medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Unreported cases for Age Dependent COVID-19 Outbreak in Japan

Quentin Griette[(a)]*, Pierre Magal[(a)*,†] and Ousmane Seydi[(b)]

[(a)] *Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France.*

CNRS, IMB, UMR 5251, F-33400 Talence, France.

[(b)] *Département Tronc Commun, École Polytechnique de Thiès, Thiès, BP A10, Sénégal*

June 9, 2020

Abstract

We investigate the age structured data for the COVID-19 outbreak in Japan. We consider a mathematical model for the epidemic with unreported infectious patient with and without age structure. In particular, we build a new mathematical model and a new computational method to fit the data by using age classes dependent exponential growth at the early stage of the epidemic. This allows to take into account differences in the response of patients to the disease according to their age. This model also allows for a heterogeneous response of the population to the social distancing measures taken by the local government. We fit this model to the observed data and obtain a snapshot of the effective transmissions occurring inside the population at different times, which indicates where and among whom the disease propagates after the start of public mitigation measures.

1 Introduction

COVID-19 disease caused by the corona virus SARS-CoV-2 first appeared in Wuhan, China, and the first cases were notified to WHO on December 31, 2019 [33, 34]. Beginning in Wuhan as an epidemic, it then spread very quickly around the world to become a global pandemic on March 11, 2020 [33]. Symptoms of this disease include fever, shortness of breath, cough, and a non-negligible proportion of infected individuals may develop severe forms of the symptoms leading to their transfer to intensive care units and, in some cases, death, see *e.g.* Guan et al [8] and Wei et al [27]. Both symptomatic and asymptomatic individuals can be infectious [22, 27, 30], which makes the control of the disease particularly challenging.

The virus is characterized by its rapid progression among individuals, most often exponential in the first phase, but also a marked heterogeneity in populations and geographic areas [31, 29, 26]. The number of reported cases worldwide exceeded 3 millions as of May 3, 2020 [32]. The heterogeneity of the number of cases and the severity according to the age groups, especially for children and elderly people, aroused the interest of several researchers [4, 20, 16, 23, 24]. Indeed, several studies have shown that the severity of the disease increases with the age and co-morbidity of hospitalized patients (see *e.g.* To et al [24] and Zhou et al [29]). Wu et al [28] have shown that the risk of developing symptoms increases by 4% per year in adults aged between 30 and 60 years old while Davies et al [5] found that there is a strong correlation between chronological age and the likelihood of developing symptoms. Since completely asymptomatic individuals can also be contagious, a higher probability of developing symptoms does not necessarily imply greater infectiousness: Zou et al [30] found that, in some cases, the viral load in asymptomatic patients was similar to that in symptomatic patients. Moreover while adults are more likely to develop symptoms, Jones et al [9] found that the viral loads in infected children do not differ significantly from those of adults.

These findings suggest that a study of the dynamics of inter-generational spread is fundamental to better understand the spread of the corona virus and most importantly to efficiently fight the COVID-19 pandemic. To this end the distribution of contacts between age groups in society (work, school,

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

ANR flashCOVID-19: MI-COVI

†Corresponding author. e-mail: pierre.magal@u-bordeaux.fr

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

home, and other locations) is an important factor to take into account when modeling the spread of the epidemic. To account for these facts, some mathematical models have been developed [1, 3, 5, 20, 23]. In Ayoub et al [1] the authors studied the dependence of the COVID-19 epidemic on the demographic structures in several countries but did not focus on the contacts distribution of the populations. In [3, 5, 20, 23] a focus on the social contact patterns with respect to the chronological age has been made by using the contact matrices provided in Prem et al [19]. While Ayoub et al [1], Chikina and Pegden [3] and Davies et al [5] included the example of Japan in their study, their approach is significantly different from ours. Indeed, Ayoub et al [1] use a complex mathematical model to discuss the influence of the age structure on the infection in a variety of countries, mostly through the basic reproduction number $\mathcal{R}_0$. They use parameter values from the literature and from another study of the same group of authors [2], where the parameter identification is done by a nonlinear least-square minimization. Chikina and Pegden [3] use an age-structured model to investigate age-targeted mitigation strategies. They rely on parameter values from the literature and do discuss using age-structured temporal series to fit their model. Finally, Davies et al [5] also discuss age-related effects in the control of the COVID epidemic, and use statistical inference to fit an age-structured SIR variant to data; the model is then used to discuss the efficiency of different control strategies. We provide a new, explicit computational solution for the parameter identification of an age-structured model. The model is based on the SIUR model developed in Liu et al [11], which accounts for a differentiated infectiousness for reported and non-reported cases (contrary to, for instance, other SIR-type models). In particular, our method is significantly different from nonlinear least-squares minimization and does not involve statistical inference.

In this article we focus on an epidemic model with unreported infectious symptomatic patients (i.e. with mild or no symptoms). Our goal is to investigate the age structured data of the COVID-19 outbreak in Japan. In section 2 we present the age structured data and in section 3 the mathematical models (with and without age structure). One of the difficulties in fitting the model to the data is that the growth rate of the epidemic is different in each age class, which lead us to adapt our early method presented in Liu et al [11]. The new method is presented in the Appendix A. In section 4 we present the comparison of the model with the data. In the last section we discuss our results.

2 Data

Patient data in Japan have been made public since the early stages of the epidemic with the quarantine of the *Diamond Princess* in the Haven of Yokohama. We used data from the website covid19japan.com[1] which is based on reports from national and regional authorities. Patients are labeled "confirmed" when tested positive to COVID-19 by PCR. Interestingly, the age class of the patient is provided for 13 660 out of 13970 confirmed patients (97.8% of the confirmed population) as of April 29. The age distribution of the infected population is represented in Figures 1 compared to the total population per age class (data from the Statistics Bureau of Japan estimate for October 1, 2019). In Figure 2 we plot the number of reported cases per 10 000 people of the same age class (*i.e.* the number of infected patients divided by the population of the age class times 10 000). Both datasets are given in Table 1 and a statistical summary is provided by Table 2. Note that the high proportion of 20-60 years old confirmed patients may indicate that the severity of the disease is lower for those age classes than for older patients, and therefore the disease transmits more easily in those age classes because of a higher number of asymptomatic individuals. Elderly infected individuals might transmit less because they are identified more easily. The cumulative number of death (Figure 5) is another argument in favor of this explanation. We also reconstructed the time evolution of the reported cases in Figure 3 and Figure 4. Note that the steepest curves precisely concern the 20-60-years old, probably because they are economically active and therefore have a high contact rate with the population.

[1] https://covid19japan.com. Accessed May 06, 2020

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

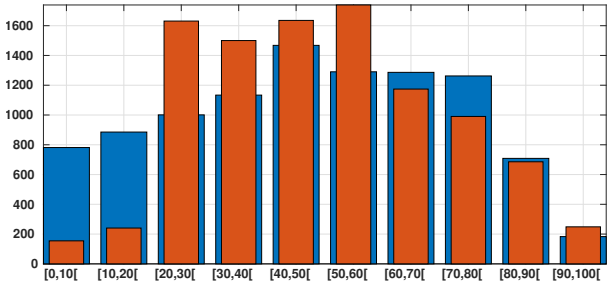

Figure 1: *In this figure we plot in blue bars the age distribution of the Japanese population for* 10 000 *people and we plot in orange bars the age distribution of the number of reported cases of SARS-CoV-2 for* 10 000 *patient on April 29 (based on the total of* 13660 *reported cases). We observe that* 77% *of the confirmed patients belong to the 20–60 years age class.*

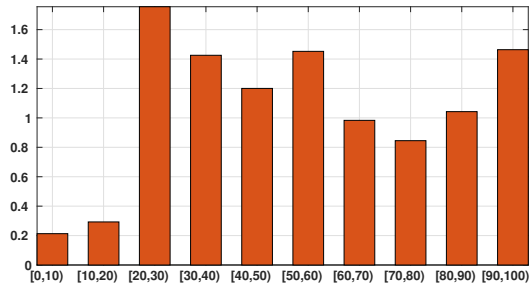

Figure 2: *In this figure we plot the number of infected patients for each age class per* 10 000 *individuals of the same age class (i.e. the number of infected individuals divided by the population of the age class times* 10 000*). The figure shows that the individuals are more or less likely to becomes infected depending on their age class. The bars describe the susceptibility of people to the SARS-CoV-2 depending on their age class.*

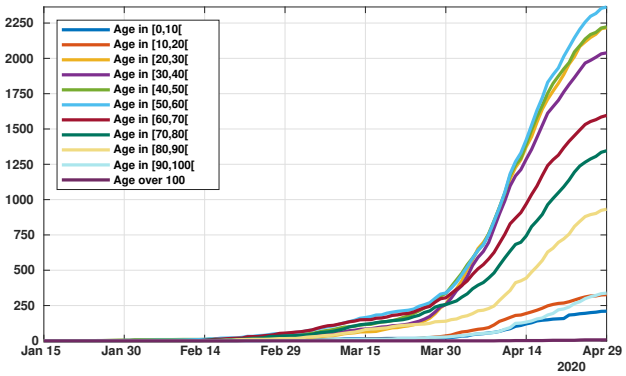

Figure 3: *Time evolution of the cumulative number of reported cases of SARS-CoV-2 per age class. The vertical axis represents the total number of cumulative reported cases in each age class.*

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

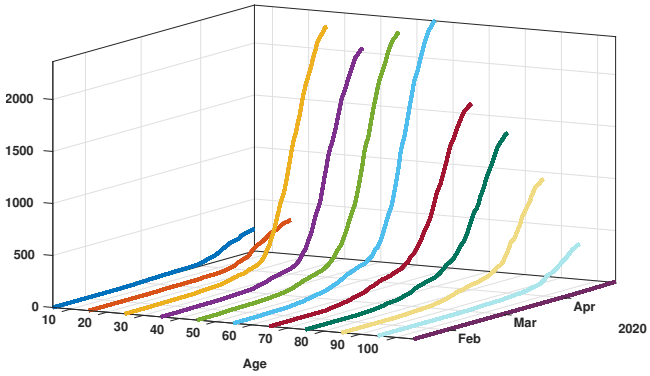

Figure 4: *Time evolution of the cumulative number of reported cases of SARS-CoV-2 per age class. The vertical axis represents the total number of cumulative reported cases in each age class*

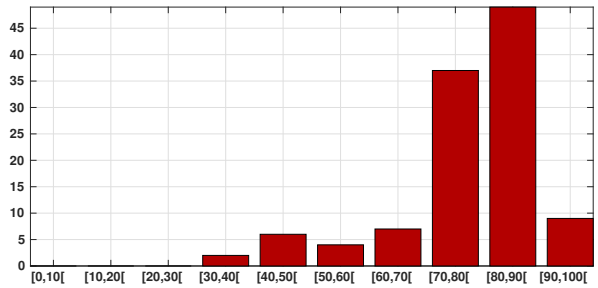

Figure 5: *Cumulative number of SARS-CoV-2-induced deaths per age class (red bars). We observe that* 83% *of death occur in between* 70 *and* 100 *years old.*

Age group	$[0, 10[$	$[10, 20[$	$[20, 30[$	$[30, 40[$	$[40, 50[$	$[50, 60[$	$[60, 70[$	$[70, 80[$	$[80, 90[$	$[90, 100[$
Age class for 2019	9859515	11171044	12627964	14303042	18519755	16277853	16231582	15926926	8939954	2309313
Age class per 10,000 people	781	885	1000	1133	1467	1290	1286	1262	709	183
Confirmed Cases	211	327	2216	2034	2220	2355	1566	1289	857	304
Death	0	0	0	2	6	4	7	37	49	9

Table 1: *The age distribution of Japan is taken from the Statistics Bureau of Japan [35]. The number of cases and the number of death the data come from Prefectural Governments and Japan Ministry of Health, Labour and Welfare.*

Dataset	Japanese population	Infected	Deceased
First Quartile	28	28	68
Median	48	44	75
Third Quartile	67	59	81

Table 2: *Statistical summary of the data from Table 1.*

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

3 Methods

3.1 SIUR Model

The model consists of the following system of ordinary differential equations:

$$\begin{cases} S'(t) = -\tau(t)S(t)\dfrac{I(t)+U(t)}{N}, \\ I'(t) = \tau(t)S(t)\dfrac{I(t)+U(t)}{N} - \nu I(t), \\ R'(t) = \nu_1 I(t) - \eta R(t), \\ U'(t) = \nu_2 I(t) - \eta U(t). \end{cases} \tag{3.1}$$

This system is supplemented by initial data

$$S(t_0) = S_0 \geq 0,\ I(t_0) = I_0 \geq 0,\ R(t_0) \geq 0 \text{ and } U(t_0) = U_0 \geq 0. \tag{3.2}$$

Here $t \geq t_0$ is time in days, t_0 is the starting date of the epidemic in the model, $S(t)$ is the number of individuals susceptible to infection at time t, $I(t)$ is the number of asymptomatic infectious individuals at time t, $R(t)$ is the number of reported symptomatic infectious individuals at time t, and $U(t)$ is the number of unreported symptomatic infectious individuals at time t. A flow chart of the model is presented in Figure 6.

Asymptomatic infectious individuals $I(t)$ are infectious for an average period of $1/\nu$ days. Reported symptomatic individuals $R(t)$ are infectious for an average period of $1/\eta$ days, as are unreported symptomatic individuals $U(t)$. We assume that reported symptomatic infectious individuals $R(t)$ are reported and isolated immediately, and cause no further infections. The asymptomatic individuals $I(t)$ can also be viewed as having a low-level symptomatic state. All infections are acquired from either $I(t)$ or $U(t)$ individuals. A summary of the parameters involved in the model is presented in Table 3.

Our study begins in the second phase of the epidemics, *i.e.* after the pathogen has succeeded in surviving in the population. During this second phase $\tau(t) \equiv \tau_0$ is constant. When strong government measures such as isolation, quarantine, and public closings are implemented, the third phase begins. The actual effects of these measures are complex, and we use a time-dependent decreasing transmission rate $\tau(t)$ to incorporate these effects. The formula for $\tau(t)$ is

$$\begin{cases} \tau(t) = \tau_0,\ 0 \leq t \leq D, \\ \tau(t) = \tau_0 \exp\left(-\mu\,(t-D)\right),\ D < t. \end{cases} \tag{3.3}$$

The date D is the first day of public intervention and μ characterises the intensity of the public intervention.

A similar model has been used to describe the epidemics in mainland China, South Korea, Italy, and other countries, and give reasonable trajectories for the evolution of the epidemic based on actual data [7, 11, 12, 13, 14, 15]. Compared with these models, we added a scaling with respect to the total population size N, for consistency with the age-structured model (3.12). This only changes the value of the parameter τ and does not impact the qualitative or quantitative behavior of the model.

Symbol	Interpretation	Method
t_0	Time at which the epidemic started	fitted
S_0	Number of susceptible at time t_0	fixed
I_0	Number of asymptomatic infectious at time t_0	fitted
U_0	Number of unreported symptomatic infectious at time t_0	fitted
$\tau(t)$	Transmission rate at time t	fitted
D	First day of public intervention	fitted
μ	Intensity of the public intervention	fitted
$1/\nu$	Average time during which asymptomatic infectious are asymptomatic	fixed
f	Fraction of asymptomatic infectious that become reported symptomatic infectious	fixed
$\nu_1 = f\,\nu$	Rate at which asymptomatic infectious become reported symptomatic	fixed
$\nu_2 = (1-f)\,\nu$	Rate at which asymptomatic infectious become unreported symptomatic	fixed
$1/\eta$	Average time symptomatic infectious have symptoms	fixed

Table 3: *Parameters of the model.*

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

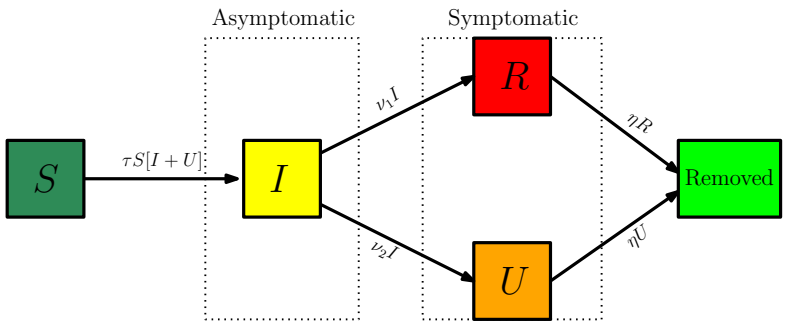

Figure 6: *Compartments and flow chart of the model.*

3.2 Comparison of the model (3.1) with the data

At the early stages of the epidemic, the infectious components of the model $I(t)$, $U(t)$ and $R(t)$ must be exponentially growing. Therefore, we can assume that

$$I(t) = I_0 \exp\left(\chi_2 \left(t - t_0\right)\right).$$

The cumulative number of reported symptomatic infectious cases at time t, denoted by $CR(t)$, is

$$CR(t) = \nu_1 \int_{t_0}^{t} I(s) ds. \tag{3.4}$$

Since $I(t)$ is an exponential function and $CR(t_0) = 0$ it is natural to assume that $CR(t)$ has the following special form:

$$CR(t) = \chi_1 \exp\left(\chi_2 t\right) - \chi_3. \tag{3.5}$$

As in our early articles [11, 12, 13, 14, 15], we fix $\chi_3 = 1$ and we evaluate the parameters χ_1 and χ_2 by using an exponential fit to

$$\chi_1 \exp\left(\chi_2 t\right) \simeq CR_{data}(t).$$

We use only early data for this part, from day $t = d_1$ until day $t = d_2$, because we want to catch the exponential growth of the early epidemic and avoid the influence of saturation arising at later stages.

Remark 3.1 *The estimated parameters χ_1 and χ_2 will vary if we change the interval $[d_1, d_2]$.*

Once χ_1, χ_2, χ_3 are known, we can compute the starting time of the epidemic t_0 from (3.5) as :

$$CR(t_0) = 0 \Leftrightarrow \chi_1 \exp\left(\chi_2 t_0\right) - \chi_3 = 0 \;\Rightarrow\; t_0 = \frac{1}{\chi_2}\left(\ln\left(\chi_3\right) - \ln\left(\chi_1\right)\right).$$

We fix $S_0 = 126.8 \times 10^6$, which corresponds to the total population of Japan. The quantities I_0, R_0, and U_0 correspond to the values taken by $I(t)$, $R(t)$ and $U(t)$ at $t = t_0$ (and in particular R_0 should not be confused with the basic reproduction number $\mathcal{R}_0$). We fix the fraction f of symptomatic infectious cases that are reported. We assume that between 80% and 100% of infectious cases are reported. Thus, f varies between 0.8 and 1. We assume that the average time during which the patients are asymptomatic infectious $1/\nu$ varies between 1 day and 7 days. We assume that the average time during which a patient is symptomatic infectious $1/\eta$ varies between 1 day and 7 days. In other words we fix the parameters f, ν, η. Since f and ν are known, we can compute

$$\nu_1 = f\nu \;\text{ and }\; \nu_2 = (1 - f)\,\nu. \tag{3.6}$$

Computing further (see below for more details), we should have

$$I_0 = \frac{\chi_1 \chi_2 \exp\left(\chi_2 t_0\right)}{f\,\nu} = \frac{\chi_3 \chi_2}{f\,\nu}, \tag{3.7}$$

$$\tau = N \frac{\chi_2 + \nu}{S_0} \frac{\eta + \chi_2}{\nu_2 + \eta + \chi_2}, \tag{3.8}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

$$R_0 = \frac{\nu_1}{\eta + \chi_2} I_0 = \frac{f\nu}{\eta + \chi_2} I_0. \tag{3.9}$$

and

$$U_0 = \frac{\nu_2}{\eta + \chi_2} I_0 = \frac{(1-f)\nu}{\eta + \chi_2} I_0. \tag{3.10}$$

By using the approach described in Diekmann et al [6], van den Driessche and Watmough [25], the basic reproductive number for model (3.1) is given by

$$\mathcal{R}_0 = \frac{\tau S_0}{\nu N}\left(1 + \frac{\nu_2}{\eta}\right).$$

By using (3.8) we obtain

$$\mathcal{R}_0 = \frac{\chi_2 + \nu}{\nu} \frac{(\eta + \chi_2)}{\nu_2 + \eta + \chi_2}\left(1 + \frac{\nu_2}{\eta}\right). \tag{3.11}$$

3.3 Model SIUR with age structure

In what follows we will denote $N_1, \ldots, N_{10}$ the number of individuals respectively for the age classes $[0, 10[, \ldots, [90, 100[$. The model for the number of susceptible individuals $S_1(t), \ldots, S_{10}(t)$, respectively for the age classes $[0, 10[, \ldots, [90, 100[$, is the following

$$\begin{cases} S_1'(t) = -\tau_1 S_1(t)\left[\phi_{1,1}\dfrac{(I_1(t) + U_1(t))}{N_1} + \ldots + \phi_{1,10}\dfrac{(I_{10}(t) + U_{10}(t))}{N_{10}}\right], \\ \vdots \\ S_{10}'(t) = -\tau_{10} S_{10}(t)\left[\phi_{10,1}\dfrac{(I_1(t) + U_1(t))}{N_1} + \ldots + \phi_{10,10}\dfrac{(I_{10}(t) + U_{10}(t))}{N_{10}}\right]. \end{cases} \tag{3.12}$$

The model for the number of asymptomatic infectious individuals $I_1(t), \ldots, I_{10}(t)$, respectively for the age classes $[0, 10[, \ldots, [90, 100[$, is the following

$$\begin{cases} I_1'(t) = \tau_1 S_1(t)\left[\phi_{1,1}\dfrac{(I_1(t) + U_1(t))}{N_1} + \ldots + \phi_{1,10}\dfrac{(I_{10}(t) + U_{10}(t))}{N_{10}}\right] - \nu I_1(t), \\ \vdots \\ I_{10}'(t) = \tau_{10} S_{10}(t)\left[\phi_{10,1}\dfrac{(I_1(t) + U_1(t))}{N_1} + \ldots + \phi_{10,10}\dfrac{(I_{10}(t) + U_{10}(t))}{N_{10}}\right] - \nu I_{10}(t). \end{cases} \tag{3.13}$$

The model for the number of reported symptomatic infectious individuals $R_1(t), \ldots, R_{10}(t)$, respectively for the age classes $[0, 10[, \ldots, [90, 100[$, is

$$\begin{cases} R_1'(t) = \nu_1^1\, I_1(t) - \eta R_1(t), \\ \vdots \\ R_{10}'(t) = \nu_1^{10}\, I_{10}(t) - \eta R_{10}(t). \end{cases} \tag{3.14}$$

Finally the model for the number of unreported symptomatic infectious individuals $U_1(t), \ldots, U_{10}(t)$, respectively in the age classes $[0, 10[, \ldots, [90, 100[$, is the following

$$\begin{cases} U_1'(t) = \nu_2^1\, I_1(t) - \eta U_1(t), \\ \vdots \\ U_{10}'(t) = \nu_2^{10}\, I_{10}(t) - \eta U_{10}(t). \end{cases} \tag{3.15}$$

In each age class $[0, 10[, \ldots, [90, 100[$ we assume that there is a fraction $f_1, \ldots, f_{10}$ of asymptomatic infectious individual who become reported symptomatic infectious (i.e. with severe symptoms) and a fraction $(1-f_1), \ldots, (1-f_{10})$ who become unreported symptomatic infectious (i.e. with mild symptoms). Therefore we define

$$\begin{gathered} \nu_1^1 = \nu f_1 \text{ and } \nu_2^1 = \nu(1 - f_1), \\ \vdots \\ \nu_1^{10} = \nu f_{10} \text{ and } \nu_2^{10} = \nu(1 - f_{10}). \end{gathered} \tag{3.16}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In this model $\tau_1, \ldots, \tau_{10}$ are the respective transmission rates for the age classes $[0, 10[, \ldots, [90, 100[$.

The matrix ϕ_{ij} represents the probability for an individual in the class i to meet an individual in the class j. In their survey, Prem and co-authors [19] present a way to reconstruct contact matrices from existing data and provide such contact matrices for a number of countries including Japan. Based on the data provided by Prem et al [19] for Japan we construct the contact probability matrix ϕ. More precisely, we inferred contact data for the missing age classes $[80, 90[$ and $[90, 100[$. The precise method used to construct the contact matrix γ is detailed in Appendix B. An analogous contact matrix for Japan has been proposed by Munasinghe, Asai and Nishiura [18]. The contact matrix γ we used is the following

$$[\gamma_{ij}] = \begin{bmatrix} 4.03 & 0.92 & 0.47 & 1.69 & 0.83 & 0.92 & 0.78 & 0.56 & 0.57 & 0.57 \\ 0.71 & 8.06 & 1.38 & 1.36 & 1.96 & 1.74 & 0.75 & 0.86 & 0.74 & 0.57 \\ 0.55 & 1.05 & 4.63 & 2.25 & 1.84 & 1.92 & 0.94 & 0.46 & 0.74 & 0.73 \\ 1.52 & 1.20 & 2.54 & 4.97 & 2.98 & 2.40 & 1.76 & 0.99 & 0.53 & 0.73 \\ 0.69 & 1.42 & 1.93 & 2.87 & 3.91 & 2.76 & 1.35 & 1.33 & 0.95 & 0.53 \\ 0.34 & 0.48 & 1.20 & 1.46 & 1.61 & 2.97 & 1.40 & 0.98 & 1.23 & 0.95 \\ 0.28 & 0.18 & 0.20 & 0.52 & 0.38 & 0.77 & 2.67 & 1.72 & 0.92 & 1.23 \\ 0.12 & 0.10 & 0.09 & 0.18 & 0.19 & 0.25 & 0.76 & 1.99 & 1.18 & 0.93 \\ 0.09 & 0.10 & 0.08 & 0.09 & 0.13 & 0.17 & 0.27 & 0.64 & 1.61 & 1.19 \\ 0.09 & 0.09 & 0.10 & 0.08 & 0.09 & 0.13 & 0.17 & 0.27 & 0.64 & 1.61 \end{bmatrix}, \tag{3.17}$$

where the ith line of the matrix γ_{ij} is the average number of contact made by an individuals in the age class i with an individual in the age class j during one day. Notice that the higher number of contacts are achieved within the same age class. The matrix of conditional probability ϕ of contact between age classes is given by (3.18) and we plot a visual representation of this matrix in Figure 7.

$$[\phi_{ij}] = \begin{bmatrix} 0.35 & 0.08 & 0.04 & 0.14 & 0.07 & 0.08 & 0.06 & 0.04 & 0.05 & 0.05 \\ 0.03 & 0.44 & 0.07 & 0.07 & 0.10 & 0.09 & 0.04 & 0.04 & 0.04 & 0.03 \\ 0.03 & 0.06 & 0.30 & 0.14 & 0.12 & 0.12 & 0.06 & 0.03 & 0.04 & 0.04 \\ 0.07 & 0.06 & 0.12 & 0.25 & 0.15 & 0.12 & 0.08 & 0.05 & 0.02 & 0.03 \\ 0.03 & 0.07 & 0.10 & 0.16 & 0.22 & 0.15 & 0.07 & 0.07 & 0.05 & 0.03 \\ 0.02 & 0.03 & 0.09 & 0.11 & 0.12 & 0.23 & 0.11 & 0.07 & 0.09 & 0.07 \\ 0.03 & 0.02 & 0.02 & 0.05 & 0.04 & 0.08 & 0.30 & 0.19 & 0.10 & 0.13 \\ 0.02 & 0.01 & 0.01 & 0.03 & 0.03 & 0.04 & 0.13 & 0.34 & 0.20 & 0.16 \\ 0.02 & 0.02 & 0.01 & 0.02 & 0.02 & 0.03 & 0.06 & 0.14 & 0.36 & 0.27 \\ 0.02 & 0.02 & 0.03 & 0.02 & 0.02 & 0.03 & 0.05 & 0.08 & 0.19 & 0.48 \end{bmatrix}. \tag{3.18}$$

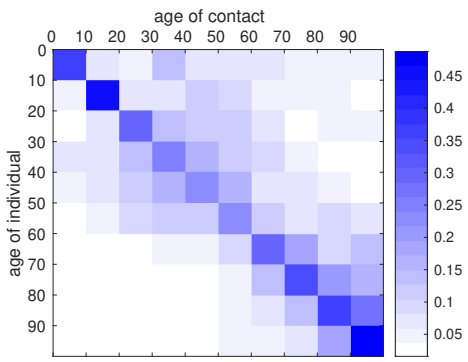

Figure 7: *Graphical representation of the contact matrix ϕ. The intensity of blue in the cell (i, j) indicates the conditional probability that, given a contact between an individual of age group i and another individual, the latter belongs to the age class j. The matrix was reconstructed from the data of Prem et al [19], with the method described in Appendix B.*

4 Results

4.1 Model without age structure

The daily number of reported cases from the model can be obtained by computing the solution of the following equation:

$$DR'(t) = \nu_1 \, I(t) - DR(t), \text{ for } t \geq t_0 \text{ and } DR(t_0) = DR_0. \tag{4.1}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In Figure 8 and Figure 9 we employ the method presented previously in Liu et al [15] to fit the data for Japan without age structure.

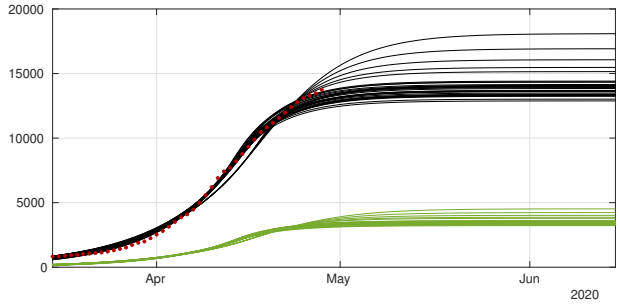

Figure 8: *Cumulative number of cases. We plot the cumulative data (reds dots) and the best fits of the model* $CR(t)$ *(black curve) and* $CU(t)$ *(green curve). We fix* $f = 0.8$, $1/\eta = 7$ *days and* $1/\nu = 7$ *and we apply the method described in Liu et al [15]. The best fit is* $d_1 =$ *April* 2, $d_2 =$ *April* 5, $D =$ *April* 27, $\mu = 0.6$, $\chi_1 = 179$, $\chi_2 = 0.085$, $\chi_3 = 1$ *and* $t_0 =$ *January* 13.

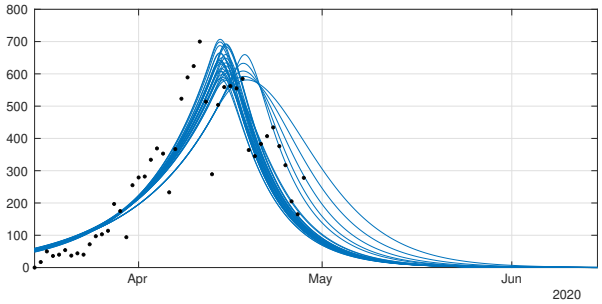

Figure 9: *Daily number of cases. We plot the daily data (black dots) with* $DR(t)$ *(blue curve). We fix* $f = 0.8$, $1/\eta = 7$ *days and* $1/\nu = 7$ *and we apply the method described in Liu et al [15]. The best fit is* $d_1 =$ *April* 2, $d_2 =$ *April* 5, $N =$ *April* 27, $\mu = 0.6$, $\chi_1 = 179$, $\chi_2 = 0.085$, $\chi_3 = 1$ *and* $t_0 =$ *January* 13.

The model to compute the cumulative number of death from the reported individuals is the following

$$D'(t) = \eta_D \, p \, R(t), \text{ for } t \geq t_0 \text{ and } D(t_0) = 0, \tag{4.2}$$

where η_D is the death rate of reported infectious symptomatic individuals and p is the case fatality rate (namely the fraction of death per reported infectious individuals).

In the simulation we chose $1/\eta_D = 6$ days and the case fatality rate $p = 0.286$ is computed by using the cumulative number of confirmed cases and the cumulative number of deaths (as of April 29) as follows

$$p = \frac{\text{cumulative number of deaths}}{\text{cumulative number of reported cases}} = \frac{393}{13744}. \tag{4.3}$$

In Figure 10 we plot the cumulative number of $D(t)$ by using the same simulations than in Figures 8-9.

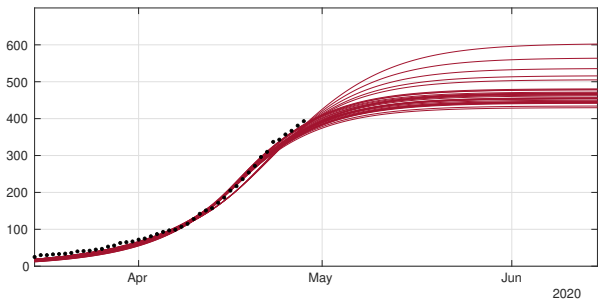

Figure 10: *In this figure we plot the data for the cumulative number of death (black dots), and our best fits for* $D(t)$ *(red curves).*

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

4.2 Model with age structure

In order to describe the confinement for the age structured model (3.12)-(3.15) we will use for each age class $i = 1, \ldots, 10$ a different transmission rate having the following form

$$\begin{cases} \tau_i(t) = \tau_i,\ 0 \leq t \leq D_i, \\ \tau_i(t) = \tau_i \exp\left(-\mu_i\left(t - D_i\right)\right),\ D_i < t. \end{cases} \tag{4.4}$$

The date D_i is the first day of public intervention for the age class i and μ_i is the intensity of the public intervention for each age class.

In Figure 11 we plot the cumulative number of reported cases as given by our model (3.12)-(3.15) (solid lines), compared with reported cases data (black dots). We used the method described in the Appendix A to estimate the parameters τ_i from the data. In Figure 12 we plot the cumulative number of *unreported* cases (solid lines) as given by our model with the same parameter values, compared to the existing data of *reported* cases (black dots).

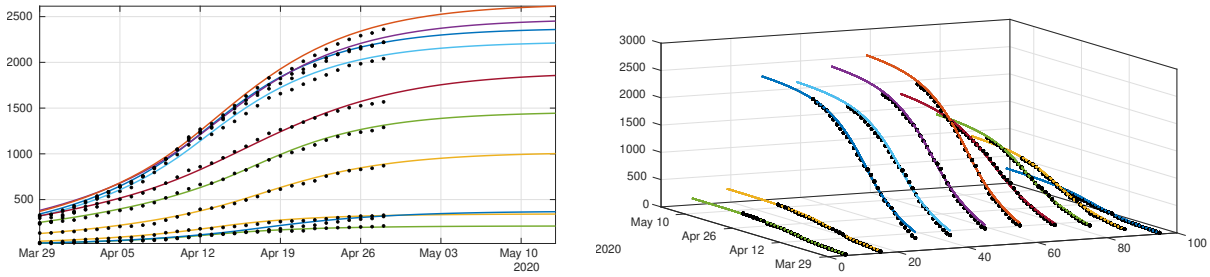

Figure 11: *We plot a comparison between the model (3.12)-(3.15) and the age structured data from Japan by age class. We took* $1/\nu = 1/\eta = 7$ *days for each age class. Our best fit is obtained for* f_i *which depends linearly on the age class until it reaches 90%, with* $f_1 = 0.1$, $f_2 = 0.2$, $f_3 = 0.3$, $f_4 = 0.4$, $f_5 = 0.5$, $f_6 = 0.6$, $f_7 = 0.7$, $f_8 = 0.8$, $f_9 = 0.9$, *and* $f_{10} = 0.9$. *The values we used for the first day of public intervention are* $D_i =$ *Apr. 13 for the 0-20 years age class* $i = 1, 2$, $D_i =$ *Apr. 11 for the age class going from* $[20, 30[$ *to* $[60, 70[$ $i = 3, 4, 5, 6, 7$, *and* $D_i =$ *Apr. 16 for the remaining age classes. We fit the data from March* 30 *to April* 20 *to derive the value of* χ_1^i *and* χ_2^i *for each age class. For the intensity of confinement we use the values* $\mu_1 = \mu_2 = 0.4829$, $\mu_3 = \mu_4 = 0.2046$, $\mu_5 = \mu_6 = 0.1474$, $\mu_7 = 0.0744$, $\mu_8 = 0.1736$, $\mu_9 = \mu_{10} = 0.1358$. *By applying the method described in Appendix A, we obtain* $\tau_1 = 0.1630$, $\tau_2 = 0.1224$, $\tau_3 = 0.3028$, $\tau_4 = 0.2250$, $\tau_5 = 0.1520$, $\tau_6 = 0.1754$, $\tau_7 = 0.1289$, $\tau_8 = 0.1091$, $\tau_9 = 0.1211$ *and* $\tau_{10} = 0.1642$. *The matrix* ϕ *is the one defined in (3.18).*

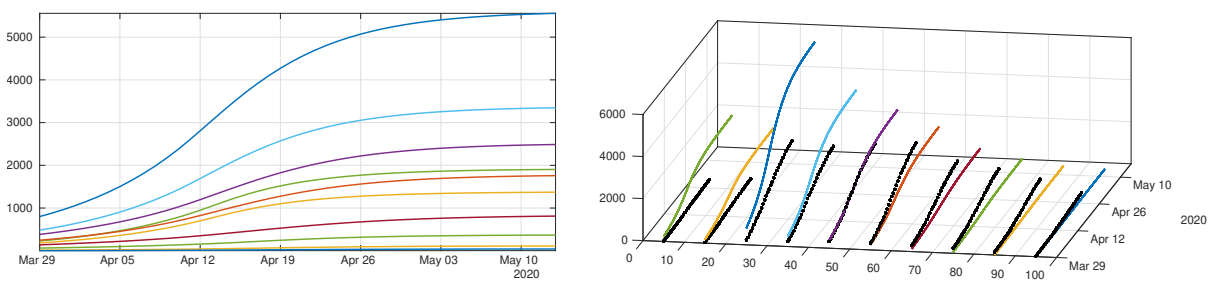

Figure 12: *Cumulative number of unreported cases as given by the fit of the model (3.12)-(3.15) to Japanese data. The solid curves represent the solution of the model and the black dots correspond to the reported cases data. Parameter are the same as in Figure 11.*

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

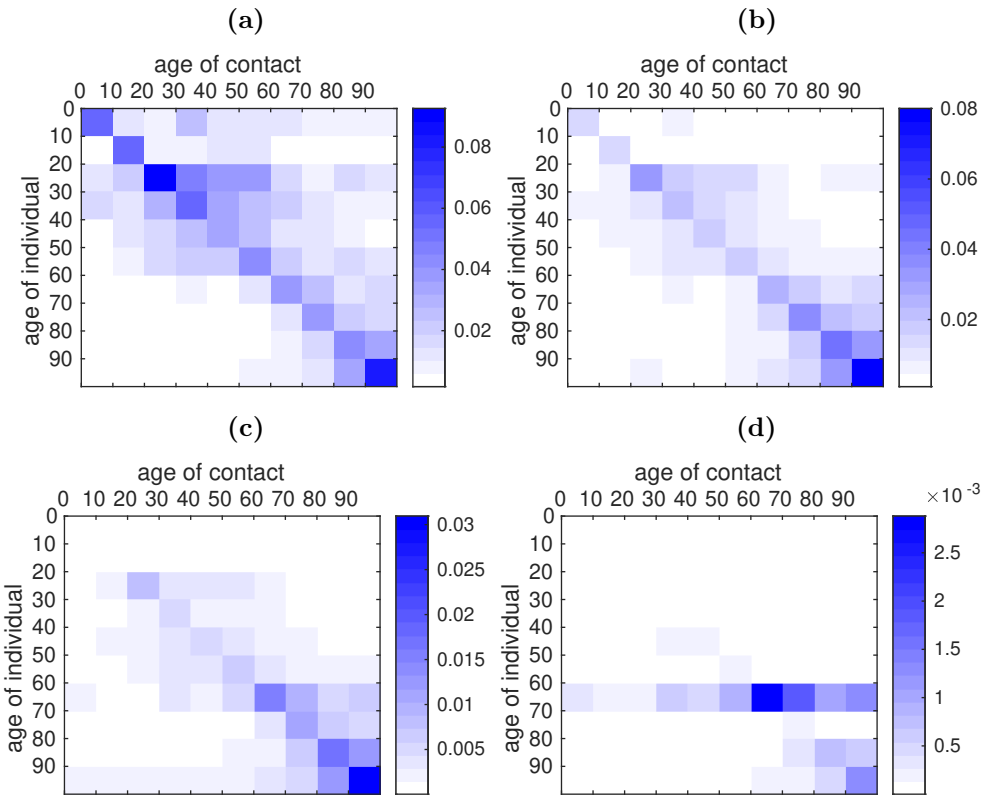

Figure 13: *Rate of contact between age classes according to the fitted data. For each age class in the y-axis we plot the rate of contacts between one individual of this age class and another individual of the age class indicated on the x-axis.* ***(a)*** *is the rate of contacts before the start of public measures (April 11).* ***(b)*** *is the rate of contacts at the date of effect of the public measures for the last age class (April 16).* ***(c)*** *is the rate of contacts one week later (April 23).* ***(d)*** *is the rate of contacts one month later (May 16). In this figure we use* $\tau_1 = 0.1630$, $\tau_2 = 0.1224$, $\tau_3 = 0.3028$, $\tau_4 = 0.2250$, $\tau_5 = 0.1520$, $\tau_6 = 0.1754$, $\tau_7 = 0.1289$, $\tau_8 = 0.1091$, $\tau_9 = 0.1211$ *and* $\tau_{10} = 0.1642$, $\mu_1 = \mu_2 = 0.4829$, $\mu_3 = \mu_4 = 0.2046$, $\mu_5 = \mu_6 = 0.1474$, $\mu_7 = 0.0744$, $\mu_8 = 0.1736$, $\mu_9 = \mu_{10} = 0.1358$, *and* $D_1 = D_2 =$ *Apr. 13,* $D_3 = D_4 = D_5 = D_6 = D_7 =$ *Apr. 11,* $D_8 = D_9 = D_{10} =$ *Apr. 16.*

In order to understand the role of transmission network between age groups in this epidemic, we plot in Figure 13 the transmission matrices computed at different times. The transmission matrix is the following

$$C(t) = \mathrm{diag}\,(\tau_1(t), \tau_2(t), \ldots, \tau_{10}(t)) \times \phi \tag{4.5}$$

where the matrix ϕ describes contacts and is given in (3.18), and the transmission rates are the ones fitted to the data as in Figure 11

$$\tau_i(t) = \tau_i^0(t) \exp(-\mu_i(t - D_i)_+).$$

During the early stages of the epidemic, the transmission seems to be evenly distributed among age classes, with a little bias towards younger age classes (Figure 13 (a)). Younger age classes seem to react more quickly to social distancing policies than older classes, therefore their transmission rate drops rapidly (Figure 13 (b) and (c)); one month after the start of social distancing measures, the transmission mostly occurs within elderly classes (60-100 years, Figure 13 (d)).

5 Discussion

The recent COVID-19 pandemic has lead many local governments to enforce drastic control measures in an effort to stop its progression. Those control measures were often taken in a state of emergency and without any real visibility concerning the later development of the epidemics, to prevent the collapse of the health systems under the pressure of severe cases. Mathematical models can precisely help see more

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

clearly what could be the future of the pandemic provided that the particularities of the pathogen under consideration are correctly identified. In the case of COVID-19, one of the features of the pathogen which makes it particularly dangerous is the existence of a high contingent of unidentified infectious individuals who spread the disease without notice. This makes non-intensive containment strategies such as quarantine and contact-tracing relatively inefficient but also renders predictions by mathematical models particularly challenging.

Early attempts to reconstruct the epidemics by using SIUR models were performed in Liu et al [11, 12, 13, 14], who used them to fit the behavior of the epidemics in many countries, by including undetected cases into the mathematical model. Here we extend our modeling effort by adding the time series of deaths into the equation. In section 4 we present an additional fit of the number of disease-induced deaths coming from symptomatic (reported) individuals (see Figure 10). In order to fit properly the data, we were forced to reduce the length of stay in the R-compartment to 6 days (on average), meaning that death induced by the disease should occur on average faster than recovery. A shorter period between infection and death (compared to remission) has also been observed, for instance, by Verity et al [26].

The major improvement in this article is to combine our early SIUR model with chronological age. Early results using age structured SIR models were obtained by Kucharski et al [10] but no unreported individuals were considered and no comparison with age-structured data were performed. Indeed in this article we provide a new method to fit the data and the model. The method extends our previous method for the SIUR model without age (see Appendix A).

The data presented in section 2 suggests that the chronological age plays a very important role in the expression of the symptoms. The largest part of the reported patients are between 20 and 60 years old (see Figure 1), while the largest part of the deceased are between 60 and 90 years old (see Figure 5). This suggests that the symptoms associated with COVID-19 infection are more severe in elderly patients, which has been reported in the literature several times (see *e.g.* Lu et al [16], Zhou et al [29]). In particular, the probability of being asymptomatic (our parameter f) should in fact depend on the age class.

Indeed, the best match for our model (see Figure 11) was obtained under the assumption that the proportion of symptomatic individual among the infected increases with the age of the patient. This linear dependency of f as a function of age is consistent with the observations of Wu et al [28] that the severity of the symptoms increase linearly with age. As a consequence, unreported cases are a majority for young age classes (for age classes less than 50 years) and become a minority for older age classes (more than 50 years), see Figure 12. Moreover, our model reveals the fact that the policies used by the government to reduce contacts between individuals have strongly heterogeneous effects depending on the age classes. Plotting the transmission matrix at different times (see Figure 13) shows that younger age classes react more quickly and more efficiently than older classes. This may be due to the fact that the number of contacts in a typical day is higher among younger individuals. As a consequence, we predict that one month after the effective start of public measures, the new transmissions will almost exclusively occur in elderly classes. The observation that younger ages classes play a major roles in the transmission of the disease has been highlighted several times in the literature, see e.g. Davies et al [5], Cao et al [4], Kucharski et al [10] for the COVID-19 epidemic, but also Mossong et al [17] in a more general context.

We develop a new model for age-structured epidemic and provided a new and efficient method to identify the parameters of this model based on observed data. Our method differs significantly from the existing nonlinear least-squares and statistical inference methods and we believe that it produces high-quality results. Moreover, we only use the initial phase of the epidemic for the identification of the epidemiological parameters, which shows that the model itself is consistent with the observed phenomenon and argues against overfitting. Yet our study could be improved in several direction. We only use reported cases which were confirmed by PCR tests, and therefore the number of tests performed could introduce a bias in the observed data – and therefore our results. We are currently working on an integration of this number of tests in our model. We use a phenomenological model to describe the response of the population in terms of number of contacts to the mitigation measures imposed by the government. This could probably be described more precisely by investigating the mitigation strategies in terms of social network. Nevertheless we believe that our study offers a precise and robust mathematical method which adds to the existing literature.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

A Appendix: Method to fit of the age structured model to the data

We first choose two days d_1 and d_2 between which each cumulative age group grows like an exponential. By fitting the cumulative age classes $[0, 10[$,$[10, 20[$, ... and $[90, 100[$ between d_1 and d_2, for each age class $j = 1, \dots 10$ we can find χ_1^j and χ_2^j

$$CR_j^{data}(t) \simeq \chi_1^j \, e^{\chi_2^j t}.$$

We choose a starting time $t_0 \leq d_1$ and we fix

$$\chi_3^j = \chi_1^j \, e^{\chi_2^j t_0}, \forall j = 1, \dots, n,$$

and we obtain

$$\begin{cases} CR_1(t) = \chi_1^1 \, e^{\chi_2^1 t} - \chi_3^1, \\ \vdots \\ CR_n(t) = \chi_1^n \, e^{\chi_2^n t} - \chi_3^n i \end{cases} \tag{A.1}$$

where

$$\chi_j^i \geq 0, \forall i = 1, \dots, n, \; \forall j = 1, 2, 3.$$

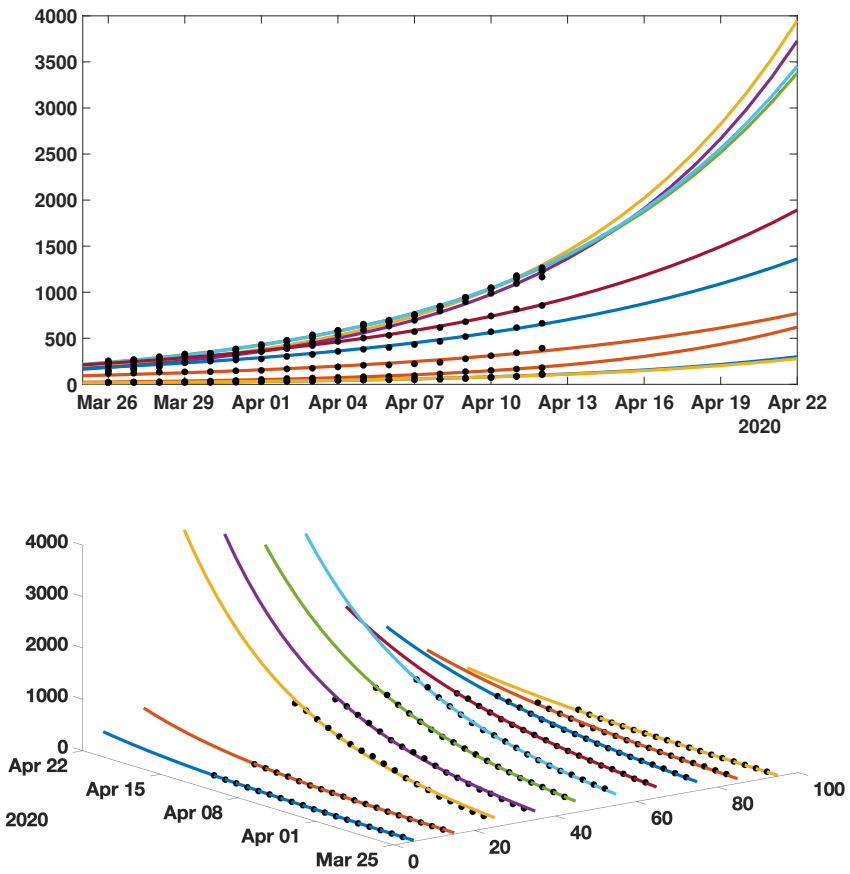

Figure 14: *We plot an exponential fit for each age classes using the data from Japan.*

We assume that

$$\begin{aligned} CR_1(t)' &= \nu_1^1 I_1(t), \\ &\vdots \\ CR_n(t)' &= \nu_1^n I_n(t), \end{aligned} \tag{A.2}$$

where

$$\nu_1^i = \nu \, f_i, \text{ and } \nu_2^i = \nu \, (1 - f_i), \forall i = 1, \dots, n.$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Therefore we obtain

$$I_j(t) = I_j^\star e^{\chi_2^j t} \tag{A.3}$$

where

$$I_j^\star := \frac{\chi_1^j \chi_2^j}{\nu_1^j}.$$

By assuming that the number of susceptible individuals remains constant we have

$$\begin{cases} I_1'(t) = \tau_1 S_1 \left[\phi_{11} \dfrac{I_1(t)+U_1(t)}{N_1} + \ldots + \phi_{1n} \dfrac{I_n(t)+U_n(t)}{N_n}\right] - \nu I_1(t), \\ \qquad\qquad \vdots \\ I_n'(t) = \tau_n S_n \left[\phi_{n1} \dfrac{I_1(t)+U_1(t)}{N_1} + \ldots + \phi_{nn} \dfrac{I_n(t)+U_n(t)}{N_n}\right] - \nu I_n(t), \end{cases} \tag{A.4}$$

and

$$\begin{cases} U_1'(t) = \nu_2^1 I_1(t) - \eta U_1(t), \\ \qquad \vdots \\ U_n'(t) = \nu_2^n I_n(t) - \eta U_n(t). \end{cases} \tag{A.5}$$

If we assume that the $U_j(t)$ have the following form

$$U_j(t) = U_j^\star e^{\chi_2^j t}, \tag{A.6}$$

then by substituting in (A.5) we obtain

$$U_j^\star = \frac{\nu_2^j I_j^\star}{\eta + \chi_2^j}. \tag{A.7}$$

The cumulative number of unreported cases $CU_j(t)$ is computed as

$$CU_j(t)' = \nu_2^j I_j(t),$$

and we used the following initial condition:

$$CU_j(0) = CU_j^* = \int_{-\infty}^{0} \nu_2^j I_j^* e^{\chi_2^j s} ds = \frac{\nu_2^j I_j^*}{\chi_2^j}.$$

We define the error between the data and the model as follows

$$\begin{cases} \varepsilon_1(t) = I_1'(t) - \tau_1 S_1 \left[\phi_{11} \dfrac{I_1(t)+U_1(t)}{N_1} + \ldots + \phi_{1n} \dfrac{I_n(t)+U_n(t)}{N_n}\right] + \nu I_1(t), \\ \qquad\qquad \vdots \\ \varepsilon_n(t) = I_n'(t) - \tau_n S_n \left[\phi_{n1} \dfrac{I_1(t)+U_1(t)}{N_1} + \ldots + \phi_{nn} \dfrac{I_n(t)+U_n(t)}{N_n}\right] + \nu I_n(t), \end{cases} \tag{A.8}$$

or equivalently

$$\begin{cases} \varepsilon_1(t) = \left(\chi_2^1 + \nu\right) I_1^\star e^{\chi_2^1 t} - \tau_1 S_1 \left[\phi_{11} \dfrac{I_1^\star + U_1^\star}{N_1} e^{\chi_2^1 t} + \ldots + \phi_{1n} \dfrac{I_n^\star + U_n^\star}{N_n} e^{\chi_2^n t}\right], \\ \qquad\qquad \vdots \\ \varepsilon_n(t) = \left(\chi_2^n + \nu\right) I_n^\star e^{\chi_2^n t} - \tau_n S_n \left[\phi_{n1} \dfrac{I_1^\star + U_1^\star}{N_1} e^{\chi_2^1 t} + \ldots + \phi_{nn} \dfrac{I_n^\star + U_n^\star}{N_n} e^{\chi_2^n t}\right]. \end{cases} \tag{A.9}$$

Let the matrix ϕ be fixed. We look for the vector $\tau = (\tau_1, \ldots, \tau_n)$ which minimizes of

$$\min_{\tau \in \mathbb{R}^n} \sum_{j=1,\ldots,n} \int_{d_1}^{d_2} \varepsilon_j(t)^2 dt.$$

Define for each $j = 1, \ldots, n$

$$K_j(t) := \left(\chi_2^j + \nu\right) I_j^\star e^{\chi_2^j t}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

and

$$H_j(t) := S_j\left[\phi_{j1}\frac{I_1^\star + U_1^\star}{N_1}e^{\chi_2^1 t} + \ldots + \phi_{jn}\frac{I_n^\star + U_n^\star}{N_n}e^{\chi_2^n t}\right],$$

so that

$$\varepsilon_j(t) = K_j(t) - \tau_j H_j(t).$$

Hence for each $j = 1, \ldots, n$

$$\int_{d_1}^{d_2} \varepsilon_j(t)^2 dt = \int_{d_1}^{d_2} K_j(t)^2 dt - 2\tau_j \int_{d_1}^{d_2} K_j(t)H_j(t)dt + \tau_j^2 \int_{d_1}^{d_2} H_j(t)^2 dt,$$

and by setting

$$0 = \frac{\partial}{\partial \tau_j}\int_{d_1}^{d_2} \varepsilon_j(t)^2 dt = -2\int_{d_1}^{d_2} K_j(t)H_j(t)dt + 2\tau_j \int_{d_1}^{d_2} H_j(t)^2 dt$$

we deduce that

$$\tau_j = \frac{\int_{d_1}^{d_2} K_j(t)H_j(t)dt}{\int_{d_1}^{d_2} H_j(t)^2 dt}. \tag{A.10}$$

Remark A.1 *It does not seem possible to estimate the matrix of contact ϕ by using similar optimization method. Indeed, if we look for a matrix $\phi = (\phi_{ij})$ which minimizes*

$$\min_{\phi \in M_n(\mathbb{R})} \sum_{j=1,\ldots,n} \int_{d_1}^{d_2} \varepsilon_j(t)^2 dt,$$

it turn out that

$$\sum_{j=1,\ldots,n} \int_{d_1}^{d_2} \varepsilon_j(t)^2 dt = 0$$

whenever ϕ is diagonal. Therefore the optimum is reached for any diagonal matrix. Moreover by using similar considerations, if several χ_j^2 are equal, we can find a multiplicity of optima (possibly with ϕ not diagonal). This means that trying to optimize by using the matrix ϕ does not yield significant and reliable information.

In the figure 15 below, we present an example of application of our method to fit the Japanese data. We use the period going from March 20 to April 15.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

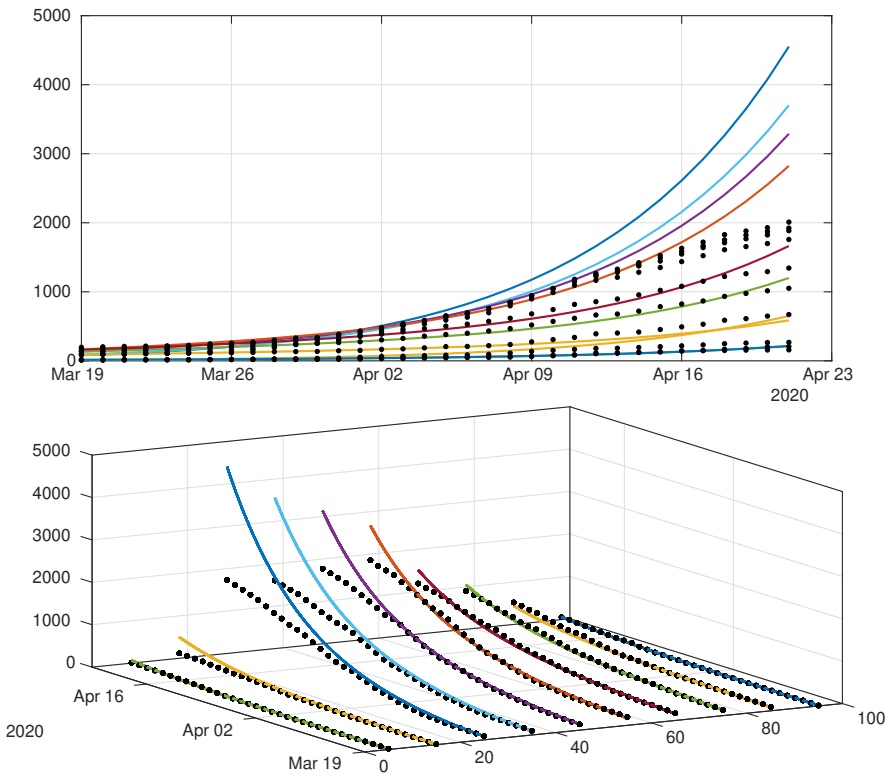

Figure 15: *We plot a comparison between the model (3.12)-(3.15) (without public intervention) and the age structured data from Japan. We set* $1/\nu = 1/\eta = 7$ *days,* f_i *which actually depends on the age class, with* $f_1 = 0.1$, $f_2 = 0.2$, $f_3 = 0.4$, $f_4 = 0.4$, $f_5 = 0.6$, $f_6 = 0.6$, $f_7 = 0.8$, $f_8 = 0.8$, $f_9 = 0.8$, *and* $f_{10} = 0.9$. *and we obtain* $\tau_1 = 0.1264$, $\tau_2 = 0.1655$, $\tau_3 = 0.3538$, $\tau_4 = 0.2966$, $\tau_5 = 0.1513$, $\tau_6 = 0.1684$, $\tau_7 = 0.1251$, $\tau_8 = 0.1168$, $\tau_9 = 0.1015$, $\tau_{10} = 0.1258$. *The matrix* ϕ *is the one defined in (3.18).*

B Appendix: Construction of the contact matrix

The survey [19] presents reconstructed contact matrices for a number of countries including Japan for the 5-years age classes $[0, 5)$, $[5, 10)$, ..., $[75, 80)$ at various locations (work, school, home, and other locations) and a compilation of those contact matrices to account for all locations. The precise description of the compilation is presented in the paper. Note that this paper is a follow-up of Mossong et al. [17] where the survey procedure is described (including the data collection protocol) for several European countries participating in the POLYMOD study.

The data is publicly available online[2] and is presented in the form of a zipped collection of spreadsheets, containing the data for several countries in columns X1 X2 ... X16. The columns stand for the average number of contact of one individual of the corresponding age class (0-5 years for X1, 5-10 years for X2, etc...), with an individual of the age class indicated by the row (first row is 0-5 years, second is 5-10 years etc...). Since the age span covered by the study stops at 80, we had to infer the number of contacts for people over the age of 80. We postulated that most people aged 80 or more are retired and that their behaviour does not significantly differs from the behavior of people in the age class $[75, 80)$. Therefore we completed the missing columns by copying the last available information and shifting it to the bottom. We repeated the procedure for lines. We believe that the introduced bias is kept to a minimum since the numerical values are relatively low compared to the diagonal.

Because we use 10-years ages classes and the data is given in 5-years age classes, we had to combine adjacent columns to recover the average number of contacts. To combine columns together, we used the weighted average

$$C'_i = \frac{N_{2(i-1)+1}C_{2(i-1)+1} + N_{2(i-1)+2}C_{2(i-1)+2}}{N_{2(i-1)+1} + N_{2(i-1)+2}},$$

[2] Prem et al [19], Supporting dataset, DOI: 10.1371/journal.pcbi.1005697.s002

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

where the column C'_i corresponds to the average number of contacts of an individual taken at random in the $[10(i-1), 10i)$ and C_i is the average number of contacts of an individual taken at random in the age class $[5(i-1), 5i)$. To combine two lines, we simply use the sum of the data

$$L'_i = L_{2(i-1)+1} + L_{2(i-1)+2}.$$

The matrix γ in (3.17) is the transpose of the array obtained by the former procedure applied to the "all locations" dataset. Then ϕ is obtained by scaling the lines of γ to 1, *i.e.*

$$\phi_{ij} = \frac{\gamma_{ij}}{\sum_{k=1}^{10} \gamma_{ik}}.$$

Acknowledgements: Data from covid19japan.com.

Conflict of Interest: None declared.

Funding: Q.G. and P.M. acknowledge the support of ANR flash COVID-19 MPCUII.

Keywords: corona virus, age-structured data, reported and unreported cases, isolation, quarantine, public closings; epidemic mathematical model

References

[1] H. H. Ayoub, *et al.*, Age could be driving variable SARS-CoV-2 epidemic trajectories worldwide, *medRxiv* (2020). https://doi.org/10.1101/2020.04.13.20059253

[2] H. H. Ayoub, *et al.*, Characterizing key attributes of the epidemiology of COVID-19 in China: Model-based estimations, *medRxiv* (2020). https://dx.doi.org/10.1101/2020.04.08.20058214

[3] M. Chikina and W. Pegden, Modeling strict age-targeted mitigation strategies for COVID-19, *arXiv* (2020). https://arxiv.org/abs/2004.04144

[4] Q. Cao, *et al.*, SARS-CoV-2 infection in children: Transmission dynamics and clinical characteristics, *Journal of the Formosan Medical Association* **119(3)** (2020), 670–673. https://dx.doi.org/10.1016%2Fj.jfma.2020.02.009

[5] N. G Davies, P. Klepac, Y. Liu, *et al.*, Age-dependent effects in the transmission and control of COVID-19 epidemics, *medRxiv*. https://doi.org/10.1101/2020.03.24.20043018.

[6] O. Diekmann, J. A. P. Heesterbeek and J. A. J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, *Journal of Mathematical Biology* **28(4)** (1990), 365-382. https://doi.org/10.1007/BF00178324

[7] Q. Griette, Z. Liu and P. Magal, Estimating the last day for COVID-19 outbreak in mainland China, *medRxiv*. https://doi.org/10.1101/2020.04.14.20064824

[8] W. Guan, Z. Ni, Y. Hu, *et al.*, Clinical Characteristics of Coronavirus Disease 2019 in China, *The New England Journal of Medicine* **382(18)** (2020), pp. 1708-1720. https://dx.doi.org/10.1056/NEJMoa2002032

[9] T. C. Jones, *et al.*, An analysis of SARS-CoV-2 viral load by patient age, *Online preprint*. https://zoonosen.charite.de/fileadmin/user_upload/microsites/m_cc05/virologie-ccm/dateien_upload/Weitere_Dateien/analysis-of-SARS-CoV-2-viral-load-by-patient-age.pdf

[10] A. Kucharski, T. W. Russell, C. Diamond, *et al.*, Early dynamics of transmission and control of COVID-19: a mathematical modelling study, *The Lancet Infectious Diseases* **20(5)** (2020), 553–558. https://doi.org/10.1016/S1473-3099(20)30144-4

[11] Z. Liu, P. Magal, O. Seydi and G. Webb, Understanding unreported cases in the 2019-nCov epidemic outbreak in Wuhan, China, and the importance of major public health interventions, *Biology* **9(3)** (2020), 50. https://doi.org/10.3390/biology9030050

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[12] Z. Liu, P. Magal, O. Seydi and G. Webb, Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data, *Mathematical Biosciences and Engineering* **17(4)** (2020), 3040-3051. https://doi.org/10.3934/mbe.2020172

[13] Z. Liu, P. Magal, O. Seydi and G. Webb, A COVID-19 epidemic model with latency period, *Infectious Disease Modelling (to appear)*. https://doi.org/10.1016/j.idm.2020.03.003

[14] Z. Liu, P. Magal, O. Seydi, and G. Webb, A model to predict COVID-19 epidemics with applications to South Korea, Italy, and Spain, *SIAM News* (2020).

[15] Z. Liu, P. Magal and G. Webb, Predicting the number of reported and unreported cases for the COVID-19 epidemic in China, South Korea, Italy, France, Germany and United Kingdom, *medRxiv*. https://doi.org/10.1101/2020.04.09.20058974

[16] X. Lu, *et al.*, SARS-CoV-2 Infection in Children, *New England Journal of Medicine* **382(17)** (2020), 1663-1665. https://doi.org/10.1056/NEJMc2005073

[17] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen *et al.*, Social contacts and mixing patterns relevant to the spread of infectious diseases, *PLoS Medicine* **5(3)** (2008), e74. https://doi.org/10.1371/journal.pmed.0050074

[18] L. Munasinghe, Y. Asai and H. Nishiura, Quantifying heterogeneous contact patterns in Japan: a social contact survey. *Theoretical biology and medical modelling*, **16(1), 6** (2019). https://doi.org/10.1186/s12976-019-0102-8

[19] K. Prem, A. R. Cook, M. Jit, Projecting social contact matrices in 152 countries using contact surveys and demographic data, *PLoS Computational Biology* **13(9)** (2017), e1005697. https://doi.org/10.1371/journal.pcbi.1005697

[20] K. Prem, Y. Liu, T. W Russell, *et al.*, 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* **5(5)** (2020). https://doi.org/10.1016/S2468-2667(20)30073-6

[21] L. Roques, E. Klein, J. Papaix, A. Sar, S. Soubeyrand, Effect of a one-month lockdown on the epidemic dynamics of COVID-19 in France, *medRxiv preprint*. https://doi.org/10.1101/2020.04.21.20074054.

[22] C. Rothe, *et al.*, Transmission of 2019-nCoV infection from an asymptomatic contact in Germany, *New England Journal of Medicine* **382(10)** (2020), 970–971. https://dx.doi.org/10.1056/NEJMc2001468

[23] R. Singh and R. Adhikari, Age-structured impact of social distancing on the COVID-19 epidemic in India, *arXiv e-print*. https://arxiv.org/abs/2003.12055.

[24] K. K-W. To, et al., Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study, *The Lancet Infectious Diseases* **20(5)** (2020), 565–574. https://doi.org/10.1016/S1473-3099(20)30196-1

[25] P. van den Driessche and J. Watmough, Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission, *Mathematical Biosciences* **180** (2002), 29-48. https://doi.org/10.1016/S0025-5564(02)00108-6

[26] R. T. Verity, *et al.*, Estimates of the severity of coronavirus disease 2019: a model-based analysis, *The Lancet Infectious Diseases* (2020). Published online March 20. https://doi.org/10.1016/S1473-3099(20)30243-7

[27] W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, , M. P. Toh, and V. J. Lee, Presymptomatic Transmission of SARS-CoV-2–Singapore, January 23–March 16, 2020, *Morbidity and Mortality Weekly Report* **69(14)** (2020), 411. http://dx.doi.org/10.15585/mmwr.mm6914e1

[28] J. T. Wu, K. Leung, M. Bushman, *et al.*, Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China, *Nature Medicine* 26 (2020), 506–510. https://doi.org/10.1038/s41591-020-0822-7

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20093807; this version posted June 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[29] F. Zhou, *et al.*, Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, *The Lancet* **395(10229)** (2020), 1054–1062. https://doi.org/10.1016/S0140-6736(20)30566-3

[30] L. Zou, *et al.*, SARS-CoV-2 viral load in upper respiratory specimens of infected patients, *New England Journal of Medicine* **382(12)** (2020), 1177-1179. https://doi.org/10.1056/NEJMc2001737

[31] World Health Organization, Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), 2020. https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19) (accessed on 06 May 2020)

[32] World Health Organization, Coronavirus disease 2019 (COVID-19): situation report, 104, 2020. https://apps.who.int/iris/handle/10665/332058 (accessed on 21 May 2020)

[33] WHO Timeline - COVID-19. https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 (accessed on 21 May 2020)

[34] World Health Organization, Pneumonia of unknown cause – China, *Disease Outbreak News*, 5 January 2020. https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/

[35] Portal Site of Official Statistics of Japan website. Reference Table for the year 2019: Computation of Population by Age (Single Years) and Sex - Total population, Japanese population. Published 2020-04-14. http://www.stat.go.jp/english/data/jinsui/index.htm (accessed on 06 May 2020)