Importance of untested infectious individuals for the suppression of COVID-19 epidemics

Francisco J. Pérez-Reche and Norval Strachan

School of Natural and Computing Sciences, University of Aberdeen, Aberdeen, Scotland

Correspondence: fperez-reche@abdn.ac.uk

Abstract

A mathematical model which accounts for tested and untested infectious individuals is calibrated during the early stages of COVID-19 outbreaks in Germany, the Hubei province, Italy, Spain and the UK. The predicted percentage of untested infected individuals depends on the specific outbreak but we found that they typically represent 50% to 80% of the infected individuals. Even when unreported cases are taken into account, we estimate that less than 8% of the population would have been exposed to SARS-CoV-2 by 09/04/2020 in the analysed outbreaks. These levels are far from the 70-85% needed to ensure herd immunity and we predict a resurgence of infection if ongoing lockdowns in the analysed outbreaks are fully lifted. We propose that partially lifted lockdowns together with fast and thorough testing allowing for quick isolation of both symptomatic and asymptomatic cases could lead to suppression of secondary waves of COVID-19 epidemics.

Introduction

Coronavirus disease (COVID-19) produced by the SARS-CoV-2 virus emerged in Wuhan, China in December 2019[1]. The virus has spread at an unprecedented rate since then, leading to 1,521,252 confirmed cases and 92,798 deaths distributed in 213 countries as of 10 April 2020[2]. The worldwide burden of the disease is still growing despite significant efforts of many countries to suppress the spread of the virus. So far, efforts have focused on non-pharmaceutical interventions which range from handwashing or social distancing to more stringent measures such as isolation of infected individuals, banning of large gatherings or severe lockdowns[3,4].

Optimising interventions to mitigate or suppress the burden of COVID-19 remains a pressing global challenge due to significant uncertainties regarding the transmissibility of SARS-CoV-2 and other factors such as possibly a large proportion of undocumented infections as well as political, social and economic considerations[5,6]. Underreporting of infections may depend on the testing ability of different countries and the presence of asymptomatic infected individuals[7–9]. There is no consensus on the proportion of unreported cases and their potential impact on the spread of SARS-CoV-2. For instance, a World Health Organization report in February suggested that "the proportion of truly asymptomatic infections is unclear but appears to be relatively rare and does not appear to be a major driver of transmission"[10]. Studies testing for SARS-CoV-2 infection in both symptomatic and asymptomatic individuals[7–9], however, suggest that asymptomatic carriers can represent 50% or more of the cases. In many countries that mostly test individuals when they have symptoms, unreported infections are likely to include at least most of the asymptomatic individuals or those with mild symptoms. Such individuals can act as silent carriers for SARS-CoV-2 and have been suggested as a key

factor promoting the rapid spread of the virus[11], similar to what has been observed in other infectious diseases[12]. On the positive side, if recovery from infection leads to immunity, one could hope that untested positive individuals could significantly contribute to the build-up of herd immunity in the population[13,14]. It is not clear to what extent this could be the case. The importance of silent carriers on interventions for mitigation and suppression[15] of the infection is not clear either.

Mathematical modelling has been very successful in epidemiology[16–18] and there is an ongoing effort to propose models to describe the dynamics of COVID-19 epidemics[3,11,14,15,19–25]. Unreported infectious individuals have been included in some models[3,11,22,26] but their influence on control strategies has not been analysed.

Here, we use data from the outbreaks in Germany, Hubei (China), Italy, Spain and UK to calibrate a mathematical model that accounts for the force of infection associated with both tested (reported) and untested infectious individuals (see a scheme of the simplest version of the model in Figure 1 and more details in Methods). In order to compare outbreaks in different regions/countries, we fit the model independently to each outbreak. The calibrated model is used to study the effect of two suppression strategies: Interventions aiming for a reduction of transmission at the population level (representing, e.g., social distancing or a lockdown) and local interventions consisting in isolation of both tested and untested infectious individuals.

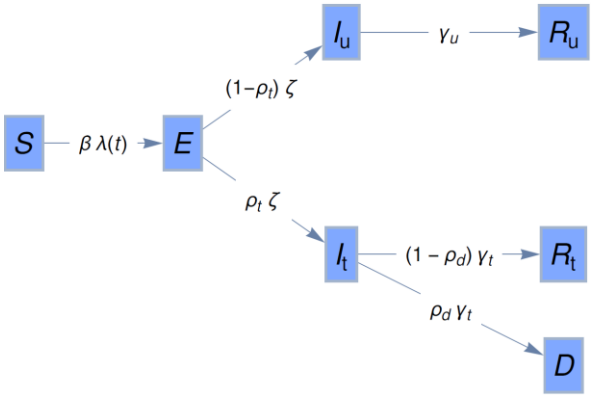

Figure 1. Simplest version of the model used to simulate the SARS-CoV-2 epidemic. At any given time, individuals can belong to one of 7 compartments: S (susceptible to SARS-CoV-2), E (exposed but not yet infectious), I_t (infectious tested), R_t (recovered from tested infected), D (dead), I_u (infected untested) and R_u (recovered from untested infected). It is assumed that all deaths are associated with tested infectious individuals. Recovered individuals are assumed to be immune to the virus. The transition rates between compartments are indicated on the arrows. β is the rate at which susceptible individuals become exposed. The force of infection, $\lambda(t)$, accounts for both tested and untested individuals. Here, N is the size of the modelled population. The mean incubation period is ζ^{-1}. The proportion of exposed individuals that become infected and are tested defines the testing rate, ρ_t. Tested infectious individuals die or recover at a rate γ_t. The parameter ρ_d gives the proportion of infected tested that die. Untested infectious individuals recover at a rate γ_u.

Model calibration, under-reporting and reporting delays

Figure 2 shows that the model accurately captures the early stages of the fitted outbreaks. Estimates for the parameters of each outbreak are given in Table 1.

The obtained values for the testing rate ρ_t reveal that during the early stage of outbreaks, Germany scored the highest in terms of testing for infection (median 53%). Hubei follows Germany in terms of testing, followed by Spain, Italy and the UK. Our prediction for Hubei is not far from the 65% reporting rate estimated by Li et al. [11] for China in the period considered here. The high testing rate predicted for Germany agrees with the known high testing capacity in this country[27]. Taking the confidence

intervals into account, we estimate that for each infected individual tested in the UK, there could have been between 2 and 10 untested infected individuals. At the other end of the testing spectrum, we estimate that for each infected individual tested in Germany, between 0.2 and 2 individuals might have not been tested at the beginning of the epidemic. A higher testing rate for Germany is in qualitative agreement with estimates given elsewhere[22,26]. Our estimates for the reporting rate, however, tend to be higher than those obtained by Jagodnik et al.[26] and the differences we found between countries are not as extreme as those given by Chicci et al.[22]

Assuming that the testing rate remains constant during the course of epidemics and no control interventions are implemented, our model predicts that the number of tested and untested infected individuals would evolve in parallel in all the studied outbreaks which would last for around 12 weeks in all cases (see Figure 3). We see, however, that the epidemic in Germany would belong to a different class in the sense that the number of untested infected individuals remains smaller than the number of tested individuals during the whole epidemic. Italy, Spain and UK exhibit the opposite behaviour with more untested than tested individuals. For Hubei, we predict similar levels of untested and tested rates. Obviously, these predictions will not be fulfilled since control interventions are imposed in all these countries and testing strategies might change during the pandemic.

Bearing in mind that 50% of infected individuals might be asymptomatic[7,9], we conclude that most of those that were infected but not tested in Germany were asymptomatic. In contrast, untested individuals in other countries might include a significant number of individuals with mild symptoms in addition to those that are asymptomatic.

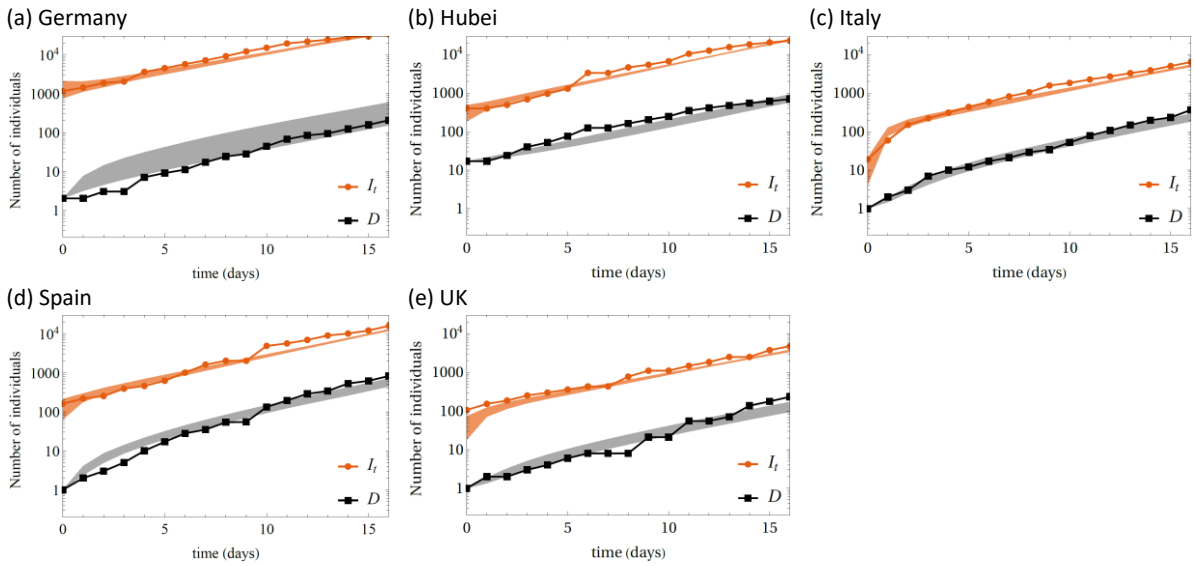

Figure 2. Number of infected (red) and dead (black/grey) individuals registered in (a) Germany, (b) Hubei, (c) Italy, (d) Spain and (e) UK. Symbols show the data and shaded regions show the 90% confidence interval of model predictions at any given time. Time given in days since the first day with a positive number of deaths in the datasets (see Table 2). Logarithmic scale is used in the vertical axis of each plot.

Table 1. Estimates of the model parameters given in terms of the 5% percentile, median and 95% percentile. β is the transmission rate, ρ_t is the proportion of tested infectious (in percentage), ρ_d is the proportion of tested infectious that die (in percentage), γ_t is the rate of recovery of tested infectious individuals, γ_u is the rate of recovery of untested infectious individuals, $E(0)$ is the initial number of exposed individuals and $\mathcal{R}_0$ is the reproduction number.

	β (day^{-1})	ρ_t (%)	ρ_d (%)	γ_t (day^{-1})	γ_u (day^{-1})	$E(0)$	$\mathcal{R}_0$
Germany	(0.6,0.8,1.0)	(31,53,87)	(0.4,1.7,3.1)	(0.05,0.15,0.24)	(0.12,0.27,0.42)	(544,3976,6939)	(2.0,3.2,6.4)
Hubei	(1.0,1.2,1.4)	(22,40,57)	(2.0,3.3,5.0)	(0.14,0.24,0.36)	(0.18,0.31,0.39)	(1626,3101,6526)	(3.2,3.9,5.8)
Italy	(0.9,1.2,1.4)	(10,25,50)	(2.7,3.7,4.9)	(0.21,0.30,0.38)	(0.17,0.31,0.38)	(1165,2271,5302)	(3.2,3.9,5.9)
Spain	(1.0,1.3,1.4)	(19,33,62)	(2.7,3.7,4.8)	(0.23,0.32,0.38)	(0.18,0.30,0.38)	(1207,2620,5590)	(3.6,4.2,6.2)
UK	(0.8,1.0,1.3)	(9,16,31)	(2.2,3.2,4.9)	(0.17,0.28,0.39)	(0.13,0.26,0.38)	(1031,2452,5179)	(3.2,4.2,5.8)

The proportion of tested infected individuals that die, ρ_d, is smaller for the outbreak in Germany than for the other outbreaks. This might be a combined effect of the fact that infected individuals in this country were relatively young at the beginning of the outbreak[28] and the high testing rate. On the one hand, the COVID-19 fatality rate is lower for the younger than for the elderly[29]. On the other hand, the higher the testing rate, the more individuals with mild symptoms will be included in the tested infected compartment of our model. The lower death rate of individuals with mild symptoms will lead to an effectively lower death rate for the whole set of infected individuals in this compartment. We note that the specific value of ρ_d for a country does not fully determine the expected fraction of deaths in the country. Indeed, we found that the predicted fraction of deaths by the end of unmitigated epidemics is not too different for different countries (medians are as follows: 0.4% for Germany, 0.9% for Hubei, 0.5% for Italy, 0.8% for Spain and 0.3% for UK).

The median of the recovery rate γ_t gives the following estimates for the time from reporting to recovery or death of tested infected individuals, γ_t^{-1}: 3 days for Spain, 3.3 days for Italy, 3.6 days for the UK, 4.2 days for Hubei and 7 days for Germany. The time for Hubei is consistent with the 3.48 days reported by Li et al.[11] for China. In general, the values we obtained are smaller than the infectious period (time from infection to death/recovery) reported elsewhere for COVID-19[3,29,30]. Our estimates thus probably reflect a reporting delay in all the studied outbreaks, in agreement with data on the onset of symptoms and reporting[28,31,32]. Our model predicts the smallest reporting delay for Germany. This is again in agreement with the high testing capacity of this country.

The recovery period for untested infected individuals, γ_u^{-1}, takes values of around 3 days for all the studied outbreaks. Comparison with the reporting-to-recovery period γ_t^{-1} and bearing in mind the reporting delays in all outbreaks, our estimates of γ_u^{-1} suggest that untested individuals remain infectious for a shorter time than tested individuals. This is in line with the lower infectivity of untested individuals proposed in a recent study that, instead of assuming different recovery rates for tested and untested individuals, assumed a lower transmission rate for unreported infectious individuals[11].

We predict that the number of exposed individuals at the beginning of our simulations, $E(0)$, is of the order of several thousand for all the countries, in qualitative agreement with estimates of a previous study for China[11].

We obtained similar values of the reproduction number $\mathcal{R}_0$ for all the studied outbreaks. To some extent this reflects our prior assumption that transmission of SARS-CoV-2 is intrinsically similar in different regions. The transmission rate, β, was derived from the estimates of ρ_t, γ_t, γ_u and $\mathcal{R}_0$, using Eq. (4) (see Methods). Values of β are statistically similar for all countries except Germany which features a smaller value. Bearing in mind that $\mathcal{R}_0$ and γ_u take similar values for all the countries, we conclude that a lower value of β is a consequence of the higher testing rate and smaller recovery rate

γ_t for this country. Indeed, according to Eq. (4), β decreases with increasing ρ_t and decreasing γ_t (or increasing period γ_t^{-1}).

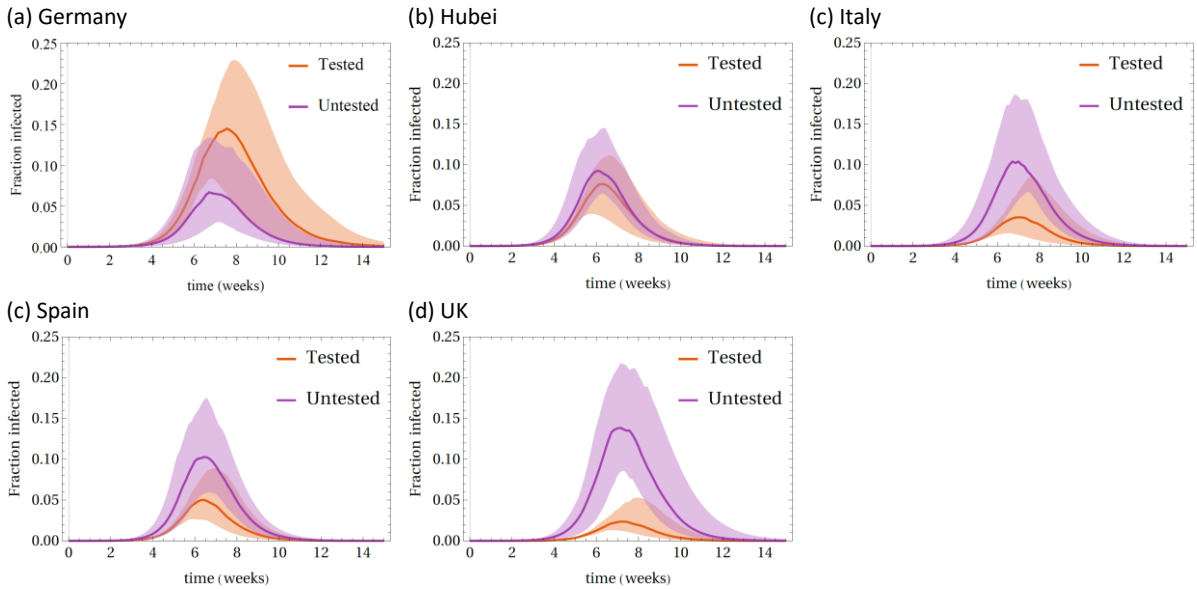

Figure 3. Fraction of the susceptible population in (a) Germany, (b) Hubei, (c) Italy, (d) Spain and (e) UK of tested (red) and untested (purple) that are infected (no control interventions implemented). Lines give the median of the model predictions and shaded areas give the 90% confidence interval of predictions at any given time. Time given in weeks since the first day considered for each country/region (see Table 2 in Methods).

Interventions focusing on the reduction of the transmission rate at the population level

Interventions such as lockdowns or social distancing can be effectively studied by reducing the transmission rate β in our model. As illustrated in Figure 4(a), the outbreak can be significantly delayed if the transmission rate is reduced from early stages in the epidemic, in agreement with other works[15]. In spite of that, the predicted number of deaths by the end of the epidemic only reduces significantly when β is reduced by a factor close to $r = 1 - 1/\mathcal{R}_0$ to ensure an early eradication of the infection[16]. Based on our estimated values for $\mathcal{R}_0$, this requires reducing the transmission rate by more than 70% in all of the studied outbreaks. This is illustrated in Figure 4(b) for the UK. As can be seen, the number of deaths would only reduce significantly if the number of contacts were reduced by approximately 80%.

Our model can be readily used to predict the effect of lockdowns of arbitrary duration and different exit strategies from such lockdowns.

Figure 5 shows predictions for the lockdowns in Hubei, Italy and Spain. Predictions for Germany and the UK are also possible but a comparison with observed effects is uncertain at present since lockdowns were implemented more recently in these countries.

For Hubei, the model reproduces well the observed daily deaths assuming a 90% reduction of the transmission rate from 6 February 2020 (see Figure 5(a)). The later is an effective date between the 23 January when the lockdown was implemented in Wuhan, the capital of the province, and 13 February when it was implemented in the whole province. Despite the spectacular reduction of daily deaths induced by the lockdown in Hubei, our model predicts that fully removing the lockdown after

60 days (i.e. approximately at the time of writing) would lead to a rapid resurge of the epidemic (see the marked increase predicted after week 12 in Figure 5(a)). In contrast, an exit strategy in which the transmission is kept reduced by a 75% is predicted to keep the number of daily deaths below 100 for many weeks (see Figure 5(c)).

We estimate that the lockdown ordered in Italy reduced the transmission by around 80%. Our prediction suggests that this will lead to a significant decrease of the number of deaths if it is kept for a long enough time. In particular, Figure 5(b) shows a scenario in which the lockdown is kept at the same level for 90 days since its implementation on 11 March 2020. In this case, we predict around 69 [90% CI: (16,568)] daily deaths at the end of the lockdown. As for Hubei, a full removal of the lockdown leads to a fast resurge of the epidemic (Figure 5(b)). For Italy, we estimate that an exit strategy from this lockdown should still keep the transmission at low values (~70% reduction) for the daily deaths to remain at a moderate value of around 100 (see Figure 5(d)).

The effectiveness of the lockdown imposed in Spain is predicted to be similar to the one in Italy (Figure 5(c)): The current lockdown managed to reduce the transmission by ~80% and resurge of infection is predicted to occur if the lockdown is completely removed after 90 days. At the end of the initial 90 days lockdown, we predict around 99 [90% CI: (22,501)] daily deaths. The number of daily deaths is predicted to remain at this level if the lockdown is partially lifted to a situation in which transmission is kept to a 72% reduced level.

Irrespective of the effectiveness of the lockdown, our model predicts that epidemics will resurge after relatively extended lockdowns in Hubei, Italy and Spain. The same is likely to occur in other countries. In fact, resurgence of the disease is predicted even for much longer lockdowns. This is due to the fact that an early lockdown delays the spread but does not lead to herd immunity. Assuming that recovered individuals are immune to SARS-CoV-2, herd immunity is only achieved when a proportion $1 - 1/\mathcal{R}_0$ of the susceptible population has been infected and died or recovered. Even if the number of untested individuals that may have recovered are taken into account, we estimate that, as of 09/04/2020, the the proportion of the susceptible population that has been exposed to the virus (i.e. the attack rate, see Methods) is 0.65% [90% CI: 0.46%-1.18%] for Germany, 0.5% [0.3%-1.0%] for Hubei, 4.6% [3.3%-7.2%] for Italy, 3.7% [2.0%-6.4%] for Spain and 4.4% [2.7%-6.9%] for the UK. These proportions are small compared to the 70-85% needed to ensure herd immunity for these epidemics. Our conclusion is in qualitative agreement with the results by Flaxman et al.[3] despite the fact that their estimates for the attack rate tend to be higher than ours. Our conclusions, however, disagree with Lourenço et al.[14] that predicted much higher attack rates that would be close to herd immunity.

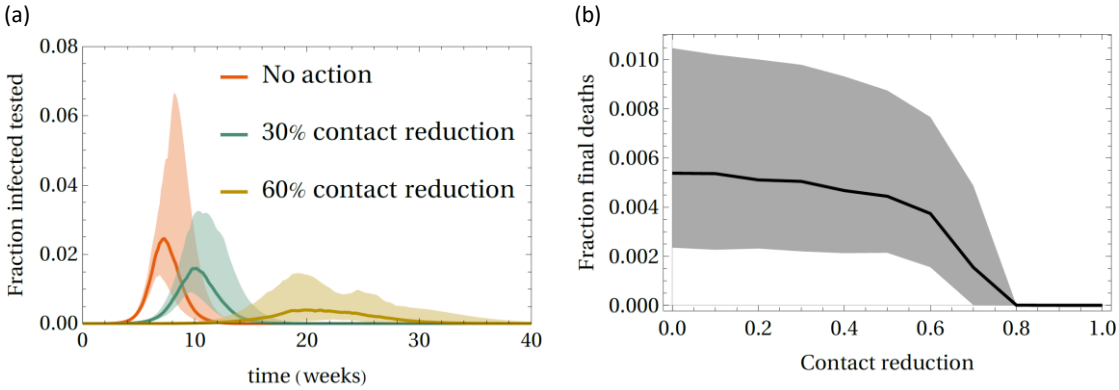

Figure 4. Predicted effect of reducing the transmission rate β on the outbreak in the UK in a hypothetical scenario in which the intervention was applied from the beginning of the epidemic and kept active until the end. (a) Proportion of tested infectious individuals as a function of time if no interventions are implemented (red) or if the transmission rate is reduced by 30% (green) or 60% (brown). (b) Effect of transmission reduction on the fraction of the susceptible population that die during the epidemic. The line gives the median and the shaded regions give the 90% confidence interval of predictions at any value of the reduction of the transmission rate.

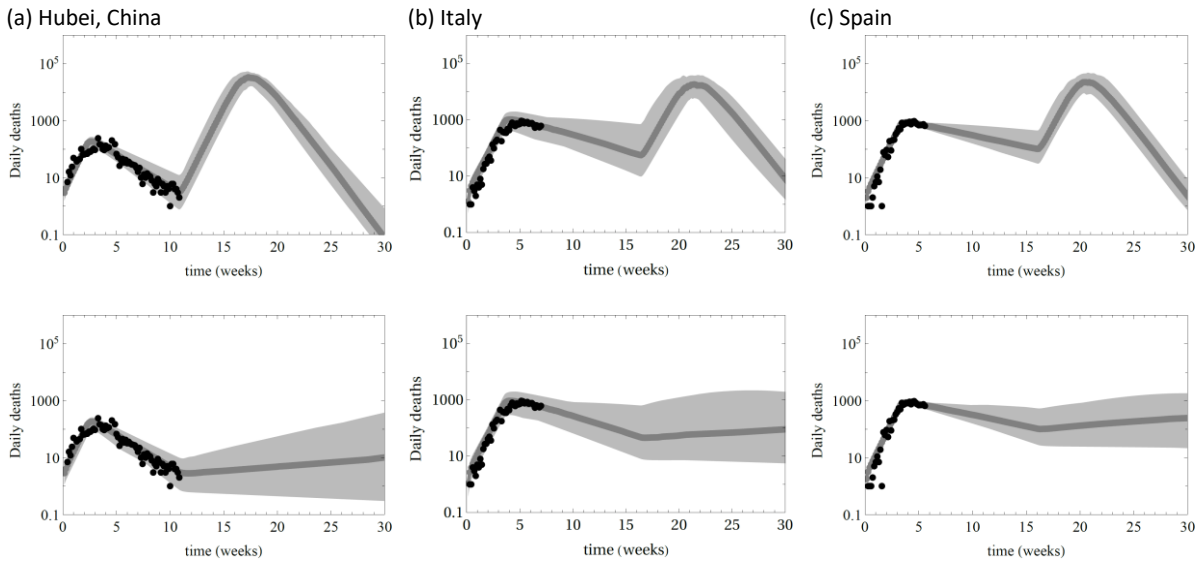

Figure 5. Prediction for the daily deaths during and after lockdowns in Hubei, Italy and Spain. (a) Upper panel: Predicted effect of a lockdown in Hubei with a 90% reduction of transmission for 60 days and return to full transmission after the lockdown. Lower panel: A 60 days lockdown as in the upper panel followed by a weaker lockdown with 75% reduction of transmission. (b) Upper panel: Predicted effect of a lockdown in Italy with transmission reduced by 80% for 90 days and return to full transmission after the lockdown. Lower panel: A 90 days lockdown as in the upper panel followed by a weaker lockdown with 70% reduction of transmission. (c) Upper panel: Predicted effect of a lockdown in Spain with transmission reduced by 80% for 90 days and return to full transmission after the lockdown. Lower panel: A 90 days lockdown as in the upper panel followed by a weaker lockdown with 72% reduction of transmission. Symbols represent data, the thick grey line gives the median of the predictions and shaded areas give the 90% confidence interval. Logarithmic scale is used in the vertical axis of all the plots.

Isolation of infected individuals

Prompt isolation of infected individuals is regarded as an effective strategy to reduce the transmission of infection and significantly reduce the size of epidemics[33]. Identification of silent carriers is difficult, however, and this makes the implementation of this strategy challenging for SARS-CoV-2.

In order to study the effect of isolating tested and untested individuals, we extended the model shown in Figure 1 to include compartments for isolated tested and untested infected individuals (see Model 2 in Methods). Interventions are parametrised by the fraction of tested (untested) infected individuals that are randomly selected to be isolated, ρ_{Qt} (ρ_{Qu}), and the rate δ at which they are isolated after testing/reporting. Isolation strategies leading to eradication of infection satisfy the condition $\mathcal{R}_{0Q} < 1$, where $\mathcal{R}_{0Q}$ is the reproduction number for Model 2 (see Eq. (6) in Methods). Given the significant reporting delays we found, isolation has to be fast after testing positive. In particular, we found that isolation after an average time of δ^{-1}= 1 day could only lead to eradication in Germany; for other countries, $\mathcal{R}_{0Q}$ remains larger than 1 for any ρ_{Qt} and ρ_{Qu}. Following this, in Figure 6 we show results for a scenario in which individuals are isolated after an average time of δ^{-1}= 1/2 days. In this case, eradication is possible in all the studied countries if *both* ρ_{Qt} and ρ_{Qu} are large enough, i.e. if enough tested and untested cases are isolated. The estimated boundaries separating the eradication region ($\mathcal{R}_{0Q} < 1$) from the epidemic region ($\mathcal{R}_{0Q} > 1$) are different for different countries but differences are statistically less marked if confidence intervals are taken into account.

Interventions that only isolate tested infected individuals (i.e. with $\rho_{Qu} = 0$) are predicted to have a minor effect on the final fraction of deaths even if they manage to isolate all tested individuals (see Figure 6(b)). Isolation of tested individuals is not effective due to the underlying transmission associated with silent carriers that are not isolated and keep $\mathcal{R}_{0Q} > 1$.

Successful eradication of infection requires isolating both tested and untested infected individuals. For instance, an intervention in the UK in which 70% of tested infected individuals were isolated in δ = 1/2 days could drastically reduce the number of deaths if 40% of untested infected were isolated at the same rate. For Germany, the percentage of untested individuals that should be isolated to ensure eradication in this situation is around 15% since testing seems to be already faster and more effective than in other countries.

Isolating infected individuals after an average time of half a day since testing is likely to be difficult to implement in practice. This time could indeed be extended by rapid identification of carriers of the virus. This highlights the importance of fast identification of infected individuals. Identification of asymptomatic cases is expected to be challenging. However, we believe that efficient tracing of the contacts of symptomatic individuals and fast testing of such contacts could facilitate the identification of asymptomatic cases.

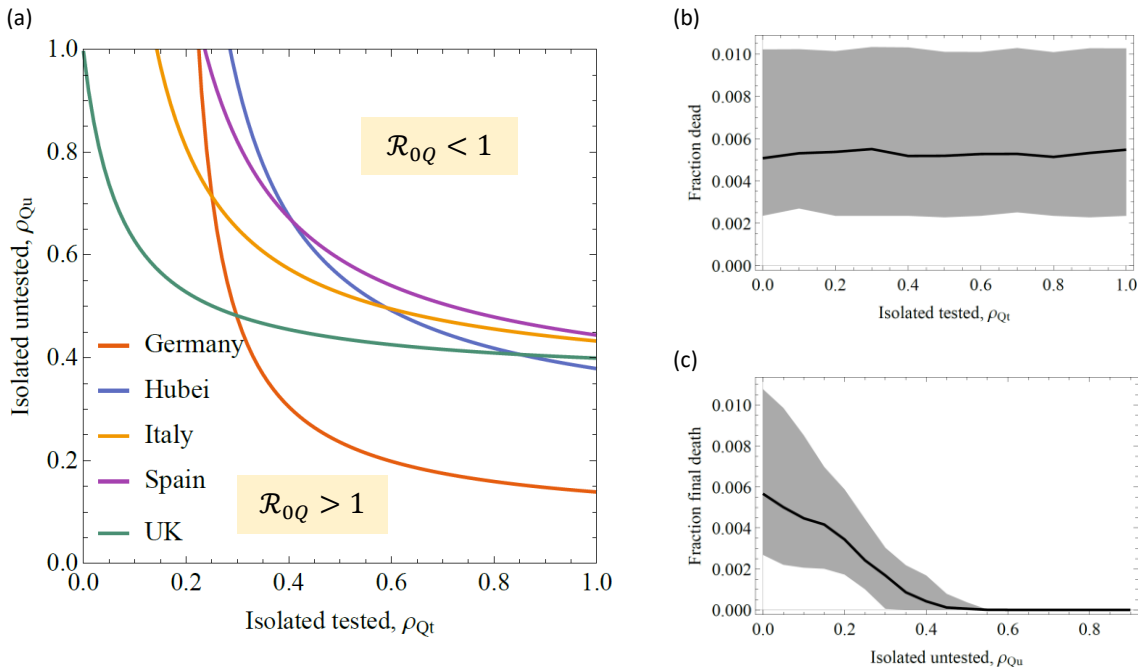

Figure 6. Interventions in which a fraction ρ_{Qt} of tested infected individuals and a fraction ρ_{Qu} of untested infected individuals are isolated in an average time of δ^{-1} = 1/2 days after testing. (a) Lines separating the regions in the space (ρ_{Qt}, ρ_{Qu}) where eradication occurs (above the line for a given country) from the regions where the epidemic grows (below the line). The lines are based on the median of $\mathcal{R}_{0Q}$; confidence intervals are not shown for clarity. (b) Interventions that only isolate tested infected individuals (i.e. cases with $\rho_{Qu} = 0$). Parameters for the outbreak in the UK are used as an example. The line gives the median of the predicted fraction of susceptible individuals that die at the end of epidemics as a function of ρ_{Qt}. (c) Interventions in which 70% of tested cases are isolated ($\rho_{Qt} = 0.7$). The line gives the median of the predicted fraction of susceptible individuals that die at the end of epidemics as a function of the proportion ρ_{Qu} of untested individuals isolated. The shaded regions in (a) and (b) give the 90% confidence interval for the predicted fraction of deaths.

Implications for policy, limitations of the models and conclusions

There are two main implications from the models which are directly relevant for policy in dealing with the outbreak. The first, involves the existence of a significant proportion of cases that are not tested and may act as silent carriers of the infection. We found that the predicted percentage of untested infected individuals may represent 50% to 80% of the cases in Germany, Hubei, Italy, Spain and the UK. The specific percentage depends on the country and we found the lowest proportion of unreported cases in Germany. Based on studies in Iceland[9] and the Diamond Princess cruise[7], we conclude that asymptomatic infected individuals are likely to be the main contribution to the untested cases in all analysed outbreaks but a fraction of cases with mild symptoms are also likely to be untested. Even when unreported cases are taken into account, we estimate that less than 8% of the population would have been exposed to SARS-CoV-2 by 09/04/2020 in the analysed outbreaks. In policy terms, our results demonstrate that the current suppression strategies being employed in Germany, Hubei, Italy, Spain and the UK will not facilitate sufficient levels of herd immunity in the population that would control and eventually eradicate the virus. This leaves the risk of re-emergence of the virus once suppression strategies are lifted, similar to second waves of infection observed in 1918 influenza epidemics[34]. We predict, however, that partial relaxation of ongoing lockdowns could keep the number of daily deaths to potentially tolerable levels.

The second implication involves the finding that unreported cases play an important role in the control of COVID-19 epidemics. In particular, unreported cases act as silent carriers and control strategies would need to account for them or be prone to the risk of re-emergence or ineffective suppression of

spread. For instance, we predict that isolation of infected individuals can have a limited impact on the suppression of spread unless it includes silent carriers that are currently missed by most countries. In line with previous suggestions[14,35], we suggest that, thorough testing combined with contact tracing[20,21], isolation of infected individuals and social distancing can be more effective to suppress SARS-CoV-2 spread than severe lockdowns. At present, however, lockdowns are probably the most effective way to delay epidemics until more effective pharmaceutical (e.g. a vaccine) or non-pharmaceutical interventions (e.g. fast and thorough testing) become feasible.

The main aim of our models is to contribute to the understanding of the epidemiological patterns of SARS-Cov-2 rather than to provide exact predictions. Hence the models should be viewed as a general guide of how the outbreak and interventions may play out rather than as an exact representation of COVID-19 epidemics[17]. We made several simplifying assumptions in the models. In particular, spatial and age heterogeneities were ignored. However, it is known that there is both spatial heterogeneity within the populations[36] (e.g. cities and rural areas) as well as differences in both susceptibility and mortality across different age and vulnerability groups[29,37]. Another limitation of our approach is that it assumes that the transition times between compartments are exponentially distributed. This memory-less assumption is usual for classical compartmental models[17]. For COVID-19, however, transitions between compartments are better described in terms of gamma distributions[3,29] and using models with memory would provide a more precise description of the dynamics[3,17,38]. In addition, our model is limited in terms of the specific ways to make interventions operational. For instance, reductions in transmission are treated at a generic level without specifying if they could be achieved by enhanced social distancing, school closure, etc. Accounting for such details would require using individual-based simulations[15].

The models assume immunity after recovery from infection but whether or not this is the case is still unclear[39]. Our model could easily be extended to account for re-infections and predictions might significantly change. There is also a growing body of evidence for pre-symptomatic transmission[40–44]. There is the potential to incorporate this in our models by varying the incubation rate parameter. Finally it is unknown what percentage of the population is actually susceptible. Here it is assumed all of the population are susceptible but it may be that a proportion are not for genetic reasons[45] or due to cross immunity[46].

Methods

Data

Data on numbers of infected and dead individuals by country or region were obtained from the Wolfram Data Repository[47]. Models were calibrated by considering data from the first available day in which the number of deaths is non-zero, as listed in Table 2. The date when lockdowns were ordered in each of the countries/regions is also given in Table 2.

Table 2. Details on the first day used to calibrate the models, number of deaths by that day, date when a lockdown was ordered in each of the countries/region analysed, and population of each country/region.

Country/Region	**First day considered**	**Number of deaths on first day considered**	**Date when lockdown ordered**[3,4]	**Population, *N***
Germany	9/3/2020	2	22/03/2020	82,114,224
Hubei	22/1/2020	17	Wuhan 23/1/2020, Hubei province 13/2/2020	58,160,000
Italy	21/2/2020	1	11/3/2020	59,359,900
Spain	3/3/2020	1	14/3/2020	46,354,321
United Kingdom (UK)	5/3/2020	1	24/3/2020	66,181,585

Models

We used extensions of the SEIR model[17] to include two types of infected individuals described by the compartments I_t and I_u (see Figure 1). The SEIR model with a single compartment for infectious individuals has already been used to describe the COVID-19 outbreak in China[19,23] and a model with two compartments for infected individuals analogous to those proposed here was used by Li et al.[11]

Model 1

The model shown in Figure 1 is run with deterministic, continuous-time dynamics given by the following differential equations:

$$\begin{aligned}
\dot{S} &= -\beta\lambda S, \\
\dot{E} &= \beta\lambda S - \zeta E, \\
\dot{I}_t &= \zeta\rho_t E - \gamma_t I_t, \\
\dot{I}_u &= \zeta(1-\rho_t)E - \gamma_u I_u, \\
\dot{R}_t &= (1-\rho_d)\gamma_t I_t, \\
\dot{R}_u &= \gamma_u I_u, \\
\dot{D} &= \gamma_t I_t.
\end{aligned} \tag{1}$$

Here,

$$\lambda = \frac{I_t + I_u}{N - D} \tag{2}$$

is the force of infection. N is the population size.

The attack rate at a given time t is defined as the fraction of individuals in a population of size N that have been exposed to the disease by that time. For the model described by Eqs. (1), we calculate the attack rate as follows:

$$\text{Attack rate}(t) = \frac{E(t) + I_t(t) + I_u(t) + R_t(t) + R_u(t) + D(t)}{N}. \tag{3}$$

The reproductive number corresponding to this model can be analytically calculated using the next generation method[18] and is given by

$$\mathcal{R}_0 = \beta\left[\frac{\rho_t}{\gamma_t} + \frac{1-\rho_t}{\gamma_u}\right]. \tag{4}$$

Model 2 – Isolation of infectious individuals

Isolation of infectious individuals is modelled by adding two more compartments, Q_t and Q_u, which contain isolated tested and untested infectious individuals, respectively (see Figure 7). The fraction of tested and untested infectious individuals are denoted as ρ_{Qt} and ρ_{Qu}, respectively. Both types of infectious individuals are assumed to become isolated with the same rate, δ. The force of infection in this model is

$$\lambda = \frac{I_t + I_u}{N - D - Q_t - Q_u}. \tag{5}$$

One can again use the next generation method[18] to obtain the following expression for the reproduction number:

$$\mathcal{R}_{0Q} = \beta \left[\frac{\rho_t}{\gamma_t(1-\rho_{Qt}) + \rho_{Qt}\,\delta} + \frac{1-\rho_t}{\gamma_u(1-\rho_{Qu}) + \rho_{Qu}\,\delta} \right]. \tag{6}$$

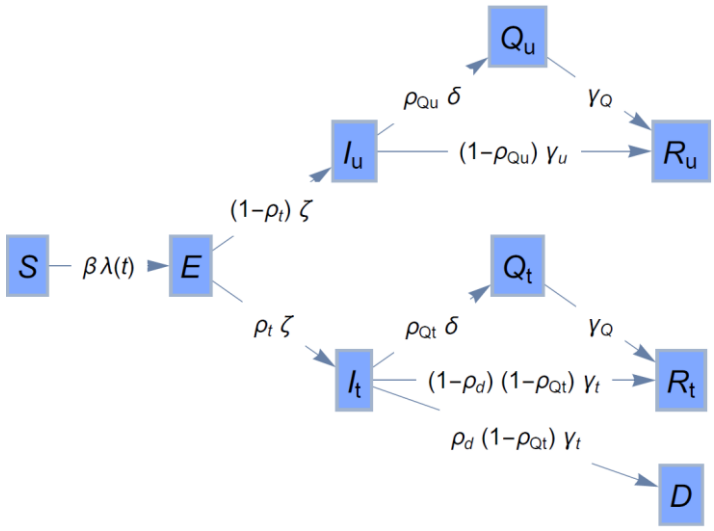

Figure 7. Extension of the basic epidemic model of Figure 1 to incorporate a compartment Q_t for isolation of a tested infectious individual and a compartment Q_u for isolation of untested infectious individuals.

Parameter estimation

We fitted the Model 1 to data. Values for the incubation rate was set to[37] $\zeta = 1/5.2$ days^{-1}. The free parameters in our fits were the rate of transmission, β, proportion of infectious that were tested, ρ_t, proportion of tested infectious that die, ρ_d, rate to recovery of tested infectious individuals, γ_t, rate of recovery of untested infectious individuals, γ_u, and initial number of exposed individuals, $E(0)$. We denote the free parameters by a vector $\boldsymbol{\theta} = \{\beta, \rho_t, \rho_d, \gamma_t, \gamma_u, E(0)\}$. The model was fit to the time series for the number of daily reported infected individuals and cumulative deaths, $\mathcal{D}_{\text{obs}} = \{i_\tau, d_\tau\}_{\tau=1}^{m}$, in a period of m days in the early stages of epidemics (here, τ is used to denote discrete time in days). In particular, we used $m = 15$ days since the first data point with a positive number of deaths (see Table 2). We used data at early stage of each outbreak to minimise the influence of suppression strategies on our parameter estimates.

Using data on deaths is important to obtain reliable descriptions of COVID-19 epidemics since data on deaths is more accurate than data on infected and recovered individuals[3,14,26,48]. In addition to deaths, we can use data on infected individuals which is represented by the tested infectious compartment, I_t, in our models.

Our fitting procedure aims at calculating the posterior probability density function for the parameters given the data, $\pi(\boldsymbol{\theta}|\mathcal{D}_{\text{obs}})$. To this end, we use an approximate Bayesian algorithm which follows the same steps as the minimum distance method proposed by Perez-Reche et al.[49]; the only difference being that here we use a likelihood function to quantify the similarity of simulated and observed time series instead of a quadratic distance.

The posterior $\pi(\boldsymbol{\theta}|\mathcal{D}_{\text{obs}})$ is approximated by the empirical distribution of a set of 500 point estimates $\widehat{\boldsymbol{\theta}}$ of the model parameters. A point estimate $\widehat{\boldsymbol{\theta}}$ is obtained by simulating $n_e = 3000$ epidemics with parameters sampled from a prior probability density $\hat{\pi}(\boldsymbol{\theta})$. In each realization, a simulation of Model 1 produces deterministic evolution functions $I_t(t)$ and $D(t)$ for the number of tested cases and cumulative deaths. The functions $I_t(t)$ and $D(t)$ are used to build a random daily time series $\mathcal{D}_{\text{sim}}(\boldsymbol{\theta}) = \left\{i_\tau^{\text{sim}}(\boldsymbol{\theta}), d_\tau^{\text{sim}}(\boldsymbol{\theta})\right\}_{\tau=1}^{m}$, where i_τ^{sim} and d_τ^{sim} are, respectively, the number of tested infected and deaths predicted at day τ. We assume that $i_\tau^{\text{sim}}(\boldsymbol{\theta}) \sim \text{Pois}(I_t(\tau))$ and $d_\tau^{\text{sim}}(\boldsymbol{\theta}) \sim \text{Pois}(D(\tau))$, i.e. the predicted number of tested infected and deaths are described as random variables obeying a Poisson distribution with mean $I_t(\tau)$ and $D(\tau)$, respectively. The point estimate $\widehat{\boldsymbol{\theta}}$ is defined as the parameter vector corresponding to the realization that gives the closest prediction, $\mathcal{D}_{\text{sim}}$, to the observations, $\mathcal{D}_{\text{obs}}$. More explicitly, the point estimate for the model parameters is given by

$$\widehat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\left\{\mathcal{L}\left(\mathcal{D}_{\text{obs}}\middle|\mathcal{D}_{\text{sim}}(\boldsymbol{\theta})\right)\right\}, \tag{7}$$

where $\mathcal{L}\left(\mathcal{D}_{\text{obs}}\middle|\mathcal{D}_{\text{sim}}(\boldsymbol{\theta})\right)$ is a loglikelihood function defined as

$$\mathcal{L}\left(\mathcal{D}_{\text{obs}}\middle|\mathcal{D}_{\text{sim}}(\boldsymbol{\theta})\right) = \sum_{\tau=1}^{m} \ln\left(\text{pois}(i_\tau|i_\tau^{\text{sim}}(\boldsymbol{\theta}))\,\text{pois}(d_\tau|d_\tau^{\text{sim}}(\boldsymbol{\theta}))\right). \tag{8}$$

Here,

$$\text{pois}(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!} \tag{9}$$

is the Poisson probability mass function.

The prior probability density is defined as the product of priors for each parameter:

$$\hat{\pi}(\boldsymbol{\theta}) = \hat{\pi}(\beta)\hat{\pi}(\rho_t)\hat{\pi}(\rho_d)\hat{\pi}(\gamma_u)\hat{\pi}(E(0)) \tag{10}$$

The priors used in our fits are summarized in Table 3. Normally distributed informative priors were used when prior information on the parameter was available. In particular, since the reproductive number has been more thoroughly studied in previous works than the transmission rate, we set an informative prior for $\mathcal{R}_0$ and derive β from Eq. (4) using the priors of $\mathcal{R}_0, \rho_t, \gamma_t$ and γ_u.

Table 3. Assumptions for the prior probability distribution of the estimated parameters. $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. $\mathcal{U}(a, b)$ denotes a uniform distribution in the interval (a, b).

Parameter	**Prior**	**Support**
Fraction tested infected, ρ_t	$\mathcal{U}(0,1)$	Uninformative for $\rho_t \in [0,1]$
Fraction of tested infected that die, ρ_d	$\mathcal{N}(0.034, 0.01^2)$	Mean set to the global estimate of WHO[50]
Recovery rate for untested infected, γ_t	$\mathcal{U}(0,0.4)$	Range assumed from manual fit exploration (contains typical values for recovery rate[19,29])
Recovery rate for untested infected, γ_u	$\mathcal{U}(0,0.4)$	Similar assumptions as for γ_t
Reproduction number, $\mathcal{R}_0$	$\mathcal{N}(4, 1^2)$	Estimates from Ref.[3]
Transmission rate, β	Derived from other parameters using Eq. (4)	
Number of exposed at the first data point, $E(0)$	$\ln E(0) \sim \mathcal{N}(8, 0.5^2)$	Refs.[11,19] and manual fit exploration

References

1. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).

2. World Health Organization. *Coronavirus disease 2019 (COVID-19) Situation Report – 81*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200410-sitrep-81-covid-19.pdf?sfvrsn=ca96eb84_2 (2020).

3. Flaxman, S., Mishra, S., Gandy, A. & al, E. *Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries*. (2020) doi:10.25561/77731.

4. Ainslie, K. E. C., Han Fu, C. W. & al, E. *Evidence of initial success for China exiting COVID-19 social distancing policy after achieving containment*. (2020) doi:10.25561/77646.

5. Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet* **395**, 931–934 (2020).

6. Collinson, S. Contagion: The Economic and Social Impacts of Covid-19 on our Region. https://www.birmingham.ac.uk/news/thebirminghambrief/items/2020/03/contagion-the-economic-and-social-Impacts-of-covid-19-on-our-region.aspx.

7. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* **25**, 2000180 (2020).
8. Day, M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *BMJ* **368**, m1165 (2020).
9. Tara John. *Iceland lab's testing suggests 50% of coronavirus cases have no symptoms*. https://edition.cnn.com/2020/03/14/health/coronavirus-asymptomatic-spread/index.html.
10. World Health Organization. *Report of the WHO-China Joint Mission on coronavirus disease 2019 (COVID-19)*. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf (2020).
11. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* eabb3221 (2020) doi:10.1126/science.abb3221.
12. Stein, R. A. Super-spreaders in infectious diseases. *Int. J. Infect. Dis.* **15**, e510–e513 (2011).
13. Clive Cookson. *Coronavirus may have infected half of UK population — Oxford study*. https://www.ft.com/content/5ff6469a-6dd8-11ea-89df-41bea055720b.
14. Lourenço, J. *et al.* Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. *medRxiv* 2020.03.24.20042291 (2020) doi:10.1101/2020.03.24.20042291.
15. Ferguson, N. M., Laydon, D., Nedjati-Gilani, G. & et al. *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. https://doi.org/10.25561/77482 (2020).
16. Anderson, R. M. & May, R. M. *Infectious diseases of humans: dynamics and control*. (Oxford University Press, 1991).
17. Keeling, M. J. & Rohani, P. *Modeling infectious diseases in humans and animals*. (Princeton UP, 2007).

18. Diekmann, O., Heesterbeek, H. & Britton, T. *Mathematical Tools for Understanding Infectious Disease Dynamics*. (Princeton University Press, 2013).
19. Wang, H. *et al.* Phase-adjusted estimation of the number of Coronavirus Disease 2019 cases in Wuhan, China. *Cell Discov.* **6**, 10 (2020).
20. Hellewell, J. *et al.* Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**, e488–e496 (2020).
21. Keeling, M. J., Hollingsworth, T. D. & Read, J. M. The Efficacy of Contact Tracing for the Containment of the 2019 Novel Coronavirus (COVID-19). *medRxiv* 2020.02.14.20023036 (2020) doi:10.1101/2020.02.14.20023036.
22. Chicchi, L., Di Patti, F., Fanelli, D., Piazza, F. & Ginelli, F. *Analysis and forecast of COVID-19 spreading*. https://www.researchgate.net/profile/Francesco_Piazza3/project/Analysis-and-forecast-of-COVID-19-spreading/attachment/5e766abfcfe4a7809f8a0ff0/AS:871607706861568@1584818879039/download/COVID19_update21.03.2020.pdf?context=ProjectUpdatesLog (2020).
23. Read, J. M., Bridgen, J. R. E., Cummings, D. A. T., Ho, A. & Jewell, C. P. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv* 2020.01.23.20018549 (2020) doi:10.1101/2020.01.23.20018549.
24. Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* **0**, (2020).
25. Arenas, A. *et al.* A mathematical model for the spatiotemporal epidemic spreading of COVID19. *medRxiv* 2020.03.21.20040022 (2020) doi:10.1101/2020.03.21.20040022.
26. Lachmann, A., Jagodnik, K. M., Giorgi, F. M. & Ray, F. Correcting under-reported COVID-19 case numbers: estimating the true scale of the pandemic. *medRxiv* 2020.03.14.20036178 (2020) doi:10.1101/2020.03.14.20036178.

27. Eckner, C. How Germany has managed to perform so many Covid-19 tests. https://www.spectator.co.uk/article/how-germany-has-managed-to-perform-so-many-covid-19-tests.
28. Robert Koch Institut. *Coronavirus Disease 2019 (COVID-19) Daily Situation Report of the Robert Koch Institute - 11/04/2020*. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-04-11-en.pdf?__blob=publicationFile.
29. Verity, R. *et al.* Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **0**, (2020).
30. Linton, N. M. *et al.* Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. doi:10.1101/2020.01.26.20018754.
31. Instituto de Salud Carlos III. *Informe sobre la situación de COVID-19 en España. Informe COVID-19 nº 21. 6 de abril de 2020*. https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/INFORMES/Informes%20COVID-19/Informe%20n%C2%BA%2021.%20Situaci%C3%B3n%20de%20COVID-19%20en%20Espa%C3%B1a%20a%206%20de%20abril%20de%202020.pdf (2020).
32. Istituto Superiore di Sanita. *Integrated surveillance of COVID-19 in Italy*. https://www.epicentro.iss.it/en/coronavirus/bollettino/Infografica_11aprile%20ENG.pdf (2020).
33. Lipsitch, M. *et al.* Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science* **300**, 1966–1970 (2003).
34. Hatchett, R. J., Mecher, C. E. & Lipsitch, M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc. Natl. Acad. Sci.* **104**, 7582 (2007).

35. World Health Organization. *WHO Director-General's opening remarks at the media briefing on COVID-19*. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—16-march-2020 (2020).
36. *How many confirmed cases are there in your area?* https://www.bbc.com/news/uk-51768274 (2020).
37. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N. Engl. J. Med.* NEJMoa2001316 (2020) doi:10.1056/NEJMoa2001316.
38. Wearing, H. J., Rohani, P. & Keeling, M. J. Appropriate Models for the Management of Infectious Diseases. *PLOS Med.* **2**, e174 (2005).
39. Lawton, G. Can you catch the coronavirus twice? We don't know yet. *New Scientist* https://institutions.newscientist.com/article/mg24532754-600-can-you-catch-the-coronavirus-twice-we-dont-know-yet/.
40. Rothe, C. *et al.* Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *N. Engl. J. Med.* **382**, 970–971 (2020).
41. Pan, X. *et al.* Asymptomatic cases in a family cluster with SARS-CoV-2 infection. *Lancet Infect. Dis.* **20**, 410–411 (2020).
42. Kimball, A. Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility — King County, Washington, March 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, (2020).
43. Bai, Y. *et al.* Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* (2020) doi:10.1001/jama.2020.2565.
44. Wei, W. E. Presymptomatic Transmission of SARS-CoV-2 — Singapore, January 23–March 16, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, (2020).
45. Tanigawa, Y. & Rivas, M. Initial Review and Analysis of COVID-19 Host Genetics and Associated Phenotypes. *Preprints* 2020030356 doi:10.20944/preprints202003.0356.v1.

46. King, A. *Possible Biological Explanations for Kids' Escape from COVID-19*. https://www.the-scientist.com/news-opinion/possible-biological-explanations-for-kids-escape-from-covid-19-67273 (2020).
47. Wolfram Research, Inc. Epidemic Data for Novel Coronavirus COVID-19" from the Wolfram Data Repository (2020). https://www.wolframcloud.com/obj/resourcesystem/published/DataRepository/resources/Epidemic-Data-for-Novel-Coronavirus-COVID-19.
48. Jombart, T. *et al.* Inferring the number of COVID-19 cases from recently reported deaths. *medRxiv* (2020) doi:10.1101/2020.03.10.20033761.
49. Pérez-Reche, F. J., Neri, F. M., Taraskin, S. N. & Gilligan, C. A. Prediction of invasion from the early stage of an epidemic. *J. R. Soc. Interface* **9**, (2012).
50. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 3 March 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---3-march-2020.