Modelling of the second (and subsequent) waves of the coronavirus epidemic. Spain and Germany as case studies.

Francisco de Castro

Agri-Food and Biosciences Institute. 18a Newforge Lane, Belfast, Northern Ireland. UK, BT9 5PX

Contact: Francisco.decastro@afbini.gov.uk

Abstract

The first wave of the coronavirus pandemic is waning in many countries. Some of them are starting to lift the confinement measures adopted to control it, but there is considerable uncertainty about if it is too soon and it may cause a second wave of the epidemic. To explore this issue, I fitted a SEIR model with time-dependent transmission and mortality rates to data from Spain and Germany as contrasting case studies. The model reached an excellent fit to the data. I then simulated the post-confinement epidemic under several scenarios. The model shows that (in the absence of a vaccine) a second wave is likely inevitable and will arrive soon, and that a strategy of adaptive confinement may be effective to control it. The model also shows that just a few days delay in starting the confinement may have caused and excess of thousands of deaths in Spain.

Keywords: coronavirus, covid-19, SEIR, second wave, confinement, infectious disease.

Introduction

The epidemic of coronavirus (SARS-CoV-2) has spread all over the world since January 2020. Most governments have taken measures in trying to limit its spread and to reduce its severity, with greater or lesser success. The most common, non-pharmaceutical measures are directed to reduce the rate of contacts between people, thus reducing the transmission rate. Among other possibilities, these measures may entail forbidding large assemblies of people, closing of schools and universities, closing of businesses, home isolation of confirmed or suspected cases and tracing their contacts, etc. In what follows I refer to all these measures collectively as confinement.

The rigor of these measures and their timing vary among countries, or even between regions. Some countries adopted measures very early in their epidemic while others took more time. At present (June 2020) in many countries the epidemic has passed its peak and is quickly decreasing, which is leading many governments to start lifting the confinement. Although this is done with a phased, gradual approach, there is still considerable uncertainty about the consequences, and if it may cause a resurgence of the virus and a second wave of cases.

It is difficult to know exactly the effect that the end of confinement will have on the transmission rate and, therefore, on the number of infections. For this reason, predictive models fitted to specific situations are useful as tools to predict the probability of a second (and subsequent) waves, their potential amplitude and to test strategies for their control before they arrive.

Here, I fitted a SEIR model with time-dependent transmission and mortality rates to data from Spain and Germany as case studies. With the fitted model, I simulated the epidemic in both countries post-confinement to show the possibility of a second wave and its potential amplitude and timing under different scenarios. Additionally, I used the model to explore the option of adaptive confinement to control the second and subsequent waves.

Methods

All the calculations were done with MATLAB (2018).

SEIR model

I used a standard SEIR model modified with time-dependent transmission and mortality rates. SEIR models classify the population into four classes: susceptible (S), exposed (E, infected but not yet infectious), infected (and infectious, I), and recovered (R). The total population is $N= S+E+I+R$. The model is represented by the equations:

$$\frac{dS}{dt} = -\frac{\beta_t SI}{N} \qquad \text{[eq 1]}$$

$$\frac{dE}{dt} = \frac{\beta_t SI}{N} - \sigma E \qquad \text{[eq 2]}$$

$$\frac{dI}{dt} = \sigma E - (\gamma + \mu_t)I \quad \text{[eq 3]}$$

$$\frac{dR}{dt} = \gamma I \quad \text{[eq 4]}$$

where: β_t is time-dependent transmission rate (eq. 7), σ is the loss of latency rate, γ is recovery rate and μ_t is time-dependent, disease-induced mortality rate (eq. 8). This formulation assumes a complete mixing of the population and no population dynamics (i.e. births and natural mortality are considered negligible within the time frame of the epidemic). It also assumes that surviving the disease confers permanent immunity.

Two additional equations are included to keep track of the accumulated number of deaths (D) and accumulated infections (denoted AI, to avoid confusion with currently infected I), which are later used to fit the model to data. These are:

$$\frac{dAI}{dt} = \sigma E \quad \text{[eq 5]}$$

$$\frac{dD}{dt} = \mu_t I \quad \text{[eq 6]}$$

Transmission rate β_t and mortality rate μ_t are not constants but a function of time, to capture the effect of measures taken to control the epidemic. Regarding transmission, this refers to the confinement of a large part of the population and the quarantine of detected cases. Regarding mortality rate, the time-dependency was introduced because a variable mortality rate achieved a better fit, and, with constant mortality, the deviation between model and data suggested a decrease of mortality rate after confinement measures were put in place. Time-dependent transmission and mortality rates are modelled with sigmoid functions, as:

$$\beta_t = \begin{cases} \beta & if\ t \leq t_i \\ \beta\left(1 - \frac{\delta_d(t-t_i)^2}{1+\delta_d(t-t_i)^2}\right) & if\ t_i < t < t_e \\ \beta_{min} + (\beta - \beta_{min})\left(\frac{\delta_i(t-t_e)^2}{1+\delta_i(t-t_e)^2}\right) & if\ t > t_e \end{cases} \quad \text{[eq 7]}$$

$$\mu_t = \begin{cases} \mu & if\ t \leq t_i \\ \mu\left(1 - \frac{\delta_m(t-t_i)^2}{1+\delta_m(t-t_i)^2}\right) & if\ t > t_i \end{cases} \quad \text{[eq 8]}$$

where β is the initial transmission rate, without any measures to limit contact between people, and μ is the same for the mortality rate. t is time (days) from the first detected case, and t_i is the initial day of confinement minus 2, and t_e is the end of confinement. δ_d is the rate of *decrease* of transmission because of the control measures, while δ_m is the same for mortality. β_{min} is the value of β_t at time t_e, and δ_i is the rate of *increase* of transmission after t_e (see fig. 3 for an example of β_t).

Fitting of the model

To fit the model, I used data on daily cases and daily deaths from the European Centre for Disease Control (ECDC 2020) and data of recovered from the John Hopkins University github (CSSE 2020), because the ECDC does not include recovery data. Accumulated time series were generated from daily values. Only values up to the end of the confinement period were used in the fitting, so only the first two lines of eq. 7 are pertinent in this case. At the time of writing there is not yet enough data after confinement to find a good fit for parameter δ_i.

The date of start of confinement is not always easy to define, as measures may have been taken incrementally along several days or weeks or may have been adopted at different times in different regions of a country. I relied on published news about the confinement to choose t_i and t_e for Spain and Germany. For Spain: t_i was the 14th of March and t_e the 10th of April. For Germany t_i was the 23rd of March and t_e the 20th of April.

When comparing the model to data we must consider the question of detectability. Most of the cases are undetected, either because they are mild or even asymptomatic or because they are not tested for the virus for different reasons. In the model all cases are known exactly, so to make a meaningful comparison with the data the model values were multiplied by a probability of detection, p_{det}. Recent extensive serological studies in Spain have shown a prevalence of 5%, or around 2.3 million people (Instituto de Salud Carlos III, 2020). This is approximately ten times the number of diagnosed cases in Spain, which suggests a value of 0.1 for p_{det}. I used the same value for Germany. Deaths, on the other hand, are assumed to be known exactly. If more deaths were due to the epidemic, the parameter

estimates could change, particularly mortality rate, but until a more accurate estimation of deaths is available, I rely on data as published.

The fitting is done by least squares, combining four time-series: accumulated infected, accumulated deaths, accumulated recovered, and currently infected. The quantity to be minimized is, then:

$$SS = \sum\left(p_{det}\, AI_m - AI_d\right)^2 + \sum\left(D_m - D_d\right)^2 + \sum\left(p_{det}\, R_m - R_d\right)^2 + \sum\left(p_{det}\, I_m - I_d\right)^2 \qquad \text{[eq 9]}$$

The minimization was done with MATLAB function *lsqcurvefit*. In eq. 9 the subscript *d* stands for data, while subscript *m* stands for model. *AI* is accumulated infected (eq. 5), *D* is accumulated deaths (eq. 6), *R* is accumulated recovered (eq. 4) and *I* are currently infected (eq. 3). The number of current infections is typically not reported, so I calculated it as:

$$I_d(t) = I_d(t-1) + NI(t) - ND(t) - NR(t) \qquad \text{[eq 10]}$$

where: $I_d(t)$ are observed current infections on day *t*, $I_d(t-1)$ are the current infections the day before and $NI(t)$, $ND(t)$, and $NR(t)$ are the new infections, new deaths, and new recoveries on day *t*, respectively.

To generate a model prediction with a specified set of parameters the model equations are integrated using MATLAB function *ode45*. This solver uses an explicit Runge-Kutta formula to do the integration. The initial values for each population class were: as susceptible, the total population of the country (World Bank 2019), no individuals exposed, one individual infected, no dead, and no recovered. The initial values for the parameters and their bounds during fitting are shown in Table 1.

After fitting the model, the basic reproductive number (R_0) is estimated (Diekmann et al. 1990, Blackwood & Childs 2018) as:

$$R_0 = \frac{r}{\gamma + \mu}$$

where *r* is the exponential rate of increase of cases at the beginning of the epidemic, which, in turn, is estimated by fitting an exponential function using only the first 50 days of data.

Once a best-fitting model is identified, confidence intervals for all the parameters (and R_0) were estimated by bootstrap, assuming a Poisson error structure and using 1000 realizations (following Chowell 2017). I fitted the model to data from two countries: Spain and Germany because they have shown contrasting epidemic trajectories, which can shed light on

the benefits or pitfalls of different control strategies. The parameter estimates and their confidence intervals for both countries are shown in Table 2.

Results and simulations

The model reached a very good fit for both countries and for all the components of the population: total infections, current infections, deaths, and recoveries (fig. 1). This good fit lends confidence in using the model as a predictive tool and for the testing of hypothesis or the consequences of control strategies. I used the model to simulate the evolution of the epidemic post-confinement in several scenarios, and the feasibility of an adaptive confinement strategy.

The second wave

Many countries, including Spain and Germany, are already lifting the confinement measures, typically in a slow, phased approach. However, there is considerable unease and uncertainty about if the lifting of confinement will provoke a resurgence of the virus and a second wave of cases, and its potential magnitude. It is very difficult to know with any exactitude the effect that the lifting of each measure will have on the transmission rate and, therefore, on the number of infections. However, it seems safe to assume that any increase of transmission rate will be slower than the decrease that happened as a consequence of the confinement. Or, in terms of the model, that $\delta_i \ll \delta_d$. I used the Spain and Germany models to simulate the possible evolution of the epidemic post-confinement in three scenarios, each characterized by a value of the rate of increase of transmission δ_i. The three values are expressed as a fraction of the rate of *decrease* of transmission δ_d. Specifically I used: $\delta_i = \delta_d/10$, $\delta_i = \delta_d/50$, and $\delta_i = \delta_d/100$ (fig. 2). It is important to keep in mind that these simulations assume that no further confinement measures would be taken in any case.

The results suggest that there will be a second wave in both countries and that their magnitude would dwarf the current wave. In fact, they would be as large as the first wave would have been if no measures at all had been taken. At the peak of the second wave the number of current infections would be between 2 and 3 million and the number of deaths would be in the hundreds of thousands, for both Spain and Germany. The timing of the

second wave, on the other hand, would be different in both countries. Spain, in the worst-case scenario, would have the peak of the second wave in mid-June, and in the best-case scenario in late July. Germany, on the other hand, would have a second-wave peak between early July and early October, depending on the scenario.

Adaptive confinement

The results of the simulation of the second wave strongly suggest that, in the absence of a vaccine or an extremely efficient track-and-trace system, further periods of confinement will be necessary in the near future. I used the Spain-fitted model to simulate a situation in which the confinement is triggered when the number of currently infected exceeds a given threshold, and it is gradually lifted when the number drops back below it (Ferguson et al. 2020). I used the number of currently infected as trigger variable and simulated two scenarios, with a threshold of 3500 and 1000, respectively. The former is the number of infected there were in Spain two days before the "estate of alert" (i.e. confinement) was declared, on the 14th of March. The later was reached approximately four days before, on the 8th of March (fig. 4).

The results show that the epidemic can be controlled by using this type of adaptive strategy, imposing or easing a lockdown depending on the state of the epidemic. This adaptive model shows three waves of decreasing amplitude (including the present one) in the first two years, with the thresholds mentioned above. Crucially, the simulations also show large differences in the number of infections and deaths depending on the threshold used to trigger the confinement. This fact, given the exponential growth of the number of cases at the start of the epidemic, stresses the importance of taking measures early in the epidemic.

Discussion

I fitted a SEIR model with time-variable transmission and mortality rates to the COVID19 epidemic data of Spain and Germany, as case studies. The model fitted the data extremely well in both cases, which allows to use it with confidence to predict the future evolution of the epidemic or the outcomes of different control strategies. The fact that the transmission rate, particularly, is not constant but time-dependent allows the model to

respond to the confinement of part of the population, which is the most commonly used strategy to control or reduce the epidemic.

The timing of control measures plays an essential role in the model, but it is not straightforward to define this timing in a way that matches reality. In the model, the confinement starts in a specific day for the whole population, but this is typically not the case. In some countries, control measures were taken incrementally as the epidemic progressed, not all in a single day. Furthermore, some measures may be applied at different times in different regions of a country, like it happened in Italy, which does not fit the assumption of homogenous mixing implicit in the model. However, despite this difficulty, the model retains its capacity to fit the observations quite well, thus making it a good tool for prediction and testing.

The estimates of the model parameters show differences between Spain and Germany which could be related to their management of the epidemic (Table 2). Given that Germany's population is almost twice that of Spain, a stronger epidemic would have been expected there. However, the total cases, deaths and peak of infections were considerably lower. Although the initial transmission rate (β) was higher in Germany, the decrease of transmission with confinement was also much faster, and this, coupled with a higher recovery rate and lower mortality, led to a much-restricted epidemic. The basic reproductive number (R_0) also offers clues on the severity of the epidemic. It was twice as high in Spain, and higher in both countries than most estimates from other studies (Liu et al. 2020), but still within the range estimated by other authors (Shen et al. 2020, Sanche el at. 2020). Further, by the end of confinement the case fatality rate was 0.5% in Germany and 1.22% in Spain; while the fatalities per hundred thousand people were 11.2 and 64.1, respectively. It would be difficult to attribute any of these differences to specific circumstances. Undoubtedly, they are the combined outcome of multiple factors, but taken together suggest that the management of the epidemic was more effective in Germany.

The simulations of the post-confinement period clearly show that a second wave is probably inevitable, no matter how gradually or carefully the confinement is lifted. The reason is the large pool of infected (mostly undetected) still present among the population when the lifting of confinement measures is started. The model estimate of the number of infected at the end of the confinement was 775×10^3 in Spain, and 149×10^3 in Germany. This is undoubtedly much larger than the initial, imported cases that started the epidemic in either country, which makes a second

wave very likely and with the potential of being larger than the first. The difference between the two countries in number of infected when confinement was lifted is also the reason for the differences in the timing of the second-wave peak, which the model predicts will happen much earlier in Spain. In the best-case scenario this difference is approximately 2.5 months. To delay a second wave as much as possible in Spain it would be necessary to prolong the confinement measures until the pool of infected has reduced considerably below its present level. However, it would be probably unfeasible to prolong it until the complete disappearance of the disease. The model predicts that the last case in Spain (maintaining confinement) would occur in late November and in Germany by the end of August. A good track-and-trace system may help to reduce the amplitude of the second wave, but it would require a very high efficacy.

Given that a second wave is likely inevitable and that is not possible to maintain the confinement measures until the disease disappears completely, other strategies will be required, like the adaptive policy suggested by Fergusson et al. (2020). The simulations of this strategy with the Spain-fitted model (fig. 4) show that it is feasible to control the epidemic by alternating confinement periods with others of non-confinement. There is a trade-off to consider, however, in using this strategy; setting the threshold higher makes the periods of confinement slightly shorter, but the epidemic waves will be larger and, consequently, there will be more fatalities. Another factor to consider is what to choose as indicator. I used here the number of active infections, while Fergusson et al. (2020) suggested the case incidence in ICU patients. This question is open to debate, but the best indicator (which may be different in each country) would be one that can be measured as accurately and with as little delay as possible.

The model predicts that there would be three waves of diminishing amplitude in two years, including the current one, with the thresholds chosen. The reason for the decrease in amplitude is the increasing percentage of the population that has had the disease, thus acquiring immunity. Permanent immunity after having the disease is an assumption of the model, but if it happens and how long it lasts is still an open question (see a review of studies in Flodgren 2020).

The simulations of adaptive confinement make it clear the importance of taking early measures against the epidemic. I simulated two scenarios, one with a threshold for confinement of 3500 current infections and another

with a threshold of 1000. The first is approximately the number of infected that there were in Spain two days before the confinement was officially declared by the government on the 14th of March, a scenario chosen to represent the actual events. The lower threshold was reached about four days before that, on the 8th of March. The differences between the two scenarios are striking (fig. 4). The peak of the first wave is approximately 108×10^3 infected with the highest threshold, but only 33×10^3 with the lower one. More importantly, the model-predicted fatalities reach approximately thirty thousand with the higher threshold (very close to the actual official number) but only *nine thousand* with the lower one; a third of the number of deaths. Thus, the delaying of the start of confinement by a mere four days may have caused a two-thirds increase in fatalities, or twenty thousand more deaths. An earlier confinement would have reduced this number even further. At the start of an epidemic the number of cases rises exponentially, meaning that every single day is important to control the epidemic.

The predictions of the model clearly show that some control strategies and implementations are better that others, even if they are not perfect. They also show that strong measures must be taken very early in the epidemic and very aggressively; that the argument that some measures "cannot be adopted too early" is unjustifiable if the goal is to control the epidemic and save lives. Hopefully, the lessons of modelling and analysis exercises like this and others will be learnt and considered when the next wave arrives and, especially when the next pandemic emerges.

References and data sources

MATLAB. 2018. version 9.5.0 (R2018b). Natick, Massachusetts: The MathWorks Inc.

ECDC. 2020. European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide

CSSE. 2020. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Github repository: https://github.com/CSSEGISandData/COVID-19

Instituto de Salud Carlos III. 2020. Estudio Nacional de Sero-Epidemiología de la Infección por SARS-COV-2 en España. Ministerio de Ciencia e

Innovación.
https://www.ciencia.gob.es/stfls/MICINN/Ministerio/FICHEROS/ENECOVID_Informe_preliminar_cierre_primera_ronda_13Mayo2020.pdf

World Bank. 2019. Population Estimates and Projections, World Bank Group. https://data.worldbank.org/indicator/SP.POP.TOTL

Diekmann, O., Heesterbeek, JAP., Metz JAJ. 1990. On the definition and the computation of the basic reproduction ratio Ro in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology 28:365-382.

Blackwood, JC, Childs, LM. 2018. An introduction to compartmental modeling for the budding infectious disease modeler. Letters in Biomathematics, 5:1, 195-221, DOI: 10.1080/23737867.2018.1509026

Chowell, G. 2017. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. Infectious Disease Modelling, 2: 379-378.

Ferguson NM, et al. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. DOI: https://doi.org/10.25561/77482

Liu Y., Gayle AA., Wilder-Smith A. Rocklöv, J. 2020. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of Travel Medicine. 1–4.

Shen M., Peng Z., Xiao Y., Zhang L. 2020. Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China. bioRxiv. doi:https://doi.org/10.1101/2020.01.23.916726

Sanche S., Lin YT., Xu C., Romero-Severson E., Hengartner N., Ke R. 2020. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerging Infectious Diseases. https://doi.org/10.3201/eid2607.200282

Flodgren GM. 2020. Immunity afterSARS-CoV-2infection, 1stupdate –a rapid review. Oslo: Norwegian Institute of Public Health.

Table 1. Initial values and bounds of the free parameters of the model. The values of σ and γ are expressed as the inverse of the latency period and infective period, respectively. The value of β is calculated from γ and μ assuming an R_0 of 3.

Parameter	Initial value	Lower bound	Upper bound
β	3 (γ+μ)	0	5
σ	1/3.5	1/10	1
γ	1/15	1/60	1/4
μ	0.01	0	0.25
δ_d	$5\ 10^{-4}$	0	0.25
δ_m	$5\ 10^{-4}$	0	0.25

Table 2. Parameter estimates and 95% confidence intervals for Spain and Germany. For a description of the parameters, see methods. R_0, while not a model parameter, is shown for its indicative value.

	Spain			Germany		
Parameter	Estimate	Lower CI	Upper CI	Estimate	Lower CI	Upper CI
β	0.6101	0.6023	0.6179	1.0779	1.0736	1.0819
σ	0.2578	0.2521	0.2637	0.0863	0.0859	0.0867
γ	0.0334	0.0333	0.0335	0.0711	0.0707	0.0715
μ	$15\ 10^{-4}$	$14\ 10^{-4}$	$16\ 10^{-4}$	$3.74\ 10^{-4}$	$3.63\ 10^{-4}$	$4.14\ 10^{-4}$
δ_d	0.0172	0.0169	0.0175	0.0717	0.0709	0.0725
δ_m	0.0016	0.00152	0.00176	$1.17\ 10^{-8}$	$2.36\ 10^{-14}$	$9.78\ 10^{-5}$
R_0	6.27	6.25	6.30	3.16	3.13	3.19

Figures

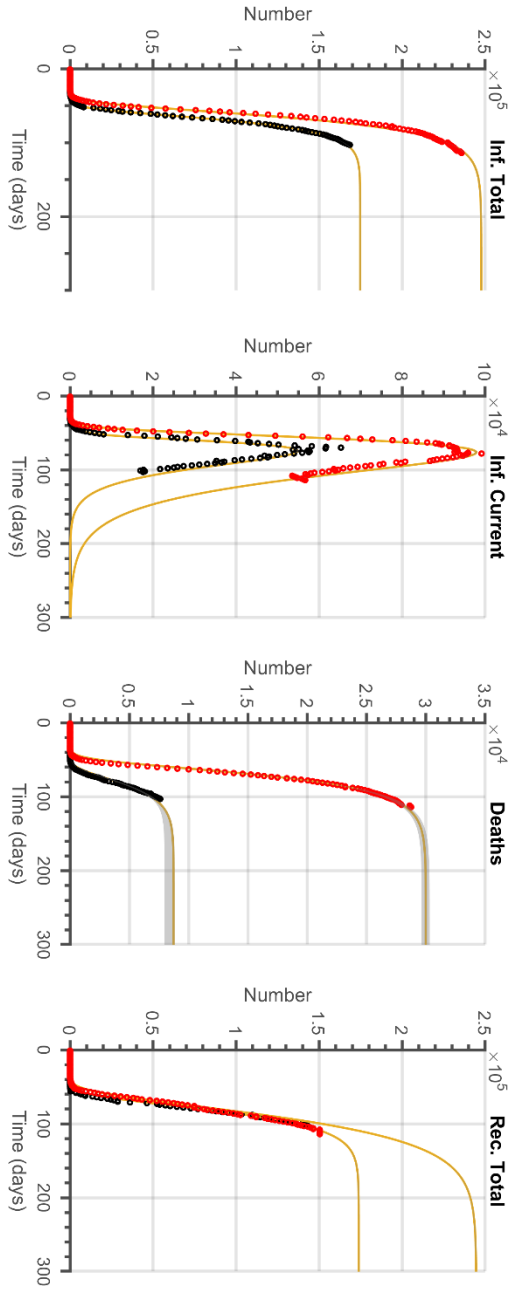

Fig. 1. Fit and forecast of the model for Spain and Germany. From left to right the panels represent: accumulated infected, currently infected, deaths, and recovered. Red dots represent Spain and black dots represent Germany. The yellow line is the best-fitting model, forecasted 300 days after the first infection (in each country). Grey areas are 95% confidence intervals for the prediction from a bootstrap procedure (see methods). Confidence intervals are not visible for the most part because they are very narrow.

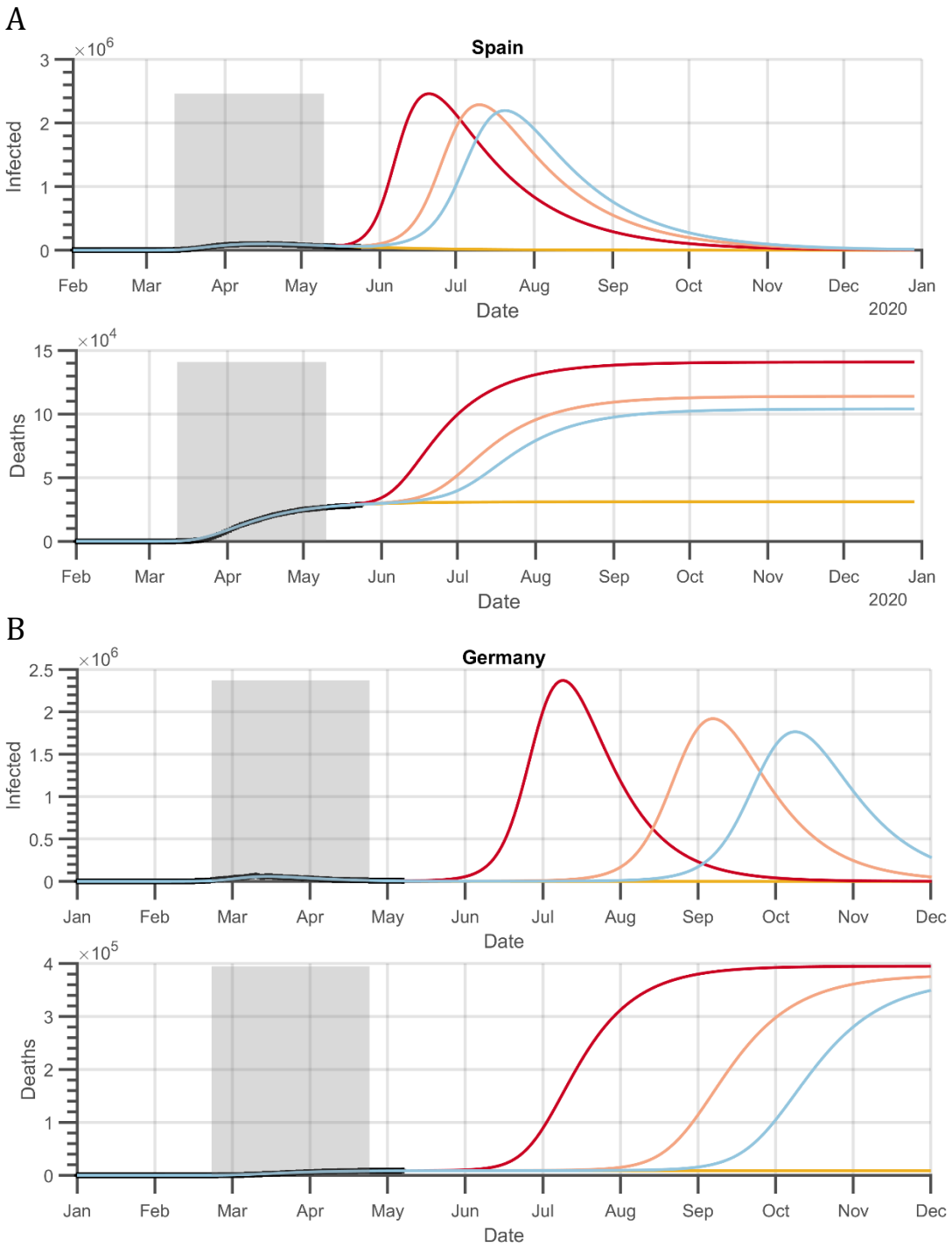

Fig. 2. Simulation of the easement of confinement. Panel A shows results for Spain and panel B for Germany. The grey area is the period of confinement. In each panel, the top figure shows current infections and the lower one number of deaths. Each line represents a different speed of confinement release, expressed as a fraction of the speed of *decrease* of transmission during confinement (δ_t). Red line: 1/10, pink line: 1/50, blue line: 1/100. The yellow line shows model prediction maintaining confinement.

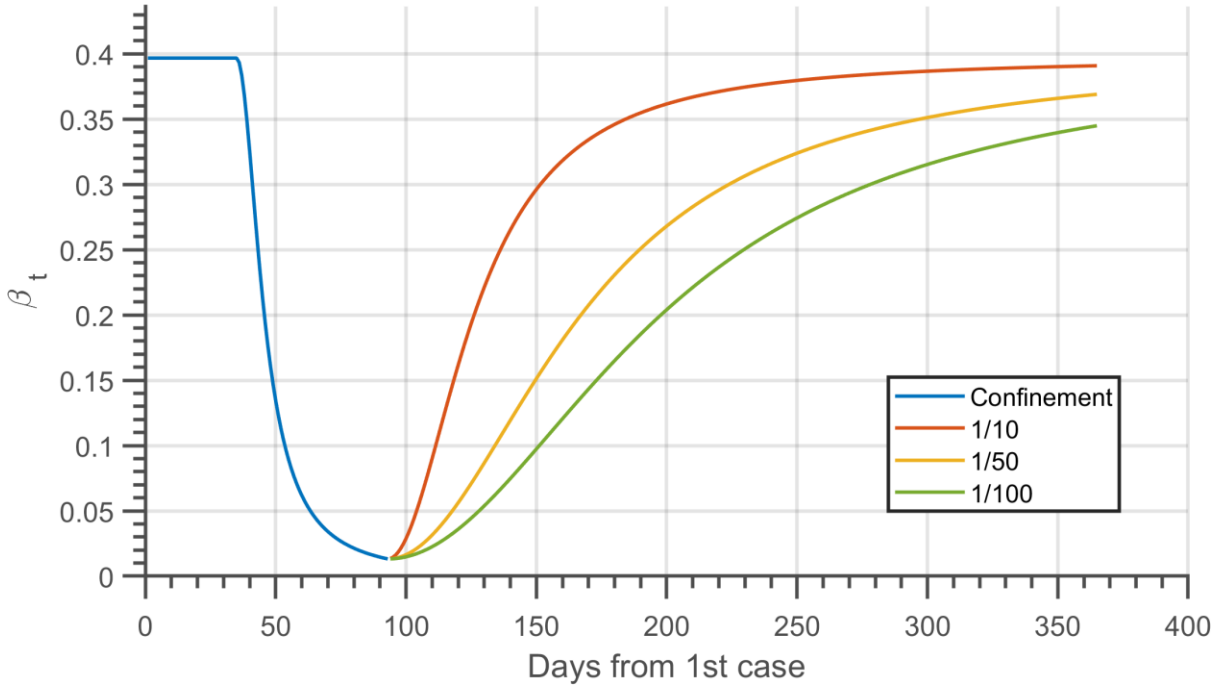

Fig. 3. Evolution of time-dependent transmission rate (β_t) for Spain before, during, and after confinement. The blue line shows the initial phase with constant β (up to day 40) and the decline of transmission with confinement, until day 93. The rate of decrease (δ_d) is estimated from model fitting. After day 93 each line represents a different speed of confinement release, as a fraction of the speed of decrease. Red: 1/10, Orange: 1/50 and Green: 1/100.

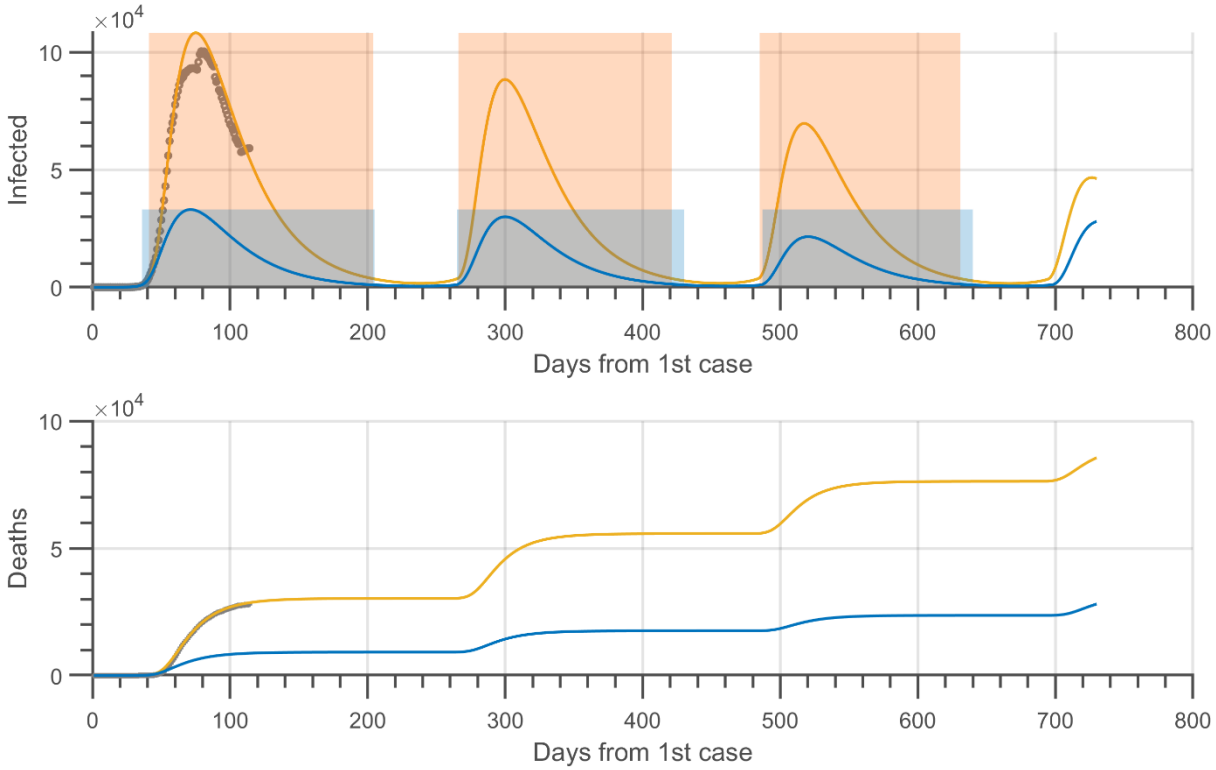

Fig. 4. Simulation of adaptive confinement with the Spain-fitted model. The confinement is triggered when the number of active infections exceeds a specified threshold and is *gradually* lifted when the number drops back below it. The threshold was 3500 for the yellow line and 1000 for the blue line. The top panel shows active infections and the lower panel shows deaths. The semi-transparent areas in the top panel mark the periods of confinement (above threshold), in orange for the higher threshold and blue for the lower one.