medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Monitoring and forecasting the SARS-Covid-19 pandemic in France

Jürgen Reingruber[1*], Andrea Papale[1], Stéphane Ruckly[2], Jean-Francois Timsit[2], David Holcman[1*]
[1]Group of data modeling and computational biology,
IBENS-PSL, Ecole Normale Superieure, INSERM U1024, Paris, France.
[2] AP-HP, Bichat Hospital, Medical and infectious diseases ICU (MI2), F-75018 Paris, France
University of Paris, IAME, INSERM, F-75018 Paris, France
INSERM IAME, U1137, Team DesCID, Paris, France

July 28, 2021

Abstract

Over the past year, many countries have resorted multiple times to drastic social restrictions to prevent saturation of their health care system, and to regain control over an otherwise exponentially increasing SARS-Covid-19 pandemic evolution. With the advent of data-sharing, computational approaches have gained a key role in evaluating future scenarios. To design optimal social, hospitalization and economical strategies that guarantee control over the pandemic progression, we developed a data-driven modelling framework with the aim to provide reliable near future predictions under constantly evolving social and pandemic conditions. The framework is flexible enough to be used at a single hospital, regional or national level. We used a variety of data such as social, serological, testing and clinical data to compute the infection dynamics and the hospital workload for France. We developed inference methods to calibrate model parameters from observed hospitalization statistics over adjustable time periods. We applied our model to study the age stratified pandemic evolution inside and outside hospitals until February 2021, and the competition between vaccinations and the novel delta variant. We obtained several predictions about hidden pandemic properties such as fractions of infected, infection hospitality and infection fatality ratios. We show that reproduction numbers and herd immunity levels are not universal but strongly depend on the underlying social dynamics. We find that with normal social interactions the present vaccination status and rate is not sufficient to prevent a new pandemic wave driven by the delta variant.

1 Introduction

The fast spreading pandemic of SARS-Covid-19 has destabilized the world during the past year, forcing most countries into alternating periods of confinement and deconfinement. These sequential measures have attenuated the disease progression that otherwise would have increased exponentially, and thus prevented health care systems from getting destabilized [1, 2, 3, 4]. For example, the age-stratified SARS-Covid-19 hospitalization data for France during the past year exhibits how the national lockdowns at March 18, 2020 and October 30, 2020 have stopped the exponential pandemic growth (Fig. 1). With the availability of vaccines, the pandemic seems under control and the economical and social life begin to return to normal after the population became progressively vaccinated. However, it remains the serious threat that current vaccines are not efficient to handle the constant

*Corresponding authors: jurgen.reingruber@ens.psl.eu and david.holcman@ens.psl.eu

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

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

emergence of new mutations, and the pandemic might get out of control again [5]. When no vaccines are available, for example at the beginning of a new pandemic, constant monitoring and modeling are essential to predict and curtail the pandemic growth together with its consequences on the health care system [6].

The past year has shown that surveillance, testing and social restrictions were successful to curb the viral spread and prevented hospitals from becoming overloaded. However, these severe restrictions not only came at a heavy economical cost [7], they also heavily disturbed social, school and academic life. Designing efficient social measures that not only curb the pandemic and exonerate the health care system, but also minimize social and economic impacts [8], remains challenging. In presence of a fast exponential pandemic growth characterized by a large reproduction number $R_0 \sim 3$, non-pharmaceutical inventions have to be taken as soon as possible because a small delay of only a few days can significantly affect the hospital load [9].

A large variety of models have been recently developed to study the Covid-19 pandemic: continuous time SIR type compartmental models based on differential equations [10, 11, 12, 13, 14, 15, 16]; SIR type models combined with Bayesian inference, SMC or MCMC methods [9, 17, 18, 19, 4, 20, 21, 22, 23, 24]; stochastic frameworks and agent based models [3, 25, 26, 2, 27, 1, 28]; or deterministic discrete time compartmental models [29, 30]. These models were applied to analyse the severity of the disease [18, 2], to estimate the impact of social measures [9, 1, 3, 25, 1, 31, 4], to study the early dynamics of the viral spread [19, 17, 29, 10], and to predict the effect of deconfinement measures [9, 3, 25, 17, 29, 32, 30].

In this work we not only devise a new model, but we also develop an automated forecasting framework based on web-data acquisition to constantly update and adjust predictions of the pandemic and hospitalisation evolution (Fig. 2). The predictions are generated in two steps: in the first step, we compute the number of new hospitalisations from social interaction matrices, while in the second step, we predict the hospital workload from the influx of new hospitalisations. We calibrate the model with data from various sources (Fig. 2A): web accessible hospitalisation statistics (Fig. 1), clinical data from the Bichat hospital in Paris (Fig. 6), serology and positive testings (Fig. S2A), and social interaction matrices (Eq. 6 in the SI). We either directly used statistics inferred from data as input parameters, or we implemented fitting procedures to extract parameter values. Although our framework can be calibrated to estimate pandemic consequences from single hospital up to national level, we present here a coarse-grained analysis at a national level. We calibrated the model to reproduce the pandemic evolution in France with multiple social interventions until February 2021, and we infer pandemic consequences inside and outside hospitals.

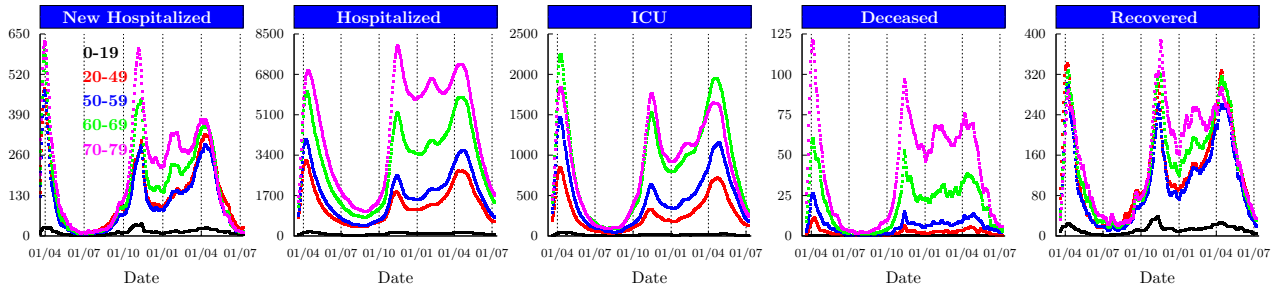

Figure 1: **Age-stratified SARS-Covid-19 hospitalisation data for France**. Daily number of patients newly hospitalised, hospitalised, in Intensive Care Units (ICU), deceased in hospitals, recovered patients that were release from hospitals. Data from March 18, 2020 up to July 12, 2021. Data sources are [33, 34].

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2 Materials and Methods

2.1 Web-available epidemiological and testing data, and hospital specific clinical data

To calibrate model parameters we combine data from several sources, as described below.

Age-stratified hospitalisation data from the French governmental website

Daily updated and age-stratified hospitalisation data is provided by the French Government (www.data.gouv.fr) since March 18th 2020. We used the databases *donnees-hospitalieres-covid19* and *donnees-hospitalieres-classe-age-covid19* to obtain the daily number of patients hospitalised $H(n)$, in intensive care (ICU), deceased in hospitals $D(n)$, and recovered from hospitalisations $R(n)$ (Fig. 1). To reduce fluctuations we smoothed the data with a Gaussian filter. We further computed the daily number of new hospitalisations as $newH(n+1) = H(n+1) - (H(n) - R(n) - D(n))$.

Age-stratified testing data from the French governmental website

Results from COVID-19 testings in laboratories and hospitals are collected in France by the "Systeme dInformations de Depistage (SI-DEP)" since May 13, 2020. The positive testing data is shown in Fig. S2 in the SI.

Age-stratified clinical data from the Bichat hospital

The Bichat hospital in Paris provided anonymized and age-stratified data collected between March 2020 and January 2021 that displays the hospitalisation history of 586 patients who all received ICU treatment. The data does not contain patients in the age group 0-19 years old. Statistics extracted from this data are shown in Fig. 6.

The data comes from the OutcomeRea database that was declared to the *Commission Nationale de l'Informatique et des Libertés* (#999,262), in accordance with French law, and this study was approved by the institutional review board of Clermont Ferrand. Informed consent was not required because the study did not modify patients management and the data were anonymously collected.

2.2 Feedback algorithm to monitor and forecast the pandemic and hospitalisation evolution

The feedback algorithm is based on a variety of data to update and calibrate model parameters (Fig. 2A): web accessible hospitalisation statistics (Fig. 1), clinical data revealing hospital procedures (Fig. 6), testing results (Fig. S2), and social contact matrices (Eq. 6 in the SI). The calibration and forecasting procedure consists of three steps (Fig. 2B-D):

1. In the first step, the model $M(t_0)$ is calibrated with the data $D(t_0)$ up to time t_0 (Fig. 2B).
2. In the second step, the calibrated model $M(t_0)$ is used to forecast the pandemic and hospitalisation evolution $P(t|t_0)$ for times $t > t_0$ (Fig. 2C).

If new social measures are put into effect at time t_0, the contact matrix is adjusted such that $P(t|t_0)$ takes into account the expected impact of these measures. As time progresses, a checkpoint procedure constantly confronts the prediction $P(t|t_0)$ to the newest data $D(t)$ to verify whether the prediction is still conform with reality (Fig. 2D). If at time t the discrepancy between prediction and data becomes

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

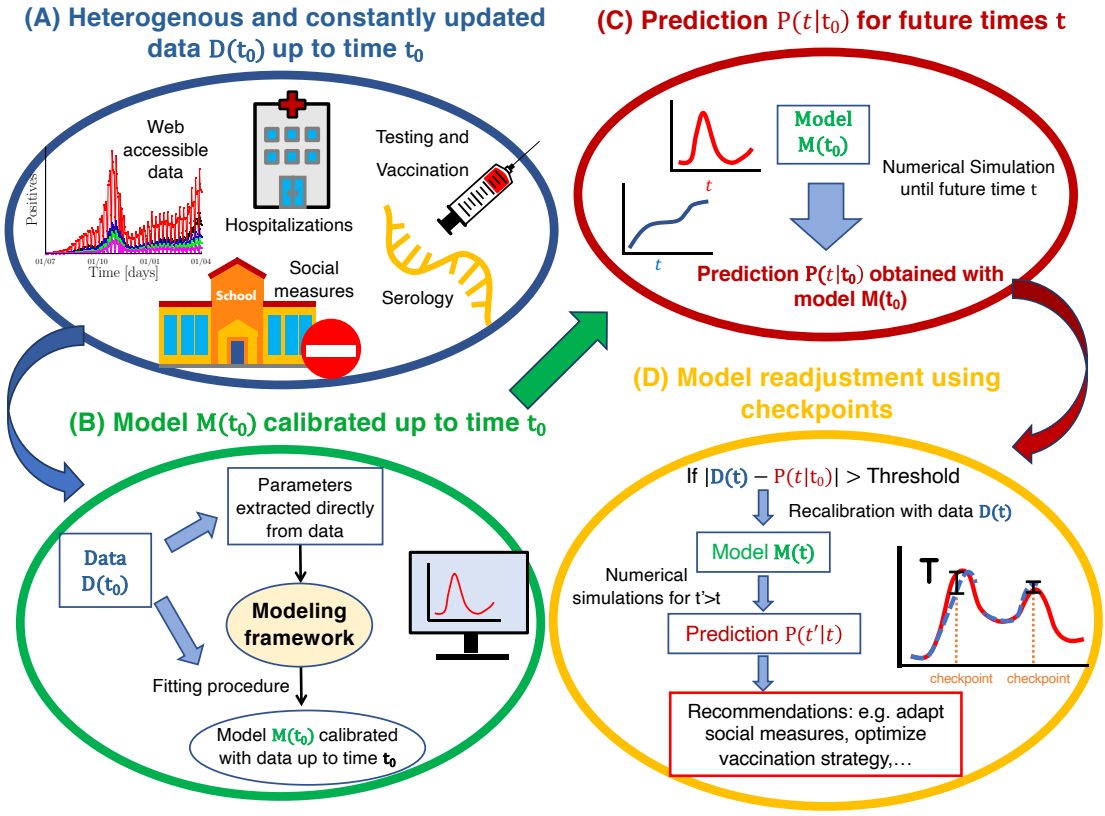

Figure 2: **Automated checkpoint procedure to monitor and forecast the pandemic and hospitalisation evolution.** (A) A variety of constantly updated data is used to calibrate the model. (B) The model $M(t_0)$ is calibrated with data $D(t_0)$ up to time t_0. (C) The calibrated model $M(t_0)$ is used to predict the pandemic and hospitalisation situation $P(t|t_0)$ at future times $t > t_0$. (D) As time progresses, a checkpoint procedure constantly confronts the predictions $P(t|t_0)$ to the data $D(t)$. If at some time t the discrepancy between model and data becomes larger than a threshold T, a recalibrated model $M(t)$ is derived using the data $D(t)$. The model $M(t)$ is then used to generate new predictions $P(t'|t)$ for times $t' > t$. Based on these predictions, social measures and strategies are adjusted.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

larger than a threshold T, $|D(t) - P(t|t_0) > T$, a recalibrated model $M(t)$ is derived with the data $D(t)$ up to time t. With the model $M(t)$ new predictions $P(t'|t)$ are computed for times $t' > t$. If the new predictions are no longer conform with the expectations, this can be used as an early warning signal to readjust social measures.

2.3 Multiscale modeling framework

The modeling framework that we developed implements the feedback algorithm specified above. Mathematical details are given in the SI. We developed a spatially homogeneous compartmental model with a discrete time-resolution of one day (Fig. 3A). We distinguish between 5 age groups (Fig. 3B) and 5 infection categories (Fig. 3C). We use transition probabilities between these categories to compute the disease progression (Fig. 3C). Social interactions are modelled with time-dependent contact matrices that account for national political interventions (Fig. 3D).

Multiple time scales to compute the disease progression

We distinguish between the calendar time n measuring the number of days since January 1, 2020, and the days l since an infected person has joined a given infection category (Fig. 3A). The total time m since infection can be computed by adding the times spent in each category. For asymptomatic persons in category $j = 1$, the times m and l are identical. Transition probabilities might not only depend on the times n and l, but also on the time m since infection. However, since m is usually unknown, we only have data for time l (see Fig. 6). We therefore use only the times n and l for the modeling.

Population divided into five age-stratified groups

We classify the population into 5 age groups labelled by $k = 1 \ldots 5$ (Fig. 3B): 0-19, 20-49, 50-59, 60-69 and 70-79 years old. The population in age group k is $N_k = N_{tot} p_k$, where $N_{tot} = 67$ millions and $p_k = (24\%, 36.3\%, 13.1\%, 11.9\%, 10.2\%)$ is from *https://www.statista.com/statistics/464032/distribution-population-age-group-france*.

Disease progression characterized by five infection categories

We consider 5 infection categories and transition probabilities that specify the disease progression (Fig. 3C). A newly infected person starts in the asymptomatic category $j = 1$. From there it can either develop symptoms and become symptomatic (switching to $j = 2$), or remain asymptomatic and eventually become recovered. For a symptomatic person the disease either further deteriorates such that hospitalisation is needed (switching to $j = 3$), or the symptomatic person will recover without hospitalisation. A hospitalised patient in category $j = 3$ can be transferred to ICU (switching to $j = 4$), or it will eventually recover and be released from hospital. In ICU, a patient can either die or be transferred back to normal hospital (switching to $j = 5$). Finally, a patient in hospital category 5 will eventually be released and become recovered. In total the model distinguishes between four types of recovered persons.

Infection dynamics leading to new hospitalisations with social interventions

Since the number of new infections is hidden and not precisely known, we used the new hospitalisation data to calibrate the infection dynamics of the model (Fig. 1). We implemented constrained fitting procedures to estimate parameter values from this data. As initial condition for the pandemic we

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

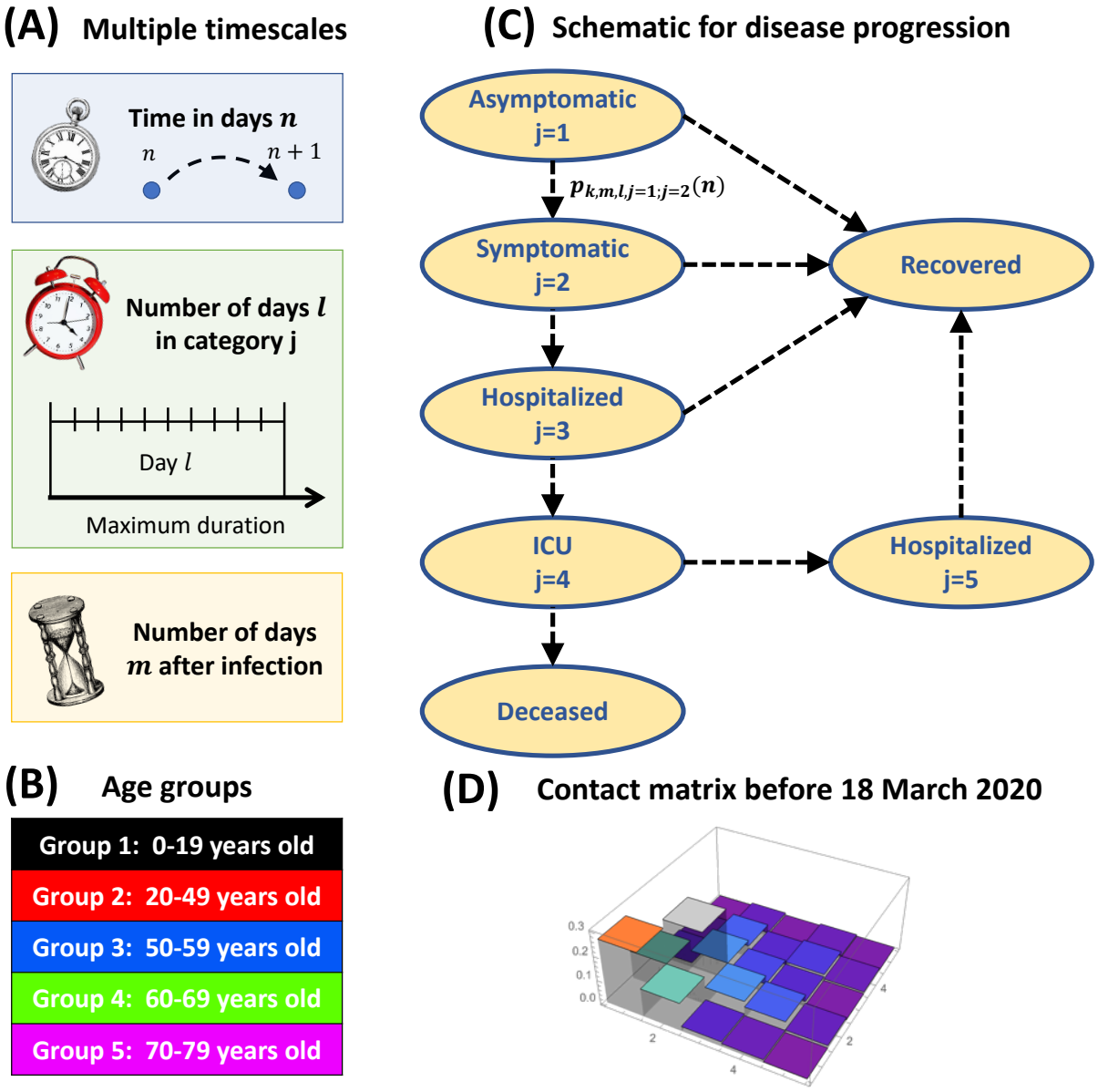

Figure 3: **Spatially coarse grained and time discrete model with a time resolution of one day**. (A) Multiple time scales: the calendar time n ($n = 1$ corresponds to January 1, 2020), the number of days l since an infected joined its current category, the number of days m since infection. (B) Definition of the five age groups considered in the model (labelled by the index k). (C) Model for disease progression with transitions between 5 infection categories (labelled by the index j). The transition probabilities $p_{k,m,l,j_{old};j_{new}}(n)$ determine the switching dynamics from category j_{old} to j_{new}. Infected persons only die in hospitals in Intensive Care Units (ICU). Recovered persons are differentiated depending on the category from which they recover. (D) Normalized contact matrix for France from [35] for the time period before the first lockdown at March 18, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

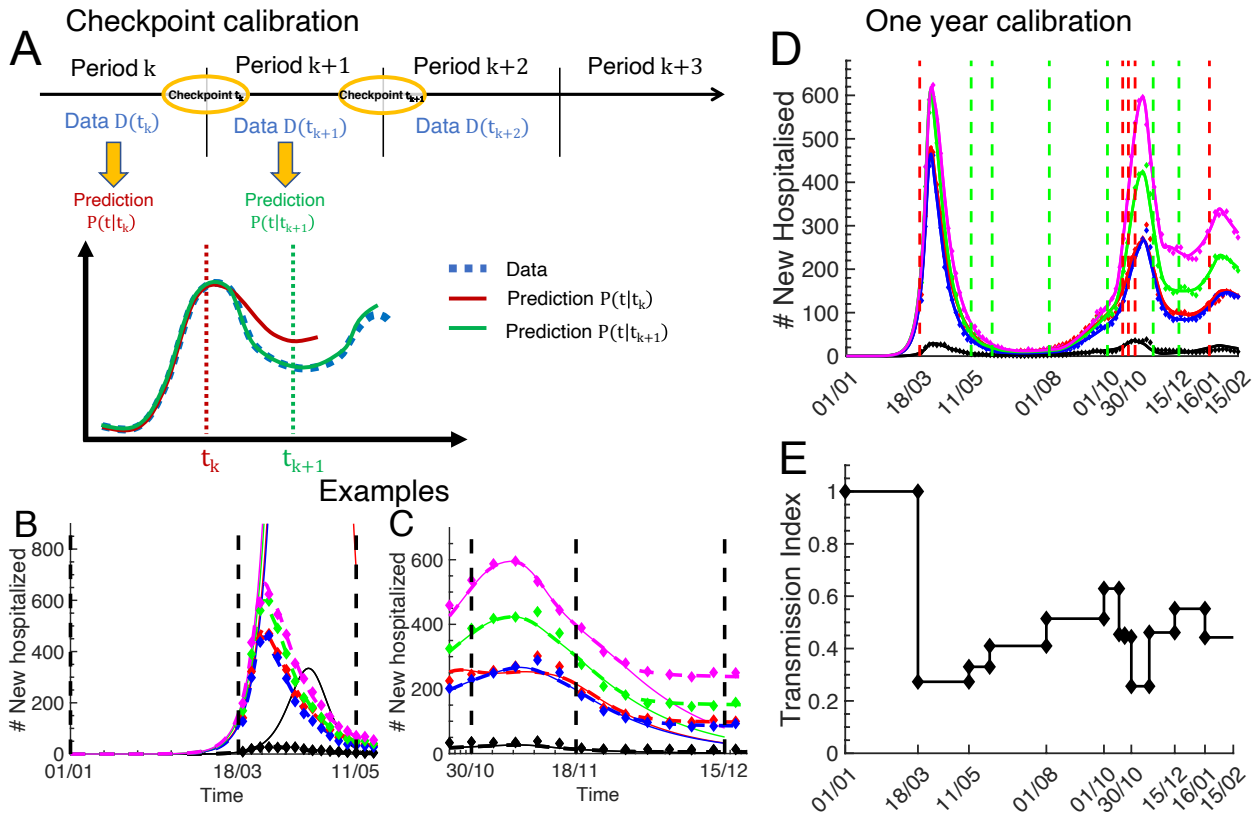

Figure 4: **Sequential checkpoint calibration procedure with new hospitalisation data.** The checkpoint times t_k correspond to social intervention times from Table 1. The new hospitalisation data (diamonds) in (B-D) is from Fig. 1. (A) Schematic of the sequential calibration procedure. The data for period k+1 ($[t_k, t_{k+1}]$) is used to fit the new contact matrix $c_{a;a'}(t_k)$ at time t_k that accounts for new social measures that are put into effect at t_k. The model prediction $P(t|t_k)$ (red curve) computed with the unchanged contact matrix before deviates from the data for $t > t_k$. The prediction $P(t|t_{k+1})$ (green curve) is computed with the new contact matrix. (B-C) Two calibration examples. The predictions (solid lines) computed with contact matrices for the times before March 18 and November 18, 2020 deviate from the data (diamonds). The predictions computed with contact matrices that were adapted at March 18 and November 18 (dashed lines) reproduce the data. (D) Result of the sequential calibration procedure until February 2021. Model simulations (solid lines) are compared to the new hospitalisation data (diamonds). (E) The normalized transmission index $t_k = \frac{\sum_{a' \geq a} c_{a;a'}(t_k)}{\sum_{a' \geq a} c_{a;a'}(1)}$ (see Material and Methods).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Date	Event
01/01/2020	First day of simulation
18/03/2020	Beginning of the first lockdown
11/05/2020	Partial deconfinement and introduction of masks
02/06/2020	Further deconfinement
01/08/2020	Beginning of vacation period
01/10/2020	Beginning of the academic year
17/10/2020	Region specific curfew
23/10/2020	National curfew
30/10/2020	Beginning of the second lockdown
18/11/2020	Beginning of deconfinement
15/12/2020	Further deconfinement
16/01/2021	Installation of curfew at 18h

Table 1: Checkpoint events where we changed the social contact matrix. Events that had a confining effect are marked with red color, deconfining events with green.

fitted a seed of new infected persons in the age-group 20-49, since person from this age group are likely to have introduce the virus in France due to travailing. We fixed January 1, 2020 as initial date for simulations. We verified that changing this date does not affect simulation results at later times because it can be compensated by changing the value of the initial seed.

Initially a new infected person is in the asymptomatic category $j = 1$. When the infection manifests itself with symptoms, the infected person switches to the symptomatic category $j = 2$. We assume that the now an alerted person takes precautions to avoid infecting others. Thus, in our model only infected persons belonging to the asymptomatic category spread the disease. We therefore consider a non-zero infectiousness only for the asymptomatic category (see Eq. 4 in the SI). With an incubation period of 5 days [36], we assume that a person is uniformly infectious within the time period $l = 5 - 11$ days after the infection (Eq. 4 in the SI). We use Eq. 16 in the SI to transform distributions which are functions of the time l into transition probabilities. With two days of pre-symptomatic transmission, we assume that an infected person can develop symptoms with a certain probability uniformly distributed within the period $l = 6 - 11$ days after the infection [37]. We fitted the probability to remain asymptomatic, which is complementary to the one for developing symptoms. To reduce the number of fitting parameters, we consider that the age groups 2-4 have the same probability to remain asymptomatic. For age group 1, we constrained the fitting range to values between 30% and 80%, for age groups 2-4 to 30% and 50%, and for age group 5 to 20% and 40%. Based on findings that the viral load is independent of age [38, 37, 39], we assumed that the infectivity is independent of age. However, since it is unclear whether children and adolescents have the same infection dynamics as adults [40], we fitted the susceptibility value of group 1 with the constrain that it is reduced by maximally 50% compared the normalized value of one for the other groups (see Eq. 7 in the SI).

We fitted age stratified probabilities for symptomatic persons to become hospitalised (Fig. S1). To constrain the fitting range, we fitted the data for new hospitalisations from the number of positive testings (Fig. S2) to extract age stratified probabilities that a positively tested person becomes hospitalised: 0.6%, 1.5%, 4.5%, 10.5%, and 22.8%. Assuming that symptomatic and new hospitalised persons have been tested positive, these values provide lower limits for the probability that a symp-

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

tomatic person becomes hospitalised. Due to reduced travelling and social interactions, we assumed that a person in the age group 70-79 is tested primarily because of symptoms. We therefore reduced fitting parameters by using the value of 22.8% for group 5 as a fixed anchor (Fig. S1A in the SI). We further reduced fitting parameters by assuming that the distributions that specify at which day l a persons becomes hospitalised after showing symptoms are age independent (Fig. S1B in the SI).

We modified contact matrices to account for multiple governmental interventions that affected social interactions. Table 1 specifies the dates when new social restrictions have been put into effect (red entries), or restrictions have been relieved (green entries). We added two additional dates beginning of August and October to account for the beginnings of summer vacation and new academic year. The dates in Table 1 correspond to checkpoint times t_k where contact matrices were modified (all other model parameters were kept unchanged). We divided the past year into consecutive time periods that are delimited by these checkpoint times. For example, period$_1$ ($[t_0, t_1]$) corresponds to the time from January 1, 2020 until March 18, 2020 when the first national lockdown was imposed, and period$_2$ ($[t_1, t_2]$) is the lockdown period until May 11, 2020. The contact matrix for period$_1$ is not fitted but extracted from [35] (Eq. 6 in the SI). At each time t_k we fitted a new contact matrix $c_{a;a'}(t_k)$ (15 fitting parameters) using the new hospitalisation data for period$_{k+1}$ ($[t_k, t_{k+1}]$) (Fig. 3A).

To constrain the fitted lockdown contact matrix, we made the following assumptions: contacts between group 1 and groups 4 and 5 are reduced by at least 90%; due to school closure, the intra-group contact of group 1 is reduced by at least 70%; for all the remaining parameters are reduced by at least 40%. Considering that new hospitalisation data is only available starting from period 2, we recursively fitted period 1 and 2. With this recursive initial fitting procedure we not only estimated the lockdown contact matrix, but also all the remaining infection parameters, that were then kept constant. Without changing the contact matrix at March 18, 2020 the pandemic would have continued to grow exponentially (Fig. 3B, solid curves). In contrast, with the adapted contact matrix the model reproduces the pandemic decline during the lockdown (Fig. 3B, dashed curves). As another example that shows the impact of contact matrix changing, we consider October 30 where a second national lockdown has been imposed that lasted officially until November 28. However, we could not satisfactorily reproduce the new hospitalisation data by assuming the same contact matrix between October 30 and November 28 (Fig. 3C, solid curves). Instead, a much better fit was obtained by adapting the contact matrix already around 10 days before the official end of the lockdown (Fig. 3C, dashed curves), suggesting that the social interaction dynamics started to change already before the official end of the lockdown.

Since masks were largely unavailable before May 11 2020, the pandemic has declined during the first lockdown entirely as a consequence of a reduction in social contacts. In contrast, with the availability of masks, the infection dynamics further depends on a reduced infectivity of social interactions due to mask wearing [41]. However, because a reduction in the infectivity parameter $\beta(n)$ due to masks has the same effect as a uniform reduction of the contacts matrix (Eq. 10 in the SI), we adopted the following fitting procedure starting from May 11: 1) we kept $\beta(n)$ constant and fitted only contact matrix parameters; 2) if the social measures had a deconfining character, we used the preceding contact matrix as lower limit, otherwise as upper limit. Consequently, the fitted contact matrices starting from May 11, 2020 have to be considered as effective matrices that also comprise the effect of mask wearing. With this method we sequentially calibrated the model to the new hospitalisation data until February 2021 (Fig. 4D; fitting results are solid lines, diamonds are data; vertical red dashed lines indicate confining events, green dashed lines deconfining events). To visualise these changes, we defined the transmission index $t_k = \frac{\sum_{a' \geq a} c_{a;a'}(t_k)}{\sum_{a' \geq a} c_{a;a'}(1)}$, and we show how this value evolved due to mask wearing and social measures (Fig. 4E). The index is normalized to one

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

for the initial uncontrolled phase of the pandemic. The strong reduction of the transmission index by almost 80% during the first lockdown after march 18, 2020 is entirely due to contact reductions, whereas the values at later times comprise the combined effect of mask wearing and contact reduction.

Summary of fitted infection parameters

The initial seed of infected in age group 2 at January 1, 2020 is 11. The fraction of asymptomatic is 80%, 46.15%, 46.15%, 46.15%, and 20%. The values 80% and 20% correspond to upper and lower limits. The susceptibility of age group 1 was reduced by 37% compared to the other groups. The value for the infectivity $\beta(n)$ was 2.63×10^7 (Eq. 7 in the SI). This value determines the exponential growth during the initial phase, and depends on the normalisation of the contact matrix. The probability that a symptomatic person becomes hospitalised is given in Fig. S1 in the SI.

3 Results

Although there is regional heterogeneity of the infection spread throughout France [30, 3, 42], we adopted a coarse grained approach to model the pandemic evolution at national level. We split the model into two parts: In the first part we model the infection dynamics leading to new hospitalisations, and in the second part we use the new hospitalisations to predict the hospital workload.

3.1 Infection dynamics with multiple social interventions.

Multiple governmental interventions during the past year affected social interactions, which we accounted for by adapting the contact matrix of the model (see Materials and Methods). The latest contact matrix accounts for the social measures from January 16, 2021. We did not derive contact matrices for the time after January 20201 because the model does not include the effect of vaccinations and viral mutations. We simulated the pandemic evolution during 2021 to show the consequences of a scenario with persisting social restrictions from January 2021 (Fig. 5; the vertical dashed lines indicate that the model has been calibrated to data until February 15, 2021).

The simulated number of new infected per age group exhibit three prominent peaks in March 2020, October 2020 and January 2021 when the pandemic was not under control (Fig. 5A). Interestingly, the various age groups contribute very differently to new infections: the dominant contributions come from age groups 0-19 and 20-49 due to their large number of interactions (contact matrix, Eq. 6 in the SI), whereas infected persons in the age group 60-79 contribute only little to spread the disease (Fig. 5A). The simulation reveals that with the social restrictions from January 2021 the pandemic would have declined during 2021 if no viral mutations had appeared.

From the number of new infections, we computed the age dependent fraction of infected population (Fig. 5B). At May 11, 2020 around 5% of the population had been infected, which increased to around 13.6% until January 16, 2021 (Fig. 5B, yellow curve). However, whereas by January 16, 2021 only around 5% of the population in the age group 70-79 has been infected (Fig. 5B, magenta), this value increases to around 18% for the age group 0-19 (black).

To characterize the pandemic growth, we computed age group dependent reproduction numbers $R_{0,k}(n)$ that represent the mean number of new infections that will be generated by a person that becomes newly infected at day n in age group k (Eq. 12 in the SI). By averaging over age groups we further defined the mean reproduction number $R_0(n)$ (Eq. 13 in the SI). We found that the reproduction numbers $R_{0,k}(n)$ are strongly age and time dependent (Fig. 5C). This reflects the impact of multiple social measures and the age stratification of interaction frequencies. For example,

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

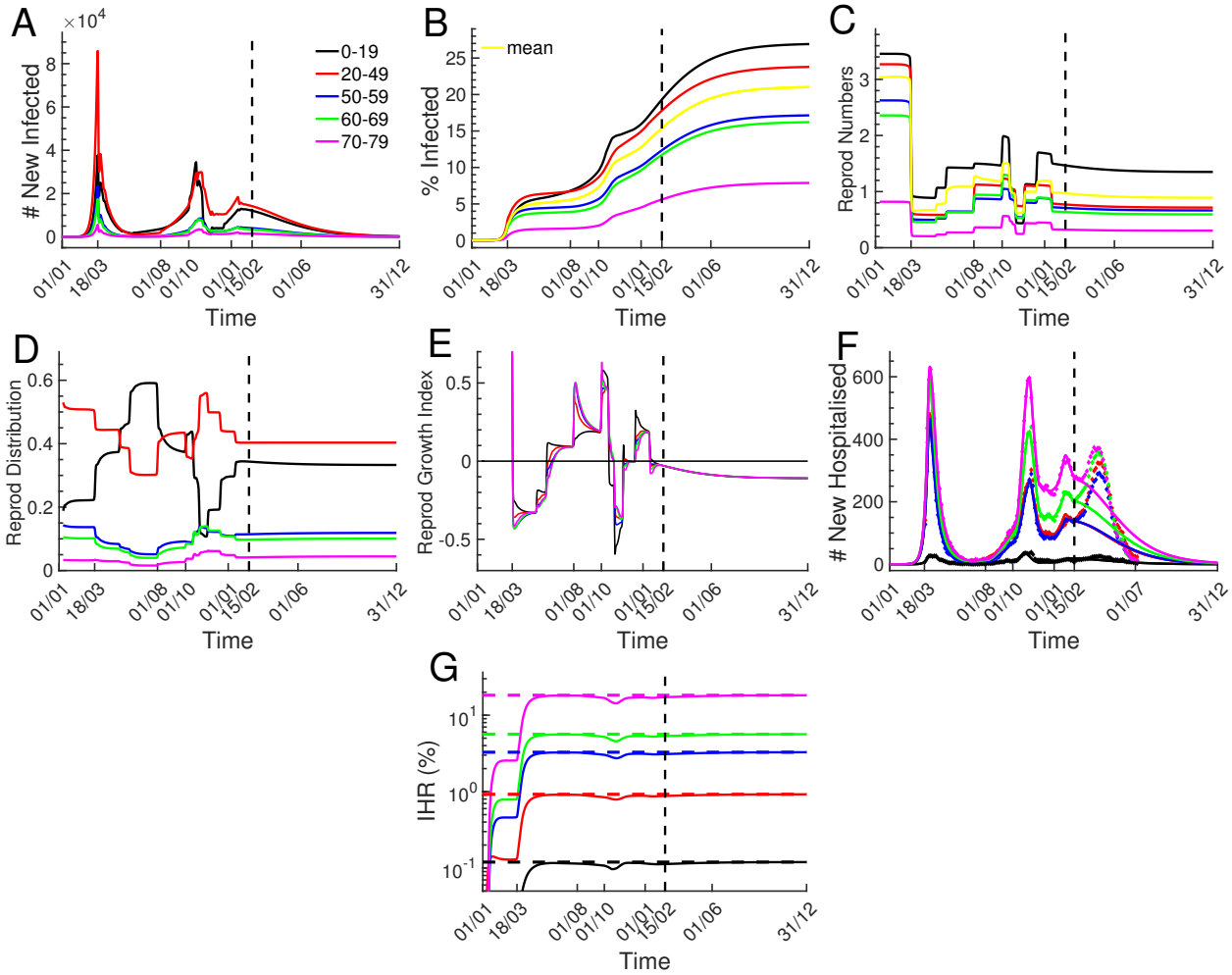

Figure 5: **Infection dynamics with multiple social interventions.** Simulations from January 1, 2020 until December 31, 2021 performed with the model that has been calibrated to data until February 15, 2021, as indicated by the vertical dashed lines. Because the model does not account for new viral strands and vaccinations, the simulations after February 2021 show the pandemic evolution for a hypothetical scenario without new viral strands and vaccinations, and social restrictions from January 2021 that persist throughout the year. (A) New infections per age group. (B) Fractions of infected population per age group, and mean fraction of infected population (yellow curve). (C) Age group specific reproduction numbers $R_{0,k}(n)$ (Eq. 12 in the SI). The value $R_{0,k}(n)$ gives the average number of new infections that will be generated by a person that becomes newly infected at day n in group k. The yellow curve represents the mean reproduction number $R_0(n)$ (Eq. 13 in the SI). (D) Reproduction distributions showing in which age group new infections will be generated (Eq. 14 in the SI). (E) Reproduction growth indices (Eq. 15 in the SI). A positive value indicates that the number of new infections will increase in the specific age group. (F) New hospitalisation simulations (solid lines) are compared to the data from Fig. 1. (G) Probability that an infected person becomes hospitalized (Infection Hospitality Ratio, IHR). The IHR values are obtained by dividing the cumulated new hospitalisations from (F) with the cumulated new infections from (A). The values 0.12%, 0.92%, 3.3% , 5.6% and 18.2% marked by the dashed lines are the probability for an asymptomatic to become symptomatic multiplied by the probability for a symptomatic to become hospitalised.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the reproduction number for age group 70-79 is always smaller than one due to the low interaction frequency, whereas it is almost always above one for age group 0-19 (Fig. 4C, magenta vs black curve).

To characterise how new infections are distributed among age groups, we defined the reproduction distribution $RD_k(n)$ that specifies the fraction of new infections generated in age group k (Eq. 14 in the SI). We found that more than 80% of new infections are generated within group 1 and 2 (Fig. 4D).

Because the reproduction numbers $R_{0,k}(n)$ provide only information about the number of infections that will be generated by an infected person from group k, they do not reveal whether the new infections in group k will decline or increase. To characterise these changes, we introduced reproduction growth indices $RGI_k(n)$ (Eq. 15 in the SI) that measure the change in new infections per age group (Fig. 4E). A positive values indicates that the pandemic grows. Interestingly, although the reproduction number for age group 0-19 is larger than one during 2021, the growth index for this group is negative. This shows that infected persons in this age group spread the disease to other groups.

In Fig. 5G we compare the new hospitalisation simulation (solid lines) to data from Fig. 1 (diamonds). Up to February 15, 2021, Fig. 5G is identical to Fig. 4D. The new hospitalisation data has a forth peak beginning of April 20201, which is probably due to the effect of more infectious viral strands. The sharp decline in April and May is probably a combined effect of vaccinations and social restrictions from March 20201. In contrast, with the contact matrix from January 2021 and without new viral strands and vaccinations, the simulation predicts that the number of new hospitalisations would have gradually declined during 2021.

Finally, we computed the Infection Hospitality Ratio (IHR) (by dividing the cumulative number of new hospitalisations with the cumulative number of new infections from (A) Fig. 5G, continuous lines). We further computed analytic IHR values (Fig. 5G, dashed lines) by multiplying the probability for infected persons to become symptomatic with the probability for symptomatic to become hospitalised (the probabilities are specified in the Materials and Methods), and found the age group dependent values 0.12%, 0.92%, 3.3% , 5.6%, and 18.2%. Thus, an infected person in age group 70-79 is around 150 times more likely to become hospitalised compared to a person in age group 0-19. Interestingly, the transient plateau values in Fig. 5G before March 18,2020 differ from the analytic predictions during the exponential growth phase because of the time delay between new infection and new hospitalisation.

3.2 Hospitalisation dynamics

Predicting the hospital workload from the number of new hospitalisations requires statistics at single hospital level. Because such data is not provided by the French governmental websites, we used age stratified clinical data from the Bichat hospital in Paris to estimate transition probabilities that define the hospitalisation dynamics for our model (Fig. 3). With this data, we derived age stratified distributions specifying the day l at which patients are transferred to ICU, leave ICU, die in ICU, and are released from hospital (Fig. 6A-D, dashed lines). We further extracted the probability that a patient dies in ICU (Fig. 6E). Due to the small sample size per age group, the distributions show large fluctuations (Fig. 6A-D, dashed lines). We fitted the data with the discrete Gamma distribution (Eq. 26 in the SI) to smooth the curves, which reveal more clearly difference between age groups (Fig. 6A-D, solid lines). However, we found that using the smooth or fluctuating distributions had only a minor impact on predicting the hospital workload (see next paragraph). Finally, we computed the average number of days patients spend in categories $j = 3, 4, 5$ before making transition to a new

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

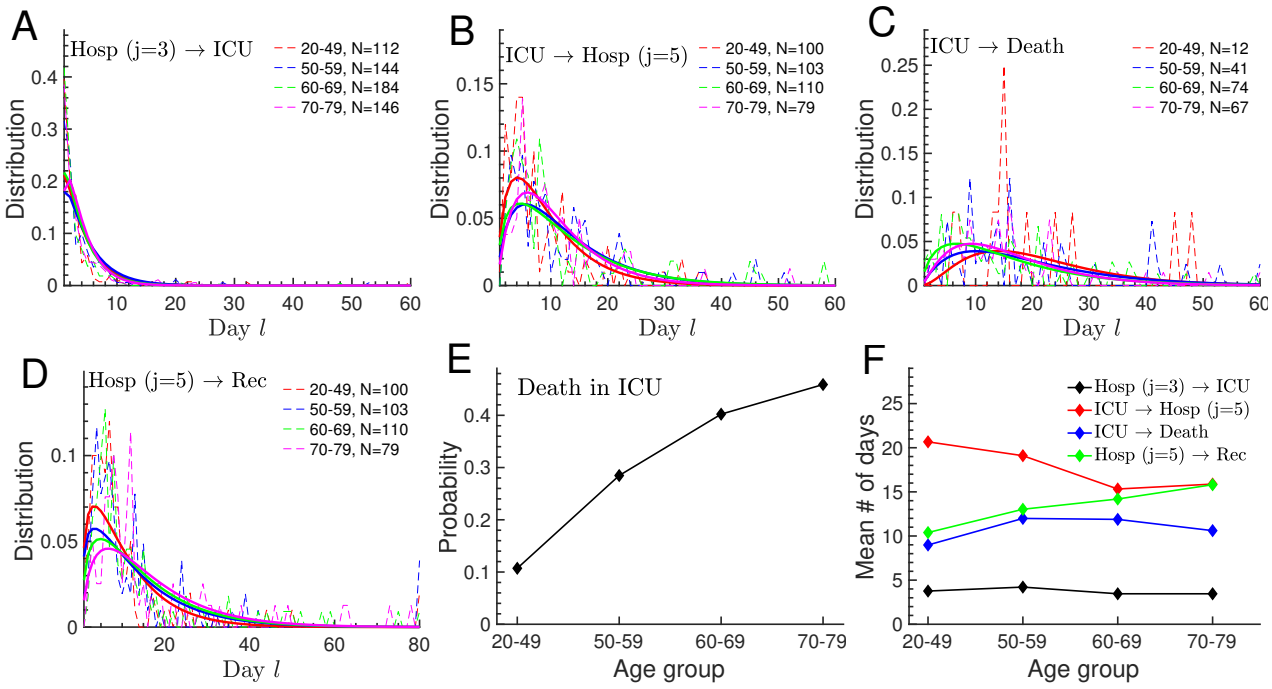

Figure 6: **Hospitalisation statistics from clinical data.** Age stratified statistics were computed with clinical data of 586 patients from the Bichat hospital in Paris (see Materials and Methods). All patients received ICU treatment and all were older than 20 years. (A-D) Switching distributions as a function of the number of days l after joining the category. We truncated l to a sufficiently large value $l_{max} = 60$ for $j = 3, 4$, and $l_{max} = 80$ for $j = 5$. Fluctuating dashed lines show the data, and smooth solid lines are fits obtained with the discrete Gamma distribution from Eq. 26 in the SI. The parameter N is the number of patients that contributed to the distribution. The individual panels display the distributions to be transferred from hospital ($j = 3$) to ICU ($j = 4$) (A), from ICU to hospital ($j = 5$) (B), to die in ICU (C), and to be released from hospital (D). (E) Probability to die in ICU. (F) Mean number of days before switching computed with the dashed distributions from (A-D).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

category (Fig. 6F).

In the subsequent step we used the transition probabilities extracted from the Bichat data in Fig. 6 to predict the hospitalisation workload in France based on the evolution of new hospitalisations (Fig. 5F). From the Bichat data we could not extract the probability to be transferred to ICU, the distribution when patients without ICU treatment are released from hospital, and there was no data for age group 0-19. To compensate for these missing parameters, we used for age group 0-19 the distributions in (Fig. 6A-D) for age group 20-49, and all the remaining parameters were estimated by fitting for each age group the hospitalisation data for France until February 15, 2021 (Fig. 1). We applied this mixed procedure to concurrently fit the evolution of patients hospitalised, in ICU, that died in ICU, and released from hospital. For example, for the age group dependent probabilities to be transferred to ICU, we fitted the values 15.2%, 19.5%, 26.0%, 35.7%, and 31.4%.

Finally, we simulated the hospital workload from January 2020 until December 2021, which we compared to data until July 2021 (Fig 7A-D, simulations (solid lines) versus data (diamonds)). The simulated new hospitalisation from Fig. 5F were used as input. We find that hospital parameters estimated from the Bichat data allow to satisfactorily reproduce the hospitalisation dynamics in France until February 2021. Because the simulated number of deceased persons in the age group 70-79 is reduced compared to the France data (Fig 7C), this could indicate that the Bichat hospital had a better success in treating elderly patients. We obtain larger discrepancies between simulation and data for hospitalised patients (Fig 7A) due to differences in the daily number of recovered persons that accumulate over time (Fig 7D). The simulations further suggest that the social contact restrictions from January 2021 would have been sufficient to end the pandemic growth had no new viral strands emerged. Lastly, we computed the age dependent Infection Fatality Ratio (IFR) by dividing the cumulated numbers of deceased persons and new infections (Fig 7E, solid lines). The analytical values 0.0005%, 0.019%, 0.24%, 0.81%, and 2.62% are obtained by multiplying the probabilities to become symptomatic with the ones for symptomatic to become hospitalised, for patients to be transferred to ICU and to die in ICU (Fig 7F, horizontal dashed lines). For the mean IFR ratio we find a value of 0.2%. Remarkably, the probability to die after infection in the age group 70-79 is more than 5000 times higher compared to the age group 0-19.

3.3 Forecasting the pandemic evolution with the delta variant

We applied our calibrated model to evaluate whether the vaccination status from July 2021 is sufficient to prevent a fourth wave driven by the new and fast expanding delta variant, in the context that all social restrictions in France have been lifted in June 2021. For this purpose, we performed simulations with the following conditions: For social interactions we used the unrestricted contact matrix prior to the first lockdown in March 2020 (Eq. 6). To account for the higher transmissibility of the delta variant compared to the initial strain , we increased the transmissibility parameter β (Eq. 7) by a factor two [43]. All other model parameters remained unchanged. To estimate the initial conditions for the number of susceptible and infected in July 2021, we considered the age stratified data for vaccinations and positive testings available from the French governmental website. By assuming that only fully vaccinated people are protected, we reduced the age group dependent pool of susceptible by 4.05%, 36.1% , 59.1%, 70.9%, and 70% , which corresponds to the published data for the population that has have been fully vaccinated by July 12, 2021. We further reduced the pool of susceptible by considering the population that had been infected. Since the number of infected people did not significantly increase since February 2021 due to the severe social restrictions, we used our estimations for the fraction of infected population at February 15, 2021 to further reduce the number of susceptible by 21.05%, 17.9%, 12.4%, 11.7% and 5.6%. We implemented a constant future

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

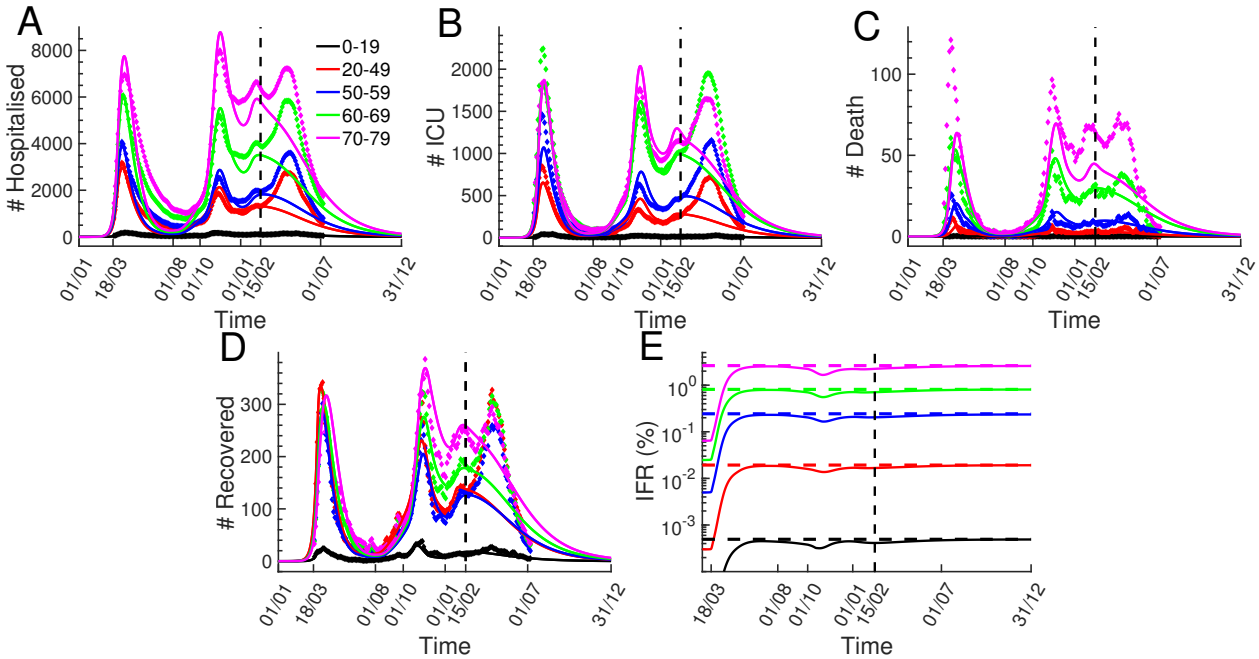

Figure 7: **Hospitalisation dynamics.** Simulations of the hospitalisation evolution (solid lines) are compared to the data from Fig. 1 (diamonds). Simulations were performed with the new hospitalisation simulations from Fig. 5F as input, and with hospital parameters extracted from the clinical data in Fig. 6). The simulations after February 2021 show the hospitalisation progression in a scenario where the social restrictions from January 2021 would have persisted throughout 2021 without the impact of new viral strands and vaccinations. (A-D) Evolution of patients hospitalised (A), in ICU (B), deceased in ICU (C), and recovered patients that were released from hospitals (D). The simulated recovered are the sum of patients released from hospital categories $j = 3$ and $j = 5$. (E) Age group dependent Infection Fatality Ratio (IFR) computed by dividing the cumulated deceased from (C) with the cumulated new infections from Fig. 5A. The dashed lines indicate the analytic values 0.0005%, 0.019%, 0.24%, 0.81%, and 2.62% computed with the probabilities for asymptomatic to become symptomatic, for symptomatic to become hospitalised, to be transferred to ICU, and to die in ICU. For the mean IFR we find 0.2%.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

vaccination rate by reducing each day the number of susceptible by $(32.07, 231.1, 60.8, 32.8, 11.9) \times 10^3$ persons, which corresponds to the number of persons that had received their second vaccination on July 12. We used zero initial conditions for the categories with $j > 1$ because they do no contribute to the viral spread. To estimate the initial conditions for the number of infected persons in category $j = 1$ as a function of the time scale l, we used the number of positive tested persons between July 1 and 12. We show predictions of the pandemic evolution for the next 5 months for two initial scenarios where 20% (solid lines) and 70% (dashed lines) of positive tested persons between July 1 to 12 had been infected by the delta variant (Fig. 8). For both scenarios, we predict a strong fourth wave that peaks around mid September to beginning of October 2021, and impacts mostly age groups 1, 2 and 5 (Fig. 8A). The predicted peak hospitalisation and deceased numbers are comparable or larger than the peak values for the first wave in March 2020 (compare Fig. 8B-C vs Fig 7A-C). Although the susceptible fraction in age group 5 is reduced below 30% due to vaccinations (Fig. 8E, magenta), this is not sufficient to prevent a strong hospitalisation and deceased wave for this group (Fig. 8A-D, magenta). The new pandemic wave will be driven mostly by infections in the age group 0-20 (Fig. 8F, black) due to the high number of susceptible (Fig. 8E, black). But despite such a high number of new infections in the age group 0-20, the predicted hospitalisation and deceased numbers remain well below the values for age group 70-79 due to much reduced probabilities to become hospitalised and to develop a severe form of the disease.

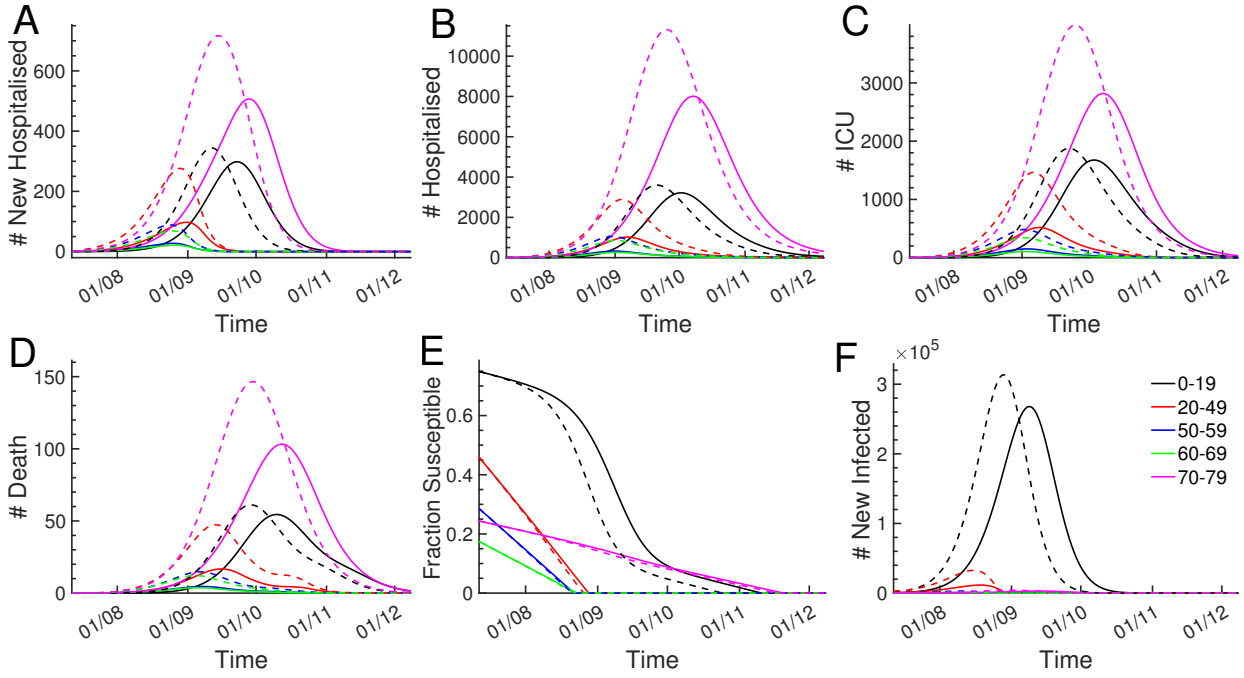

Figure 8: **Forecasting scenarios with the delta variant.** Simulations of the age stratified pandemic evolution after July 12, 2021 with normal social interactions and a constant vaccination rate ($(32.07, 231.1, 60.8, 32.8, 11.9) \times 10^3$ per day) for two scenarios where 20% (solid lines) and 70% (dashed lines) of positive tested persons between July 1 to 12 had been infected by the delta variant (see text for more details). (A) Number of new hospitalised. (B) Number of hospitalised. (C) Number of hospitalised in ICU. (D) Number of deceased in hospital. (E) Fraction of susceptible population per age group. (F) Number of new infections.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

4 Discussion

During the early phase of a pandemic when vaccines are not available, social restrictions are key to prevent an exponential growth and a possible collapse of the health care system. In such circumstances forecasting models allow to test the consequences of social interventions on the pandemic evolution. However, providing long-term predictions is unrealistic due to incomplete knowledge, constant changes in social behaviour and unexpected events. To provide reliable predictions in such circumstances, a flexible data-driven modeling framework is needed that constantly adapts to reality.

To account for changes in the pandemic evolution, we developed an age-stratified modeling framework that periodically updates parameters at checkpoint times using the most recent data (Fig. 2). To reliably forecast infection dynamics and consequences for hospitals, we integrated a large variety of data to calibrate parameters, such as hospitalisation (Fig. 1), clinical (Fig. 6), serological and testing data (Fig. S2 in the SI) and social interaction matrices (Eq. 6 in the SI). To handle missing data, we implemented fitting procedures to extract parameter values. To provide flexibility, we structured the modeling framework into two parts: First, we compute the infection dynamics leading to new hospitalisations. Second, we compute the hospital workload depending on the evolution of new hospitalisations.

To show the relevance of our framework, we studied the Covid-19 pandemic in France with multiple social interventions. To account for the age stratified Covid-19 disease progression, we considered 5 age groups, 5 infection categories, one category for deceased and 4 for recovered persons (Fig. 3). Besides the calendrical time, we also count the number of days l after an infected person joins its current category (Fig. 3). We model changes in disease properties depending on the times l, and we do not introduce categories like exposed, paucisymptomatic or prodromic necessary in SIR type of models based only on the calendrical time. For example, instead of implementing an exposed category we set infectivity parameters to zero for the time $l < 5$ after infection (Eq. 4 in the SI). We classified non-hospitalised infected people into asymptomatic and symptomatic, similar to the dual classification from [44]. The difference between asymptomatic and symptomatic persons is that the latter are aware of their infection and take precautions to avoid infecting others. Infected persons with unnoticed symptoms therefore belong to the asymptomatic category. Consequently, we consider that only persons belonging to the asymptomatic category spread the disease, in agreement with the prevalent role of undocumented and silent transmissions [45, 46, 27, 22].

For the age groups 20-69 and 70-79 we estimated that 46% and 20% of the infections remain asymptomatic, respectively, values that are within the published range of 15% to 60% [47, 48, 37, 3, 25, 37, 29]. In contrast, for the age group 0-19 we found a surprisingly high value around 80% due to the following reasons: 1) the age group 0-19 is involved in around 42% of all contacts, comparable to 57% for age group 20-49 (Eq. 6 in the SI); 2) the probability for a symptomatic person in the age group 0-19 to become hospitalised is around 0.6%, which is reduced only by a factor 3 compared to the age group 20-49 (Fig. S1); 3) new hospitalisations in the age group 0-19 are reduced by a large factor 16 compared to 20-49 (Fig. 1). As a consequence, to obtain such a drastic reduction in the number of new hospitalisations, the probability to remain asymptomatic has to be much higher in the age group 0-19 compared to 20-49. A much higher asymptomatic fraction for children and adolescents is in agreement with more asymptomatic and mild infections, and fewer symptoms [49, 50, 51, 52, 53]. We further obtained a 37% reduced susceptibility to infections in age group 0-19, compatible with a reduced susceptibility for children and adolescents [54, 55, 50]. We tested scenarios where the susceptibility to infections is independent of age, and found that the fraction of asymptomatic in the age group 0-19 needs to be around 90% (not shown). To conclude, our results point out profound differences in the disease progression between children and adults.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We found that by May 11, 2020 around 5% of the French population had been infected, a value that increased to around 13.6% by January 16, 2021 (Fig. 5B, yellow curve). These values are consistent with 5.7% and 14.9% obtained by a deconvolution procedure [42]. We estimated that the fraction of infected population is strongly age dependent: by January 16, 2021 only 5% of the population has been infected in the age group 70-79, contrary to 17% in the age group 0-19, similar to the values from [42]. Our analysis further showed that also reproduction numbers are strongly age dependent (Fig. 5C), which is often ignored by focusing on the classical reproduction number R_0 [8] (Fig. 5C, yellow curve). Interestingly, because reproduction numbers depend on the social interaction dynamics, there is no absolute value for the level of herd immunity [30, 12]. For example, with the contact matrix from January 2021 the herd immunity level would be around 21% (Fig. 5B, yellow curve). The reproduction numbers $R_{0,k}$ count the number of infections that a person in age group k will generate (Eq. 12 in the SI), but they do not specify how these infections are distributed among age groups, and whether the infection level in a specific group will decline or increase. To quantify these properties, we introduced a reproduction distribution RD_k (Eq. 14 in the SI) and a reproduction growth index RGI_k (Eq. 15 in the SI). By computing RD_k we found that more than 80% of new infections are generated within age groups 0-19 and 20-49 (Fig. 4D). The RGI_k values revealed that the infection level can decline in an age group (Fig. 5E, black curve) although the reproduction number is above one (Fig. 5C, black curve). We computed infection hospitality ratios (IHR) and found the age group dependent values 0.12%, 0.92%, 3.3%, 5.6%, and 18.2% (Fig. 5G), comparable to [18, 42]. These IHR values reveal that a person in age group 70-79 is around 150 times more likely to become hospitalised compared to person in age group 0-19.

We addressed the challenging task of predicting the hospital workload from the number of new hospitalisation (Fig. 5E). For this purpose we developed mixed procedures that combine clinical data from one or many hospitals with national hospitalisation data. This combination compensates for missing data at single hospital or at a national level. For example, we used the clinical data from the Bichat hospital to extract most of the parameters (Fig. 6), while the missing ones were estimated by fitting hospitalisation data at a national level (Fig. 1). We then applied these parameters together with the new hospitalisation evolution (Fig. 5E) to simulate a scenario about the hospital workload until the end 2021 (Fig 7A-D). In addition, we computed the age group specific IFR values of 0.0005%, 0.019%, 0.24%, 0.81%, and 2.67%, with a mean 0.2%, showing that the Covid-19 mortality is highly age dependent, consistent with [18, 45, 56, 57, 48, 21]. Thus, an infected person in the age group 70-79 has a more than 5000 times higher decease probability compared to a person in the age group 0-19.

Finally, we applied our model to simulate future pandemic scenarios with the new and highly contagious delta variant (Fig. 8). The simulations showed that with unrestricted social interactions and the current vaccination status and rate a strong fourth wave is inevitable, in agreement with modelling results for the UK [16]. In contrast, by performing the same simulations with the less contagious initial viral strain, we found that vaccinations would have been sufficient to prevent a new wave (not shown). To reduce the hospitalisation impact of the new wave there are several possible interventions: 1) social restrictions; 2) increase of the vaccination rate in age group 70-79, which is most affected by the new wave despite of the already large vaccination status in this group; 3) increase of the vaccination rate in group 1, which is responsible for most of the spreading.

To conclude, we developed a modeling framework that integrates the pandemic dynamics from infection to hospitalisation. We implemented automatic procedures to constantly calibrate the framework to a continually changing pandemic evolution, and fitting methods to infer unknown or hidden parameters from data, such as contact matrices or fraction of asymptomatic infections. Our parameter inference goes much beyond classical sensitivity analysis where variations around a given

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

parameter value are explored by numerical simulations. For example, with our fitting methods we can infer social contact matrices that are necessary to avoid an overload of the health care system. The reliability of our inferred parameters will augment as more precise data will become available that connects social measures to contact matrices. Improved future models might as well integrate more specificity in contact matrices, the effect of vaccinations and new viral variants, and a spatially resolved social dynamics [58, 24]. The final goal is to provide a comprehensive ready-to-use modelling template and software that can be activated in case of a new pandemic, or new COVID-19 variants, in order to provide periodic predictions of the pandemic evolution at national, regional or single hospital levels.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Information

A Supplementary Figures

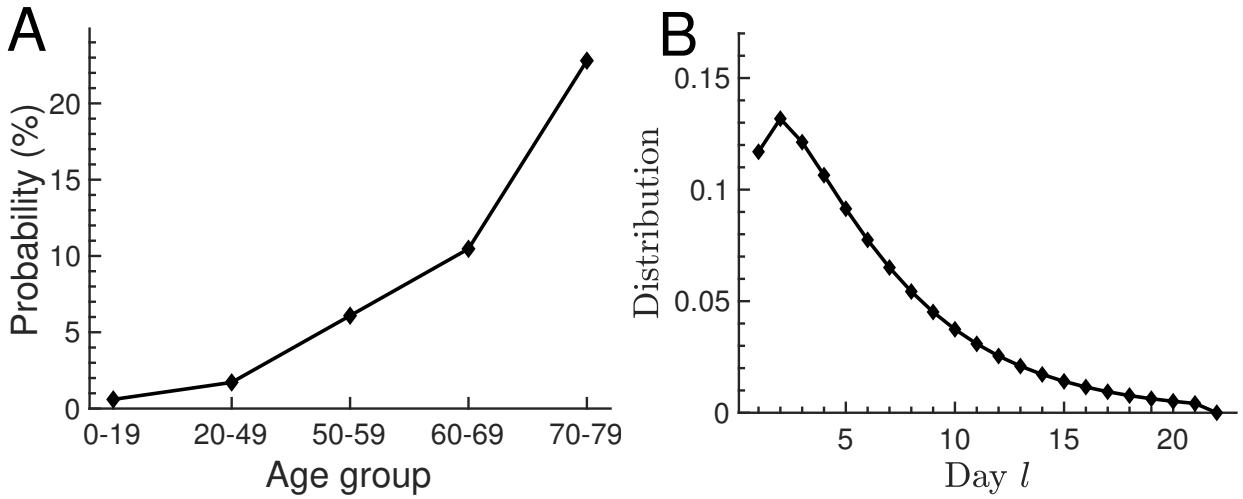

Figure S1: **Probability and time distribution for a symptomatic person to become hospitalised.** The mathematical expression for the fitting distribution is given by Eq. 26 with $l_{max} = 22$ days. (A) The fitting results for the age stratified probability that a symptomatic person becomes hospitalised are 0.6%, 1.7%, 6.1%, 10.5% and 22.8%. (B) Fitting result for the distribution that specifies at which day l after showing symptoms the hospitalisation occurs. We assumed the same distribution for all age groups. The mean value is 5.9 days.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

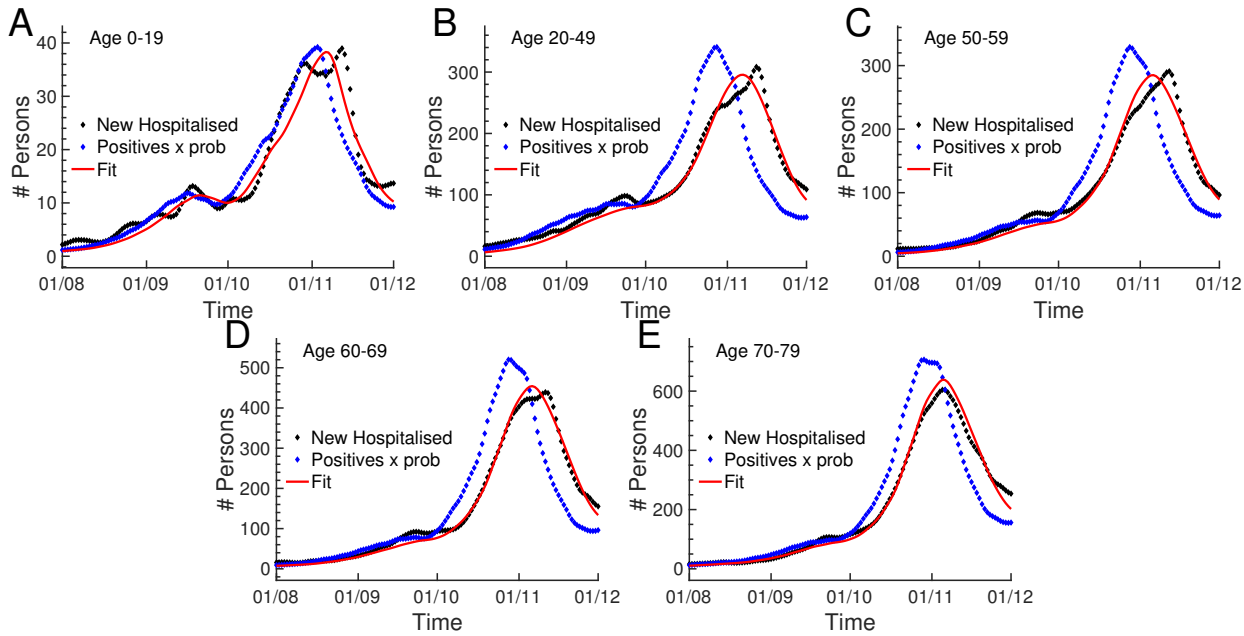

Figure S2: **Correlation between the data for positive testings and new hospitalisations.** For each age group we fitted the probability that a positive tested person becomes hospitalised using the data for the number of new hospitalisations (black dots) and the data for the number of positive tested persons (blue dots; the data is multiplied by the fitted probability). The fitting result is shown in red. The mathematical expression for the fitting distribution is given by Eq. 26 with $l_{max} = 21$ days. The fitted probabilities are 0.6%, 1.5%, 4.5%, 10.5% and 22.8%.

.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

B Modeling framework

B.1 Model structure

The parameters, variables and probability distributions of the model are:

- k: Index for the age stratified groups.
- $1 \leq j \leq 5$: Index of infection categories.
  j=1 asymptomatic; j=2 symptomatic; j=3 in hospital; j=4 in ICU; j=5 in hospital after ICU.
- Time scales.
  n: Time measured as number of days since 01/01/2020.
  l: Number of days since an infected person joined its current category j.
  h: Vector that specifies the previous infection history. For example, the infection history of a hospitalised person that is currently in category $j = 5$ since l days comprises the number of days l_1 that this person has been asymptomatic, the number of days l_2 that it has been symptomatic, the number of hospitalisation days l_3 before ICU, and the number of days l_4 in ICU, such that $h = (l_1, l_2, l_3, l_4)$. In this case the number of days since infection are $m = l_1 + l_2 + l_3 + l_4 + l$. The vector h is updated each time the infection category is changed. However, in the current implementation we consider only the time scales n and l.
- $S_k(n)$: Number of susceptible persons in group k at time n.
- $\Phi_k(n)$: Number of new infected in group k at time n.
- $I_{k,h,j,l}(n)$: Number of infected at time n belonging to age group k, infection history h, current category j at day l after joining this category . For $j = 1$ the infection history is fully determined by l, and here we can omit the index h.
- $p_{k,h,j,l;j'}(n)$: Probability to transition from category j at day l to category $j' \neq j$ for an infected in age group k with history h. The index n indicates that these probabilities might change over time due to social measures or modified hospital procedures.
- $p_{k,h,j,l;rec}(n)$: Probability to recover from category j.
- $p_{k,h,j,l;dec}(n)$: Probability to die in category j.
- $p_{k,h,j,l;j}(n) = 1 - \sum_{j' \neq j} p_{k,h,j,l;j'}(n) - p_{k,h,j,l;rec}(n) - p_{k,h,j,l;dec}(n)$: Probability to remain in category j.
- $R_{k,h,j,l}(n) = p_{k,h,j,l;rec}(n) I_{k,h,j,l}(n)$: Number of person that recover at time n from category j.
- $D_{k,h,j,l}(n) = p_{k,h,j,l;dec}(n) I_{k,h,j,l}(n)$: Number of persons that die in category j at time n. We consider that infected only die in ICU ($j = 4$).

The recurrence relations for the time evolution are

$$\begin{aligned}
S_k(n+1) &= S_k(n) - \Phi_k(n) \quad \text{(reduction of susceptible)} \\
I_{k,1,1}(n+1) &= \Phi_k(n) \quad \text{(new infections)} \\
I_{k,h,j,l+1}(n+1) &= p_{k,h,j,l;j}(n) I_{k,h,j,l}(n) \quad \text{(evolution within the same infection category)} \\
I_{k,h',j,1}(n+1) &= \sum_{j' \neq j,h,l} p_{k,h,j',l;j}(n) I_{k,h,j',l}(n) \quad \text{(switching to a new infection category)}
\end{aligned} \tag{1}$$

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

For each age group k we have the conservation equation

$$\begin{aligned} S_k(n+1) + \sum_{j,h,l} I_{k,h,m,l}(n+1) &= S_k(n) + \sum_{j,h,l} I_{k,h,j,l}(n) - \sum_{j,h,l} \left(R_{k,h,j,l}(n) + D_{k,h,j,l}(n)\right) \\ &= S_k(1) + \sum_{j,h,l} I_{k,j,h,l}(1) - \sum_{n' \leq n,j,h,l} \left(R_{k,h,j,l}(n') + D_{k,h,j,l}(n')\right) \\ &= N_k - \sum_{n' \leq n,j,h,l} \left(R_{k,h,j,l}(n') + D_{k,h,j,l}(n')\right) \end{aligned} \tag{2}$$

where N_k is the initial population. If $n = 1$ corresponds to the very beginning of the pandemic, we have $N_k = S_k(1) + \Phi_k(0)$, where $\Phi_k(0)$ is the initial seed of new infections.

B.2 New infections

We assume that the number of new infections $\Phi_k(n)$ in group k at day n is proportional to the daily number of contacts $\mathcal{C}_{k;k',j,m,l}(n)$ that susceptible persons in group k make with infected persons from groups k' in category j,

$$\Phi_k(n) = \sum_{k',h,j,l} \mathcal{I}_{k;k',h,j,l}(n) \mathcal{C}_{k;k',h,j,l}(n) \,. \tag{3}$$

The proportionality constants are the infectivity parameters $\mathcal{I}_{k;k',h,j,l}(n)$. We simplify and assume that as soon as the infection manifests itself by showing symptoms (switching to category $j \geq 2$) an infected person takes precautions to avoid infecting others, in which case new infections only occur in interactions with asymptomatic persons ($j = 1$). With this assumption we have (note that $h = l$ for $j = 1$, and we omit h in the following)

$$\mathcal{I}_{k;k',j,l}(n) = \beta(n) \hat{\beta}_{k;k'} \xi_{k',l} \delta_{j,1} \,, \tag{4}$$

where $\delta_{i,j}$ is the Kronecker delta. The matrix $\hat{\beta}_{k;k'}$ models group specificities in the infection dynamics (we choose a normalization $\hat{\beta}_{k;k'} \leq 1$; the matrix $\hat{\beta}_{k;k'}$ does not have to be symmetric). For example, with $\hat{\beta}_{k;1} < 1$ and $\hat{\beta}_{1;k'} < 1$ we consider the possibility that age group 1 is less infectious and less susceptible to infections. The parameters $\xi_{k',l}$ describe how the contagiousness evolves as a function of the time l after infection (we normalize and use $\xi_{k',l} \leq 1$). The time dependent infectivity parameter $\beta(n)$ accounts for changes that equally affect all groups, e.g. the introduction of masks. The initial value of $\beta(n)$ depends on the normalisations of the other parameters. Moreover, if the affinity for mask wearing would be different among age groups, one would have to consider a time dependent matrix $\hat{\beta}_{k;k'}(n)$.

To approximate the interactions between susceptible and asymptomatic persons we use

$$\mathcal{C}_{k;k',1,l}(n) \approx c_{k;k'}(n) \frac{S_k(n)}{N_k} \frac{I_{k',1,l}(n)}{N_{k'}} \tag{5}$$

where $c_{k;k'}(n)$ is the total daily number of contacts between groups k and k', and $\frac{S_k(n)}{N_k} \frac{I_{k',1,l}(n)}{N_{k'}}$ is the fraction of these contacts that correspond to encounters between susceptible and asymptomatic. The contact matrix $c_{k;k'}(n)$ is symmetric by definition, $c_{k;k'}(n) = c_{k';k}(n)$. The matrix is time dependent

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

to account for social measures. Before lockdown we use [35]

$$c_{k;k'} = \begin{pmatrix} & gp_1 & gp_2 & gp_3 & gp_4 & gp_5 \\ gp_1 & 0.2656 & 0.1182 & 0.0202 & 0.0129 & 0.0044 \\ gp_2 & 0.1182 & 0.3196 & 0.0745 & 0.0505 & 0.0147 \\ gp_3 & 0.0202 & 0.0745 & 0.0394 & 0.0225 & 0.0056 \\ gp_4 & 0.0129 & 0.0505 & 0.0225 & 0.0323 & 0.0128 \\ gp_5 & 0.0044 & 0.0147 & 0.0056 & 0.0128 & 0.0065 \end{pmatrix} . \tag{6}$$

The contact matrix in Eq. 6 is normalized such that the total number of contacts is one, $\sum_{k,k'\geq k} c_{k;k'} = 1$.

Finally, with Eqs. 3-5 the number of new infections at day n generated in group k by infectious persons is

$$\Phi_k(n) = I_{k,1,1}(n+1) = \frac{S_k(n)}{N_k} \sum_{k'} \beta(n)\hat{\beta}_{k;k'} c_{k;k'}(n) \frac{\sum_l \xi_{k',l} I_{k',1,l}(n)}{N_{k'}} . \tag{7}$$

With

$$I_{k,1,l}(n) = I_{k,1,1}(n-l+1)\pi_{k,l} = \Phi_k(n-l)\pi_{k,l} , \tag{8}$$

where $\pi_{k,l}$ is the probability that an infected person is still asymptomatic after l days, and

$$S_k(n) = S_k(0) - \sum_{n'=1}^{n} I_{k,1,1}(n') = S_k(0) - \sum_{n'=0}^{n-1} \Phi_k(n') \tag{9}$$

we obtain the recurrence relations

$$\Phi_k(n) = \left(\frac{S_k(0)}{N_k} - \sum_{n'=0}^{n-1} \frac{\Phi_k(n')}{N_k} \right) \sum_{k'} \beta(n)\hat{\beta}_{k;k'} c_{k;k'}(n) \sum_l \xi_{k',l}\pi_{k',l} \frac{\Phi_{k'}(n-l)}{N_{k'}} . \tag{10}$$

B.3 Reproduction numbers

To connect to reproduction numbers, we consider the number of new infections that will be generated in age group k by a newly infected person in age group k'. By neglecting changes during the short contagious period, we obtain from Eq. 10 the reproduction matrix

$$R_{0,k;k'}(n) = \frac{S_k(n)}{N_k} \beta(n)\hat{\beta}_{k;k'} c_{k;k'}(n) \frac{\sum_l \xi_{k',l}\pi_{k',l}}{N_{k'}} . \tag{11}$$

By considering the total number of new infections that will be generated by this infected person in group k', we obtain the group dependent reproduction numbers

$$R_{0,k'}(n) = \sum_k R_{0,k;k'}(n) . \tag{12}$$

Finally, with the new infected persons $\Phi_k(n)$ that are present at day n, we define the average reproduction number

$$R_0(n) = \sum_{k'} R_{0,k'}(n) \frac{\Phi_{k'}(n)}{\sum_{k'} \Phi_{k'}(n)} = \sum_{k,k'} R_{0,k;,k'}(n) \frac{\Phi_{k'}(n)}{\sum_{k'} \Phi_{k'}(n)} . \tag{13}$$

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The pandemic grows for $R_0(n) > 1$ because the generated new infections $\sum_{k'} R_{0,k'}(n)\Phi_{k'}(n)$ is larger than the current value $\sum_{k'} \Phi_{k'}(n)$.

To characterize how the generated numbers of new infections will be distributed over the age groups, we define the reproduction distribution

$$RD_k(n) = \frac{\sum_{k'} R_{0,k;k'}(n)\Phi_{k'}(n)}{\sum_{k,k'} R_{0,k;k'}(n)\Phi_{k'}(n)}\,, \tag{14}$$

such that $\sum_k RD_k(n) = 1$. To characterize whether the infection grows or declines in age group k, we define the reproduction growth index

$$RGI_k(n) = \frac{\sum_{k'} R_{0,k;k'}(n)\Phi_{k'}(n)}{\Phi_k(n)} - 1\,. \tag{15}$$

The pandemic grows in group k for $RGI_k(n) > 0$.

B.4 Transition probabilities

The model probabilities $p_{k,h,j,l;i}(n_0)$ to transition from category j to category i (the label i also comprises the deceased and recovered category) are computed from the overall probabilities $\tilde{P}_{k,h,j;i}$ to transition from category j to category i, and the distributions $\tilde{p}_{k,h,j,l;i}(n_0)$ that specify at which day l this transition occurs. The time n_0 indicates that these probabilities might change at some time n_0. We simplify the notation and omit the labels k, h and n_0 for the following calculations.

The probabilities $p_{j,l;i}$ satisfy the normalisation conditions $\sum_i p_{j,l;i} = 1$. In contrast, $\tilde{P}_{j;i}$ and $\tilde{p}_{j,l;i}$ satisfy the normalization conditions $\sum_{i\neq j} \tilde{P}_{j;i} = 1$ and $\sum_{l=1}^{l_{max}} \tilde{p}_{j,l;i} = 1$. The probability $\tilde{P}_{j;j}$ to remain in a category j is zero, since all infected will eventually recover or die. For numerical reasons, we truncate $\tilde{p}_{k,j,l;i}$ at a value l_{max} that is chosen sufficiently large. We chose $l_{max} = 60$ for $j = 3, 4$ and $l_{max} = 80$ for $j = 5$, since it takes longer for hospitalized to recover after ICU. As a function of $\tilde{P}_{j;i}$ and $\tilde{p}_{j,l;i}$ the probabilities $p_{j,l;i}$ can be computed as

$$p_{j,l;i} = \frac{\tilde{P}_{j;i}\tilde{p}_{j,l;i}}{1 - \sum_i \sum_{l'=1}^{l-1} \tilde{P}_{j;i}\tilde{p}_{j,l';i}} \quad (j \neq i)\,. \tag{16}$$

and $p_{j,l;j} = 1 - \sum_{i\neq j} p_{j,l;i}$. Because the probability to be found in category j after l_{max} days is zero, we have $p_{j,l_{max};j} = 0$.

B.5 Time evolution of infected within a category

We compute the time evolution of the infected in a category j with a given initial condition and a prescribed influx. The probability to be found in category j after l days is

$$\pi_{j,l} = \prod_{l'=1}^{l-1} p_{j,l';j}\,, \quad 1 \leq l \leq l_{max}\,. \tag{17}$$

The occupancy of category j at time n consists of persons that originate from an initial condition, and persons that joined this category via influx from other categories.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

B.5.1 Occupancy due to influx

The occupancy $O_j(n)$ due to the influx $I_{j,1}(n)$ from other categories is given by the convolution

$$O_j(n) = \sum_{l=1}^{l_{max}} I_{j,l}(n) = \sum_{l=1}^{\min(n,l_{max})} I_{j,1}(n+1-l)\pi_{j,l}\,. \tag{18}$$

The number of persons $C_{j;i}(n)$ that change to a different category i is

$$C_{j;i}(n) = \sum_{l=1}^{l_{max}} I_{j,l}(n)p_{j,l;i} = \sum_{l=1}^{\min(n,l_{max})} I_{j,1}(n+1-l)\pi_{j,l}p_{j,l;i}\,. \tag{19}$$

The number of new persons in category i that originate from category j is $I_{i,1;j}(n+1) = C_{j;i}(n)$.

B.5.2 Occupancy due to initial conditions

We now consider the occupancy that originates from an initial condition $I_{j,l}(n_0)$ at initial day n_0 (without loss of generality we consider $n_0 = 1$). The number of infected at a future day n that are still in category j is

$$I_{j,l}(n) = I_{j,l-n+1}(1) \prod_{l'=l-n+1}^{l-1} p_{j,l';j}\,, \quad n \le l \le l_{max}\,. \tag{20}$$

The occupancy $\hat{O}_j(n)$ due to the initial condition therefore is

$$\hat{O}_j(n) = \sum_l I_{j,l}(n) = \sum_{l=1}^{l_{max}} M_{n,l} I_{j,l}(1) \tag{21}$$

where

$$M_{n,l} = \prod_{l'=l}^{l+n-2} p_{j,l';j} \tag{22}$$

For $n + l > l_{max} + 2$ we define $M_{n,l} = 0$.

The number of persons that change to a different category i is

$$\hat{C}_{j;i}(n) = \sum_l I_{j,l}(n)p_{j,l;i} = \sum_{l=1}^{l_{max}} M^i_{n,l} I_{j,l}(1) \tag{23}$$

with

$$M^i_{n,l} = p_{j,l+n-1;i} M_{n,l}\,. \tag{24}$$

For $l + n > l_{max} + 1$ we define $M^i_{n,l} = 0$.

B.6 Method to implement parameter changes

To implement parameter changes from P_{old} to P_{new} at day n_0 we use

$$P(n) = P_{old} + \frac{1}{2}\left(1 + \tanh\left(\frac{n-n_0}{\Delta}\right)\right)(P_{new} - P_{old}) \tag{25}$$

with $\Delta = 1$.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

B.7 Discrete Gamma distribution

To fit distributions we use the discrete Gamma distribution defined as [59]

$$\tilde{p}_l = \frac{F(l) - F(l-1)}{F(l_{max})}\,, \quad l \leq 1 \leq l_{max}\,, \tag{26}$$

with

$$F(x) = \int_0^x \frac{1}{\Gamma(a)b} \left(\frac{y}{b}\right)^{a-1} e^{-\frac{y}{b}} dy\,. \tag{27}$$

The distribution depends on two parameters $a > 0$ and $b > 0$.

B.8 Fitting procedure

For the fitting procedure we used the constrained optimization procedure based on MATLAB's *fminsearch* procedure [60].

Author contributions

JR and DH designed research. JR designed the model, performed analysis, developed fitting procedures and implemented the model in MATLAB. SR and JFT provided clinical data. AP collected, processed and analysed data. JR and DH wrote the manuscript.

Competing interests

The authors declare no competing interests.

Acknowledgements

AP received funding from FRM (SPF201909009284), DH is supported by INSERM Plan Cancer and a Computational Neuroscience NIH-ANR grant. JR is supported by an ANR grant. We thank Dr. Dan Longrois for discussions.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

[1] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, M. Monod, P. N. Perez-Guzman, N. Schmit, L. Cilloni, K. E. C. Ainslie, M. Baguelin, A. Boonyasiri, O. Boyd, L. Cattarino, L. V. Cooper, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, B. Djaafara, I. Dorigatti, S. L. van Elsland, R. G. FitzJohn, K. A. M. Gaythorpe, L. Geidelberg, N. C. Grassly, W. D. Green, T. Hallett, A. Hamlet, W. Hinsley, B. Jeffrey, E. Knock, D. J. Laydon, G. Nedjati-Gilani, P. Nouvellet, K. V. Parag, I. Siveroni, H. A. Thompson, R. Verity, E. Volz, C. E. Walters, H. Wang, Y. Wang, O. J. Watson, P. Winskill, X. Xi, P. G. T. Walker, A. C. Ghani, C. A. Donnelly, S. Riley, M. A. C. Vollmer, N. M. Ferguson, L. C. Okell, S. Bhatt, and I. C. C.-. R. Team, "Estimating the effects of non-pharmaceutical interventions on covid-19 in europe," *Nature*, vol. 584, no. 7820, pp. 257–261, 2020.

[2] H. Salje, C. Tran Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, Y. Le Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boelle, and S. Cauchemez, "Estimating the burden of sars-cov-2 in france," *Science*, vol. 369, no. 6500, pp. 208–211, 2020.

[3] L. Di Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, and V. Colizza, "Impact of lockdown on covid-19 epidemic in île-de-france and possible exit strategies," *BMC Medicine*, vol. 18, no. 1, p. 240, 2020.

[4] K. Leung, J. T. Wu, D. Liu, and G. M. Leung, "First-wave covid-19 transmissibility and severity in china outside hubei after control measures, and second-wave scenario planning: a modelling impact assessment," *The Lancet*, vol. 395, no. 10233, pp. 1382–1393, 2020.

[5] T. N. Starr, A. J. Greaney, A. Addetia, W. W. Hannon, M. C. Choudhary, A. S. Dingens, J. Z. Li, and J. D. Bloom, "Prospective mapping of viral mutations that escape antibodies used to treat covid-19," *Science*, vol. 371, no. 6531, pp. 850–854, 2021.

[6] R. N. Thompson, "Epidemiological models are important tools for guiding covid-19 interventions," *BMC medicine*, vol. 18, no. 1, pp. 1–4, 2020.

[7] M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Iosifidis, M. Agha, and R. Agha, "The socio-economic implications of the coronavirus pandemic (covid-19): A review," *International Journal of Surgery*, vol. 78, pp. 185–193, 2020.

[8] "Key questions for modelling covid-19 exit strategies," *Proceedings of the Royal Society B: Biological Sciences*, vol. 287, no. 1932, 2020.

[9] J. Dehning, J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek, and V. Priesemann, "Inferring change points in the spread of covid-19 reveals the effectiveness of interventions," *Science*, vol. 369, no. 6500, 2020.

[10] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study," *The Lancet*, vol. 395, pp. 689–697, 2021/04/14 2020.

[11] R. C. Reiner, R. M. Barber, J. K. Collins, P. Zheng, C. Adolph, J. Albright, C. M. Antony, A. Y. Aravkin, S. D. Bachmeier, B. Bang-Jensen, M. S. Bannick, S. Bloom, A. Carter, E. Castro,

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

K. Causey, S. Chakrabarti, F. J. Charlson, R. M. Cogen, E. Combs, X. Dai, W. J. Dangel, L. Earl, S. B. Ewald, M. Ezalarab, A. J. Ferrari, A. Flaxman, J. J. Frostad, N. Fullman, E. Gakidou, J. Gallagher, S. D. Glenn, E. A. Goosmann, J. He, N. J. Henry, E. N. Hulland, B. Hurst, C. Johanns, P. J. Kendrick, A. Khemani, S. L. Larson, A. Lazzar-Atwood, K. E. LeGrand, H. Lescinsky, A. Lindstrom, E. Linebarger, R. Lozano, R. Ma, J. Mansson, B. Magistro, A. M. M. Herrera, L. B. Marczak, M. K. Miller-Petrie, A. H. Mokdad, J. D. Morgan, P. Naik, C. M. Odell, J. K. O'Halloran, A. E. Osgood-Zimmerman, S. M. Ostroff, M. Pasovic, L. Penberthy, G. Phipps, D. M. Pigott, I. Pollock, R. E. Ramshaw, S. B. Redford, G. Reinke, S. Rolfe, D. F. Santomauro, J. R. Shackleton, D. H. Shaw, B. S. Sheena, A. Sholokhov, R. J. D. Sorensen, G. Sparks, E. E. Spurlock, M. L. Subart, R. Syailendrawati, A. E. Torre, C. E. Troeger, T. Vos, A. Watson, S. Watson, K. E. Wiens, L. Woyczynski, L. Xu, J. Zhang, S. I. Hay, S. S. Lim, C. J. L. Murray, and I. H. C.-. F. Team, "Modeling covid-19 scenarios for the united states," *Nature Medicine*, vol. 27, no. 1, pp. 94–105, 2021.

[12] T. Britton, F. Ball, and P. Trapman, "A mathematical model reveals the influence of population heterogeneity on herd immunity to sars-cov-2," *Science*, vol. 369, no. 6505, pp. 846–849, 2020.

[13] Q. Griette and P. Magal, "Clarifying predictions for covid-19 from testing data: The example of new york state," *Infectious Disease Modelling*, vol. 6, pp. 273–283, 2021.

[14] Z. Liu, P. Magal, and G. Webb, "Predicting the number of reported and unreported cases for the covid-19 epidemics in china, south korea, italy, france, germany and united kingdom," *Journal of theoretical biology*, vol. 509, p. 110501, 2021.

[15] Z. Liu, P. Magal, O. Seydi, and G. Webb, "Understanding unreported cases in the covid-19 epidemic outbreak in wuhan, china, and the importance of major public health interventions," *Biology*, vol. 9, no. 3, p. 50, 2020.

[16] S. Moore, E. M. Hill, M. J. Tildesley, L. Dyson, and M. J. Keeling, "Vaccination and non-pharmaceutical interventions for covid-19: a mathematical modelling study," *The Lancet Infectious Diseases*, vol. 21, no. 6, pp. 793–802, 2021.

[17] M. J. Keeling, E. Hill, E. Gorsich, B. Penman, G. Guyver-Fletcher, A. Holmes, T. Leng, H. McKimm, M. Tamborrino, L. Dyson, and M. Tildesley, "Predictions of covid-19 dynamics in the uk: short-term forecasting and analysis of potential exit strategies," *PLOS Computational Biology*, vol. 17, pp. 1–20, 01 2021.

[18] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, and N. M. Ferguson, "Estimates of the severity of coronavirus disease 2019: a model-based analysis," *The Lancet Infectious Diseases*, vol. 20, no. 6, pp. 669–677, 2020.

[19] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, N. Davies, A. Gimma, K. van Zandvoort, H. Gibbs, J. Hellewell, C. I. Jarvis, S. Clifford, B. J. Quilty, N. I. Bosse, S. Abbott, P. Klepac, and S. Flasche, "Early dynamics of transmission and control of covid-19: a mathematical modelling study," *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 553–558, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[20] L. Roques, E. K. Klein, J. Papaïx, A. Sar, and S. Soubeyrand, "Impact of lockdown on the epidemic dynamics of covid-19 in france," *Frontiers in medicine*, vol. 7, p. 274, 2020.

[21] L. Roques, E. K. Klein, J. Papaix, A. Sar, and S. Soubeyrand, "Using early data to estimate the actual infection fatality ratio from covid-19 in france," *Biology*, vol. 9, no. 5, p. 97, 2020.

[22] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, and A. Rinaldo, "Spread and dynamics of the covid-19 epidemic in italy: Effects of emergency containment measures," *Proceedings of the National Academy of Sciences*, vol. 117, no. 19, pp. 10484–10491, 2020.

[23] L. Mari, R. Casagrandi, E. Bertuzzo, D. Pasetto, S. Miccoli, A. Rinaldo, and M. Gatto, "The epidemicity index of recurrent sars-cov-2 infections," *Nature communications*, vol. 12, no. 1, pp. 1–12, 2021.

[24] E. Bertuzzo, L. Mari, D. Pasetto, S. Miccoli, R. Casagrandi, M. Gatto, and A. Rinaldo, "The geography of covid-19 spread in italy and implications for the relaxation of confinement measures," *Nature communications*, vol. 11, no. 1, pp. 1–11, 2020.

[25] L. Di Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, and V. Colizza, "Modelling safe protocols for reopening schools during the covid-19 pandemic in france," *Nature communications*, vol. 12, no. 1, pp. 1–10, 2021.

[26] G. Pullano, L. Di Domenico, C. E. Sabbatini, E. Valdano, C. Turbelin, M. Debin, C. Guerrisi, C. Kengne-Kuetche, C. Souty, T. Hanslik, T. Blanchon, P.-Y. Boëlle, J. Figoni, S. Vaux, C. Campèse, S. Bernard-Stoecklin, and V. Colizza, "Underdetection of cases of covid-19 in france threatens epidemic control," *Nature*, vol. 590, no. 7844, pp. 134–139, 2021.

[27] F. Pinotti, L. Di Domenico, E. Ortega, M. Mancastroppa, G. Pullano, E. Valdano, P.-Y. Boëlle, C. Poletto, and V. Colizza, "Tracing and analysis of 288 early sars-cov-2 infections outside china: A modeling study," *PLOS Medicine*, vol. 17, no. 7, pp. 1–13, 2020.

[28] N. Hoertel, M. Blachier, C. Blanco, M. Olfson, M. Massetti, M. S. Rico, F. Limosin, and H. Leleu, "A stochastic agent-based model of the sars-cov-2 epidemic in france," *Nature medicine*, vol. 26, no. 9, pp. 1417–1421, 2020.

[29] F. Balabdaoui and D. Mohr, "Age-stratified discrete compartment model of the covid-19 epidemic with application to switzerland," *Scientific Reports*, vol. 10, no. 1, p. 21306, 2020.

[30] J. Reingruber, A. Papale, and D. Holcman, "Monitoring and predicting sars-cov-2 epidemic in france after deconfinement using a multiscale and age-dependent model," *medRxiv*, 2020.

[31] P. Bosetti, C. T. Kiem, Y. Yazdanpanah, A. Fontanet, B. Lina, V. Colizza, and S. Cauchemez, "Impact of mass testing during an epidemic rebound of sars-cov-2: a modelling study using the example of france," *Eurosurveillance*, vol. 26, no. 1, 2021.

[32] Q. Richard, S. Alizon, M. Choisy, M. T. Sofonea, and R. Djidjou-Demasse, "Age-structured non-pharmaceutical interventions for optimal control of covid-19 epidemic," *PLoS computational biology*, vol. 17, no. 3, p. e1008776, 2021.

[33] "https://geodes.santepubliquefrance.fr."

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[34] "https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19."

[35] G. Béraud, S. Kazmercziak, P. Beutels, D. Levy-Bruhl, X. Lenne, N. Mielcarek, Y. Yazdanpanah, P.-Y. Boëlle, N. Hens, and B. Dervaux, "The french connection: The first large population-based contact survey in france relevant for the spread of infectious diseases," *PLOS ONE*, vol. 10, no. 7, pp. 1–22, 2015.

[36] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, "The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application.," *Annals of Internal Medicine*, vol. 172, no. 9, pp. 577–582, 2020.

[37] E. Lavezzo, E. Franchin, C. Ciavarella, G. Cuomo-Dannenburg, L. Barzon, C. Del Vecchio, L. Rossi, R. Manganelli, A. Loregian, N. Navarin, D. Abate, M. Sciro, S. Merigliano, E. De Canale, M. C. Vanuzzo, V. Besutti, F. Saluzzo, F. Onelia, M. Pacenti, S. G. Parisi, G. Carretta, D. Donato, L. Flor, S. Cocchio, G. Masi, A. Sperduti, L. Cattarino, R. Salvador, M. Nicoletti, F. Caldart, G. Castelli, E. Nieddu, B. Labella, L. Fava, M. Drigo, K. A. M. Gaythorpe, A. R. Brazzale, S. Toppo, M. Trevisan, V. Baldo, C. A. Donnelly, N. M. Ferguson, I. Dorigatti, A. Crisanti, K. E. C. Ainslie, M. Baguelin, S. Bhatt, A. Boonyasiri, O. Boyd, H. L. Coupland, Z. Cucunubá, B. A. Djafaara, C. A. Donnelly, S. L. van Elsland, R. FitzJohn, S. Flaxman, K. A. M. Gaythorpe, W. D. Green, T. Hallett, A. Hamlet, D. Haw, N. Imai, B. Jeffrey, E. Knock, D. J. Laydon, T. Mellan, S. Mishra, G. Nedjati-Gilani, P. Nouvellet, L. C. Okell, K. V. Parag, S. Riley, H. A. Thompson, H. J. T. Unwin, R. Verity, M. A. C. Vollmer, P. G. T. Walker, C. E. Walters, H. Wang, Y. Wang, O. J. Watson, C. Whittaker, L. K. Whittles, X. Xi, N. M. Ferguson, and I. C. C.-. R. Team, "Suppression of a sars-cov-2 outbreak in the italian municipality of vo'," *Nature*, vol. 584, no. 7821, pp. 425–429, 2020.

[38] T. C. Jones, B. Mühlemann, T. Veith, G. Biele, M. Zuchowski, J. Hofmann, A. Stein, A. Edelmann, V. M. Corman, and C. Drosten, "An analysis of sars-cov-2 viral load by patient age," *medRxiv*, 2020.

[39] D. Cereda, M. Tirani, F. Rovida, V. Demicheli, M. Ajelli, P. Poletti, F. Trentini, G. Guzzetta, V. Marziano, A. Barone, M. Magoni, S. Deandrea, G. Diurno, M. Lombardo, M. Faccini, A. Pan, R. Bruno, E. Pariani, G. Grasselli, A. Piatti, M. Gramegna, F. Baldanti, A. Melegaro, and S. Merler, "The early phase of the covid-19 outbreak in lombardy, italy," *arxive*, 2020.

[40] A. Fontanet, R. Grant, L. Tondeur, Y. Madec, L. Grzelak, I. Cailleau, M.-N. Ungeheuer, C. Renaudat, S. F. Pellerin, L. Kuhmel, *et al.*, "Sars-cov-2 infection in primary schools in northern france: A retrospective cohort study in an area of high transmission," *MedRxiv*, 2020.

[41] K. K. Cheng, T. H. Lam, and C. C. Leung, "Wearing face masks in the community during the covid-19 pandemic: altruism and solidarity," *The Lancet*, 2020.

[42] N. Hozé, J. Paireau, N. Lapidus, C. T. Kiem, H. Salje, G. Severi, M. Touvier, M. Zins, X. de Lamballerie, D. Lévy-Bruhl, *et al.*, "Monitoring the proportion of the population infected by sars-cov-2 using age-stratified hospitalisation and serological data: a modelling study," *The Lancet Public Health*, 2021.

[43] K. Kupferschmidt and M. Wadman, "Delta variant triggers new phase in the pandemic," *Science*, vol. 372, no. 6549, pp. 1375–1376, 2021.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[44] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman, "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2)," *Science*, vol. 368, no. 6490, pp. 489–493, 2020.

[45] A. T. Levin, W. P. Hanage, N. Owusu-Boaitey, K. B. Cochran, S. P. Walsh, and G. Meyerowitz-Katz, "Assessing the age specificity of infection fatality rates for covid-19: systematic review, meta-analysis, and public policy implications," *European Journal of Epidemiology*, vol. 35, no. 12, pp. 1123–1138, 2020.

[46] S. M. Moghadas, M. C. Fitzpatrick, P. Sah, A. Pandey, A. Shoukat, B. H. Singer, and A. P. Galvani, "The implications of silent transmission for the control of covid-19 outbreaks," *Proceedings of the National Academy of Sciences*, vol. 117, no. 30, pp. 17513–17515, 2020.

[47] K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell, "Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the diamond princess cruise ship, yokohama, japan," *Euro Surveill*, vol. 25, no. 10, 2020.

[48] H. Streeck, B. Schulte, B. M. Kümmerer, E. Richter, T. Höller, C. Fuhrmann, E. Bartok, R. Dolscheid-Pommerich, M. Berger, L. Wessendorf, M. Eschbach-Bludau, A. Kellings, A. Schwaiger, M. Coenen, P. Hoffmann, B. Stoffel-Wagner, M. M. Nöthen, A. M. Eis-Hübinger, M. Exner, R. M. Schmithausen, M. Schmid, and G. Hartmann, "Infection fatality rate of sars-cov2 in a super-spreading event in germany," *Nature Communications*, vol. 11, no. 1, p. 5829, 2020.

[49] P. Zimmermann and N. Curtis, "Coronavirus infections in children including covid-19: an overview of the epidemiology, clinical features, diagnosis, treatment and prevention options in children," *The Pediatric infectious disease journal*, vol. 39, no. 5, p. 355, 2020.

[50] N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, and R. M. Eggo, "Age-dependent effects in the transmission and control of covid-19 epidemics," *Nature medicine*, vol. 26, no. 8, pp. 1205–1211, 2020.

[51] C. Jiehao, X. Jin, L. Daojiong, Y. Zhi, X. Lei, Q. Zhenghai, Z. Yuehua, Z. Hua, J. Ran, L. Pengcheng, *et al.*, "A case series of children with 2019 novel coronavirus infection: clinical and epidemiological features," *Clinical Infectious Diseases*, vol. 71, no. 6, pp. 1547–1551, 2020.

[52] C. Chen, C. Zhu, D. Yan, H. Liu, D. Li, Y. Zhou, X. Fu, J. Wu, C. Ding, G. Tian, *et al.*, "The epidemiological and radiographical characteristics of asymptomatic infections with the novel coronavirus (covid-19): A systematic review and meta-analysis," *International Journal of Infectious Diseases*, 2021.

[53] L. A. Nikolai, C. G. Meyer, P. G. Kremsner, and T. P. Velavan, "Asymptomatic sars coronavirus 2 infection: Invisible yet invincible," *International Journal of Infectious Diseases*, 2020.

[54] R. M. Viner, O. T. Mytton, C. Bonell, G. Melendez-Torres, J. Ward, L. Hudson, C. Waddington, J. Thomas, S. Russell, F. Van Der Klis, *et al.*, "Susceptibility to sars-cov-2 infection among children and adolescents compared with adults: a systematic review and meta-analysis," *JAMA pediatrics*, vol. 175, no. 2, pp. 143–156, 2021.

[55] J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani, *et al.*, "Changes in contact patterns shape the dynamics of the covid-19 outbreak in china," *Science*, vol. 368, no. 6498, pp. 1481–1486, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.28.21260870; this version posted July 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[56] M. ODriscoll, G. R. Dos Santos, L. Wang, D. A. Cummings, A. S. Azman, J. Paireau, A. Fontanet, S. Cauchemez, and H. Salje, "Age-specific mortality and immunity patterns of sars-cov-2," *Nature*, vol. 590, no. 7844, pp. 140–145, 2021.

[57] J. P. Ioannidis, "Reconciling estimates of global spread and infection fatality rates of covid-19: An overview of systematic evaluations," *European journal of clinical investigation*, vol. 51, no. 5, p. e13554, 2021.

[58] S. Riley, K. Eames, V. Isham, D. Mollison, and P. Trapman, "Five challenges for spatial epidemic models," *Epidemics*, vol. 10, pp. 68–71, 2015.

[59] Chakraborty, "Discrete gamma distributions: Properties and parameter estimations," *Communications in Statistics - Theory and Methods*, vol. 41, no. 18, pp. 3301–3324, 2012.

[60] J. D'Errico, "fminsearchbnd, fminsearchcon (https://www.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd-fminsearchcon)," *MATLAB Central File Exchange*, 2021.