HETEROGENEOUS SOCIAL INTERACTIONS AND THE COVID-19 LOCKDOWN OUTCOME IN A MULTI-GROUP SEIR MODEL

JEAN DOLBEAULT[1] AND GABRIEL TURINICI[1]

Abstract. We study variants of the SEIR model for interpreting some qualitative features of the statistics of the Covid-19 epidemic in France. Standard SEIR models distinguish essentially two regimes: either the disease is controlled and the number of infected people rapidly decreases, or the disease spreads and contaminates a significant fraction of the population until herd immunity is achieved. After lockdown, at first sight it seems that social distancing is not enough to control the outbreak. We discuss here a possible explanation, namely that the lockdown is creating social heterogeneity: even if a large majority of the population complies with the lockdown rules, a small fraction of the population still has to maintain a normal or high level of social interactions, such as health workers, providers of essential services, *etc.* This results in an apparent high level of epidemic propagation as measured through re-estimations of the basic reproduction ratio. However, these measures are limited to averages, while variance inside the population plays an essential role on the peak and the size of the epidemic outbreak and tends to lower these two indicators. We provide theoretical and numerical results to sustain such a view.

Résumé. Nous étudions des variantes du modèle SEIR afin d'interpréter certaines caractéristiques qualitatives des statistiques de l'épidémie de Covid-19 en France. Les modèles SEIR standards distinguent deux régimes: soit la maladie est contrôlée et le nombre de personnes infectées diminue rapidement, soit la maladie se propage et contamine une fraction importante de la population jusqu'à ce que l'immunité collective soit atteinte. Après le confinement, il semble à première vue que la distanciation sociale soit insuffisante pour contrôler l'épidémie. Nous avançons ici une explication possible, à savoir que le confinement crée de l'hétérogénéité sociale: même si une grande majorité de la population obéit aux règles de confinement, une petite fraction doit continuer à maintenir un niveau normal ou élevé d'interactions sociales, comme les personnels médicaux, les prestataires de services essentiels, *etc.* Cela se traduit par un niveau de propagation élevé de l'épidémie, mesuré par des ré-estimations du taux de reproduction de base. Ces mesures se limitent toutefois à des moyennes alors que la variance au sein de la population joue un rôle essentiel sur le pic et la taille de l'épidémie et contribue à abaisser ces deux indicateurs. Nous apportons des arguments théoriques et numériques pour développer ce point de vue.

Mathematics Subject Classification. 92C60, 92D30, 34C60.

June 23, 2020.

---

[1] CEREMADE (CNRS UMR n° 7534), PSL university, Université Paris-Dauphine, Place de Lattre de Tassigny, 75775 Paris 16, France. E-mail (J.D): dolbeaul@ceremade.dauphine.fr. E-mail (G.T.): turinici@ceremade.dauphine.fr.

Introduction

Although widely used in practice, compartmental models of epidemic spread (as for instance the celebrated SIR [26] model) rely on various simplifying assumptions, like a limited number of compartments, the homogeneity of the population inside a compartment, or some well defined parameters. The advantage of such models is that the impact of the variation of a single parameter on various qualitative properties can easily be studied, but the risk lies in an outrageous simplification of the representation of a complex system, with the additional difficulty that some parameters can be difficult to quantify from the available statistical data. Here we numerically study how the separation of the population into two sub-groups of individuals with different intensities of social interaction can qualitatively explain some observed features of the current pandemic of Covid-19, and provide some theoretical explanations which also apply to more realistic models.

New characteristics of the pandemic of Covid-19 are unveiled every day and reveal various interesting features. One of the issues is that curves showing the number of new cases in European countries stabilize very slowly after the beginning of the lockdown. Many questions have been raised on the methods for collecting the data and on their quality, that we do not address, not to mention variations between countries that deserve further investigations. As a simple explanation, it is has been suggested that the main reason is that a significant fraction of the population does not respect lockdown. This ignores the nonlinear properties of simple epidemic models. We show here that, if a rather small fraction of the population cannot reduce its social interactions, eventually for very good reasons as, *e.g.*, the health workers or some other key actors of our societies, the epidemic keeps spreading, stabilizes at a much slower rate than what one would expect, and finally affects a significant fraction of the population.

In this paper we shall use the SEIR model rather than the simpler SIR model because of the significant period of incubation in the Covid-19 disease. The SEIR model has various properties that can be understood, at least for the order of magnitude of its effects, using simple objects like the well known *basic reproduction ratio* $\mathcal{R}_0$. However, it also has very nonlinear features which are defying common sense and require rigorous mathematical analysis. In the accompanying numerical examples, we take values for the parameters which are compatible with the data collected during the pandemic of Covid-19. These examples are intended to understand qualitative features of the epidemic but cannot and should not be implemented for direct predictive use, as our model is too crude and oversimplified to reflect the Covid-19 epidemic in a quantitative manner.

We start by reviewing known results concerning the SEIR model applied to a homogeneous population in Section 1. A model with a large majority of the population under lockdown and a small minority which does not or cannot implement the *social distancing* is then considered. The *factor of reduction of social interactions q* quantifies the effect of the lockdown. It is certainly very difficult to measure this parameter in real life applications and structured population models should be considered, in order to reflect the variety of social interactions. Our goal is to understand how, by varying q, one eventually triggers the nonlinearity of the SEIR model and, as a consequence, how this drives the system into a *stable equilibrium* which is far from the usual target of a lockdown policy, that is, the control of the disease, but is also far from the dynamics that would develop in absence of lockdown.

1. Basic mathematical properties of the SEIR model

1.1. The SEIR model

Let us consider the SEIR (*Susceptible*, *Exposed*, *Infected*, *Recovered*) model [4] defined by the set of equations

$$\frac{dS}{dt} = -\beta S \frac{I}{N} \,, \tag{1}$$

$$\frac{dE}{dt} = \beta S \frac{I}{N} - \alpha E \,, \tag{2}$$

$$\frac{dI}{dt} = \alpha E - \gamma I \,, \tag{3}$$

$$\frac{dR}{dt} = \gamma I \,, \tag{4}$$

which is a variant of the SIR model [26] of Kermack and McKendrick. Here we neglect birth and death rates, *i.e.*, we consider a model without vital dynamics. The *average incubation period* is $1/\alpha$, the parameter β is the product of the average number of contacts per person and per unit time by the probability of disease transmission in a contact between a susceptible and an infectious individual, γ is a *transition rate* so that $1/\gamma$ measures the duration of the infection of an individual and N is the total population size. In the Covid-19 pandemy, the average incubation period is of several days and this is why an SEIR model has to be preferred to a simple SIR model. Many qualitative features are the same in the two models but the compartment E of exposed individuals makes the analysis slightly more delicate. Unreported cases or asymptomatic individuals are not taken into account here: this is an important aspect of the Covid-19 epidemic, see for instance [9,34,35], with important consequences on the epidemic size, but probably not so much on the qualitative issues that are discussed in this paper. Other factors, like delays for the transmission of the information studied in [7], certainly also play a role in the current outbreak.

The SEIR model is a *compartmental model* used to understand the mathematical modelling of infectious diseases in a large population, with enough individuals in each compartment so that stochastic effects can be neglected. Spatial effects are also neglected, which is a rather crude approximation. However, such a simplified model allows us to perform an analysis of the sensibility to the parameters which are of interest for a qualitative description of the outbreak of an epidemic disease.

The system (1)-(4) is homogeneous so that we can simply consider the *fractions*

$$\mathsf{s} := \frac{S}{N} \,, \quad \mathsf{e} := \frac{E}{N} \,, \quad \mathsf{i} := \frac{I}{N} \,, \quad \mathsf{r} := \frac{R}{N}$$

of the *Susceptible*, *Exposed*, *Infected* and *Recovered* individuals among the whole population.

1.2. Conservations and large time asymptotics

Here we perform a simple analysis of the model, which is done in the spirit of [22]. By conservation of the total number of individuals, we have the relation

$$\mathsf{s}(t) + \mathsf{e}(t) + \mathsf{i}(t) + \mathsf{r}(t) = 1 \tag{5}$$

4

for any time $t \geq 0$, which is easily proved by summing the equations in the system, while the evolution is now governed by the system

$$\frac{d\mathsf{s}}{dt} = - \beta \, \mathsf{s} \, \mathsf{i} \,, \tag{6}$$

$$\frac{d\mathsf{e}}{dt} = \beta \, \mathsf{s} \, \mathsf{i} - \alpha \, \mathsf{e} \,, \tag{7}$$

$$\frac{d\mathsf{i}}{dt} = \alpha \, \mathsf{e} - \gamma \, \mathsf{i} \,, \tag{8}$$

$$\frac{d\mathsf{r}}{dt} = \gamma \, \mathsf{i} \,. \tag{9}$$

An important and classical observation is that the domain

$$\mathsf{s} \geq 0 \,, \quad \mathsf{e} \geq 0 \,, \quad \mathsf{i} \geq 0 \,, \quad \mathsf{s} + \mathsf{e} + \mathsf{i} \leq 1$$

is stable under the action of the flow (6)-(9), and it is straightforward to check that

$$\frac{d}{dt} \left(\gamma \, \log \mathsf{s} - \beta \, (\mathsf{s} + \mathsf{e} + \mathsf{i}) \right) = 0 \,. \tag{10}$$

Since s is monotone non-increasing by (6), this means that (6)-(9) has a global solution for any $t \geq 0$ and

$$\lim_{t \to +\infty} \mathsf{s}(t) = \mathsf{s}^\star > 0 \,.$$

By an elementary analysis, we find that any solution $(\mathsf{s}, \mathsf{e}, \mathsf{i}, \mathsf{r})$ of (6)-(9) with initial data $(\mathsf{s}_0, \mathsf{e}_0, \mathsf{i}_0, \mathsf{r}_0)$ such that $\mathsf{i}_0 + \mathsf{e}_0 > 0$ converges as $t \to +\infty$ to a stationary solution $(\mathsf{s}^\star, \mathsf{e}^\star, \mathsf{i}^\star, \mathsf{r}^\star)$ with

$$\mathsf{e}^\star = \mathsf{i}^\star = 0 \,, \quad \mathsf{s}^\star + \mathsf{r}^\star = 1 \quad \text{and} \quad \gamma \, \log \mathsf{s}^\star - \beta \, \mathsf{s}^\star = \gamma \, \log \mathsf{s}_0 - \beta \, (\mathsf{s}_0 + \mathsf{e}_0 + \mathsf{i}_0) \,, \tag{11}$$

according to (5) and (10). Note that the solution is unique if $\gamma - \beta > 0$ but there are two solutions if $\gamma - \beta < 0$ and $\gamma \, \log \mathsf{s}_0 + \beta \, \mathsf{r}_0 = \gamma \, \log \mathsf{s}_0 - \beta \, (\mathsf{s}_0 + \mathsf{e}_0 + \mathsf{i}_0) + \beta > 0$. In our numerical applications, we shall assume that $\gamma \, \log \mathsf{s}_0 + \beta \, \mathsf{r}_0 = \gamma \, \log \mathsf{s}_0 - \beta \, (\mathsf{s}_0 + \mathsf{e}_0 + \mathsf{i}_0) + \beta < 0$ if $\gamma - \beta < 0$, so that $\mathsf{s}^\star$ is uniquely defined.

The *epidemic size* ζ is defined as the fraction of individuals that are affected by the epidemic, here $\mathsf{s}_0 - \mathsf{s}^\star$. Next, let us consider more specifically the case of an initial datum which is a perturbation of the constant in time *DFE solution*, or *Disease Free Equilibrium* (see [45]), corresponding to

$$(\mathsf{s}_0, \mathsf{e}_0, \mathsf{i}_0, \mathsf{r}_0) = (1, 0, 0, 0) \,. \tag{12}$$

1.3. Stable equilibrium, epidemic size and phase transition

Let us assume that $\mathsf{e}_0 + \mathsf{i}_0 + \mathsf{r}_0 = \varepsilon > 0$ is small and $\mathsf{r}_0 = \vartheta \, \varepsilon$ for some given $\vartheta \in [0, 1)$. This initial condition, in the limit as $\varepsilon \to 0_+$, is a perturbation of the DFE solution. For any $\varepsilon > 0$, we know by (5) that $\mathsf{s}_0 = 1 - \varepsilon$. The solution of (6)-(9) converges as $t \to +\infty$ to $(\mathsf{s}^\star, 0, 0, \mathsf{r}^\star = 1 - \mathsf{s}^\star)$, the stationary solution $(\mathsf{s}^\star, 0, 0, \mathsf{r}^\star)$ is stable and this is why we call it the *stable equilibrium* solution. See [36, Section 7.3] for a discussion of the stability. In a model with birth and death rates, the solution is usually called the *endemic equilibrium*, but we shall prefer to call it simply the *stable equilibrium* solution as we neglect birth and death issues. The following discussion is given here in preparation for the next two sections. We do not claim originality and refer for instance to [22] for a detailed study motivated by very explicit examples. Depending whether the DFE solution is stable or not, we have three regimes for $\mathsf{s}^\star = \mathsf{s}^\star(\varepsilon)$:

▷ **Control of the epidemic.** If $\gamma > \beta$, we find that

$$\mathsf{s}^\star(\varepsilon) = \mathsf{s}_0 - \frac{\beta\,\varepsilon}{\gamma - \beta} + o(\varepsilon) = 1 - \frac{\gamma\,\varepsilon}{\gamma - \beta} + o(\varepsilon) \tag{13}$$

as $\varepsilon \to 0_+$. In other words, the stable equilibrium is a perturbation of the DFE solution and the *epidemic size* ζ is of order ε.

▷ **Epidemic spreading and herd immunity.** If $\gamma < \beta$, then we note that $s \mapsto \gamma \log s - \beta s$ achieves a maximum point on $(0, 1)$ at $s = \gamma/\beta$ and, as a consequence, that

$$\mathsf{s}^\star(\varepsilon) < \frac{\gamma}{\beta} \tag{14}$$

even for an arbitrarily small value of $\varepsilon > 0$. Hence we find that $\mathsf{s}^\star(\varepsilon)$ is not only a perturbation of order ε of s_0 but also that the value of $\mathsf{s}^\star(\varepsilon)$ is of order $\mathsf{s}^\star(0)$, the unique root in $(0, 1)$ of $\gamma \log s + \beta\,(1 - s) = 0$. Since the constant solution $\left(\mathsf{s}^\star(\varepsilon), 0, 0, 1 - \mathsf{s}^\star(\varepsilon)\right)$ is a stable equilibrium, *herd immunity* is always granted in the sense that no outbreak can occur. Let $\mathcal{R}_0 = \beta/\gamma$.

With $\mathcal{R}_0 > 1$, the epidemic spreads in the sense that $\mathsf{s}^\star(0)$ is the solution of

$$\log s - \mathcal{R}_0\,(s - 1) = 0\,. \tag{15}$$

This means that $\mathsf{s}^\star(0) = -\frac{1}{\mathcal{R}_0}\,W\left(-\mathcal{R}_0\,e^{-\mathcal{R}_0}\right)$ where the function W is known as the *Lambert function* (see for instance [8]) and defined as the inverse of $w \mapsto w\,e^w$. Note that $r = 1 - s$ is given for $s = \mathsf{s}^\star(0)$ by the equation

$$e^{-\mathcal{R}_0\,r} + r = e^{-\frac{\beta\,r}{\gamma}} + r = 1 \tag{16}$$

and that $r = 1 - \mathsf{s}^\star(\varepsilon)$ solves

$$\mathsf{s}_0\,e^{-\frac{\beta}{\gamma}\,(r - \mathsf{r}_0)} + r = 1\,, \quad \mathsf{s}_0 = 1 - \varepsilon\,, \tag{17}$$

so that $\mathsf{s}^\star(\varepsilon)$ also depends on ϑ (but this dependence disappears in the limit as $\varepsilon \to 0_+$). Altogether, if $\gamma < \beta$, that is, for $\mathcal{R}_0 > 1$, the *epidemic size* ζ is of order 1 as $\varepsilon \to 0_+$. See Fig. 2 for an illustration of the dependence of ζ in $\mathcal{R}_0$.

▷ In the **threshold case** $\beta = \gamma$, *i.e.*, $\mathcal{R}_0 = 1$, which is typical of a phase transition, we have to solve

$$\log \mathsf{s}^\star(\varepsilon) - \mathsf{s}^\star(\varepsilon) = \log(1 - \varepsilon) + \mathsf{r}_0 - 1 \tag{18}$$

and find that

$$\mathsf{s}^\star(\varepsilon) = 1 - \sqrt{2\,(1 - \vartheta)\,\varepsilon} + o\!\left(\sqrt{\varepsilon}\right) \tag{19}$$

as $\varepsilon \to 0_+$. We recall that $0 \le \mathsf{r}_0 < \varepsilon$ and observe that the *epidemic size* ζ is of order $\sqrt{\varepsilon}$ as $\varepsilon \to 0_+$. Scale invariance is reflected by the fact that there is no dependence neither on β nor on γ.

Summarizing, there are two phases and a threshold case corresponding to the *phase transition*. If $\gamma > \beta$, the epidemic size ζ is close to zero. The disease does not spread in the population and simply vanishes exponentially fast. On the opposite, if $\gamma < \beta$, the *Disease Free Equilibrium* is unstable, the diseases quickly spreads with an exponential growth, the system converges for large times to a stable equilibrium far away from the DFE solution and the epidemic size ζ is a significant fraction of the total population, irrespective of how small the fraction of initially infected individuals is. Whether $\mathcal{R}_0 > 1$ or $\mathcal{R}_0 < 1$ determines the asymptotic stable equilibrium starting from the DFE solution. In the literature, $\mathcal{R}_0$ is called the *basic reproduction ratio*, or the *basic reproduction number*. So far we simply consider it here as the *order parameter* of the phase transition, in the usual sense in physics: see for instance [30, p. 449] or [44, p. 3]. In the next section, we shall explain

the role it plays in the initial dynamics of the model and refer to Section 1.6 for more considerations on the epidemiologic interpretation of $\mathcal{R}_0$.

Note that the SIR model has exactly the same stationary states and the same phase transition as the SEIR model, as the SIR model is obtained by replacing (7) by $\beta\,\mathsf{s}\,\mathsf{i} = \alpha\,\mathsf{e}$, so that, in the SIR model, the equation for i becomes

$$\frac{d\mathsf{i}}{dt} = \beta\,\mathsf{s}\,\mathsf{i} - \gamma\,\mathsf{i}. \tag{20}$$

The order parameter is also $\mathcal{R}_0 = \beta/\gamma$ and whether $\mathcal{R}_0 > 1$ or $\mathcal{R}_0 < 1$ determines if the epidemic is spreading or if the disease is controlled.

1.4. **Linearization and the basic reproduction ratio**

In the case of the SIR model, (20) can be rewritten as

$$\frac{d\mathsf{i}}{dt} = (\beta\,s - \gamma)\,\mathsf{i} = \big(\mathcal{R}_0\,\mathsf{s} - 1\big)\,\gamma\,\mathsf{i} \tag{21}$$

and it is elementary to observe that when $\mathsf{s} \sim 1$, whether $\mathcal{R}_0 > 1$ or $\mathcal{R}_0 < 1$ determines the initial dynamics of the model: the interpretation of $\mathcal{R}_0$ is clear from the above equation. After this digression on the SIR model, let us come back to the SEIR model. Understanding the role of $\mathcal{R}_0$ is a little bit more subtle than in the SIR model. We recall that $\mathsf{r}(t) = 1 - \mathsf{s}(t) - \mathsf{e}(t) - \mathsf{i}(t)$ plays no role in the stability analysis of the DFE solution. At any time t, the linearized dynamics of $t \mapsto \big(\mathsf{s}(t), \mathsf{e}(t), \mathsf{i}(t)\big)$ is described by the matrix

$$\mathcal{M}(\mathsf{s}, \mathsf{i}) := \begin{pmatrix} -\beta\,\mathsf{i} & 0 & -\beta\,\mathsf{s} \\ \beta\,\mathsf{i} & -\alpha & \beta\,\mathsf{s} \\ 0 & \alpha & -\gamma \end{pmatrix} \tag{22}$$

and we may notice that the largest eigenvalue of $\mathcal{M}(1,0)$, corresponding to the linearization around the DFE solution, is

$$\lambda(\alpha, \beta, \gamma) := \frac{1}{2}\left(\sqrt{(\alpha - \gamma)^2 + 4\,\alpha\,\beta} - \alpha - \gamma\right), \tag{23}$$

so that $\lambda(\alpha, \beta, \gamma)$ is positive if and only if $\mathcal{R}_0 = \beta/\gamma > 1$. Moreover, the eigenspace is compatible with the nonlinear dynamics so that there are perturbations of the DFE solution which are exponentially growing with a rate $\lambda(\alpha, \beta, \gamma)$ if and only if $\mathcal{R}_0 > 1$. However, the basic reproduction ratio $\mathcal{R}_0$ is not anymore directly connected with the linearized growth mode of i. In fact, we can observe that (21) is replaced by

$$\frac{d}{dt}(\mathsf{e} + \mathsf{i}) = (\beta\,s - \gamma)\,\mathsf{i} \sim \big(\mathcal{R}_0 - 1\big)\,\gamma\,\mathsf{i} \tag{24}$$

if $\mathsf{s} \sim 1$, which is indeed the correct way of estimating the growth of the epidemic. We will come back on the interpretation of $\mathcal{R}_0$ in Section 1.6 and note that $(\mathsf{e} + \mathsf{i})$ is known in the literature as the population in the *infectious compartments*.

1.5. **Social distancing and the factor of reduction of social interactions**

The goal of a lockdown policy is to replace the system (6)-(9) by

$$\frac{d\mathsf{s}}{dt} = -\frac{\beta}{q}\,\mathsf{s}\,\mathsf{i} \tag{25}$$

$$\frac{d\mathsf{e}}{dt} = \frac{\beta}{q}\,\mathsf{s}\,\mathsf{i} - \alpha\,\mathsf{e} \tag{26}$$

$$\frac{d\mathsf{i}}{dt} = \alpha\,\mathsf{e} - \gamma\,\mathsf{i} \tag{27}$$

$$\frac{d\mathsf{r}}{dt} = \gamma\,\mathsf{i} \tag{28}$$

for some factor $q > 1$ which measures the *reduction of social interactions* of each individual. Of course, what we obtain is exactly (6)-(9) with the parameter β replaced by β/q. The point is that the basic reproduction ratio becomes

$$\mathcal{R}_0^{(1)}(q) = \frac{\beta}{\gamma\,q} = \frac{\mathcal{R}_0}{q} \tag{29}$$

and the goal is either to fix q to a value large enough so that the epidemic is controlled, that is $q > \mathcal{R}_0$, or at least to make $\mathcal{R}_0 - q > 0$ small in order to *flatten the curve*, *i.e.*, to have an epidemic going at slower pace. See Fig. 1 for an illustration. Here the exponent (1) in the notation $\mathcal{R}_0^{(1)}$ points to the assumption that we consider a population with a single group of susceptible individuals or, in other words, a socially homogeneous population.

In a SIR model with social distancing, with (25), (28) on the one hand, but (26) and (27) replaced by

$$\frac{d\mathsf{i}}{dt} = \frac{\beta}{q}\,\mathsf{s}\,\mathsf{i} - \gamma\,\mathsf{i} \tag{30}$$

on the other hand, it is possible to compute the epidemic peak. This is a classical result, see for instance [36, Section 2.1.2]. The epidemic peak is defined as the maximum $\mathsf{i}(t_p)$ of $t \mapsto \mathsf{i}(t)$ after noticing that, at $t = t_p$, we have the system of equations

$$\frac{\beta}{q}\,\mathsf{s}\,\mathsf{i} - \gamma\,\mathsf{i} = 0\,, \quad \gamma\,\log\left(\frac{\mathsf{s}}{\mathsf{s}_0}\right) + \frac{\beta}{q}\,(\mathsf{r} - \mathsf{r}_0) = 0\,, \quad \mathsf{s} + \mathsf{i} + \mathsf{r} = 1\,, \tag{31}$$

which provides us with the value

$$\mathsf{i}(t_p) = 1 - \mathsf{r}_0 - \frac{1}{\mathcal{R}_0^{(1)}(q)}\left(1 + \log\left(\mathcal{R}_0^{(1)}(q)\,\mathsf{s}_0\right)\right) \tag{32}$$

for any $q < 1/\mathcal{R}_0$. Note that $\mathsf{i}(t_p) \sim 1 - \mathcal{R}_0^{(1)}(q)^{-1}\left(1 + \log(\mathcal{R}_0^{(1)}(q))\right)$ as $(\mathsf{s}_0, \mathsf{r}_0) \to (1, 0)$, *i.e.*, in the limit of a DFE initial datum. We obtain the same expression in the SEIR model if we replace i by $\mathsf{e} + \mathsf{i}$. See [15] as a source of inspiration for such considerations, and also Property (2) of Theorem 2, in a much more general framework.

1.6. **The basic reproduction ratio and the method of the next generation matrix**

According to [12], the basic reproduction ratio $\mathcal{R}_0$ is the *expected number of secondary cases produced, in a completely susceptible population, by a typical infected individual during its entire period of infectiousness.* In [6], Blackwood and Childs provide us with a comprehensive introduction to the computation of $\mathcal{R}_0$ using the method of the *next generation matrix* in the case of the SEIR model. Such a computation goes back to [12], the standard method is exposed in [45] and we can also refer to [13] for an application to the SEIR model. As for a

more general presentation of the method in compartmental models and more formal mathematical treatments, one can refer to [11, 36], and to [23] and [4, p. 17] for early considerations on endemic and stable equilibria. Alternative definitions of the basic reproduction ratio in compartmental models are also available: see [36] for an overview, [22] for a more historical account and [21, 23] for considerations on models which are more directly linked to our interests (see Section 2).

For sake of completeness, let us give a brief summary of the method of the *next generation matrix*. First of all, one restricts the analysis to the *infectious compartments*, $\mathbf{x} = (\mathsf{e}, \mathsf{i})$ in case of (25)-(28), and consider the linearized evolution equation around the DFE solution, that is,

$$\frac{d\mathbf{x}}{dt} = (\mathrm{F} - \mathrm{V})\,\mathbf{x} \tag{33}$$

where F and V respectively denote the matrices associated with the rate of new infections and the rates of transfer between compartments, *i.e.*,

$$\mathrm{F} = \begin{pmatrix} 0 & \frac{\beta}{q} \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathrm{V} = \begin{pmatrix} -\alpha & 0 \\ \alpha & -\gamma \end{pmatrix} \tag{34}$$

See [45] for details. According to [45, Lemma 1], we observe that the matrix F is non-negative and the matrix V is non-singular. In this framework, the *basic reproduction ratio* is defined as the largest eigenvalue of $\mathrm{F}\,\mathrm{V}^{-1}$. It is an elementary computation to check that

$$\mathrm{F}\,\mathrm{V}^{-1} = \begin{pmatrix} \frac{\beta}{q\gamma} & \frac{\beta}{q\gamma} \\ 0 & 0 \end{pmatrix} \tag{35}$$

has two eigenvalues, 0 and $\frac{\beta}{q\gamma} = \mathcal{R}_0^{(1)}(q)$. This proves that $\mathcal{R}_0^{(1)}(q)$ is the basic reproduction ratio, as defined by the method of the next generation matrix. In this framework, it is known from [45, Theorem 2] that the DFE solution is stable if $\mathcal{R}_0^{(1)}(q) < 1$ and unstable if $\mathcal{R}_0^{(1)}(q) > 1$.

2. A heterogenous model of social distancing

2.1. A simple model with two groups

Let us consider a population divided in two groups indexed by $k = 1$, 2, in which the *Susceptible* individuals have a factor of reduction of social interactions q_k which differ in the two groups. We shall assume that each of these groups gather a fixed fraction of the population p_k with $p_2 = p$ small and $p_1 = 1 - p$. While the group corresponding to $k = 1$ observes a lockdown and has a factor $q_k > 1$, we are interested in the situation in which the other group has no reduction of social interactions: $q_2 = 1$, or eventually has more social interactions than average before lockdown, corresponding to some $q_2 < 1$. The typical example is the case of health workers in a period of epidemic disease or supermarket cashiers, who have contacts with a much larger number of people than an average individual. It is of course very difficult to estimate q_2 and one should take into account the efficiency of barrier procedures. Instead of trying to make rough guesses for the value of q_2, we will vary it in order to see what is the impact on the solutions.

With a straightforward notation, let us split the population of *Susceptible* individuals in two groups

$$\mathsf{s} = \mathsf{s}_1 + \mathsf{s}_2$$

and consider for $k = 1$, 2 the system

$$\frac{\mathsf{s}_k'}{\mathsf{s}_k} = -\beta_k\,\mathsf{i} \quad \text{with} \quad \beta_k = \frac{\beta}{q_k}, \quad \mathsf{e}' = (\beta_1\,\mathsf{s}_1 + \beta_2\,\mathsf{s}_2)\,\mathsf{i} - \alpha\,\mathsf{e}, \quad \mathsf{i}' = \alpha\,\mathsf{e} - \gamma\,\mathsf{i}, \quad \mathsf{r}' = \gamma\,\mathsf{i}. \tag{36}$$

There are multiple possible variants and it would make sense, for instance, to distinguish i_1 and i_2 in the above equations, with detailed contamination rules. The above system has striking properties. It is for instance straightforward to see that the linearized system around the DFE solution, *i.e.*, the matrix $\mathcal{M}(1,0)$, has a largest eigenvalue given by

$$\lambda(\alpha, \beta, \gamma) = \frac{1}{2}\left(\sqrt{(\alpha - \gamma)^2 + 4\,\alpha\,\gamma\,\mathcal{R}_0^{(2)}(q_1, q_2, p)} - \alpha - \gamma\right), \qquad (37)$$

where the basic reproduction ratio $\mathcal{R}$, as defined in the method of the *next generation matrix*, is given by

$$\mathcal{R}_0^{(2)}(q_1, q_2, p) = \frac{(1 - p)\,\beta_1 + p\,\beta_2}{\gamma}. \qquad (38)$$

One can indeed apply the method of Section 1.6 and observe that the only change lies in the matrix F, where the coefficient β/q has to be replaced by $(1-p)\,\beta_1 + p\,\beta_2$, which establishes (38). It is easy to deduce from (37) that $\lambda(\alpha, \beta, \gamma)$ is positive if and only if $\mathcal{R}_0^{(2)}(q_1, q_2, p) > 1$.

Modeling heterogeneous mixing in infectious disease dynamics when the population is subdivided by characteristics other than those that are disease-related, such as risk status or age, is not new. This has been considered for instance from the dynamical point of view in [28] or in the case of sexually transmitted diseases and particularly in HIV/AIDS models, with groups that are not all defined by disease related properties. In this perspective, *contact matrices* have been considered, which involve a detailed analysis of the transmission mechanisms. We can refer to [2, 23–25, 37] for various considerations in this direction and to [21, Section 3] for a discussion of the *homogeneous mixing fallacy* in the application to successful vaccination policies. The present paper ignores a number of issues like symmetry in transmission between groups and density-dependent transmission questions in order to focus on simple qualitative questions: any serious study with quantitative goals should of course address these issues with care: see for instance [6] for a warning. For sake of simplicity, we have chosen to consider that the origin of the infected individuals (group 1 or 2) plays no role in the transmission. This has the simple consequence that the basic reproduction ratio $\mathcal{R}$ is in the end exactly the *average of the ratios independently computed for each group*, as shown by (38). In the regime corresponding to p much smaller than $1 - p$ and whatever the details are, our model is anyway good enough to show that what matters concerning the basic reproduction ratio is the average of the ratios[1]. However, when one considers the epidemic peak and the large time asymptotics, the message is not only the average and this is what we explain next.

2.2. Conservations and large time asymptotics

System (36) inherits of the properties of the standard SEIR model. The conservation of mass

$$\mathsf{s}_1(t) + \mathsf{s}_2(t) + \mathsf{e}(t) + \mathsf{i}(t) + \mathsf{r}(t) = 1 \quad \forall\, t \geq 0 \qquad (39)$$

guarantees that all quantities are bounded by 1 as long as they are nonnegative. For a solution of (36), let us observe that

$$\frac{d}{dt}\left(\log \mathsf{s}_k\right) = -\beta_k\,\mathsf{i} = -\frac{\beta_k}{\gamma}\,\mathsf{r}', \qquad (40)$$

so that

$$\mathsf{s}_k(t) = \mathsf{s}_k^0\, e^{-\frac{\beta_k\,\mathsf{r}(t)}{\gamma}} \quad \text{with} \quad \mathsf{s}_k^0 = \mathsf{s}_k(0)\, e^{\frac{\beta_k\,\mathsf{r}(0)}{\gamma}}. \qquad (41)$$

We shall moreover assume that

$$\mathsf{s}_k(0) = p_k\,\mathsf{s}(0) \qquad (42)$$

---

[1] For instance, it is not the average of the factors of reduction of social interactions which matters, as it is usually done when considering homogeneous populations represented as a single group.

as the population can be considered, in the initial phase of the outbreak (for $t \leq 0$, in our setting), as a single group. This point could be reconsidered and studied as in [5] if one is interested in the dynamics of the epidemics, but it has a no significant impact on the stable equilibrium in the uncontrolled case. An analysis of the trajectories of (36) as in Section 1.3 shows that solutions globally exist and that there is a unique stable attractor $(\mathsf{s}_1^\star, \mathsf{s}_2^\star, 0, 0, \mathsf{r}^\star)$. We deduce from the conservation of mass

$$\mathsf{s}_1^\star + \mathsf{s}_2^\star + \mathsf{r} = 1 \tag{43}$$

that the stable stationary solution is given as the unique solution with $\mathsf{r}^\star = r > \mathsf{r}(0)$ of

$$1 = \sum_{k=1,2} \mathsf{s}_k^0 \, e^{-\frac{\beta_k \, r}{\gamma}} + r \,, \quad \mathsf{s}_k^\star = \mathsf{s}_k^0 \, e^{-\frac{\beta_k \, r}{\gamma}} \,, \quad k = 1,\, 2 \,. \tag{44}$$

Under the condition that $\big(\mathsf{e}_k(0), \mathsf{i}_k(0)\big) \neq (0,0)$, we have that

$$\lim_{t \to +\infty} \big(\mathsf{s}_1(t), \mathsf{s}_2(t), \mathsf{e}(t), \mathsf{i}(t), \mathsf{r}(t)\big) = (\mathsf{s}_1^\star, \mathsf{s}_2^\star, 0, 0, \mathsf{r}^\star) \,. \tag{45}$$

2.3. Stable equilibrium and phase transition

The same discussion as in Section 1.3 can be done. To fix ideas, let us assume for instance, as a simplifying assumption, that $\mathsf{s}(0) = 1 - \varepsilon$, $\mathsf{e}(0) + \mathsf{i}(0) + \mathsf{r}(0) = \varepsilon$, and $\mathsf{r}_0 = \vartheta \, \varepsilon$ for some given $\vartheta \in [0, 1)$, so that $\mathsf{s}_k^0 = \mathsf{s}_k(0) = p_k \, (1 - \varepsilon)$. The equation for the equilibrium (44) can be rewritten as

$$(1 - \varepsilon) \left((1 - p) \, e^{-\frac{\beta_1 \, (r - \vartheta \, \varepsilon)}{\gamma}} + p \, e^{-\frac{\lambda(\alpha \, (r - \vartheta \, \varepsilon))}{\gamma}} \right) + \mathsf{r} = 1 \tag{46}$$

and the Taylor expansion

$$(1 - \varepsilon) \left(1 - \mathcal{R}_0^{(2)}(q_1, q_2, p) \, \big(1 + o(\varepsilon)\big) \, \mathsf{r} \right) + \mathsf{r} = 1 + o(\mathsf{r}) \tag{47}$$

for $\mathsf{r} > 0$, small, shows that there is a solution with $\mathsf{r} > 0$, small, as $\varepsilon \to 0_+$ if and only if $\mathcal{R}_0^{(2)}(q_1, q_2, p) < 1$, thus giving a solution of order ε, which turns out to be the unique solution. Otherwise, the only positive solution of (46) corresponds to some $\mathsf{r} >> \varepsilon$ and we have exactly the same phase transition as in Section 1.3, with $\mathcal{R}_0^{(2)}(q_1, q_2, p)$ playing the role of an order parameter. When the disease spreads, we find that $\mathsf{r}^\star$ is of the order of the solution of (46) with $\varepsilon = 0$, that is, of the solution $\mathsf{r} = r$ of

$$(1 - p) \, e^{-\frac{\beta_1 \, r}{\gamma}} + p \, e^{-\frac{\beta_2 \, r}{\gamma}} + r = 1 \,. \tag{48}$$

By convexity, we know that

$$(1 - p) \, e^{-\frac{\beta \, r}{\gamma \, q_1}} + p \, e^{-\frac{\beta \, r}{\gamma \, q_2}} \geq e^{-\mathcal{R}_0^{(2)}(q_1, q_2, p) \, r} \,. \tag{49}$$

If we choose q such that $\mathcal{R}_0^{(1)}(q) = \mathcal{R}_0^{(2)}(q_1, q_2, p)$, which means $\frac{1}{q} = \frac{1-p}{q_1} + \frac{p}{q_2}$, then it is clear that the solution of

$$(1 - p) \, e^{-\frac{\beta \, r}{\gamma \, q}} + r = 1 \tag{50}$$

is larger than $\mathsf{r}^\star$. In other words, the *epidemic size* is reduced if we replace (25)-(28) by (36) with q, q_1 and q_2 as above.

3. General heterogeneous distribution

In this section we extend the previous framework to the situation of an arbitrary number of different population classes. Each class, or group, contains the individuals that share a given value of the transmission rate β/q. More precisely we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathfrak{S}(\omega, t)$, $\mathfrak{B}(\omega)$ random variables on this space designating respectively the state of a random individual $\omega \in \Omega$ and its β/q parameter. We denote by $\mathbb{E}[\cdot]$ the average operator.

For instance the case of Section 2 corresponds to the situation when $\mathfrak{B}$ has only two values, $\beta_1 = \beta/q_1$ and $\beta_2 = \beta/q_2$, and $\mathsf{s}_k(t) = \mathbb{P}[\mathfrak{S}(t) = \text{``Susceptible''}, \mathfrak{B} = \beta/q_k]$, with $k = 1, 2$. In order to keep notation simple, we will suppose in the following that the conditional law $\mathbb{P}[\mathfrak{B}|\mathfrak{S}(0) = \text{``Susceptible''}]$ of $\mathfrak{B}$ relative to being susceptible (at $t = 0$) only takes a finite number of values $\beta_1, ..., \beta_K$, i.e., is of the form $\sum_{k=1}^{K} p_k \, \delta_{\beta_k}$, with $p_k = \mathbb{P}[\mathfrak{B} = \beta_k | \mathfrak{S}(0) = \text{``Susceptible''}]$, but all results given below extend to the general case. Note that $\sum_{k=1}^{K} p_k = 1$. We adopt the notation $\mathfrak{R} = \mathfrak{B}/\gamma$, $R_k = \beta_k/\gamma$ for any $k = 1, ..., K$ and define

$$\mathsf{s}_k(t) = \mathbb{P}\left[\mathfrak{S}(t) = \text{``Susceptible''}, \mathfrak{B} = \beta_k\right], \qquad \mathsf{e}(t) = \mathbb{P}\left[\mathfrak{S}(t) = \text{``Exposed''}\right],$$
$$\mathsf{i}(t) = \mathbb{P}\left[\mathfrak{S}(t) = \text{``Infected''}\right], \qquad\qquad \mathsf{r}(t) = \mathbb{P}\left[\mathfrak{S}(t) = \text{``Recovered''}\right], \tag{51}$$

and

$$\mathsf{s}(t) = \mathbb{P}\left[\mathfrak{S}(t) = \text{``Susceptible''}\right] = \sum_{k=1}^{K} \mathsf{s}_k(t). \tag{52}$$

With these notations $p_k = \mathsf{s}_k(0)/\mathsf{s}(0)$. The evolution of $(\mathsf{s}_k, \mathsf{e}, \mathsf{i}, \mathsf{r})$ is governed by the system of equations[2]:

$$\frac{d\mathsf{s}_k}{dt} = -\beta_k \, \mathsf{s}_k \, \mathsf{i}, \quad \frac{d\mathsf{e}}{dt} = \left(\sum_{k=1}^{K} \beta_k \, \mathsf{s}_k\right) \mathsf{i} - \alpha \, \mathsf{e}, \quad \frac{d\mathsf{i}}{dt} = \alpha \, \mathsf{e} - \gamma \, \mathsf{i}, \quad \frac{d\mathsf{r}}{dt} = \gamma \, \mathsf{i}. \tag{53}$$

Let us start by a simple observation.

Lemma 1. *The solution* $(\mathsf{s}_k, \mathsf{e}, \mathsf{i}, \mathsf{r})$ *of* (53) *satisfies:*

$$\frac{d\mathsf{s}}{dt} = -\bar{\beta} \, a(t) \, \mathsf{s}(t) \, \mathsf{i}(t), \quad \frac{d\mathsf{e}}{dt} = \bar{\beta} \, a(t) \, \mathsf{s}(t) \, \mathsf{i}(t) - \alpha \, \mathsf{e}(t), \quad \frac{d\mathsf{i}}{dt} = \alpha \, \mathsf{i}(t) - \gamma \, \mathsf{i}(t), \quad \frac{d\mathsf{r}}{dt} = \gamma \, \mathsf{i}(t), \tag{54}$$

where $\bar{\beta} = \sum_k p_k \, \beta_k = \mathbb{E}[\mathfrak{B} \,|\, \mathfrak{S}(0) = \text{``Susceptible''}]$ *and* $a(t)$ *is a positive nonincreasing function of* t *with* $a(0) = 1$.

Proof. System (54) is satisfied with

$$a(t) = \frac{\sum_k \beta_k \, \mathsf{s}_k(t)}{\bar{\beta} \, \mathsf{s}(t)}. \tag{55}$$

By definition of $\bar{\beta}$, we know that $a(0) = 1$. Using the equation for $d\mathsf{s}_k/dt$, we can compute the derivative of a and conclude that

$$\frac{da}{dt} = -\frac{\mathsf{i}(t)}{\bar{\beta} \, \mathsf{s}(t)^2} \left(\sum_k \mathsf{s}_k(t) \sum_k \beta_k^2 \, \mathsf{s}_k(t) - \left(\sum_k \beta_k \, \mathsf{s}_k(t)\right)^2\right) \leq 0 \tag{56}$$

by the Cauchy-Schwarz inequality. $\qquad\square$

Theorem 2. *Suppose that* $\mathsf{r}(0) = 0$. *Then* (53) *possesses the following properties:*

  (1) *The basic reproduction ratio is the average of the reproduction ratios, i.e.,* $\overline{\mathfrak{R}} = \bar{\beta}/\gamma$. *In probabilistic notation,* $\overline{\mathfrak{R}} = \mathbb{E}[\mathfrak{R} \,|\, \mathfrak{S}(0) = \text{``Susceptible''}]$ [3].

---

[2] The mathematically rigorous formulation of the equations involves defining a continuous time Markov chain for any individual and the associated infection and recovering probabilities, see [29] for details.

[3] Here we will only consider the value of $\overline{\mathfrak{R}}$ for the event $\{\mathfrak{S}(0) = \text{``Susceptible''}\}$. Although the notation designates a random variable, we will consider it, when there is no ambiguity, as a real number.

(2) *If $\overline{\mathfrak{R}} > 1$, the peak (defined as the maximum value attained by $\mathsf{e} + \mathsf{i}$) is smaller than the peak obtained with a Dirac mass distribution (i.e., having only one group) with the same basic reproduction ratio.*

(3) *If $\overline{\mathfrak{R}} > 1$, the total epidemic size ζ is the unique solution of*

$$1 - \zeta = \mathsf{s}(0) \, \mathbb{E}\left[e^{-\mathfrak{R}\,\zeta} \,\middle|\, \mathfrak{S}(0) = \text{``Susceptible''} \right]. \tag{57}$$

For any distribution, ζ is smaller than the total epidemic size of a Dirac mass distribution having the same average.

By imposing the condition $\mathsf{r}(0) = 0$, what we have in mind as initial datum is a perturbation of the DFE solution, or a solution with initial values for which the *infectious compartments* are non-empty at $t = 0$, *i.e.*, $\mathsf{e}(0) + \mathsf{i}(0) > 0$, eventually small (but this is not even mandatory). As we shall see, (57) follows by convexity as in Section 2.3.

Proof. Property (1) is obtained by linearization using the next generation method. The proof is the same as for (38) when $K = 2$.

Next we prove Property (2). The peak is the value of $(\mathsf{e} + \mathsf{i})$ at some time t_p such that $(\mathsf{e} + \mathsf{i})'(t_p) = 0$, or, using Lemma 1, for the unique $t = t_p$ such that $\bar{\beta}\, a(t_p)\, \mathsf{s}(t_p) = \gamma$. In particular, since $a(t) \leq 1$, we know that $\mathsf{s}(t_p) \geq 1/\overline{\mathfrak{R}}$, which is precisely the value of s at the peak for the homogeneous case $a = 1$. The value of the peak can be written as:

$$(\mathsf{e} + \mathsf{i})(t_p) = (\mathsf{e} + \mathsf{i})(0) + \int_0^{t_p} \left(\bar{\beta}\, a(t)\, \mathsf{s}(t)\, \mathsf{i}(t) - \gamma\, \mathsf{i}(t) \right) dt = (\mathsf{e} + \mathsf{i})(0) + \int_0^{t_p} \left(\frac{(\overline{\mathfrak{R}})^{-1}}{a(t)\, \mathsf{s}(t)} - 1 \right) \frac{d\mathsf{s}}{dt}(t)\, dt$$

$$\leq (\mathsf{e} + \mathsf{i})(0) + \int_0^{t_p} (\overline{\mathfrak{R}})^{-1} \frac{d}{dt} \log \mathsf{s}(t)\, dt + \mathsf{s}(0) - \mathsf{s}(t_p) = (\mathsf{e} + \mathsf{i})(0) + h(\mathsf{s}(t_p)) - h(\mathsf{s}(0)), \quad (58)$$

where $h(x) = (\overline{\mathfrak{R}})^{-1} \log x - x$. Since the function $x \mapsto h(x)$ is monotone decreasing on $[1/\overline{\mathfrak{R}}, 1]$, we obtain

$$(\mathsf{e} + \mathsf{i})(t_p) \leq (\mathsf{e} + \mathsf{i})(0) + h(1/\overline{\mathfrak{R}}) - h(\mathsf{s}(0)) = 1 - \frac{1}{\overline{\mathfrak{R}}} \left[1 + \log\left(\overline{\mathfrak{R}}\, \mathsf{s}(0) \right) \right]. \tag{59}$$

Property (3) is obtained as follows. Denote by ζ the epidemic size, which satisfies the two identities: $\zeta = \mathsf{r}(\infty) = \mathsf{r}(\infty) - \mathsf{r}(0) = \mathsf{s}(0) + \mathsf{e}(0) + \mathsf{i}(0) - \sum_k \mathsf{s}_k(\infty)$. Thus $\zeta = \int_0^\infty \mathsf{r}'(t)\, dt = \gamma \int_0^\infty \mathsf{i}(t)\, dt$. For any $k \leq K$: $s_k(\infty) = s_k(0) \exp\left(-\int_0^\infty \beta_k\, \mathsf{i}(t)\, dt \right) = \mathsf{s}(0)\, p_k\, e^{-\zeta\, \beta_k/\gamma}$. Combining the above relations we obtain (57). In particular if the distribution is reduced to a Dirac mass centered at the mean value $\overline{\mathfrak{R}}$, the epidemic size, denoted by ζ_0, satisfies: $\zeta_0 = 1 - \mathsf{s}(0)\, e^{-\overline{\mathfrak{R}}\,\zeta_0}$ as expected. We use now the Jensen inequality for the convex function $x \mapsto e^{-x}$:

$$1 - \zeta = \mathsf{s}(0)\, \mathbb{E}[e^{-R_\omega\,\zeta} \mid \mathfrak{S}(0) = \text{``Susceptible''}] \geq \mathsf{s}(0)\, e^{-\overline{\mathfrak{R}}\,\zeta} = \mathsf{s}(0) \left(e^{-\overline{\mathfrak{R}}\,\zeta_0} \right)^{\zeta/\zeta_0} = \mathsf{s}(0) \left(\frac{1 - \zeta_0}{\mathsf{s}(0)} \right)^{\zeta/\zeta_0}. \tag{60}$$

We obtained thus $\left(\frac{1-\zeta}{\mathsf{s}(0)} \right)^{1/\zeta} \geq \left(\frac{1-\zeta_0}{\mathsf{s}(0)} \right)^{1/\zeta_0}$, which implies, after recalling that $\left(\frac{1-x}{\mathsf{s}(0)} \right)^{1/x}$ is a decreasing function on $]0, \mathsf{s}(0)[$ that $\zeta \leq \zeta_0$, our conclusion. $\qquad \square$

Note that, by assuming that $(\mathsf{e} + \mathsf{i})(0)$ is small, it is clear that the peak can be made small if $\overline{\mathfrak{R}} - 1 > 0$ is small.

4. Numerical results

4.1. **Choice of a set of parameters**

It is not the purpose of this paper to discuss the values of the parameters in a SEIR model of the pandemic of Covid-19 and we shall simply choose values that can be found in the literature, for a purely illustrative purpose. However, for the interested reader, we list our sources and some entries in the rapidly growing literature on the topic.

Concerning the initial data for the SEIR model, we shall assume that the size of the French population is $N = 67.8 \times 10^6$. The following values correspond to the situation for the Covid-19 in France on March 15, 2020, according to [5] and data released on a daily basis by SPF at [43]. We shall take as initial data the values

$$S(0) = N - \mathcal{R}_0\, R(0)\,, \quad E(0) = 5970\,, \quad I(0) = 1278\,, \quad R(0) = \frac{E(0) + I(0)}{\mathcal{R}_0 - 1} \approx 5450$$

with $\mathcal{R}_0 = 2.33$. Here we provide the formulae used in [5] to infer the numbers, which are based on an asymptotic analysis of the SEIR model during the initial phase of the epidemics (without lockdown). Such values are important for controlled epidemics but play essentially no role if the disease is spreading in the population, which is the case under investigation. Note that it is very likely that the numbers will be revised in the future, to account for non well documented cases at the time this study was done. In any case, for numerical computations, we take the following initial data, as fractions of the total population

$$\mathsf{s}(0) = 0.99981\,, \quad \mathsf{e}(0) = 8.81 \times 10^{-5}\,, \quad \mathsf{i}(0) = 1.88 \times 10^{-5}\,, \quad \mathsf{r}(0) = 8.04 \times 10^{-5}\,, \tag{61}$$

in all our examples.

Now comes the issue of estimating the parameters β, γ and α of the SEIR model, and the factor of reduction of social interactions q, at least as an average, in a socially homogeneous model and also for the majority of the population in the model with two groups. The methodology is out of the scope of this paper and we will not comment it. We refer to [35] for a recent discussion. However, in order to fix the order of magnitude and give an idea of the uncertainties, let us review some numbers which recently appeared. An estimate based on the statistics of the known cases of Covid-19 at the beginning of the outbreak has been proposed in [5], with the following values

$$\beta = 2.33\,, \quad \alpha = 0.25\,, \quad \gamma = 1\,, \quad \mathcal{R}_0 = 2.33 \tag{62}$$

before lockdown. By fitting the values for the first two weeks of lockdown, N. Bacaër came up with $q \approx 1.7$ for an estimate of the factor of reduction of social interactions. These numbers are the ones used for our choice (61). We learn from [10] that the parameters in the SEIR model can be estimated by: $\beta = 2$, $\alpha = 1/3.7 \approx 0.27$, $\gamma = 1/1.5 \approx 0.67$ and $\mathcal{R}_0 = 3$ at regional level (Ile-de-France) and it was suggested that $q \approx 2.94$ so that $\mathcal{R}_0/q = 0.68$. This is consistent with the estimate 0.67 of [42].

Before going further, let us list some limitations of our model and features which have to be taken into account in more realistic models. In [35], it is argued that *Unreported* cases (U) should be taken into account in a S(E)IRU model for correctly accounting the Covid-19 outbreaks (also see [32]). There is certainly an important point here, although we did not introduce it for sake of simplicity. Note that such models produce a very high $\mathcal{R}_0$ (at least on the basis of the data used for fitting the curves) ranging from 4.45 to 4.49. In [41], the basic reproduction ratio is adjusted to $\mathcal{R}_0 = 2.8$. In [42], the authors find that the basic reproduction ratio, which was $\mathcal{R}_0 \approx 3.41$ before lockdown, has been reduced to 0.52, which corresponds to $q \approx 6.6$. This is consistent with the factor $q = 7$ found in [40]. Of course important factors like risk status or age (see for instance [16,18]) have been put in evidence, but they are not easy to study from the point of view of transmission rates. In our aproach, it is clear that we take oversimplifying assumptions, but this is probably the price to pay to prove results and put them numerically in evidence.

Recent papers have emphasized that the heterogeneity of the transmissions rates is important in the current pandemic and that it lowers herd immunity thresholds in various modeling frameworks. During the revision of this paper, we became aware of [14, 16, 17]. Superspreaders and superspreading events are known to play an essential role in the propagation of Covid-19 according to [1,3,19,20,27,38,39]. Incorporating these considerations in compartmental models will anyway require further studies.

So far it remains difficult to choose a set of parameters, although one can hope that a better understanding of the dynamics of the pandemic will emerge out of the various studies that are currently done. What we learn after the end of the lockdown is still to be studied. Now comes a very empirical observation, in the framework of SEIR models, with parameters that are supposed to be constant in time, a crude assumption that is definitely not valid after the end of the lockdown. This empirical observation is the starting point for this paper. Situations with a basic reproduction ratio $\mathcal{R}_0 < 1$ exhibit an exponential decay of the number of cases while the data of [43] (at the beginning of the lockdown) were emblematic of a situation with $\mathcal{R}_0 > 1$. Several phases for a population under lockdown are illustrated in [31] by the computation of the *effective reproduction ratio* $\mathcal{R}_t$ (which is of the order of $\mathcal{R}_0$ as only a small proportion of the total population was concerned) in Wuhan, based on the reported daily Covid-19 infections: when there was no intervention (before January 23, 2020): $\mathcal{R}_t \approx 3.88$; under lockdown with traffic ban and many confined at home (between January 23-February 1st, 2020): $\mathcal{R}_t \approx 1.25$; using *centralized confinement* (after February 1st, 2020): $\mathcal{R}_t \approx 0.32$. Also see [33] for a detailed analysis of the data in Wuhan based on a SIR type model. The curves in France and southern Europe during the first month of lockdown look more like to what happened in Wuhan by the end of january, with an $\mathcal{R}_0 > 1$, than to what happened later. This is what we intend to explain in France, at least partially, by social heterogeneity. For the sake of simplicity, we shall retain the values of (61), and one of the reasons is that $q \approx 1.7 < \mathcal{R}_0$ is compatible with the above remarks, but our models of Sections 2 and 3 show that there might be some subtleties when heterogeneities are taken into account.

4.2. SEIR model with a single group

We start by considering the standard SEIR model (6)-(9) or its variant (25)-(28), where q is the factor of reduction of social interactions. The basic reproduction ratio in the second system is $\mathcal{R}_0^{(1)}(q) = \mathcal{R}_0/q$, which allows us to reduce both problems to (6)-(9) with various values of the basic reproduction ratio depending on the q factor. Here we choose the initial data according to (61) and the set of parameters (62). See Fig. 1 for the epidemic curves, and Fig. 2 for the epidemic size.
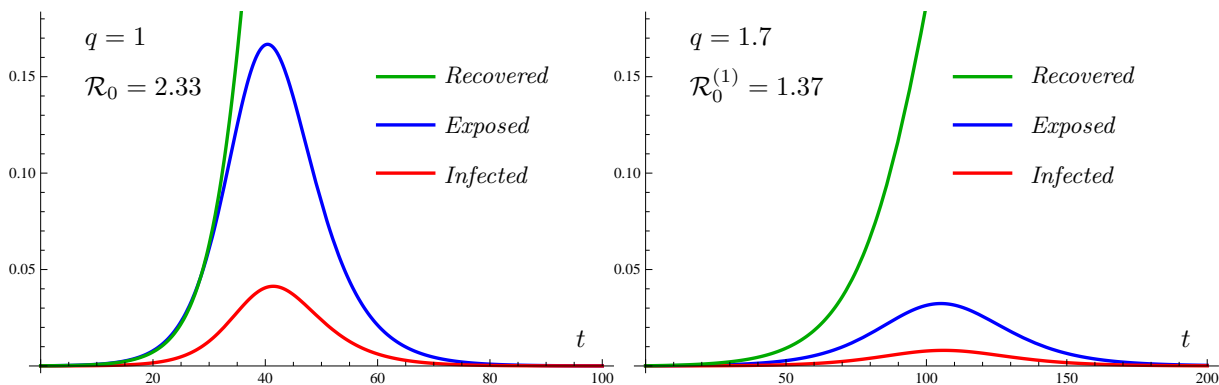

FIGURE 1. The peak of the outbreak in the SEIR model. The time t is counted in days. The vertical axis represents the fraction of the population. The basic reproduction ratio is either $\mathcal{R}_0 = 2.33$ (left) or 1.37 (right) corresponding to a reduction of social interactions by a factor $q = 1.7$ as in [5]. This illustrates the *flattening of the curves*.
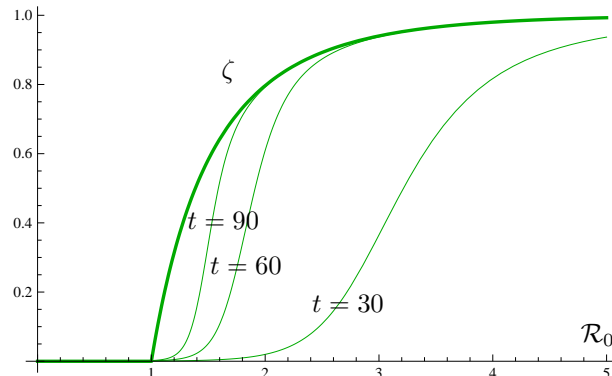
FIGURE 2. The epidemic size ζ as a function of the basic reproduction ratio $\mathcal{R}_0$ in the simplest SEIR model corresponding to (6)-(9) exhibits a clear phase transition at $\mathcal{R}_0 = 1$ (thick curve). The other curves represent $\mathsf{r}(t)$ taken for $t = 30$, 60, 90, for which equilibrium is not yet achieved. In practice, varying $\mathcal{R}_0$ is achieved by acting on the q factor in (25)-(28).

4.3. SEIR model with two groups having different factors of reduction of social interactions

One of the most disturbing results in [5] is that fitting the data of the cumulated number of cases by model (25)-(28) shows that $q \approx 1.7$, which is far below $\mathcal{R}_0 = 2.33$ and suggest that lockdown is inefficient for controlling the outbreak. However, we have seen that there is another possible model, as it is illustrated by Fig. 3, with two groups. From the figure, it is clearly impossible to distinguish between the two scenarios at the early stage of the outbreak.
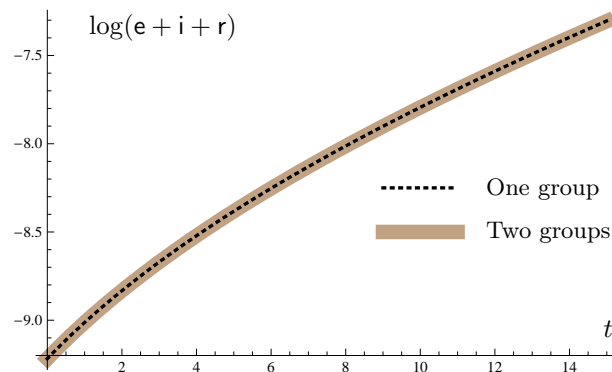
FIGURE 3. Plot of $t \mapsto \mathsf{e}(t) + \mathsf{i}(t) + \mathsf{r}(t)$ for a solution of (25)-(28) with $q = 1.7$ (black dotted line) *versus* a plot of $t \mapsto \mathsf{e}(t) + \mathsf{i}(t) + \mathsf{r}(t)$ for a solution of (36) in logarithmic scale with $p = 0.02$, $q_1 = 2.35$, and $q_2 = 0.117$ (brown). In both cases the basic reproduction ratio is 1.37.

With two groups arises the question of choosing the initial data. As a crude and simplifying assumption, we make the choice to consider that before lockdown there was a single population and that the two categories of the population had the same exposure to the disease. We can then use (61) as initial condition and take $\mathsf{s}_1(0) = (1 - p)\,\mathsf{s}(0)$ and $\mathsf{s}_2(0) = p\,\mathsf{s}(0)$. With this choice, we recover the results of (25)-(28) if $q_1 = q_2$ and $\mathsf{s} = \mathsf{s}_1 + \mathsf{s}_2$. In order to fix ideas, we also make the arbitrary choice of choosing $q_1 = 2.35$ so that the epidemic disease would extinguish by itself after affecting 0.81% of the population if $q_2 = q_1$. We can illustrate the role of the two parameters p and q_2 by showing that they completely change the picture and bring us back to a regime

with an epidemic size corresponding to some inefficient lockdown, however with lower epidemic peak and size: see Fig. 4, and Fig. 5 for the epidemic size.
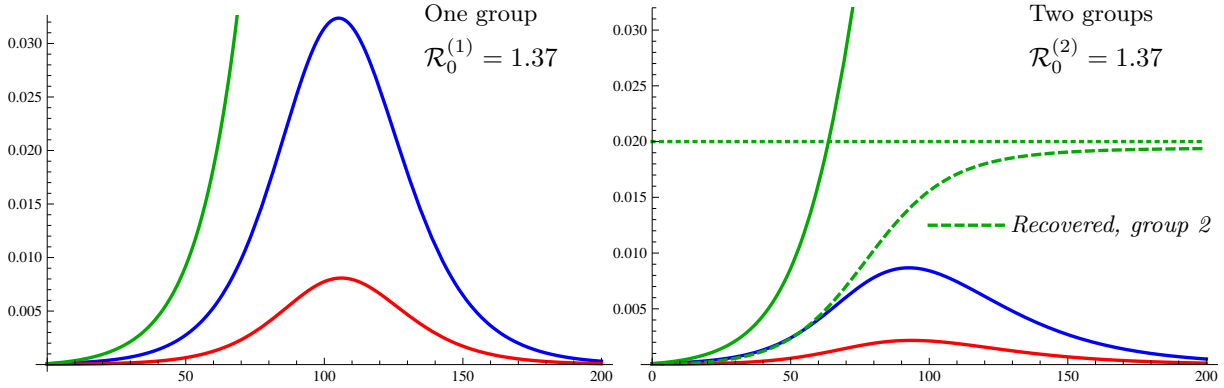

FIGURE 4. Model with a single group (left) and $q = 1.7$ corresponding to a basic reproduction ratio of 1.37, and two groups (right) with $q_1 = 2.35$, $q_2 = 0.117$, and $p = 0.02$ as in Fig. 3, with same basic reproduction ratio. Note that the figure on the left is the same as in Fig. 1 (right), on a different scale. The straight dotted line is the level p. In the case with two groups, note that almost all individuals of the second group get infected during the propagation of the disease.
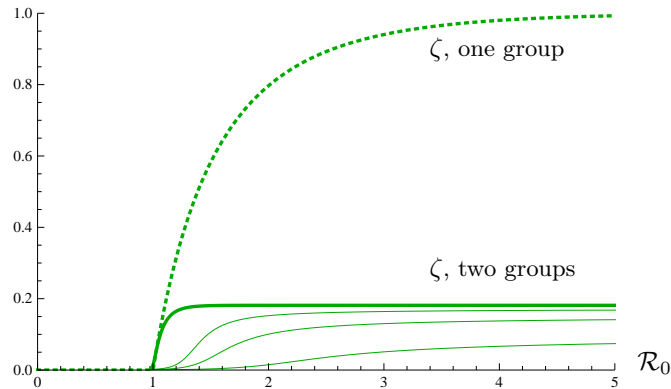
FIGURE 5. The epidemic size in a population with one group (dotted curve, same as in Fig. 2) or two groups (plain, thick curve) with $q_1 = 2.4 > \mathcal{R}_0 = 2.33$ and $p = 0.02$ as a function of the basic reproduction ratio (obtained by varying q in the first case and q_2 in the second case). The other lines correspond to simulations of $r(t)$ for $t = 30, 60, 90$ in the model with two groups. With two groups, we recover a phase transition as in Fig. 2, corresponding to $q_2 \approx 0.96$.

5. DISCUSSION

In the SEIR model with a basic reproduction ratio $\mathcal{R}_0 > 1$, the stable equilibrium does not depends much on the initial data: it is almost entirely determined by $\mathcal{R}_0$. The goal of lockdown is to decrease $\mathcal{R}_0$ by dividing it by a reduction factor for social interactions, q. With $\mathcal{R}_0/q < 1$, the disease is under control. With $\mathcal{R}_0/q > 1$, but close to 1, the epidemic disease spreads, the final state depends little on the initial data, but the curve

is flattened: qualitatively, this is the scenario that one can observe in the Covid-19 outbreak in France under lockdown.

Measuring the q factor is difficult. The aim of this article is to show that the crucial information cannot be reduced to the knowledge of an average factor q: if the population is divided into two groups, with a group for which $q > \mathcal{R}_0$ (the majority) and another group (the minority) that keeps a small q factor, the disease may continue to spread. If the q factor of the majority is larger than $\mathcal{R}_0$ but close to $\mathcal{R}_0$, so that $1 - \mathcal{R}_0/q > 0$ is small, the impact of the minority becomes extremely important as it eventually triggers the nonlinearity. However, the equilibrium asymptotic state in a two-group model is not the same as when considering a single group with an averaged basic reproduction ratio. The dynamics of the outbreak, for instance the height of the epidemic peak, is also changed. A two-group model is of course extremely simplistic, but shows the importance of understanding the distribution of the q factors in a population.

Our observations are not limited to a population divided into two groups. In our model of heterogeneous social interactions, with a whole distribution of q factors, we have shown that an average of q is not relevant. While the basic reproduction ratio behaves as a plain average across the homogeneous q categories, neither the peak nor the total epidemic size do the same. In particular the presence of heterogeneity is beneficial for both the peak and total epidemic size. Or, put it otherwise, a model with only one group and fitting the observed data in the initial phase of the outbreak will be more pessimistic concerning the epidemic outcomes than a heterogeneous model; this is even more true after lockdown when social distancing measures have been enforced, the lockdown being by its nature a creator of heterogeneity. In terms of public health, this also underlines the importance of targeting prevention measures on individuals with a high level of social interactions.

References

[1] D. Adam, P. Wu, J. Wong, E. Lau, T. Tsang, S. Cauchemez, G. Leung, and B. Cowling, *Clustering and super-spreading potential of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong*, Research Square, Doi: 10.21203/rs.3.rs-29548/v1, (2020).

[2] F. R. Adler, *The effects of averaging on the basic reproduction ratio*, Mathematical Biosciences, 111 (1992), pp. 89–98.

[3] B. M. Althouse and et al., *Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2*. https://covid.idmod.org/.

[4] R. M. Anderson, B. Anderson, and R. M. May, *Infectious diseases of humans: dynamics and control*, Oxford university press, 1992.

[5] N. Bacaër, *Un modèle mathématique des débuts de l'épidémie de coronavirus en France*, Mathematical Modelling of Natural Phenomena, 15 (2020), p. 29.

[6] J. C. Blackwood and L. M. Childs, *An introduction to compartmental modeling for the budding infectious disease modeler*, Letters in Biomathematics, 5 (2018), pp. 195–221.

[7] B. Buonomo, A. d' Onofrio, and D. Lacitignola, *Global stability of an SIR epidemic model with information dependent vaccination*, Mathematical Biosciences, 216 (2008), pp. 9–16.

[8] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, *On the Lambert W function*, Advances in Computational Mathematics, 5 (1996), pp. 329–359.

[9] A. Danchin, T. W. P. Ng, and G. Turinici, *A new transmission route for the propagation of the SARS-CoV-2 coronavirus*, medRxiv Doi: 10.1101/2020.02.14.20022939, (2020).

[10] G. Di Domenico, Laura Pullano, C. E. Sabbatini, P.-Y. Boëlle, and V. Colizza, *Expected impact of lockdown in Île-de-France and possible exit strategies.* https://www.epicx-lab.com.

[11] O. Diekmann and J. A. P. Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, vol. 5, John Wiley & Sons, 2000.

[12] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, *On the definition and the computation of the basic reproduction ratio $\mathcal{R}_0$ in models for infectious diseases in heterogeneous populations*, J. Math. Biol., 28 (1990), pp. 365–382.

[13] O. DIEKMANN, J. A. P. HEESTERBEEK, AND M. G. ROBERTS, *The construction of next-generation matrices for compartmental epidemic models*, Journal of The Royal Society Interface, 7 (2009), pp. 873–885.

[14] A. ENDO, S. ABBOTT, A. J. KUCHARSKI, AND S. FUNK, *Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China*, Wellcome Open Research, 5 (2020), p. 67.

[15] Z. FENG, *Final and peak epidemic sizes for SEIR models with quarantine and isolation*, Mathematical Biosciences & Engineering, 4 (2007), p. 675.

[16] A. GERASIMOV, G. LEBEDEV, M. LEBEDEV, AND I. SEMENYCHEVA, *Reaching collective immunity for COVID-19: an estimate with a heterogeneous model based on the data for Italy*, medRxiv Doi: 10.1101/2020.05.24.20112045, (2020).

[17] M. G. M. GOMES, R. M. CORDER, J. G. KING, K. E. LANGWIG, C. SOUTO-MAIOR, J. CARNEIRO, G. GONCALVES, C. PENHA-GONCALVES, M. U. FERREIRA, AND R. AGUAS, *Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold*, medRxiv Doi: 10.1101/2020.04.27.20081893, (2020).

[18] Q. GRIETTE, P. MAGAL, AND O. SEYDI, *Unreported cases for age dependent COVID-19 outbreak in Japan*, medRxiv Doi: 10.1101/2020.05.07.20093807, (2020).

[19] G. GROSSMANN, M. BACKENKOEHLER, AND V. WOLF, *Importance of interaction structure and stochasticity for epidemic spreading: A COVID-19 case study*, medRxiv Doi: 10.1101/2020.05.05.20091736, (2020).

[20] L. HÉBERT-DUFRESNE, B. M. ALTHOUSE, S. V. SCARPINO, AND A. ALLARD, *Beyond R0: Heterogeneity in secondary infections and probabilistic epidemic forecasting*, medRxiv Doi: 10.1101/2020.02.10.20021725, (2020).

[21] H. W. HETHCOTE, *Modeling heterogeneous mixing in infectious disease dynamics*, in Models for infectious human diseases: their structure and relation to data, Cambridge: Cambridge University Press, 1996, pp. 215–238.

[22] H. W. HETHCOTE, *The mathematics of infectious diseases*, SIAM Rev., 42 (2000), pp. 599–653.

[23] H. W. HETHCOTE AND J. W. VAN ARK, *Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs*, Mathematical Biosciences, 84 (1987), pp. 85–118.

[24] J. M. HYMAN AND E. STANLEY, *Using mathematical models to understand the aids epidemic*, Mathematical Biosciences, 90 (1988), pp. 415–473.

[25] J. A. JACQUEZ, C. P. SIMON, J. KOOPMAN, L. SATTENSPIEL, AND T. PERRY, *Modeling and analyzing HIV transmission: the effect of contact patterns*, Mathematical Biosciences, 92 (1988), pp. 119–199.

[26] W. O. KERMACK AND A. G. MCKENDRICK, *A contribution to the mathematical theory of epidemics*, Proc. R. Soc. Lond., Ser. A, 115 (1927), pp. 700–721.

[27] A. J. KUCHARSKI, P. KLEPAC, A. CONLAN, S. M. KISSLER, M. TANG, H. FRY, J. GOG, AND J. EDMUNDS, *Effectiveness of isolation, testing, contact tracing and physical distancing on reducing transmission of SARS-CoV-2 in different settings*, medRxiv Doi: 10.1101/2020.04.23.20077024, (2020).

[28] M. LACHIANY AND Y. LOUZOUN, *Effects of distribution of infection rate on epidemic models*, Physical Review E, 94 (2016).

[29] L. LAGUZET AND G. TURINICI, *Individual Vaccination as Nash Equilibrium in a SIR Model with Application to the 2009–2010 Influenza A (H1N1) Epidemic in France*, Bulletin of Mathematical Biology, 77 (2015), pp. 1955–1984.

[30] L. LANDAU AND E. LIFSHITZ, *Statistical Physics: Volume 5*, no. vol. 5 in 3rd edition, revised and enlarged, Elsevier Science, 1980.

[31] X. LIN, *Analysis of 25,000 lab-confirmed Covid-19 cases in Wuhan: Epidemiological characteristics and non-pharmaceutical intervention effects.* Department of Biostatistics and Department of Statistics, Harvard University and Broad Institute, https://cdn1.sph.harvard.edu/wp-content/uploads/sites/21/2020/03/COVID-19-03-16-2020-Lin.pdf, 2020.

[32] Z. LIU, P. MAGAL, O. SEYDI, AND G. WEBB, *A covid-19 epidemic model with latency period*, Infectious Disease Modelling, 5 (2020), pp. 323–337.

[33] 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 (2020), pp. 3040–3051.

[34] ———, *Understanding unreported cases in the Covid-19 epidemic outbreak in Wuhan, China, and the importance of major public health interventions*, Biology, 9 (2020), p. 50.

[35] P. MAGAL AND G. WEBB, *Predicting the number of reported and unreported cases for the Covid-19 epidemic in South Korea, Italy, France and Germany*, medRxiv Doi: 10.1101/2020.03.21.20040154, (2020).

[36] M. MARTCHEVA, *An introduction to mathematical epidemiology*, vol. 61 of Texts in Applied Mathematics, Springer, New York, 2015.

[37] R. M. MAY AND R. M. ANDERSON, *Spatial heterogeneity and the design of immunization programs*, Mathematical Biosciences, 72 (1984), pp. 83–111.

[38] K. PREM, Y. LIU, T. W. RUSSELL, A. J. KUCHARSKI, R. M. EGGO, N. DAVIES, M. JIT, P. KLEPAC, S. FLASCHE, S. CLIFFORD, AND 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 (2020), pp. e261–e270.

[39] J. RIOU AND C. L. ALTHAUS, *Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020*, Eurosurveillance, 25 (2020).

[40] L. ROQUES, E. K. KLEIN, J. PAPAÏX, A. SAR, AND S. SOUBEYRAND, *Impact of lockdown on the epidemic dynamics of COVID-19 in France*, Frontiers in Medicine, 7 (2020).

[41] J. Roux, C. Massonnaud, and P. Crépey, *Covid-19: One-month impact of the french lockdown on the epidemic burden.* Publications de l'équipe REPERES (Recherche en Pharmaco-épidémiologie et recours aux soins, UPRES EA-7449). https://www.ea-reperes.com/wp-content/uploads/2020/04/ImpactConfinement-EHESP-20200322v1.pdf.

[42] H. Salje, C. Tran Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, and et al., *Estimating the burden of SARS-CoV-2 in France*, Science, (2020), pp. eabc3517, Doi: 10.1126/science.abc3517.

[43] Santé Publique France, *French national public health agency.* https://dashboard.covid19.data.gouv.fr, 2020.

[44] H. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, The International Series of Monographs on Physics Oxford University Press Inc., Oxford and New York, 1971.

[45] P. van den Driessche and J. Watmough, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Mathematical Biosciences, 180 (2002), pp. 29–48.