other words, $\theta = \mathcal{R}_m/\mathcal{R} \equiv k_m^{j,i}/k^{j,i}$. Furthermore, the fraction of reported cases that are traced in the week $j + i$ is

$$q_{j+i} = \mathcal{T}_{j+i}/\mathcal{C}_{j+i}. \quad (3.8)$$

We can now write the expectation value of the effective reproduction number, given in week j , as

$$\begin{aligned}\mathbb{E}(\mathcal{R}_e^{(j)}) &= \sum_i p_i \left(k^{j,i} \left(\frac{1 - q_{j+i}}{q_{j+i}} \right) + k_m^{j,i} \right) \sum_{\sigma=1}^n C_{\lambda\sigma} \\ &= \frac{\sum_{\sigma=1}^n C_{\lambda\sigma}}{C_j - (1 - \theta)T_j} \sum_i p_i (C_{j+i} - (1 - \theta)T_{j+i}).\end{aligned}\quad (3.9)$$

Above estimates $\mathcal{R}_e^{(j)}$ as the ratio of the sum of untraced and fraction of traced reported cases (θ), weighted by probabilities p_i . The probabilities p_i are calculated using the serial interval of the COVID-19 that is estimated to be a normal distribution with mean of $\mu = 3.96$ and standard deviation of $\sigma = 4.75$, [26], therefore on weekly basis we have: $p_1 = \mathbb{P}(\mathcal{N} < 7)$, $p_2 = \mathbb{P}(7 < \mathcal{N} < 14)$ and so on. In perfect tracing (i.e. $\theta = 0$) only weekly untraced reported cases ($C_i - T_i$) play role in the calculation of $\mathcal{R}_e$ while for imperfect tracing (i.e. $\theta = 1$) all the traced and untraced reported cases are weighted equally. As a result, one does not require specific data about the traced reported cases per week. Utilising θ in this manner could serve as a useful tool for sensitivity analysis and in the countries where obtaining the contact tracing data is difficult.

We are now equipped to calculate the expected value of the effective reproduction number for London. We begin by using the social mixing matrices given in Fig. 2, where there are six age groups in participants ($n = 6$) reporting their contact pattern with eight different age groups. We note that each matrix element is the average number of contacts between different age groups. As a result, the normalised sum of each matrix's row gives the value of C for the particular age class and particular setup. Using the cumulative reported cases from [27] for London and Eq. (3.9) we estimated the expected value of $\mathcal{R}_e$ - for nine weeks beginning from the 22nd of February 2020 - for different age groups in different setups in Fig. 3. Our results show that before the lockdown the mean of the effective reproduction number is 2.55 and after the lockdown is imposed the mean of the effective reproduction reduces to 0.77 over the weeks four to nine for all age groups and setups, this is comparable to results of [20] and [25] (when averaged on weekly basis).

Figure 3: The expected value of the reproduction number as calculated by the Eq. (3.9) for London over nine weeks beginning from the 22nd of February 2020 for different age groups in different social setups. The dashed grey line on each plot shows the 23rd of March 2020 when the lockdown was imposed. Additionally, the black dashed line is the value of $\mathbb{E}(\mathcal{R}_e)$ when there is only contact tracing involved and no social mixing (i.e. assuming $C_{\lambda\sigma} = 1$) in Eq. (3.9).

In estimation of $\mathbb{E}(\mathcal{R}_e^{(j)})$ we assumed that the average number of contacts between different age groups in London is roughly the same as the whole UK. We acknowledge that Londoners

form only 16.5% of the participants in CoMIX survey yet for the simplicity we assume the contact behaviour of Londoners is the same with the rest of the country. We also assumed perfect tracing in this study, the estimation of θ is the subject of upcoming work.

4 Conclusions

In this study we proposed an SEIR model with contact tracing and social mixing. We have shown that the effective reproduction number is proportional to dominant eigenvalue of the contact matrix, a result which is also derived in [21] using another method. In Eq.(2.17) we have derived contact tracing via parameters ϕ (probability of contacts of a reported being traced) and α (fraction of untraced cases who are reported) and showed that the critical value of the required tracing to keep the effective reproduction number below one is proportional to $\alpha\phi$. We also derived the effective reproduction number when there is higher order tracing in Eq. (3.4), in which case there is a linear contribution to the effective reproduction number given by $\phi\beta_m T_m$ where β_m is the transmissibility of those infected individuals who are monitored and T_m is the time between symptom onset and isolation for the monitored infectious individuals. In the case of the COVID-19 we can assume that $\beta_m \approx \beta$, yet to have an effective contact tracing and lower values of $\mathcal{R}_e$ it is important to keep T_m at its minimum - this of course requires an effective human capital to isolate the monitored cases immediately upon developing symptoms.

We furthermore gave a formula in Eq. (3.9) to estimate the expected value of the effective reproduction number $\mathbb{E}(\mathcal{R}_e)$ on a given week. This formula estimates the average reproduction number by using the serial interval of the COVID-19 and the fraction of traced and untraced reported cases. We used this formula to estimate the $\mathbb{E}(\mathcal{R}_e)$ in London, UK over nine weeks beginning from the 22nd of February 2020. We used the proportionality of the effective reproduction number to social mixing matrices to calculate the $\mathbb{E}(\mathcal{R}_e)$ for six setups and six age groups. While our method in estimating the $\mathbb{E}(\mathcal{R}_e)$ is accurate and matches with the values of other studies (such as [25] for London) it is important to improve this calculation by using the formula presented in Eq. (3.4). This equation depends on the infectivity parameters, such as β , T_s and T_a , and they can be estimated by methods such as Markov chain Monte Carlo (MCMC) (see for instance [13]). This task is the subject of our upcoming work.

From our model and analysis it is evident that contact tracing and social mixing are both playing important roles in controlling the value of the effective reproduction number (see also [22, 23, 24, 28, 29, 30]). While there is an increasing demand to open up the society since the value of $\mathcal{R}_e$ is below the critical value of one it is important for policymakers to note that the value of $\mathcal{R}_e$ can rebound to higher values should the easing of the restrictions cause higher number of contacts between individuals. In fact, in opening up the society one must suppress the opportunities for transmission. From our model this can be done by firstly, putting an effective contact tracing measure in place, that is to trace, monitor and isolate the contacts of reported cases without time-wasting and secondly, monitoring the number of contacts between different age groups and the fraction of the reported cases classified by their age α_λ to prioritise contacts tracing by age groups when there is a lack of health workers.

Acknowledgements

The author would like to thank Mikhail Goykhman for useful feedbacks and discussion.

References

- [1] Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S., Lau, E.H., Wong, J.Y. and Xing, X., 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus?infected pneumonia. *New England Journal of Medicine*.

- [2] Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R. and Niu, P., 2020. A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*.
- [3] Lin, G., Strauss, A.T., Pinz, M., Martinez, D.A., Tseng, K.K., Schueller, E., Gatalo, O., Yang, Y., Levin, S.A. and Klein, E.Y., 2020. Explaining the Bomb-Like Dynamics of COVID-19 with Modeling and the Implications for Policy. *medRxiv*.
- [4] Dong, E., Du, H. and Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*, 20(5), pp.533-534.
- [5] Browne, C., Gulbudak, H. and Webb, G., 2015. Modeling contact tracing in outbreaks with application to Ebola. *Journal of theoretical biology*, 384, pp.33-49.
- [6] Scala, A., Flori, A., Spelta, A., Brugnoli, E., Cinelli, M., Quattrociocchi, W. and Pammolli, F., 2020. Time, space and social interactions: exit mechanisms for the COVID-19 epidemics. *arXiv preprint arXiv:2004.04608*.
- [7] Lin, Q., Zhao, S., Gao, D., Lou, Y., Yang, S., Musa, S.S., Wang, M.H., Cai, Y., Wang, W., Yang, L. and He, D., 2020. A conceptual model for the outbreak of Coronavirus disease 2019 (COVID-19) in Wuhan, China with individual reaction and governmental action. *International journal of infectious diseases*.
- [8] Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W. and Shaman, J., 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490), pp.489-493.
- [9] Prem, K., Liu, Y., Russell, T.W., Kucharski, A.J., Eggo, R.M., Davies, N., Flasche, S., Clifford, S., Pearson, C.A., Munday, J.D. and Abbott, S., 2020. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*.
- [10] Champredon, D., Dushoff, J. and Earn, D.J., 2018. Equivalence of the Erlang SEIR epidemic model and the renewal equation. *BioRxiv*, p.319574.
- [11] Krylova, O. and Earn, D.J., 2013. Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *Journal of The Royal Society Interface*, 10(84), p.20130098.
- [12] World Health Organization, 2020. *Contact tracing in the context of COVID-19: interim guidance, 10 May 2020*.
- [13] Lekone, P.E. and Finkenstädt, B.F., 2006. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, 62(4), pp.1170-1177.
- [14] Wang, C.J., Ng, C.Y. and Brook, R.H., 2020. Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. *Jama*, 323(14), pp.1341-1342.
- [15] Tang, B., Xia, F., Bragazzi, N.L., Wang, X., He, S., Sun, X., Tang, S., Xiao, Y. and Wu, J., 2020. Lessons drawn from China and South Korea for managing COVID-19 epidemic: insights from a comparative modeling study. *medRxiv*.
- [16] Legido-Quigley, H., Asgari, N., Teo, Y.Y., Leung, G.M., Oshitani, H., Fukuda, K., Cook, A.R., Hsu, L.Y., Shibuya, K. and Heymann, D., 2020. Are high-performing health systems resilient against the COVID-19 epidemic?. *The Lancet*, 395(10227), pp.848-850.
- [17] Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., Jung, S.M., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A.R. and Linton, N.M., 2020. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *International journal of infectious diseases*, 94, p.154.

- [18] Liu, K., Chen, Y., Lin, R. and Han, K., 2020. Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *Journal of Infection*.
- [19] Diekmann, O., Heesterbeek, J.A.P. and Roberts, M.G., 2010. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7(47), pp.873-885.
- [20] Jarvis, C.I., Van Zandvoort, K., Gimma, A., Prem, K., Klepac, P., Rubin, G.J. and Edmunds, W.J., 2020. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC medicine*, 18, pp.1-10.
- [21] Wallinga, J., Teunis, P. and Kretzschmar, M., 2006. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American journal of epidemiology*, 164(10), pp.936-944.
- [22] Kucharski, A.J., Klepac, P., Conlan, A., Kissler, S.M., Tang, M., Fry, H., Gog, J., Edmunds, J. and CMMID COVID-19 Working Group, 2020. Effectiveness of isolation, testing, contact tracing and physical distancing on reducing transmission of SARS-CoV-2 in different settings. *medRxiv*.
- [23] Nicholas, G.D., Petra, K., Yang, L., Kiesha, P., Mark, J., Rosalind, M. and CMMID COVID-19 working group, Age-dependent Effects in the Transmission and Control of COVID-19 Epidemics. *Nature medicine*.
- [24] Chang, S.L., Harding, N., Zachreson, C., Cliff, O.M. and Prokopenko, M., 2020. Modelling transmission and control of the COVID-19 pandemic in Australia. *arXiv preprint arXiv:2003.10218*.
- [25] P. Birrell, J. Blake, E. van Leeuwen, MRC Biostatistics Unit COVID-19 Working Group, D. De Angeli., 2020. COVID-19: nowcast and forecast. <https://www.mrc-bsu.cam.ac.uk/now-casting/>
- [26] Du, Z., Xu, X., Wu, Y., Wang, L., Cowling, B.J. and Meyers, L.A., 2020. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerging infectious diseases*, 26(6), p.1341.
- [27] Mayor of London, Coronavirus (COVID-19) Cases, Greater London Authority. <https://data.london.gov.uk/dataset/coronavirus-covid-19-cases>
- [28] Hellewell, J., Abbott, S., Gimma, A., Bosse, N.I., Jarvis, C.I., Russell, T.W., Munday, J.D., Kucharski, A.J., Edmunds, W.J., Sun, F. and Flasche, S., 2020. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*.
- [29] Adam, J.K., Petra, K., Andrew, J.K., Stephen, M.K., Maria, L.T., Hannah, F., Julia, R. and CMMID COVID-19 working group, Effectiveness of Isolation, Testing, Contact Tracing, and Physical Distancing on Reducing Transmission of SARS-CoV-2 in Different Settings: A Mathematical Modelling Study. *The Lancet. Infectious diseases*, pp.S1473-3099.
- [30] MacIntyre, C.R., 2020. Case isolation, contact tracing, and physical distancing are pillars of COVID-19 pandemic control, not optional choices. *The Lancet Infectious Diseases*.
- [31] Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S.A., Zhang, T. and Gao, W., 2020. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*.
- [32] Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G. and Lessler, J., 2020. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9), pp.577-582.