Research

Behavioural response to heterogeneous severity of COVID-19 explains temporal variation of cases among different age groups

Benjamin Steinegger[1], Lluís Arola-Fernández[1], Clara Granell[1], Jesús Gómez-Gardeñes[2,3] and Alex Arenas[1]

[1]Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona 43007, Spain
[2]Department of Condensed Matter Physics, University of Zaragoza, Zaragoza E-50009, Spain
[3]GOTHAM Lab—BIFI, University of Zaragoza, Zaragoza E-50018, Spain

JGG, 0000-0001-5204-1937; AA, 0000-0003-0937-0334

Together with seasonal effects inducing outdoor or indoor activities, the gradual easing of prophylaxis caused second and third waves of SARS-CoV-2 to emerge in various countries. Interestingly, data indicate that the proportion of infections belonging to the elderly is particularly small during periods of low prevalence and continuously increases as case numbers increase. This effect leads to additional stress on the health care system during periods of high prevalence. Furthermore, infections peak with a slight delay of about a week among the elderly compared to the younger age groups. Here, we provide a mechanistic explanation for this phenomenology attributable to a heterogeneous prophylaxis induced by the age-specific severity of the disease. We model the dynamical adoption of prophylaxis through a two-strategy game and couple it with an SIR spreading model. Our results also indicate that the mixing of contacts among the age groups strongly determines the delay between their peaks in prevalence and the temporal variation in the distribution of cases.

This article is part of the theme issue 'Data science approaches to infectious disease surveillance'.

1. Introduction

The study of the interplay between human behaviour and the spreading of epidemics has received great attention since the early 2000s [1–3] and is fundamental in many modelling approaches to the spread of SARS-CoV-2 [4]. Particular attention has been paid to how individuals decide to adopt protective measures. Pioneer results from the study of vaccine uptake have shown that the voluntary adoption may cause subsequent outbreaks to occur [5]. Similar findings have been obtained by coupling the epidemic dynamics with the spread of awareness, awareness kernels, or payoff-based considerations regarding the adoption of prophylactic measures [6–10].

Consequently, already from the first epidemic waves caused by the entry of SARS-CoV-2 in different regions of the world, it was clear that changes in behavioural patterns would determine the course of subsequent waves. Actually, early analysis that incorporated the effect of human behaviour and tailored their models for the characteristics of SARS-CoV-2 showed an equivalent phenomenology [11–15]. Finally, the eventuality of a second and subsequent waves fulfilled itself in a series of countries. It was empirically confirmed that, together with seasonality, human behaviour was a substantial factor that caused the emergence of second waves. In particular, individuals continuously relaxed their prophylactic behaviour after the first wave, as can be observed from increased mobility levels [16].

While the emergence of subsequent waves is well established through modelling, we observe additional phenomena that are yet to be captured from a theoretical perspective. Figure 1 shows the evolution of the distribution of reported cases over age groups for various countries and regions. Interestingly, we observe a substantial increase of cases attributed to older age groups with the arrival of a subsequent wave. The fraction of cases attributed to the older age groups then peaks substantially after the peak in the aggregated number of reported cases. An equivalent temporal variation can be observed when considering the fraction of fatalities attributed to nursing homes [17]. Furthermore, figure 2 shows that in various cases also the absolute number of infections in the older age groups peaks with a delay of one week or more with respect to the younger age groups. We observe a clear trend that the prevalence in the elder age groups peaks later. In particular, with the exception of Catalunya, the age group 80+ exhibits a significant delay with respect to the younger age groups across all countries and regions. The same delay was previously observed for fatalities instead of case numbers [25].

In this study, we provide a simple mathematical model which exhibits all the aforementioned phenomenology by considering different age groups, i.e. groups at different risk of a severe course of the disease. We model the individual adoption of prophylactic measures, such as social distancing, increased hygiene or the wearing of face masks, using a game-theoretic approach. Importantly, we explicitly include the heterogeneity regarding the severity of a SARS-CoV-2 infection and how this heterogeneity induces a more cautious behaviour in the risk group. The behavioural model is then coupled to the evolution of a spreading dynamics. This framework allows us to qualitatively understand and mechanistically explain the emergence of the temporal variation over age groups as well as the observed delayed peak in prevalence.

2. Model

For the sake of simplicity, we describe the epidemic spreading by a standard SIR model, i.e. considering only susceptible, infected and recovered epidemic compartments. Accordingly, we disregard the possibility of reinfection. In addition, since we do not consider how contact tracing and isolation may prevent the spread of SARS-CoV-2, it is not necessary to differentiate among pre-symptomatic, asymptomatic and symptomatic infectiousness.
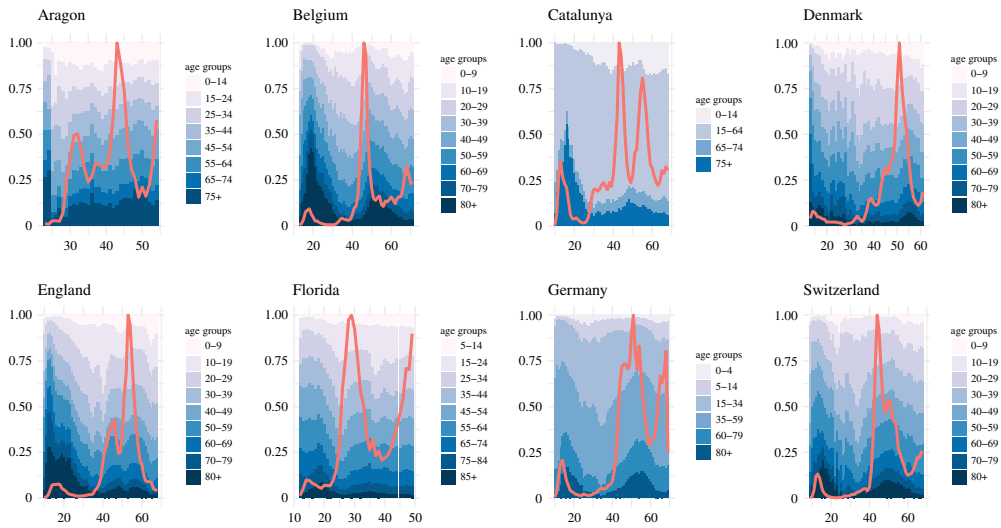
Figure 1. Weekly aggregated distribution of reported cases among the age groups for a variety of regions and countries [17–24]. The *x*-axis indicates time by numerating the weeks of 2020 and continuing after 2021. The red line shows the evolution of the average weekly incidence, which serves as a proxy for prevalence, normalized to the peak incidence. (Online version in colour.)
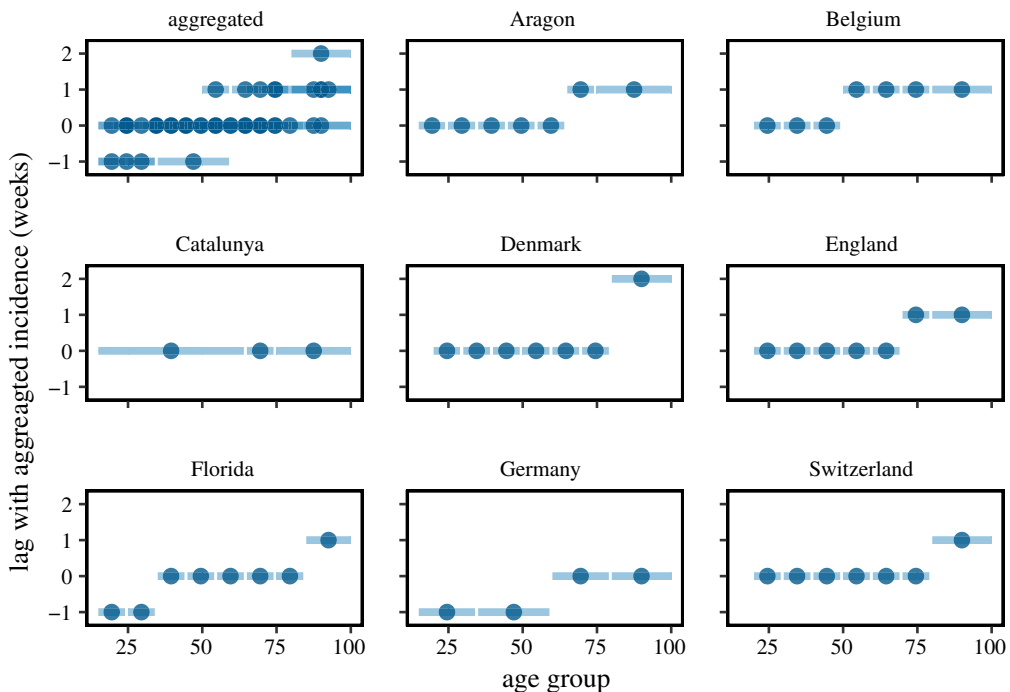

Figure 2. Temporal difference between the peak in the absolute number of infections of a particular age group and the aggregation over all age groups. Lines show the range of the age groups, while dots indicate the mean between the limits. For all countries and regions, we show the second peak. An exception is Germany, whose second peak coincides with the Christmas holidays which made reporting unreliable, wherefore we show the delay for the first wave. (Online version in colour.)

Behaviourally, the response to SARS-CoV-2 has been, and continues to be, varied. At the individual level, the adoption of prophylactic measures largely determines the course of the epidemic. Also, central actors, such as public health authorities [26–28], and external factors, for

example seasonality [29,30], influence the spread of the epidemic. In addition, all these levels depend on each other. Modelling the interplay among central actors, seasonality and individual actions from a behavioural perspective is very challenging. Nevertheless, here we show that focusing solely on the behavioural response at the individual level proves to be sufficient to explain the relevant phenomenology observed from data.

The response at the individual level mainly consists of the adoption of prophylactic measures such as social distancing, reduction of contacts, adoption of hygiene, attachment to ventilation or the use of face masks. From a mathematical perspective, in a continuous time mean field model, all these measures simply reduce the transmission rate either through a reduction of contacts or the transmission probability and are thus essentially equivalent [31]. This equivalence enables us to subsume the different measures into one category: the adoption of prophylactic measures. In our model, we assume that the adoption of prophylactic measures is driven by the risk of infection and the potentially severe consequences of such an infection, personally or in one's private sphere.

In this sense, we describe the decision to adopt prophylactic measures as a two-strategy game, reducing, for simplicity, the behavioural response to a binary choice: individuals either adopt prophylactic measures (P) or not (NP). The associated payoffs of each strategy express the trade-off between the infection cost, T, modulated by the infection risk and the adoption cost of prophylactic measures, c. Due to the existence of individuals at higher risk of a severe course of COVID-19, we divide the population into G different groups with different infection costs T_i. For the sake of generality, we will also consider all other behavioural parameters specific to each group. Accordingly, we define the payoffs of group i as follows:

$$P_i^{NP} = -T_i \sum_{j=1}^{G} f_j(I_j^{NP} + I_j^P) \tag{2.1}$$

and

$$P_i^P = -c_i - T_i(1 - \gamma_i) \sum_{j=1}^{G} f_j(I_j^{NP} + I_j^P). \tag{2.2}$$

The sum expresses the global infection risk, i.e. the aggregated fraction of infected individuals, with f_i representing the fraction of the population belonging to group i. The variables I_i^P and I_i^{NP} refer to the fraction of infectious individuals in group i that adopt and do not adopt prophylactic measures, respectively. For protected individuals, we modulate the infection risk by the efficacy of the prophylactic measures, γ_i. Given the payoffs, individuals belonging to group i adopt the more beneficial strategy according to the transition probabilities

$$\Gamma_i^{P \to NP} = \Theta(P_i^{NP} - P_i^P) \frac{P_i^{NP} - P_i^P}{c_i + T_i} \tag{2.3}$$

and

$$\Gamma_i^{NP \to P} = \Theta(P_i^P - P_i^{NP}) \frac{P_i^P - P_i^{NP}}{c_i + T_i}. \tag{2.4}$$

The letter Θ represents the Heaviside function and guarantees that only the more successful strategy is adopted. The denominator, which corresponds to the maximal difference in payoff, serves to bound the probabilities. When the dynamics evolve, individuals update their strategy according to the adoption—selection—rate α_i with the probabilities $\Gamma_i^{P \to NP}$, $\Gamma_i^{NP \to P}$. Finally, by coupling the behavioural dynamics to the SIR spreading model, the system of differential equations can be written as follows:

$$\dot{S}_i^{NP} = -\beta_i S_i^{NP} \sum_{j=1}^{G} C_{ij}(I_j^{NP} + (1 - \gamma_j)I_j^P) + \alpha_i S_i^P \Gamma_i^{P \to NP} - \alpha_i S^{NP} \Gamma_i^{NP \to P} \tag{2.5}$$

$$\dot{S}_i^P = -\beta_i S_i^P (1 - \gamma_i) \sum_{j=1}^{G} C_{ij}(I_j^{NP} + (1 - \gamma_j)I_j^P) + \alpha_i S^{NP} \Gamma_i^{NP \to P} - \alpha_i S^P \Gamma_i^{P \to NP} \tag{2.6}$$
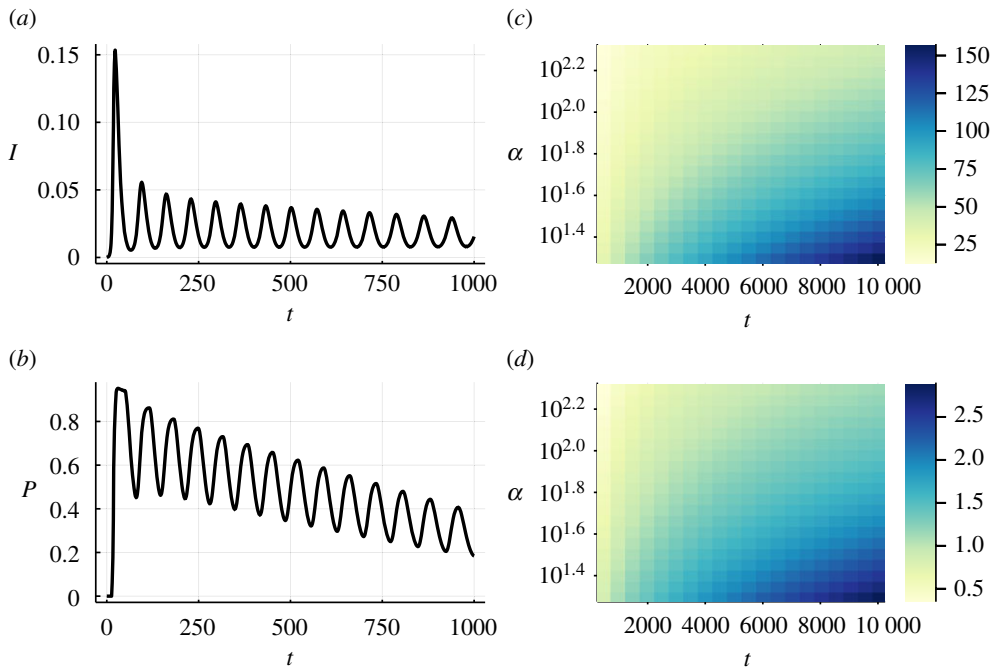
Figure 3. (a) Prevalence or fraction of infected individuals, $I = I^{NP} + I^{P}$ and (b) fraction of protected individuals $P = S^{P} + I^{P}$ evolving in time for a single realization of the process, with initial conditions $I^{NP}(0) = 10^{-5}$, and the remaining fraction is attributed to the compartment $S^{NP}(0)$, with fixed $\alpha = 50$ and $T = 2000$. (c) Average time span and (d) amplitude of the epidemic waves depending on the infection cost, T, and adoption rate, α. All results are obtained by integrating equations (2.5)–(2.9) with fixed $\Delta t = 0.005$ and parameters $\beta = 0.6$, $\mu = 0.2$, $\gamma = 0.4$ and $c = 1$. (Online version in colour.)

$$\dot{I}_i^{NP} = \beta_i S_i^{NP} \sum_{j=1}^{G} C_{ij}(I_j^{NP} + (1 - \gamma_j)I_j^{P}) - \mu_i I_i^{NP} \tag{2.7}$$

$$\dot{I}_i^{P} = \beta_i S_i^{P}(1 - \gamma_i) \sum_{j=1}^{G} C_{ij}(I_j^{NP} + \Gamma_i I_j^{P}) - \mu_i I_i^{P} \tag{2.8}$$

and $$\dot{R}_i = \mu_i(I_i^{NP} + I_i^{P}). \tag{2.9}$$

In the aforementioned equations, the parameters β_i and μ_i refer to the infection and recovery rate, respectively. The matrix **C** represents the contact matrix and controls the mixing among the different groups. The variables S_i^{P} and S_i^{NP} refer to the fraction of susceptible individuals of group i that adopt and do not adopt prophylactic measures, respectively. The variable R_i subsumes the fraction of recovered individuals in group i.

3. General phenomenology for a homogeneous population

Before focusing on the effect of different risk groups, we analyze the general phenomenology for a homogeneous population. In this scenario, we consider only one group of well-mixed individuals with constant infection cost T and the same epidemiological parameters. We study the dynamics of the system by numerically integrating the differential equations [32]. In figure 3a, we present an exemplary trajectory for the prevalence, I. We note how subsequent waves arise as occurs for other models that couple disease dynamics with the behavioural response [1,3,5–7,9]. These subsequent waves emerge from a constant strengthening and weakening of the prophylactic measures (figure 3b), which cause the reproduction number to oscillate around one. Actually, a

series of previous studies have shown as to how the reproduction number oscillates around one, whenever the prophylactic behaviour depends on the global infection risk [5,6,9], which is also the case here.

The emergence of the observed oscillatory behaviour is conditioned by the activation of a behavioural response in the population. While it is well understood that the epidemic threshold is not affected by the behavioural response [9], the protection threshold at which prophylaxis starts to be adopted strongly depends on the reach of the epidemic outbreak. In particular, there only exists a behavioural response if $P^P > P^{NP}$, i.e. if the benefits of adopting prophylactic measures are greater than not adopting them. To satisfy this condition, a critical prevalence, I^*, is required. From equations (2.1) and (2.2), we obtain that

$$I^* = \frac{c}{\gamma T}. \tag{3.1}$$

In other words, the maximum prevalence, I^{max}, during an outbreak must exceed I^* to observe prophylactic behaviour. Since we consider a homogeneous population, the maximal prevalence is given by the one of the standard SIR models [31]

$$I^{max} = 1 - \frac{1}{R_0}(1 + \ln(R_0)). \tag{3.2}$$

The variable $R_0 = \beta/\mu$ refers to the basic reproduction number of the disease. Combining equations (3.1) and (3.2), we find the critical ratio between the protection and infection cost such that a behavioural response emerges:

$$\frac{c}{T} < \gamma \left[1 - \frac{1}{R_0}(1 + \ln(R_0)) \right]. \tag{3.3}$$

If one had several groups with different infection/protection costs but with the same transmission rates, β_i, and random mixing, the same condition would hold. In this case, the group with the lowest ratio c_i/T_i would initiate the behavioural response.

Further inspection of figure 3a,b shows that, while the average prevalence stays relatively constant over time, the protection level slowly decreases. The average prevalence is almost constant since the system oscillates around the behavioural equilibrium $P^{NP} = P^P$. For $P^P > P^{NP}$, prophylaxis increases and vice-versa. However, due to the increasing immunity—intrinsic to the SIR model—less prophylaxis is needed to maintain the prevalence level, wherefore the average protection level decreases over time.

In addition, we note that the first wave leads to the highest prevalence. However, this is not always the case and strongly depends on the initial condition. When the initial fraction of protected individuals is low, the amplitude of subsequent waves continuously decreases as shown in figure 3a. However, if initially prophylaxis is widespread, the second wave leads to a higher peak in prevalence than the first one and the amplitude subsequently decreases. Furthermore, for an even higher initial protection level, the amplitude can rise for subsequent waves before the decrease starts. This dependence on the initial condition might explain cases such as Germany or Switzerland, where the second wave of SARS-CoV-2 reached higher infection levels than the first one. In analogy to the initial condition of the model, this may be explained by relatively high protection levels at the start of the first wave due to the experience from Italy and Spain. However, in this study, we focus on the regime where the first wave is the dominant one and leave a detailed study on the initial conditions for future work.

To conclude the study of the homogeneous population, we analyze the effect of the parameters related to behavioural aspects of the model. In particular, we study the influence of the adoption rate, α, and the infection cost, T, on the epidemic dynamics. For a detailed analysis of the remaining parameters, we refer the reader to reference [9]. Figure 3c,d show that both a larger infection cost and a lower adoption rate increase the time span between subsequent waves (the period) as well as their amplitude. Intuitively, the slower or more delayed the reaction is, the more time it will take to get subsequent waves under control, wherefore their amplitude also increases.

Overall, this is the basic phenomenology in the absence of any risk groups, which we will explore in the next section.

4. Distinct risk groups

Although we have so far formulated the model in general terms, i.e. by considering all the parameters inherent to every group G, we will focus on the case of two groups in the following. The reduction in the number of groups simplifies the model while preserving the key phenomenology at hand, namely the existence of a diverse risk group for which the course of COVID-19 is highly probable to be severe [33,34]. In addition, the presence of risk factors correlates widely with age, resulting in a strong and monotonous age dependence of the infection fatality ratio (IFR), which justifies our modelling in terms of young and elderly individuals [35–37].

The higher mortality for the elderly is also reflected in their behaviour. Even though the perceived and actual risk do not always match with human behaviour, a series of studies have shown that the elderly were more cautious, i.e. showed intensified social distancing through a stronger reduction in mobility [38–40], contacts [41,42] or credit card expenses [43] than the rest of the population. This is also reflected in a higher seroprevalence in the younger age groups [36].

The higher IFR for elderly combined with the observed stronger adherence to prophylactive measures supports the introduction of different infection costs, T_i. Ideally, one could work with a distribution of infection costs or one for each age strata. However, such an approach would come at the cost of simplicity and could mask the key ingredients that contribute to the basic phenomenology. For this reason, we will categorize the population into low- and high-risk groups, referred to as young (Y) and old (O), that have associated infection costs T_Y and T_O, respectively. Later in the article, we will comment on how the division into more subgroups could affect the dynamics. Similar to the infection cost, one could imagine distinguishing the adoption cost, c_i, the transmission rate, β_i, or the efficacy of the prophylactic measures γ_i between the two groups. However, for the sake of simplicity, we will consider the parameters c_i, β_i and γ_i equal in the two groups, as well as the recovery probability μ_i and the adoption rate α_i. We parametrize the interaction between the two groups, i.e. the 2×2 contact matrix $\mathbf{C}$, with a mixing rate v. To be more precise, given the fraction of old, f_O and young individuals, f_Y, in the population, the four elements of the contact matrix read as follows:

$$C_{YO} = v f_O \tag{4.1}$$

$$C_{OY} = v f_Y \tag{4.2}$$

$$C_{YY} = 1 - v f_O \tag{4.3}$$

and
$$C_{OO} = 1 - v f_Y. \tag{4.4}$$

For $v = 0$, the two groups do not interact with each other, while $v = 1$ corresponds to random mixing. In the following, we will assume that 20% of the population belong to the group at risk, i.e. $f_O = 0.2$ and $f_Y = 0.8$.

In figure 4a,b, we see that, similarly to a homogeneous population, both groups exhibit recurrent epidemic waves driven by a varying prophylaxis level. Due to the higher infection cost, the old group shows more intense prophylaxis and thus lower prevalence. In terms of the temporal evolution, we observe that the elderly population also starts to increase prophylaxis earlier (dashed lines) as prevalence increases and continues to do so for longer as prevalence decreases. This is shown in figure 4c. Furthermore, the variation in the prophylaxis level of the young is more pronounced and thus mainly drives the evolution of the prevalence. As a consequence, in figure 4d, we observe how the prevalence of the young group increases more rapidly and spills over to the old group. As a matter of fact, due to the increased prevalence tolerance caused by a lower infection cost of the young group, the reproduction number among the old is always below one. Accordingly, prevalence in the old group solely increases due to the
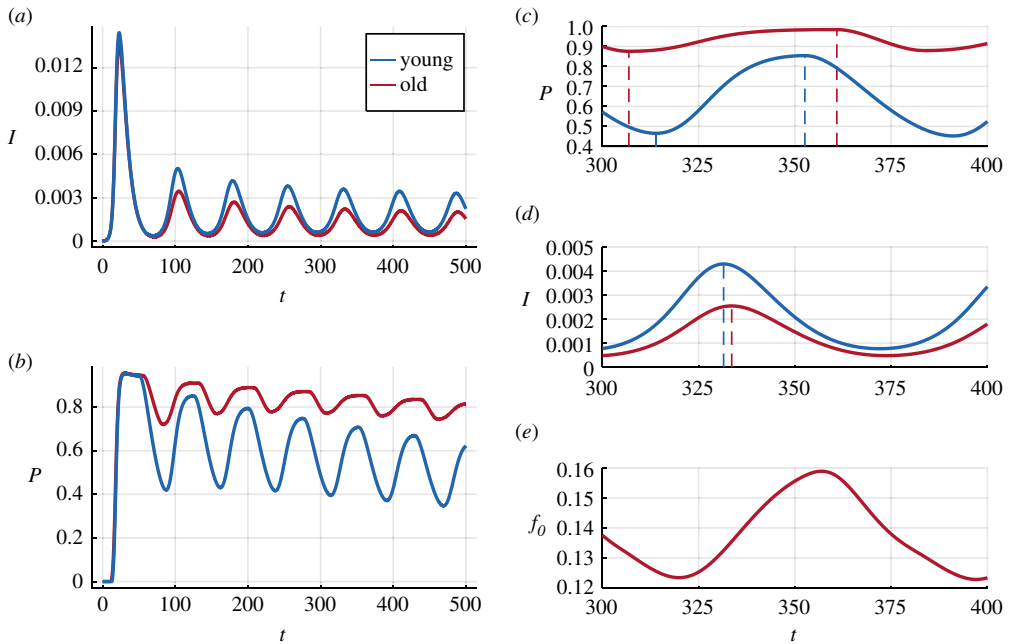
Figure 4. (*a*) Prevalence or fraction of infected individuals, $I_i = I_i^{NP} + I_i^{P}$ and (*b*) fraction of protected individuals $P_i = S_i^{P} + I_i^{P}$ in the old (red) and young (blue) groups, evolving in time, with initial conditions fixed as in figure 3 for both groups. The adoption rate is fixed at $\alpha = 50$ and the infection costs are $T_O = 2000$ and $T_Y = 500$, respectively. (*c*) Fraction of protected individuals and (*d*) fraction of infected individuals in both groups displayed in a shorter time window. (*e*) The fraction of cases belonging to the old group in time. All results are obtained by numerically integrating equations (2.5)–(2.9) using a standard ODE solver, with fixed $\Delta t = 0.005$, and in both groups, $\beta = 0.6$, $\mu = 0.2$, $\gamma = 0.4$ and $c = 1$. (Online version in colour.)

infections from the young group. This mechanism leads to a slight delay in the peak of the old group compared to the young group (dashed lines) as one can see in figure 4*d*. The model exhibits the same delay among the peaks as observed from empirical data, which we presented in figures 1 and 2.

Furthermore, due to the different growth rates, in figure 4*e*, we see a temporal variation in the case distribution, i.e. in the fraction of cases belonging to the old group. This phenomenology was also observed in the data displayed in figure 1. If we look across figure 4*c*–*e*, we see that the fraction of cases belonging to the old group (figure 4*e*) is minimal when prevalence is low (figure 4*d*), at the point where the prophylaxis level differs substantially between the two groups (figure 4*c*). However, as prevalence increases (figure 4*d*), the prophylaxis level between the two groups becomes closer (figure 4*c*), wherefore also the growth rates of the epidemic in both groups differ less (figure 4*d*) and the fraction of cases belonging to the old group approaches the population distribution (figure 4*e*). This dynamics leads to the temporal variation of the case distribution.

Overall, results in figure 4 show how the varying infection risk among the two groups, which translates into a heterogeneous behavioural reaction, can explain the phenomenology—the time varying case distribution as well as the delay among prevalence peaks—that we observed in the data. Furthermore, we checked that imposing different contact rates, β_i, when the infection cost are the same in each group, does not lead to any of the two phenomena. On the other hand, the phenomenology holds also when contact rates differ together with the infection cost. In this sense, at least under the framework of this model, the behavioural heterogeneity is truly the key factor that reproduces qualitatively the real-world observations.

As pointed out earlier, we simplified the model and separated the population into low- and high-risk groups. However, one could also introduce additional divisions or directly make use of
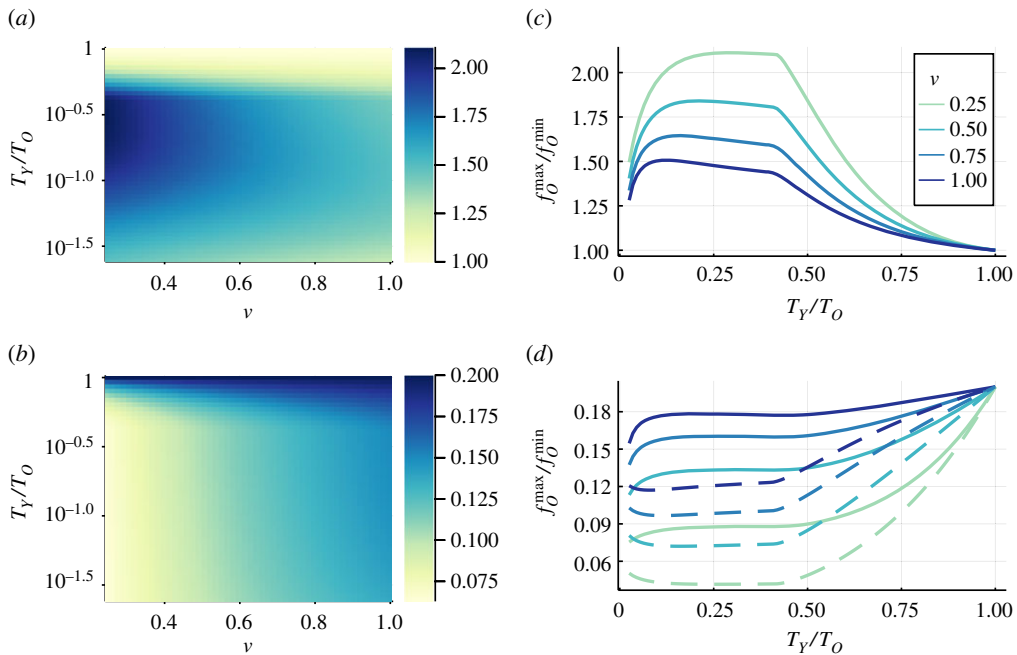
Figure 5. Top: Ratio between maximum, f_0^{max}, and minimum, f_0^{max}, in the variation of the case distribution, i.e. in the fraction of infected individuals belonging to the old group, f_0, in time depending on the mixing rate between the groups, ν, and the ratio of infection costs T_Y/T_O (a). We also show the dependence on the infection cost in the young group, when the infection cost for the old group is fixed, $T_O = 2000$, for four choices of the mixing rate (c). Bottom: Average between maximum, f_0^{max}, and minimum, f_0^{min}, in the variation of the case distribution, f_0 depending on the mixing rate between the groups, ν, and the ratio of infection costs T_Y/T_O (b). We also show how the maximal (solid) and minimal (dashed) fraction of cases belonging to the old group depends again on the infection cost of the young group T_Y, when the infection cost for the old group is fixed, $T_O = 2000$, for the same choices of the mixing rate (d). All results are obtained by numerically integrating equations (2.5)–(2.9) using a standard ODE solver, with fixed $\Delta t = 0.005$, and, in both groups, $\alpha = 0.50$, $\beta = 0.6$, $\mu = 0.2$, $\gamma = 0.4$ and $c = 1$, and initial conditions $I^{NP}(0) = 0.0015$ in both groups. (Online version in colour.)

the nine age groups as in figure 1. If one assumes that the infection cost progressively increases with age the general phenomenology is very similar. Groups will reach their epidemic peak subsequently according to their infection cost. Similarly, the fraction of infections in the groups with high infection cost increases towards the peak and is particularly small during times of low prevalence. Since the phenomenology is robust regarding the division into additional groups but hinders clear and concise visualization, we refrained from presenting it.

To which extent the phenomenology described earlier is pronounced depends on the parameters. In the following, we analyze the effect of the infection cost and the mixing rate between the groups and also the dependence on the adoption rate, unveiling a few interesting effects.

In figure 5a,c, we observe that the variation in the case distribution non-trivially depends on the ratio between the infection costs of the two groups for the whole range of the mixing rate ν. If the infection cost of the young group is too low, the protection level does not significantly vary. Accordingly, the case distribution stays pretty stable. On the other hand, if the infection cost of the young approaches that of the old, we return to the homogeneous case, where the case distribution does not vary either. The competition between those two processes causes the maximum in the relative variations of the case distribution in time.

Interestingly, the lower the mixing rate, the more pronounced is the case variation over time. In contrast, as we show in figure 5b,d, a lower mixing rate decreases prevalence among the old
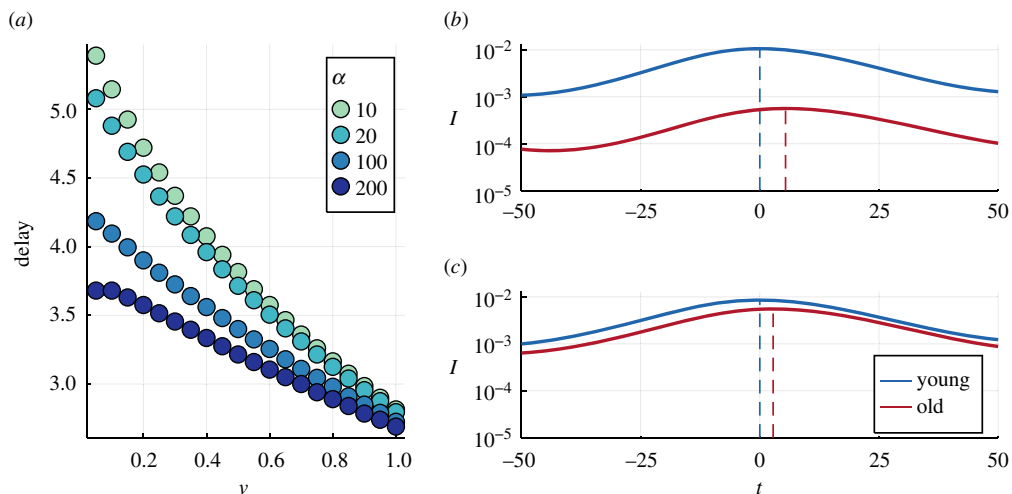
Figure 6. (a) Dependence on the mixing rate for four choices of the adoption rate α. All results are obtained by numerically integrating equations (2.5)–(2.9) using a standard ODE solver, with same initial conditions as in figure 5, with fixed $\Delta t = 0.005$, and infection costs $T_0 = 2000$ and $T_Y = 500$. In both groups, $\alpha = 10$, $\beta = 0.6$, $\mu = 0.2$, $\gamma = 0.4$ and $c = 1$. Right: Zoom into the temporal evolution of the fraction of infected individuals for (b) $\nu = 0.05$ and (c) random mixing, $\nu = 1.0$. Dashed lines indicate the position of the maximum. The adoption rate was fixed as $\alpha = 10$. Time, t, was shifted such that the peak of the age group young is located at zero. (Online version in colour.)

due to the better structural shielding. This is particularly relevant since a good shielding reduces prevalence among the risk group but leads to substantial variation of the case distribution. These variations then may cause an increase in IFR as well as in the infection hospitalization rate towards the peak of infection, which would put additional stress on the health care system. Finally, again here, we observe how the case distribution approaches the population structure as the infection cost of the old approaches that of the young.

Finally, in figure 6a, we see how the delay between the peaks monotonously decreases with mixing rate as well as with the adoption rate. In particular, a more modular contact matrix (lower ν) delays the increasing prevalence in the young group from propagating into the risk group. Figure 6b,c show the temporal evolution, for two values of the mixing rate between the young and old population, ν. For low mixing (figure 6b), we observe a larger delay in the epidemic peak than in the case of a high mixing (figure 6c). This means that a better overall shielding of the risk group (lower mixing) (figure 6b) may lead to an unforeseen, late peak in prevalence among the risk group after the arrival of a wave. As previously shown in figure 3c, a higher adoption rate, α, decreases the intervals at which waves pass and thus leaves less time for the prevalence to propagate into the risk groups. Consequently, the delay between the peaks shrinks as individuals adopt their behaviour faster.

5. Conclusion

By analyzing the available information of confirmed SARS-CoV-2 infections, we have shown how the case distribution among age groups temporally varies through time. Furthermore, we highlighted how the cases among the older age groups—in relative as well as absolute terms—peak later than the total number of cases. This phenomenology is observed for subsequent waves as well as across a diverse set of countries and regions.

We proposed a minimal, mechanistic model that exhibits the same phenomenology. The disease is controlled through the voluntary adoption of prophylactic measures, which we modelled as a two-strategy game. Within this framework, the key factor is the heterogeneous risk

of a severe course of a disease that widely correlates with age. The varying severity of the disease translates into a heterogeneous behavioural reaction. To be more precise, the group at lower risk drives the infections as their adoption of prophylaxis is not as widespread and occurs later. The changing difference in prophylaxis between the high- and low-risk group leads to a temporal variation of the case distribution. Furthermore, as the low-risk group drives the infections towards their peak, maximal prevalence is reached with a delay among the risk group. Both phenomena match the observations from the data.

There is an interest in this study from a theoretical point of view, since a minimal model is able to exhibit the same phenomenology as the reported data. More importantly, the study highlights various points relevant for the public health policies of the pandemic. Our results indicate that an initial low fraction of infections among the risk group can lead to an underestimation of the potential future stress on the health care system. In fact, we observed such an underestimation during the summer months of 2020. In this period, the reported cases were mainly among younger people, leading to a reduced stress on health care systems. Consequently, the public discourse advocated tolerating higher prevalence levels, i.e. restoring social fabric while coexisting with virus transmission. However, along with seasonal effects, the resulting behavioural relaxation led to an increase in prevalence among the at-risk group, resulting in the emergence of severe second waves in several countries around the world.

There are several limitations to this study. We did not include the impact of seasonal effects and government policies on the behavioural response to the epidemic. Such an approach would require more detailed, large-scale behavioural data and a deeper understanding of human behaviour as well as the specifities of the virus. Nevertheless, already minimal data, such as age-stratified case numbers, allowed us to observe regularities in the infections patterns. In addition to the data analysis, modelling efforts are necessary to unveil the mechanisms behind the observations. In this sense, more publicly available data regarding the behavioural response would allow us to include the behavioural response dynamically into epidemic models. Such a refined approach from a behavioural point of view is crucial to advance from conditional scenarios to actual predictions.

References

1. Funk S, Salathé M, Jansen VAA. 2010 Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R Soc. Interface* **7**, 1247–1256. (doi:10.1098/rsif.2010.0142)

2. Granell C, Gómez S, Arenas A. 2013 Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Phys. Rev. Lett.* **111**, 128701. (doi:10.1103/PhysRevLett.111.128701)

3. Wang Z, Bauch CT, Bhattacharyya S, Manfredi P, Perc M, Perra N, Salathé M, Zhao D. 2016 Statistical physics of vaccination. *Phys. Rep.* **664**, 1–113. (doi:10.1016/j.physrep.2016.10.006)

4. Arenas A, Cota W, Gómez-Gardeñes J, Gómez S, Granell C, Matamalas JT, Soriano-Paños D, Steinegger B. 2020 Modeling the spatiotemporal epidemic spreading of COVID-19 and the impact of mobility and social distancing interventions. *Phys. Rev. X* **10**, 041055.

5. Bauch CT. 2005 Imitation dynamics predict vaccinating behaviour. *Proc. R. Soc. B* **272**, 1669–1675. (doi:10.1098/rspb.2005.3153)

6. Funk S, Gilad E, Watkins C, Jansen VAA. 2009 The spread of awareness and its impact on epidemic outbreaks. *Proc. Natl Acad. Sci. USA* **106**, 6872–6877. (doi:10.1073/pnas.0810762106)

7. d'Onofrio A, Manfredi P, Salinelli E. 2007 Vaccinating behaviour, information, and the dynamics of SIR vaccine preventable diseases. *Theor. Popul. Biol.* **71**, 301–317.

8. Steinegger B, Cardillo A, Rios PDL, Gómez-Gardeñes J, Arenas A. 2018 Interplay between cost and benefits triggers nontrivial vaccination uptake. *Phys. Rev. E* **97**, 032308. (doi:10.1103/PhysRevE.97.032308)

9. Steinegger B, Arenas A, Gómez-Gardeñes J, Granell C. 2020 Pulsating campaigns of human prophylaxis driven by risk perception palliate oscillations of direct contact transmitted diseases. *Phys. Rev. Res.* **2**, 023181. (doi:10.1103/PhysRevResearch.2.023181)

10. Zhao S, Bauch CT, He D. 2018 Strategic decision making about travel during disease outbreaks: a game theoretical approach. *J. R. Soc. Interface* **15**, 20180515. (doi:10.1098/rsif.2018.0515)

11. Weitz JS, Park SW, Eksin C, Dushoff J. 2020 Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations. *Proc. Natl Acad. Sci. USA* **117**, 32 764–32 771.

12. Pedro SA, Ndjomatchoua FT, Jentsch P, Tchuenche JM, Anand M, Bauch CT. 2020 Conditions for a second wave of COVID-19 due to interactions between disease dynamics and social processes. *Front. Phys.* **8**, 167. (doi:10.3389/fphy.2020.574514)

13. Amaral MA, de Oliveira MM, Javarone MA. 2021 An epidemiological model with voluntary quarantine strategies governed by evolutionary game dynamics. *Chaos Solitons Fractals* **143**, 110616. (doi:10.1016/j.chaos.2020.110616)

14. Zhao S, Stone L, Gao D, Musa SS, Chong MKC, He D, Wang MH. 2020 Imitation dynamics in the mitigation of the novel coronavirus disease (COVID-19) outbreak in Wuhan, China from 2019 to 2020. *Ann. Transl. Med.* **8**, 448. (doi:10.21037/atm.2020.03.168)

15. Meidan D, Schulmann N, Cohen R, Haber S, Yaniv E, Sarid R, Barzel B. 2021 Alternating quarantine for sustainable epidemic mitigation. *Nat. Commun.* **12**, 220. (doi:10.1038/s41467-020-20324-8)

16. Google COVID-19 Community Mobility Reports. www.google.com/covid19/mobility/?hl=en (accessed 19 April 2021).

17. Dades COVID. https://dadescovid.cat/descarregues (accessed 19 April 2021).

18. Covid-19 Schweiz. www.covid19.admin.ch/de/overview (accessed 19 April 2021).

19. Florida COVID19 Case Line Data. https://open-fdoh.hub.arcgis.com/datasets/florida-covid19-case-line-data (accessed 19 April 2021).

20. RKI COVID19. https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0 (accessed: 19 April 2021.

21. STATISTICS DENMARK. www.statistikbanken.dk/statbank5a/SelectVarVal/Define.asp?Maintable=SMIT2&PLanguage=1 (accessed 19 April 2021).

22. ARAGÓN OPEN DATA. https://opendata.aragon.es/datos/catalogo/dataset/publicaciones-y-anuncios-relacionados-con-el-coronavirus-en-aragon (accessed 19 April 2021).

23. COVID-19 in Belgium. https://data.world/liz-friedman/covid-19-in-belgium (accessed 19 April 2021).

24. GOV.UK Coronavirus (COVID-19) in the UK. https://coronavirus.data.gov.uk/details/download (accessed 19 April 2021).

25. Mena GE, Martinez PP, Mahmud AS, Marquet PA, Buckee CO, Santillana M. 2021 Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. *Science* **372**, eabg5298. (doi:10.1126/science.abg5298)

26. Flaxman S *et al.* 2020 Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261. (doi:10.1038/s41586-020-2405-7)

27. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, Pinior B, Thurner S, Klimek P. 2020 Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat. Human Behav.* **4**, 1303–1312. (doi:10.1038/s41562-020-01009-0)

28. Brauner JM *et al.* 2021 Inferring the effectiveness of government interventions against COVID-19. *Science* **371**, 46. (doi:10.1126/science.abd9338)

29. Merow C, Urban MC. 2020 Seasonality and uncertainty in global COVID-19 growth rates. *Proc. Natl Acad. Sci. USA* **117**, 27 456–27 464. (doi:10.1073/pnas.2008590117)

30. Carlson CJ, Gomez ACR, Bansal S, Ryan SJ. 2020 Misconceptions about weather and seasonality must not misguide COVID-19 response. *Nat. Commun.* **11**, 4312. (doi:10.1038/s41467-020-18150-z)

31. Keeling MJ, Rohani P. 2011 *Modeling infectious diseases in humans and animals*. New Jersey: Princeton University Press.

32. Rackauckas C, Nie Q. 2017 DifferentialEquations.jl—a performant and feature-rich ecosystem for solving differential equations in Julia. *J. Open Res. Softw.* **5**, 125. Exported from https://app.dimensions.aion2019/05/05. (doi:10.5334/jors.151)

33. Williamson EJ *et al.* 2020 Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436. (doi:10.1038/s41586-020-2521-4)

34. Liu D *et al.* 2020 Risk factors for developing into critical COVID-19 patients in Wuhan, China: a multicenter, retrospective, cohort study. *EClinicalMedicine* **25**, 100471 (doi:10.1016/j.eclinm.2020.100471)

35. O'Driscoll M, Dos Santos GR, Wang L, Cummings DA, Azman AS, Paireau J, Fontanet A, Cauchemez S, Salje H. 2021 Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* **590**, 140–145. (doi:10.1038/s41586-020-2918-0)

36. Pastor-Barriuso R *et al.* 2020 Infection fatality risk for SARS-CoV-2 in community dwelling population of Spain: nationwide seroepidemiological study. *BMJ* **371**, m4509.

37. Davies NG *et al.* 2020 Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26**, 1205–1211. (doi:10.1038/s41591-020-0962-9)

38. Pullano G, Valdano E, Scarpa N, Rubrichi S, Colizza V. 2020 Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. *Lancet Digit. Health* **2**, e638–e649. (doi:10.1016/S2589-7500(20)30243-0)

39. Gauvin L *et al.* 2020 Socioeconomic determinants of mobility responses during the first wave of COVID-19 in Italy from provinces to neighbourhoods. *J. R. Soc. Interface* **18**, 20210092. (doi:10.1098/rsif.2021.0092)

40. Chan HF, Skali A, Savage DA, Stadelmann D, Torgler B. 2020 Risk attitudes and human mobility during the COVID-19 pandemic. *Sci. Rep.* **10**, 19931. (doi:10.1038/s41598-020-76763-2)

41. Pullano G *et al.* 2021 Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature* **590**, 134–139. (doi:10.1038/s41586-020-03095-6)

42. Santé publique France. Covid-19: une enquête pour suivre l'évolution des comportements et de la santé mentale pendant l'épidémie. www.santepubliquefrance.fr/etudes-et-enquetes/covid-19-une-enquete-pour-suivre-l-evolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemies (accessed 19 April 2021).

43. Eichenbaum M, Godinho de Matos M, Lima F, Trabandt M. 2020 How do people respond to small probability events with large, negative consequences? CEPR Discussion Paper No. DP15373.