



A stochastic agent-based model of the SARS-CoV-2 epidemic in France

Nicolas Hoertel^{1,2,3}✉, Martin Blachier⁴, Carlos Blanco⁵, Mark Olfson⁶, Marc Massetti⁴, Marina Sánchez Rico^{1,7}, Frédéric Limosin^{1,2,3} and Henri Leleu⁴

Many European countries have responded to the COVID-19 pandemic by implementing nationwide protection measures and lockdowns¹. However, the epidemic could rebound when such measures are relaxed, possibly leading to a requirement for a second or more, repeated lockdowns². Here, we present results of a stochastic agent-based microsimulation model of the COVID-19 epidemic in France. We examined the potential impact of post-lockdown measures, including physical distancing, mask-wearing and shielding individuals who are the most vulnerable to severe COVID-19 infection, on cumulative disease incidence and mortality, and on intensive care unit (ICU)-bed occupancy. While lockdown is effective in containing the viral spread, once lifted, regardless of duration, it would be unlikely to prevent a rebound. Both physical distancing and mask-wearing, although effective in slowing the epidemic and in reducing mortality, would also be ineffective in ultimately preventing ICUs from becoming overwhelmed and a subsequent second lockdown. However, these measures coupled with the shielding of vulnerable people would be associated with better outcomes, including lower mortality and maintaining an adequate ICU capacity to prevent a second lockdown. Benefits would nonetheless be markedly reduced if most people do not adhere to these measures, or if they are not maintained for a sufficiently long period.

The COVID-19 pandemic is a major global health threat. Owing to lack of a vaccine or an effective treatment, many European countries have responded with non-pharmaceutical interventions (NPIs) to diminish viral transmission by reducing contact rates¹. These measures include physical distancing; wearing a face covering when outside the home; closing schools, churches, bars and other social venues, as well as all stores except grocers and pharmacies; screening symptomatic people and their contacts; and full population lockdown. Countries in which these interventions were implemented early in the epidemic have experienced diminished incident case numbers and reduced peak healthcare demand and deaths. However, the epidemic could rebound when these measures are relaxed², as only a limited proportion of the European population will have been infected³, leading to the possibility of second or even multiple repeated lockdowns. Such measures impose harmful burdens on the population and the global economy⁴, and are difficult to tolerate for extended periods⁵. Therefore, the evaluation of alternative NPIs that could be implemented when lockdown is lifted to potentially avoid a second epidemic peak and lockdown is urgently needed^{6,7}.

On 17 March 2020, France ordered a general lockdown, asking all non-essential retailers and services to close, and the general population to stay confined at home and to adhere to physical distancing when outside the home. These measures reduced case counts and strain on the healthcare system, and the lockdown was lifted on 11 May. In France, as in many other countries where lockdowns have been ordered, there is intense debate over which post-lockdown strategies should be implemented to avoid an epidemic rebound. Model-based predictions of the public-health impact of competing NPIs on the epidemic course can help inform policy decisions.

We present results of a stochastic agent-based microsimulation (ABM) model^{8,9} of the COVID-19 epidemic in France. We projected the potential impact of competing NPIs on the cumulative incidence of the disease and mortality, and on ICU-bed occupancy. Specifically, we evaluated lockdown extension from 8 to 16 weeks and post-lockdown measures, including physical distancing, mask-wearing and shielding of the population at risk for severe COVID-19 infection. Because of several uncertainties that determine virus-transmission risk, such as number of undiagnosed asymptomatic cases and contamination risk⁶, the analysis followed recommendations for improving predictive mathematical models of the COVID-19 pandemic¹⁰ and was based on a calibration that accounts for several disease-transmission parameters within constraints defined by the contact matrix and known disease parameters. A summary of the main findings, limitations and implications of the model is shown in Table 1. Detailed information on study design is included in the Reporting Summary.

The model calibrated well, based on a visually good fit between observed and model-predicted daily ICU admissions, ICU-bed occupancy, daily mortality and cumulative mortality, as confirmed by R^2 and Pearson's R estimates, which were higher than 0.87 and 0.93, respectively, except for daily ICU admissions. Similarly, the model validated well prospectively, based on a visually good fit between observed and model-predicted daily ICU admissions, ICU-bed occupancy, daily mortality and cumulative mortality between 15 April and 15 May, as confirmed by R^2 and Pearson's R estimates, which were higher than 0.80 and 0.90, respectively (Fig. 1 and Supplementary Table 1). The Cochran–Mantel–Haenszel test comparing observed and model-predicted age distribution of deceased people was non-significant ($\chi^2=0.34$, $P=0.55$). The R_0 of COVID-19 observed in the model at the date of the lockdown was 3.1, consistent with findings of a review¹¹ suggesting that R_0 estimates would fall between 1.40 and 6.49, with a median of 2.79. Furthermore, we projected about 10.9% (uncertainty range (first to

¹AP-HP, Centre, DMU Psychiatrie et Addictologie, Corentin Celton Hospital, Issy-les-Moulineaux, France. ²INSERM U1266, Paris, France. ³Paris University, Paris, France. ⁴Public Health Expertise, Paris, France. ⁵National Institute on Drug Abuse, Bethesda, MD, USA. ⁶Columbia University/New York State Psychiatric Institute, New York, NY, USA. ⁷Universidad Complutense de Madrid, Campus de Somosaguas, Pozuelo de Alarcón, Spain.

✉e-mail: nico.hoertel@yahoo.fr

Table 1 | Policy summary

Background	Many European countries have responded to the COVID-19 threat by nationwide implementation of protection measures and lockdown; however, even assuming that population immunity will build up through the epidemic, rebound could occur once these measures are relaxed, possibly leading to a second or multiple repeated lockdowns. We built a stochastic ABM model of the COVID-19 epidemic in France to assess the potential impact of different post-lockdown measures to avoid an epidemic rebound, including physical distancing, mask-wearing and shielding those who are most vulnerable to severe COVID-19 infection.
Main findings and limitations	While lockdown is effective in containing the viral spread, once lifted, regardless of duration, it would be unlikely to prevent a rebound. We show that although physical distancing and mask-wearing are effective in slowing the epidemic and in reducing mortality, these measures alone would be insufficient to prevent overwhelming ICUs and a second lockdown. However, when coupled with shielding of vulnerable people, implementation of these measures would be associated with better outcomes, including lower mortality and maintaining an adequate ICU capacity. Benefits would nonetheless be reduced if these measures were not adhered to by most people or not maintained for a sufficiently long period. Our model relies on existing knowledge and current assumptions, as all modeling studies do. Specifically, it is likely not capturing all components driving viral transmission, which might behave differently than before lockdown.
Policy implications	Our findings indicate that post-lockdown physical distancing and wearing of masks for the whole population, coupled with continued shielding of vulnerable people, could lower mortality and prevent the need for a second lockdown in France.

last quartile), 9.4–12.5%) would have been infected by 11 May, in line with, albeit slightly higher than, estimates from a prior study conducted in France³.

Lockdown is highly effective at containing viral spread, but we projected that it would be unlikely by itself to prevent a second epidemic peak once lifted, regardless of its duration. On the basis of our model, the duration of lockdown alone (8 or 16 weeks) was not associated with a reduced cumulative COVID-19 incidence or mortality, resulting in a similar, albeit delayed, overwhelming of ICUs, likely to lead to a second lockdown (Extended Data Fig. 1 and Supplementary Table 2).

We found that maintaining physical distancing after the end of the lockdown would be associated with substantial slowing of the epidemic, as shown by a flattening of the cumulative incidence curve, and a 33% decrease in cumulative mortality, after 8-week lockdown (Fig. 2 and Supplementary Table 2). Combining physical distancing and mask-wearing further flattened the epidemic curve and was associated with a 47% reduction of mortality compared with that associated with the absence of post-lockdown protection measures. However, although effective in slowing the epidemic and in reducing mortality, we found that this combination of measures might also be ineffective in preventing a second epidemic peak, likely to result in ICU-bed capacity being exceeded and lead to a second lockdown.

We projected that shielding at-risk people until the end of the epidemic (estimated in our model at 38 weeks after the lifting of lockdown with this scenario), in addition to post-lockdown physical distancing and mask-wearing, would be associated with substantial slowing of the epidemic, as shown by a strong flattening of the cumulative incidence curve and a substantial decrease in mortality of 62% compared with that with post-lockdown physical distancing and mask-wearing only, and of 80% compared with that with absence of specific post-lockdown measures (Fig. 3 and Supplementary Table 2).

Combining these three interventions would prevent overwhelming of ICU capacity and would substantially reduce mortality, but only if the interventions were maintained for a sufficiently long period and were adhered to by most people. For example, interrupting the shielding of vulnerable people at 16 weeks after the lifting of lockdown or having only 50% of the vulnerable population shielded would result in a substantially lower decrease in mortality and an increased risk of ICU overwhelming than would a continued shielding of all vulnerable people (Fig. 4 and Supplementary Table 2).

Individually varying each model parameter value by $\pm 20\%$ would change the cumulative incidence by at most 4,000 per 100,000 people, mortality by 10 per 100,000 and ICU-bed occupancy by 25 per

100,000, suggesting the robustness of our results (Extended Data Figs. 2 and 3).

Sensitivity analyses examining the impact of varying values of parameters with the greatest uncertainty, including contamination risk, proportion of undiagnosed cases, and the efficacy of masks showed that coupling physical distancing and mask-wearing with the shielding of vulnerable people would lead to better outcomes than without shielding. These include decreased incidence, mortality and ICU-bed occupancy (Extended Data Figs. 4–6 and Supplementary Table 3).

We have proposed an ABM of the epidemic of COVID-19 in France to predict the potential impact of different post-lockdown measures on the course of the epidemic. The model calibrated and validated well, and variation of the values of individual model parameters by $\pm 20\%$ had little impact on outcome estimates. While lockdown is a highly effective means of containing viral spread, it would by itself regardless of duration be unlikely to prevent a rebound and the need for a second lockdown. Adhering to both physical distancing and mask-wearing would be effective in slowing down the epidemic and in reducing mortality, but might be insufficient to prevent an overwhelming of ICU-bed capacity. However, these measures, when coupled with shielding vulnerable people, could be associated with better outcomes than such measures without shielding, including a lower cumulative incidence, mortality and number of ICU beds needed, that are sufficient to prevent an overwhelming of ICU-bed capacity and a second lockdown. Benefits would nonetheless be substantially reduced if the measures were not applied by most people or not maintained for a sufficiently long period. When varying values of parameters with the greatest uncertainty, such as contamination risk, proportion of undiagnosed cases, or efficacy of masks in multiple sensitivity analyses, this strategy was quite effective in reducing incidence, mortality and ICU-bed occupancy in all scenarios.

Coupling NPIs with shielding vulnerable people would be associated with better outcomes than implementing physical distancing and the wearing of masks alone. Improved outcomes include a lower mortality and number of ICU beds needed, such that enough beds would be available to avoid triggering a second lockdown. Specifically, mortality would be reduced up to 80% if this scenario were applied by 100% of vulnerable people until the end of the epidemic compared with mortality following the absence of specific measures after lifting the lockdown.

In our study, shielding implied that vulnerable individuals, although having to stay home except for grocery shopping, could see their family at home but with protection measures, including physical distancing and mask-wearing. Because prolonged lockdowns

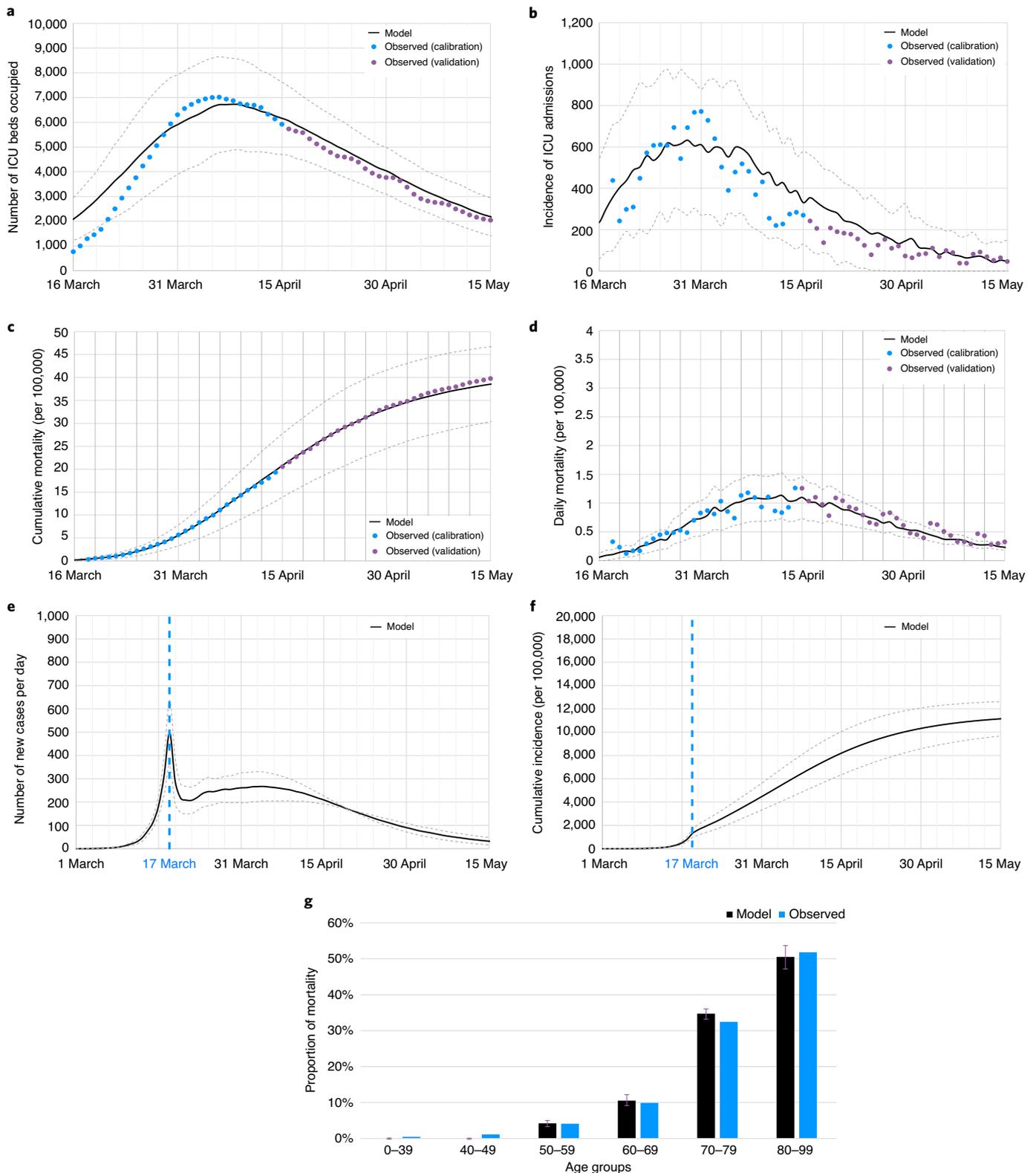


Fig. 1 | Model-predicted and observed ICU-bed occupancy, daily ICU admissions, cumulative mortality, daily mortality and age distribution of deceased people, and model-predicted daily number of new cases and cumulative incidence. **a–f**, Model-predicted and observed ICU-bed occupancy (**a**), daily ICU admissions (**b**), cumulative mortality (**c**), daily mortality (**d**) and age distribution of deceased people (**g**), and model-predicted daily number of new cases (**e**) and cumulative incidence (**f**). The dotted lines in panels **a–f** and error bars in panel **g** represent the uncertainty range (first to last quartile) stemming from the uncertainty in the parameter values. On 17 March, France ordered a general lockdown.

are associated with increased risks of psychological and medical complications^{5,12}, we believe that our less-stringent strategy would have increased likelihood of being followed by most individuals,

while ensuring there is adequate protection of the vulnerable population, whose proportion is estimated at more than three out of ten people in France. However, it is crucial to provide a clear rationale

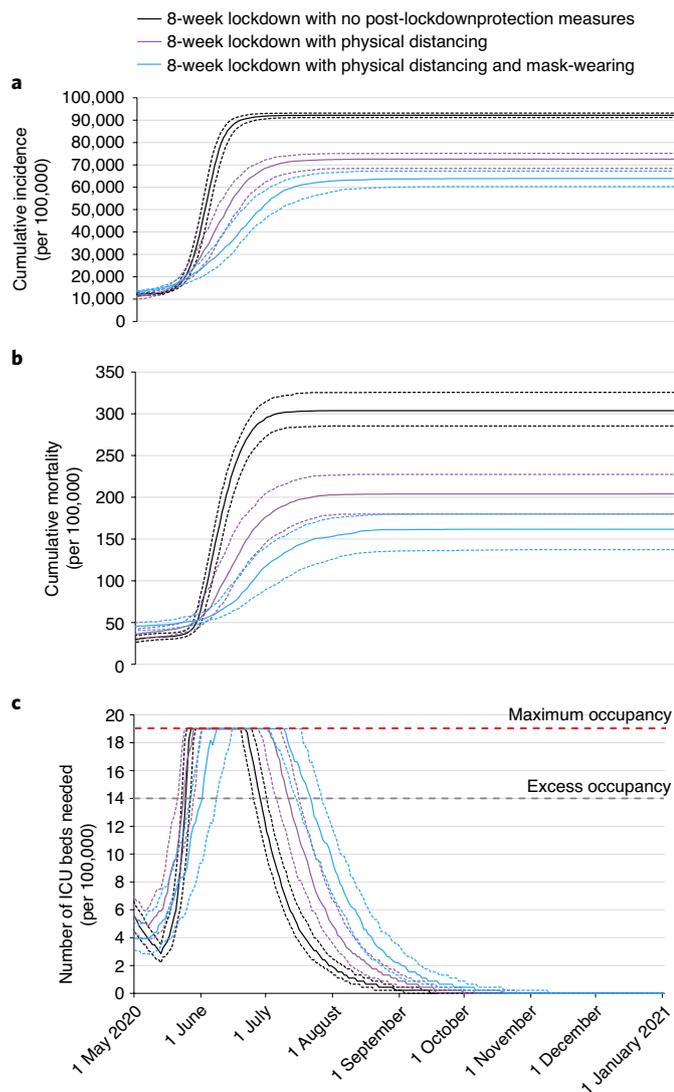


Fig. 2 | Model-predicted cumulative incidence, mortality, and number of ICU beds needed associated with post-lockdown physical distancing and mask-wearing for the general population. a–c, Model-predicted cumulative incidence (**a**), cumulative mortality (**b**) and number of ICU beds needed (**c**) associated with post-lockdown physical distancing and mask-wearing for the general population. The dotted lines represent the uncertainty range (first to last quartile) stemming from the uncertainty in the parameter values.

for and information about these measures, to appeal to altruism by reminding people of the benefits to wider society and individual responsibility and to provide sufficient medical resources and adequate healthcare access.

By slowing down viral spread down in the vulnerable population to a greater extent than in the healthier population, our results suggest that most infected people would likely be individuals who are at lower risk of developing severe or critical symptoms¹³, and would be adequately treated since ICU capacity would not be overwhelmed, even during the peak incidence. In the absence of an effective and safe vaccine, an emphasis should be placed on policies that protect the most-vulnerable population in the hopes that herd immunity¹⁴ will be achieved in the less vulnerable people as a byproduct of such measures, although this is not the ultimate goal. Indeed, the consequences of herd immunity are serious, as a substantial fraction of the less-vulnerable population would need to become infected

before reaching the end of the epidemic, estimated by our model at 38 weeks after the lockdown lifting if all individuals adhere to physical distancing and mask-wearing.

Our study has several limitations. First, as with all modeling studies, we rely on existing knowledge and current assumptions. Specifically, our model is likely not capturing all components driving viral transmission, which might behave differently after than before lockdown. For example, changes in the force of virus transmission due to seasonality, which might not depend only on direct effect of climate variables (that is, humidity and temperature), are not included in the model^{7,15}. Second, the model was calibrated on the diagnosis and mortality rates available from Santé Publique France and Institut Pasteur. However, these parameters might be underestimates, as asymptomatic undiagnosed patients are likely responsible for a large hidden epidemic. Nevertheless, the observed differences across scenarios remained consistent across sensitivity analyses, supporting the robustness of our conclusions.

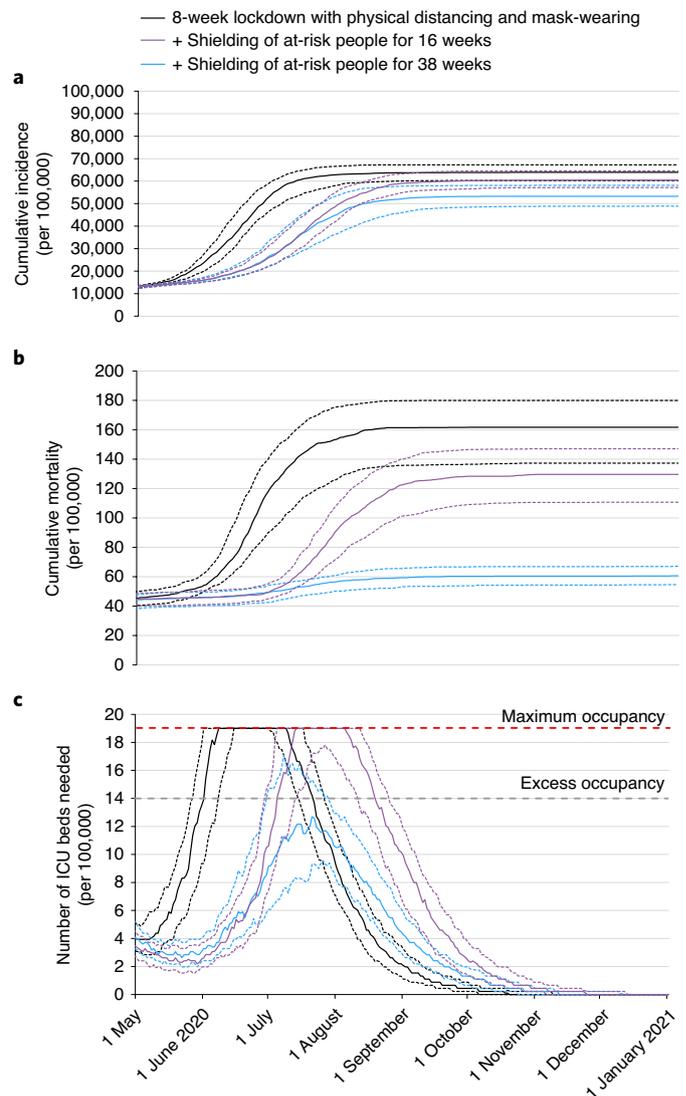


Fig. 3 | Model-predicted cumulative incidence, mortality, and number of ICU beds needed associated with post-lockdown shielding of individuals at risk for severe infection. a–c, Model-predicted cumulative incidence (**a**), cumulative mortality (**b**), and number of ICU beds needed (**c**) associated with post-lockdown shielding of individuals at risk for severe infection. The dotted lines represent the uncertainty range (first to last quartile) stemming from the uncertainty in the parameter values.

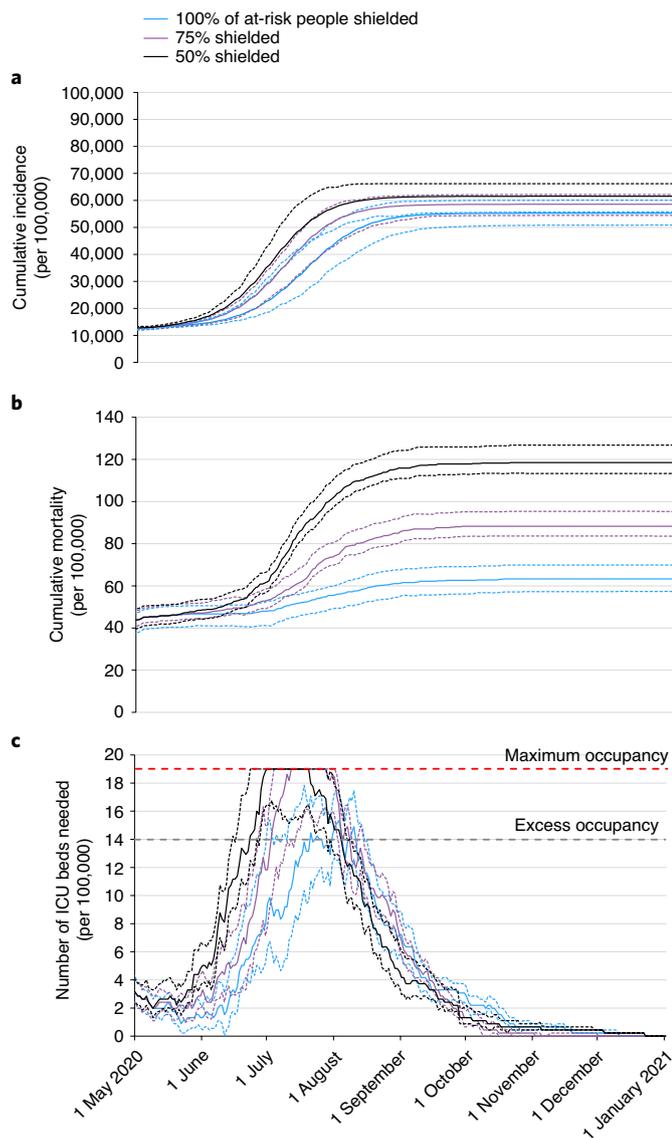


Fig. 4 | Model-predicted cumulative incidence, mortality, and number of ICU beds needed according to the proportion of at-risk people shielded.

a–c. Model-predicted cumulative incidence (**a**), cumulative mortality (**b**) and number of ICU beds needed (**c**) according to the proportion of at-risk people shielded. The dotted lines represent the uncertainty range (first to last quartile) stemming from the uncertainty in the parameter values.

Third, following standard assumptions, we considered that infected people could develop immunity for at least several months. However, post-COVID-19 immunity length remains unknown. Fourth, the impact of many mitigation measures depends on how people react and adhere to them, which is likely to vary across segments of the populations. Finally, our results should not be interpreted as absolute numbers but rather as differences in trends according to the mitigation measures applied¹⁶.

SARS-CoV-2 represents a major public-health threat in France and worldwide. Post-lockdown physical distancing and wearing of masks for the whole population, coupled with shielding of vulnerable people, could substantially lower mortality and prevent a second lockdown. If these measures are adhered to by most people and maintained for a sufficiently long period, they might provide time for herd immunity to become progressively established in the less-vulnerable population, in the absence of an effective vaccine.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-020-1001-6>.

Received: 2 May 2020; Accepted: 29 June 2020;

Published online: 14 July 2020

References

- Bootsma, M. C. & Ferguson, N. M. The effect of public health measures on the 1918 influenza pandemic in US cities. *Proc. Natl Acad. Sci. USA* **104**, 7588–7593 (2007).
- Adam, D. Special report: the simulations driving the world's response to COVID-19. *Nature* **580**, 316–318 (2020).
- Salje, H. et al. Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020).
- Nicola, M. et al. The socio-economic implications of the coronavirus pandemic (COVID-19): a review. *Int. J. Surg.* **78**, 185–193 (2020).
- Brooks, S. K. et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet* **395**, 912–920 (2020).
- Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* **395**, 931–934 (2020).
- Kissler, S., Tedijanto, C., Goldstein, E., Grad, Y. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the post-pandemic period. *Science* **368**, 860–868 (2020).
- Perez, L. & Dragicevic, S. An agent-based approach for modeling dynamics of contagious disease spread. *Int. J. Health. Geogr.* **8**, 50 (2009).
- Venkatramanan, S. et al. Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics* **22**, 43–49 (2018).
- Jewell, N. P., Lewnard, J. A. & Jewell, B. L. Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. *JAMA* **323**, 1893–1894 (2020).
- Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel. Med.* **27**, taaa021 (2020).
- Chevance, A. et al. Ensuring mental health care during the SARS-CoV-2 epidemic in France: a narrative review. *Encephale* **46**, 193–201 (2020).
- Williamson, E. et al. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. Preprint at *medRxiv* <https://doi.org/10.1101/2020.05.06.20092999> (2020).
- John, T. J. & Samuel, R. Herd immunity and herd effect: new insights and definitions. *Eur. J. Epidemiol.* **16**, 601–606 (2000).
- Baker, R. E., Yang, W., Vecchi, G. A., Metcalf, C. J. E. & Grenfell, B. T. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science* <https://doi.org/10.1126/science.abc2535> (2020).
- Hoertel, N. et al. Facing the COVID-19 epidemic in NYC: a stochastic agent-based model of various intervention strategies. Preprint at *medRxiv* <https://doi.org/10.1101/2020.04.23.20076885> (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Agent-based model (ABM). Following previously described methods^{8,16}, we built a stochastic ABM model of the COVID-19 epidemic in France. Agent-based models place equal emphasis on real data and modeling⁹. ABM models designed for infectious diseases include three key components: (1) a realistic synthetic population generated with demographic characteristics and household structure representative of the studied population, (2) a social contact network among the individuals in the population, each with a geolocalized activity sequence over the day, taking into account co-location probability and duration, and (3) a disease model, which translates the edge weights in the social contact network into infection probability of the edge over the day⁹. This framework includes a flexible individual-based approach that can capture an emergent phenomenon with complex interactions between individuals in an heterogeneous population⁹. It allows policymakers to define measures at the individual and societal level, especially when they have to deal with limited understanding of the disease and the uncertainty of the effects of the measures, as is the case for the COVID-19 epidemic, and epidemiologists to describe the characteristics of the disease pathogen and simulate the infectious-disease evolution on a realistic synthetic population⁹.

The model included 194 parameters related to French population characteristics ($n=140$), social contacts ($n=33$) and SARS-CoV-2 characteristics ($n=21$), summarized in Supplementary Table 4. Parameter values on population characteristics were based on data from the French National Statistical Institute (INSEE) and Santé Publique France. Parameters related to social contacts were based on prior studies ($n=11$) or assumptions when no data were available ($n=22$). Finally, parameters on disease characteristics were based on data from Institut Pasteur and London Imperial College, except for two unknown key parameters of the epidemic: contamination risk and proportion of undiagnosed COVID-19 cases, which were simultaneously estimated through model calibration. Source code for the model has been deposited in a recognized public source code repository (GitHub, <https://github.com/henrileu/covid19>).

Individuals' characteristics. The model was built to reproduce household composition and age structures of the French general population. Households were distributed on a square grid representing a geographical area approximating France. Based on age- and sex-stratified national estimates, all individuals were attributed a probability of having one or multiple conditions increasing the risk of severe SARS-CoV-2 infection¹³, including obesity, diabetes, chronic cardiac diseases and chronic respiratory diseases. Individuals aged over 65 years or those with a least one of these conditions were considered to be at higher risk of developing severe COVID-19 infection¹³. On the basis of the age distribution and national estimates of these conditions in France included in our model, this group would represent 36.4% of the French general population, including 21.0% aged over 65 years and 15.4% of younger people having at least one of these conditions (Supplementary Fig. 1).

Social contacts. Given the complexity of modeling social contacts, we simplified contacts at both individual and household levels to model different types of social contacts experienced during the day¹⁷. Social contacts included close contacts for a prolonged duration with a small number of individuals, such as intrafamilial contacts, or people met at school or work. They also comprised less-frequent and less-prolonged contacts with a finite set of individuals, such as friends or extended family members. Finally, they included brief contacts with individuals in centralized locations, such as grocery shopping, or in more remote locations, such as when using public transport. Contacts were defined by their average duration (in minutes), their average distance (in meters), their frequency and the number of individuals involved^{18–21}.

For intrafamilial contacts, it was assumed that their average duration was 6 h per day at a 1-m distance every day for all household members. For contacts at school, outside the lockdown period during which these contacts were considered null, average duration was 6 h at an average 2-m distance, 5 d per week, for all classmates. Classmates were identified as children of the same age living in a similar location to represent the geographic clustering of schools. It was assumed that the average class size was 30 students.

For contacts at work, outside the lockdown period during which these contacts were considered null, average contact duration with colleagues was assumed to be 7.5 h at a 2-m distance, 5 d per week. Only employed individuals aged 20 to 65 years had work-related contacts. We distinguished between small companies with ten or fewer employees and regular or large ones. Individuals working in small companies had two colleagues on average, while employees of regular or large companies had an average of ten colleagues. The number of colleagues was randomly drawn from a Poisson distribution. Work colleagues were identified at random within the city grid.

For friends and family contacts, outside the lockdown period during which these contacts were considered null, it was assumed that the average duration was 180 min at a 1-m distance, with 1 meeting each week on average. Outside the lockdown period, it was also considered that friend and family contacts occurred between households, for example, a couple with children could visit a friend's or grandparent's household.

Social networks were created on the basis of methods described by Gilbert et al.²² with a distance of 22 (Poisson distributed) in order to incorporate key aspects of social networks, such as the different sizes of personal networks, high clustering, positive assortment of degree of connectivity and low density.

Individuals were considered to visit the closest grocery store from their location 1.2 times per week, and to meet an average of 5 people (Poisson distributed). Grocery stores were uniformly distributed throughout the city grid on the basis of grocery stores' density in France, estimated at 29.3 per 100,000 inhabitants. Outside the lockdown period, contacts when going out of the home were limited to cultural activities such as museum, sport, music or cinema events. It was assumed that contacts in restaurants or bars were captured through the friend and family contacts. The average number of times the family went out per year (Poisson distributed) was based on ticket sales' from the French Ministry of Culture. Attendance at any public event was associated with a 120-min duration at a 2-m distance with an average of 5 individuals (Poisson distributed) randomly identified in the city grid. Finally, for public transport, we considered that all individuals used public transport 1.7 times per week for shopping or seeing family or friends. Workers were assumed to use public transport five times a week, twice a day (Poisson distributed). During public transport, a 30-min average duration at a 1-m distance from a mean number of 3–5 individuals (Poisson distributed) randomly identified in the city grid was assumed.

It was also considered that the first patients with COVID-19 were individuals infected via international travel. Thus, outside the lockdown period, individuals could become infected through international contacts over time at a rate based on the frequency of infected patients that were initially diagnosed in France.

SARS-CoV-2 characteristics. SARS-CoV-2 characteristics were based on reports from Santé Publique France, Institut Pasteur and the London Imperial College.

Two key uncertainties about COVID-19 are the proportion of infected individuals that are not diagnosed and the risk of contamination. These parameters were simultaneously estimated through model calibration, based on daily mortality data until 15 April, the cumulative number of cases diagnosed on 15 April and the assumption of a cumulative incidence (diagnosed + undiagnosed) of 1 in 100 diagnosis rate.

That assumption was based on a recent study³ suggesting a high number of undiagnosed infections and projecting 2.8 million (range, 1.8–4.7) people, that is 4.4% (range: 2.8–7.2) of the general population, will have been infected by 11 May in France.

The risk of contamination was assumed to be highest at the onset of symptoms and to decrease with time. To take into account the risk of transmission before developing symptoms²³, it was assumed that infected individuals were contagious starting one day after infection, albeit with a contagiousness that decreases exponentially the further away from onset. An exponential function was chosen because it fitted well with the dynamics of viral replication, based on these assumptions. Individuals who recovered were assumed to have acquired immunity against the virus and were no longer at risk of infection.

It was assumed that individuals with no or light symptoms (such as stomach pain or nausea) were not diagnosed, except if they were traceable contacts (that is, intrafamilial, work, school) of diagnosed patients, and that all individuals with mild, severe or critical symptoms were diagnosed. To reflect these assumptions, among infected individuals, the probability of being asymptomatic or lightly symptomatic in the model was set at 95% in children below age 10, since very few children have been diagnosed with COVID-19, and was assumed to decrease linearly with age. The slope of this decrease was calibrated to show a cumulative incidence (diagnosed + undiagnosed) of 1 in 100 diagnosis rate.

The probabilities of hospital admission (in case of severe symptoms), ICU admission (in case of critical symptoms) and death were obtained from data from Institut Pasteur and a recent large cohort study¹³. The probabilities of ICU admission and death were stratified by age and adjusted for comorbidities, including obesity, diabetes, chronic cardiac diseases and chronic respiratory diseases, based on hazard ratios calculated using data from Institut Pasteur and a recent large cohort study.

Delays between infection, symptom onset, hospital admission, ICU admission, death and recovery are detailed in Supplementary Table 4. Delays were randomly assigned based on the Weibull distribution.

The number of ICU beds needed over time was compared with the number of ICU beds available in France, estimated at about 5,300 beds before the epidemic. However, following healthcare system reorganization, the number of ICU beds has reached a total of 14,000 ICU beds in April. Because patients may require intensive care for other reasons than COVID-19 (for example, stroke, myocardial infarction), we considered that excess ICU-bed occupancy corresponds to full occupancy of newly created ICU beds (that is, 8,700 beds or 14 ICU beds per 100,000 people), and that the maximum number of ICU beds available for patients with critical symptoms of COVID-19 is 19 ICU beds per 100,000 inhabitants, assuming that they occupy 100% of newly created ICU beds and 75% of pre-existing ICU beds (that is, 12,675 ICU beds). Patients requiring ICU-level care with no available beds were assumed to have a 100% probability of dying.

Outcomes. Outcomes included cumulative incidence, cumulative mortality and number of ICU beds needed.

Interventions. In all scenarios, we reproduced protection measures that have been implemented in France from the beginning of the epidemic until 17 May. These measures included a general lockdown period from 17 March to 11 May, during which individuals had no contacts with other people, except prolonged contacts with intrafamilial members living in the same house and brief contacts with individuals at random in grocery stores and in streets. Restrictions for school, work and public transport were lifted on 11 May. Based on opening statements from the French government, we considered that (1) restaurants and bars, and museums will remain closed until 2 June, and (2) attendance at cinemas and public events will be authorized on 22 July. We took into account efforts to trace contacts of diagnosed patients. All intrafamilial, friend and family, work and school contacts of a diagnosed patient in the previous days were assumed to have been tested with reverse transcription with PCR (RT-PCR) after an average delay of 2 d, representing the delay of the investigation. During this delay, infected contacts could further spread the infection. On the basis of prior work²⁴, the sensitivity of RT-PCR to detect COVID-19 cases was considered to be 71%. People met in grocery stores or in public transports were assumed to be untraceable. Finally, given the uncertainty of the efficacy of masks for preventing COVID-19 (ref. ²⁵), we assumed that risk of transmission would be decreased by 47% if all individuals were wearing masks, based on a recent meta-analysis on the efficacy of masks in preventing respiratory virus transmission among non-healthcare workers²⁶. On the basis of the transmission equation used in the model (that is, risk of being infected = $1 - (1 - \text{probability}_{\text{transmission}} / \text{distance}^2)^{\text{duration of contact}}$), we estimated that physical distancing (that is, keeping at least 2 m apart) would decrease transmission risk by an average of 30%, and that both measures would reduce risk multiplicatively by 63% (ref. ²⁶).

We successively examined the following scenarios, using a 'stepped care' approach:

- (1) The natural course of the epidemic if no lockdown had been ordered.
- (2) Two durations of lockdown: the duration in France (8 weeks) and 16 weeks.
- (3) Post-lockdown protection measures for all individuals, including physical distancing and mask-wearing.
- (4) Post-lockdown shielding of individuals at risk of severe SARS-CoV-2 infection, that is, individuals aged over 65 years or having a comorbidity, including obesity, diabetes, chronic cardiac diseases and chronic respiratory diseases. Shielding implied that individuals always stayed home except for grocery shopping, could see their family at home but with protection measures (that is, physical distancing and masks), and did not attend any gatherings. We also evaluated effects of the duration of this intervention and the proportion of vulnerable people shielded on medical outcomes.

Calibration. Of 194 model parameters, 192 were estimated from the literature ($n = 170$) or by assumption ($n = 22$), and 2 parameters, contamination risk (per min per m² of contact) and proportion of undiagnosed cases, could not be estimated and were therefore calibrated. Calibration fits the value of the calibrated parameters to best reproduce the observed epidemic. We used the Nelder and Mead simplex method⁹ to calibrate these two parameters simultaneously, based on daily mortality data until 15 April with a loss function based on the Kolmogorov–Smirnov statistic, the cumulative number of cases diagnosed on 15 April and the assumption of a ratio of diagnosed/undiagnosed cases of 1 in 100, with a loss function based on the geometric distance. The algorithm was stopped after 300 iterations or if the loss function fell under a minimum value. The calibration process was repeated 30 times with different initial values to check for local minimums. The calibration process included strategies in place in France prior to 15 April, including contact tracing followed by a general lockdown on 17 March. Because the initial phase of an epidemic is largely dependent on stochastic factors, calendar dates were aligned to the date of the lockdown by using the cumulative number of diagnosed cases at this date as a reference.

Statistical analyses. The stochastic ABM was run for 360 d on 500,000 individuals. The results were based on an average of 200 simulations. Analyses were performed on 17 May using data for model parameters until 15 April. Results were extrapolated to the French population of 67 million people. We provided uncertainty measures by using 200 bootstrap samples based on the random variation of all non-calibrated parameters simultaneously, either within a 95% confidence interval for parameters estimated from the literature or within a $\pm 20\%$ interval if the parameter was assumed.

First, we examined whether the model had adequate calibration, that is, whether it was able to adequately reproduce retrospectively the 15 April daily ICU admissions, ICU-bed occupancy, daily mortality and cumulative mortality. Next, we examined whether the model had adequate validation. This involved whether it was able to predict prospectively, daily from 15 April until 15 May, ICU admissions, ICU-bed occupancy and mortality, and cumulative mortality, using data from before 15 April for estimating model parameters. We also tested the difference between model-predicted and observed age distribution among deceased people, a parameter not included in the model during the calibration. Finally, we compared

the proportion of the population infected on 11 May and the approximated value for the basic reproduction number (R_0) of COVID-19 observed in the model on 17 March, with estimates from prior published reports^{3,11}.

In our model, the value for the R_0 COVID-19 was approximated using the observed average number of new individuals infected by each single infected individual from the beginning of the epidemic until 17 March (representing about 30 d in the model). The numbers of infected and infecting people were estimated using the model.

For both calibration and validation, model-predicted and observed daily ICU admissions, ICU-bed occupancy, daily mortality and cumulative mortality were compared using R^2 , Pearson's R and visual comparison of the curves. For validation, model-predicted and observed age distribution among deceased people were compared using a Cochran–Mantel–Haenszel test and visual comparison.

We examined the robustness of our results by evaluating the impact on outcomes of varying successively individual parameter values by $\pm 20\%$, without recalibrating the model. Given the uncertainty of the two calibrated parameters and of the efficacy of masks, we performed sensitivity analyses and evaluated the impact of varying values of these parameters on the predicted epidemic course. The choice of the parameter values is based on the results of a recent meta-analysis on the efficacy of masks in preventing respiratory virus transmission among non-healthcare workers²⁶. We used the numbers at the upper and lower end of the 95% confidence interval reported by that meta-analysis, corresponding to 36% and 79% risk reduction. These analyses are detailed for the intervention 'post-lockdown physical distancing and mask-wearing for the general population' and 'shielding of at-risk individuals in addition of physical distancing and mask-wearing'.

Data were collected using Microsoft Excel. The model was performed using C++ and statistical analyses were conducted using SAS software version 9.4. The threshold for statistical significance was a priori fixed at two-sided $P < 0.05$.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Population demographics, data on social contacts and on SARS-CoV-2 characteristics and observed data for COVID-19 incidence, mortality and ICU-bed occupancy in France are available from the publicly available data of French National Statistical Institute (INSEE) (<https://insee.fr/fr/accueil>), Santé Publique France (<https://www.santepubliquefrance.fr/>), Institut Pasteur (<https://www.pasteur.fr/fr>), London Imperial College (<https://www.imperial.ac.uk/>) and Ministère des Solidarités et de la Santé (<https://drees.solidarites-sante.gouv.fr/etudes-et-statistiques/>) for the year 2020, Direction Générale des Entreprises (DGE) (<https://www.entreprises.gouv.fr/>) for the year 2018 and Ministère de la Culture for the year 2015 (<https://www.culture.gouv.fr/>).

Code availability

The source code of the model has been deposited in a recognized public source code repository (GitHub, <https://github.com/henrileu/covid19>).

References

- Petri, G. & Barrat, A. Simplicial activity driven model. *Phys. Rev. Lett.* **121**, 228301 (2018).
- Eubank, S. et al. Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004).
- Riley, S. Large-scale spatial-transmission models of infectious disease. *Science* **316**, 1298–1301 (2007).
- Epstein, J. M. Modelling to contain pandemics. *Nature* **460**, 687 (2009).
- Lacopini, I., Petri, G., Barrat, A. & Latora, V. Simplicial models of social contagion. *Nat. Commun.* **10**, 2485 (2019).
- Gilbert, E., Karahalios, K. & Sandvig, C. The network in the garden: designing social media for rural life. *Am. Behav. Sci.* **53**, 1367–1388 (2010).
- Kucharski, A. J. et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet. Infect. Dis.* **20**, 553–558 (2020).
- Ai, T. et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* <https://doi.org/10.1148/radiol.2020200642> (2020).
- Cheng, K. K., Lam, T. H. & Leung, C. C. Wearing face masks in the community during the COVID-19 pandemic: altruism and solidarity. *Lancet* [https://doi.org/10.1016/S0140-6736\(20\)30918-1](https://doi.org/10.1016/S0140-6736(20)30918-1) (2020).
- Liang, M. et al. Efficacy of face mask in preventing respiratory virus transmission: a systematic review and meta-analysis. *Travel. Med. Infect. Dis.* <https://doi.org/10.1016/j.tmaid.2020.101751> (2020).

Acknowledgements

We thank M. Wall and Y. Wang for their helpful comments on early versions of the methods.

Author contributions

N.H. drafted the manuscript, and contributed to the model design and analyses; H.L. designed the model, performed the analyses and contributed to the writing of the manuscript; M.B. critically revised the model and manuscript for scientific content; C.B., M.O., M.M., M.S.R. and F.L. critically revised the manuscript for scientific content. All authors contributed to and have approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

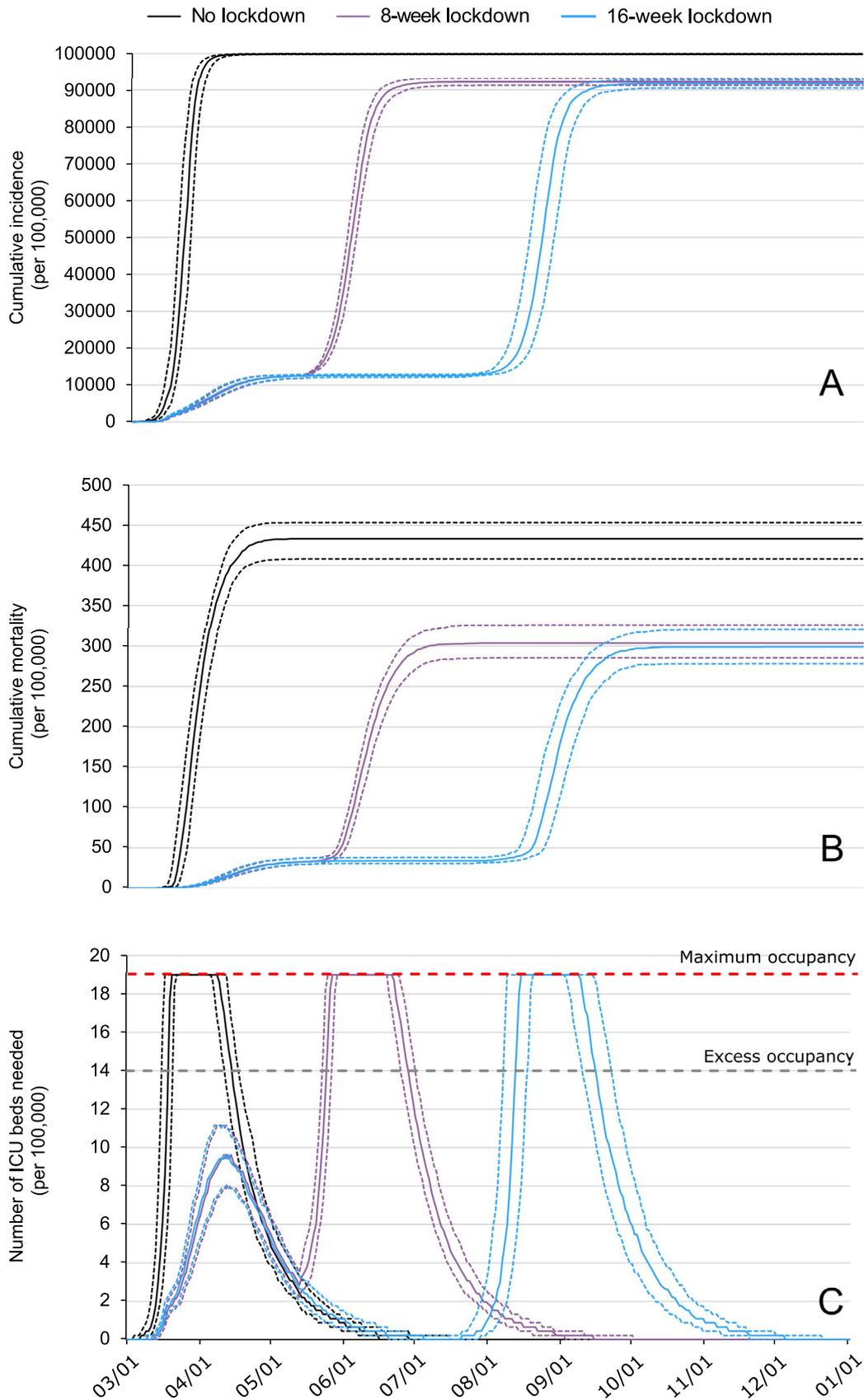
Extended data is available for this paper at <https://doi.org/10.1038/s41591-020-1001-6>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41591-020-1001-6>.

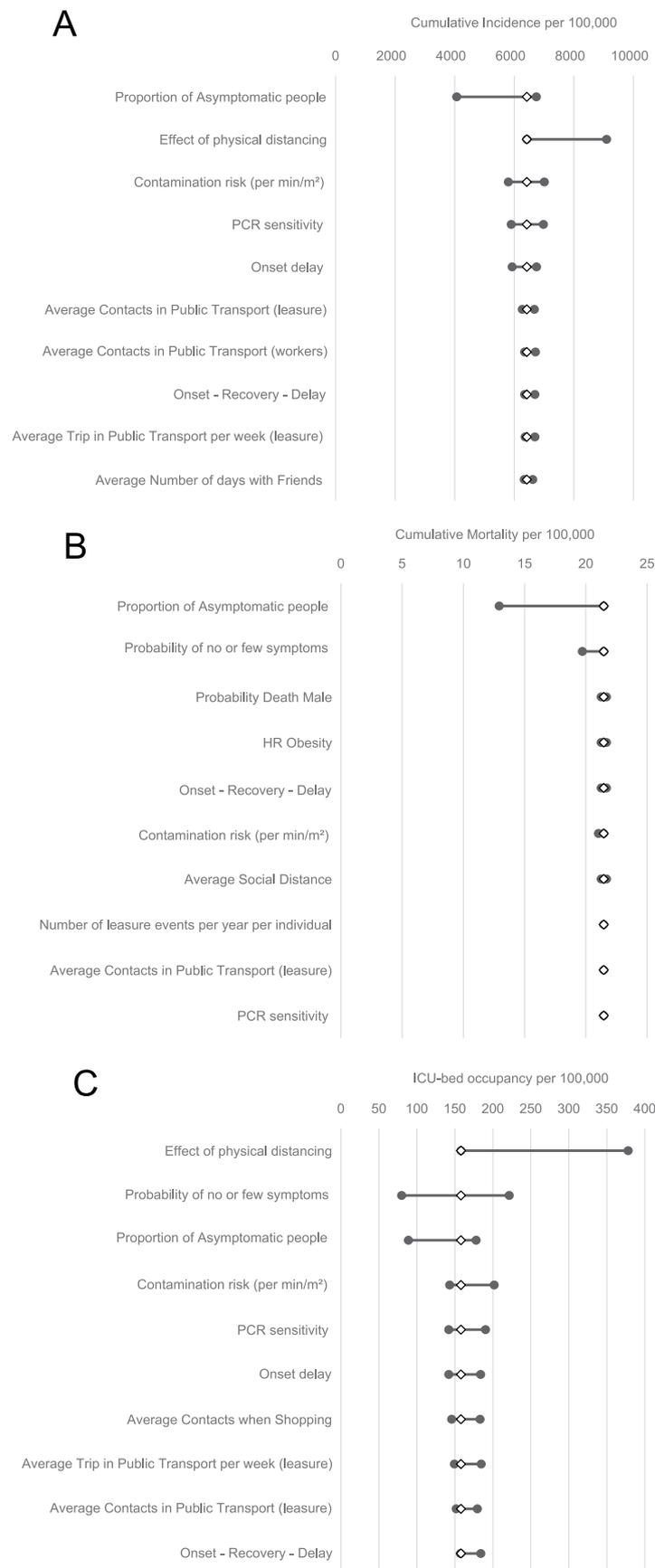
Correspondence and requests for materials should be addressed to N.H.

Peer review information Jennifer Sargent was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

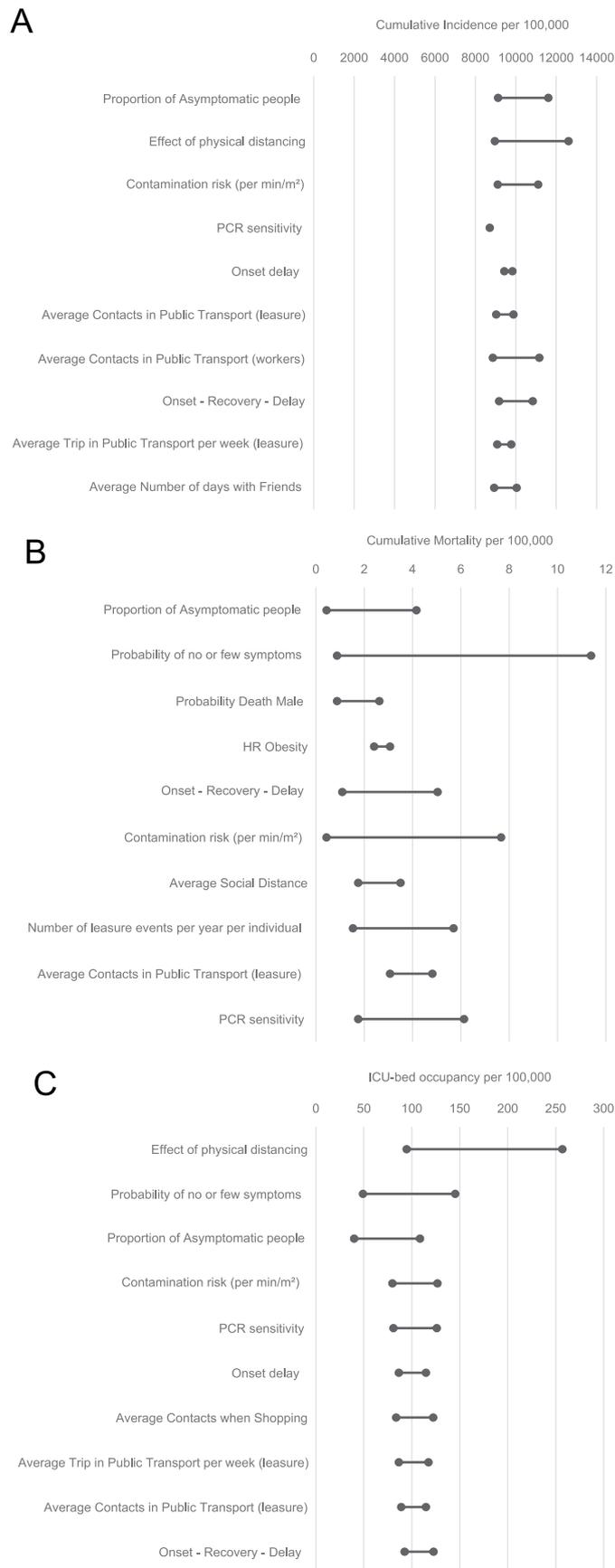
Reprints and permissions information is available at www.nature.com/reprints.



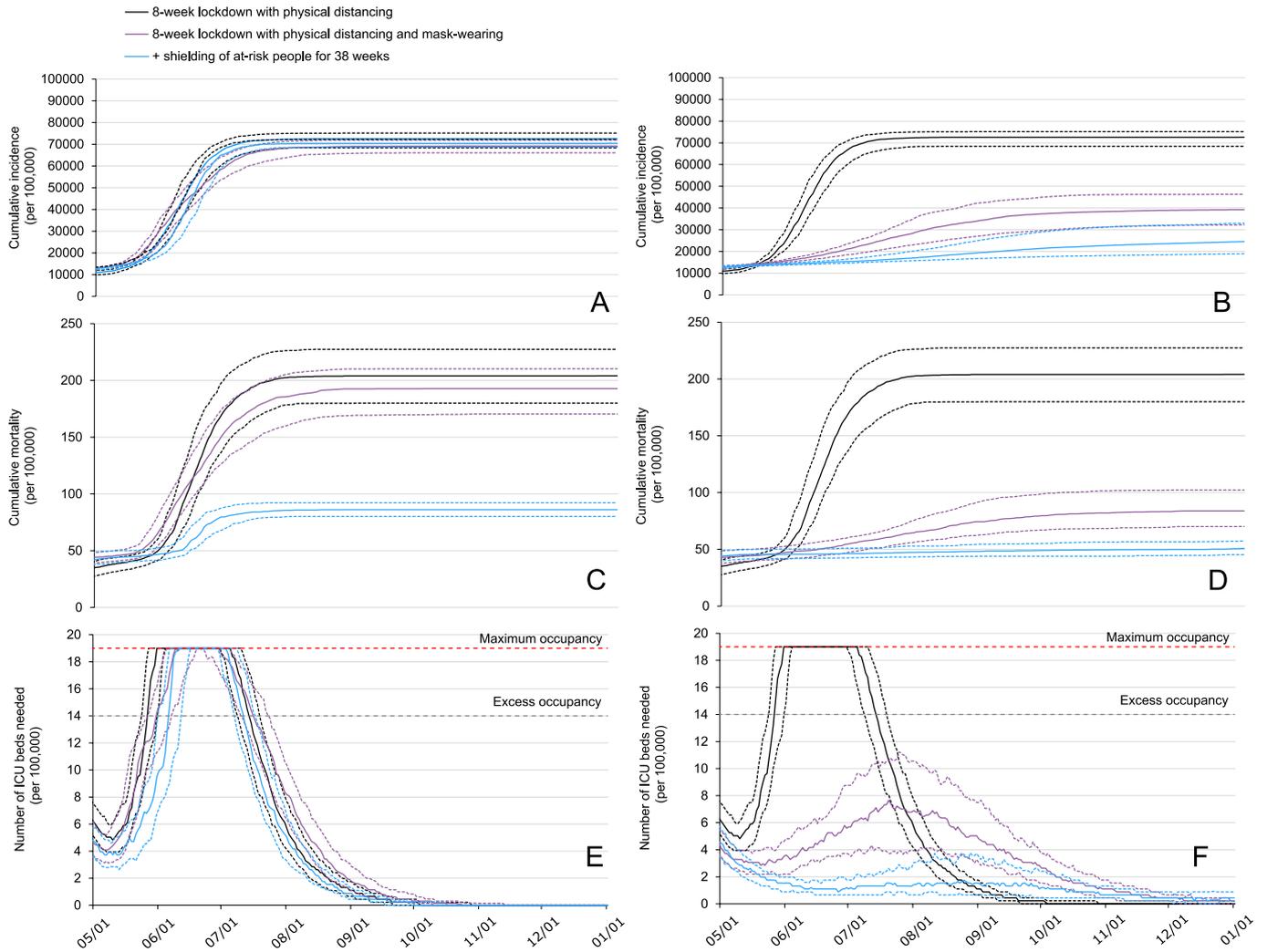
Extended Data Fig. 1 | Model-predicted cumulative incidence **a**, mortality **b**, and number of ICU beds needed **c**, by lockdown duration. The dotted lines represent the uncertainty range (first to last quartile) stemming from the uncertainty in the parameter values.



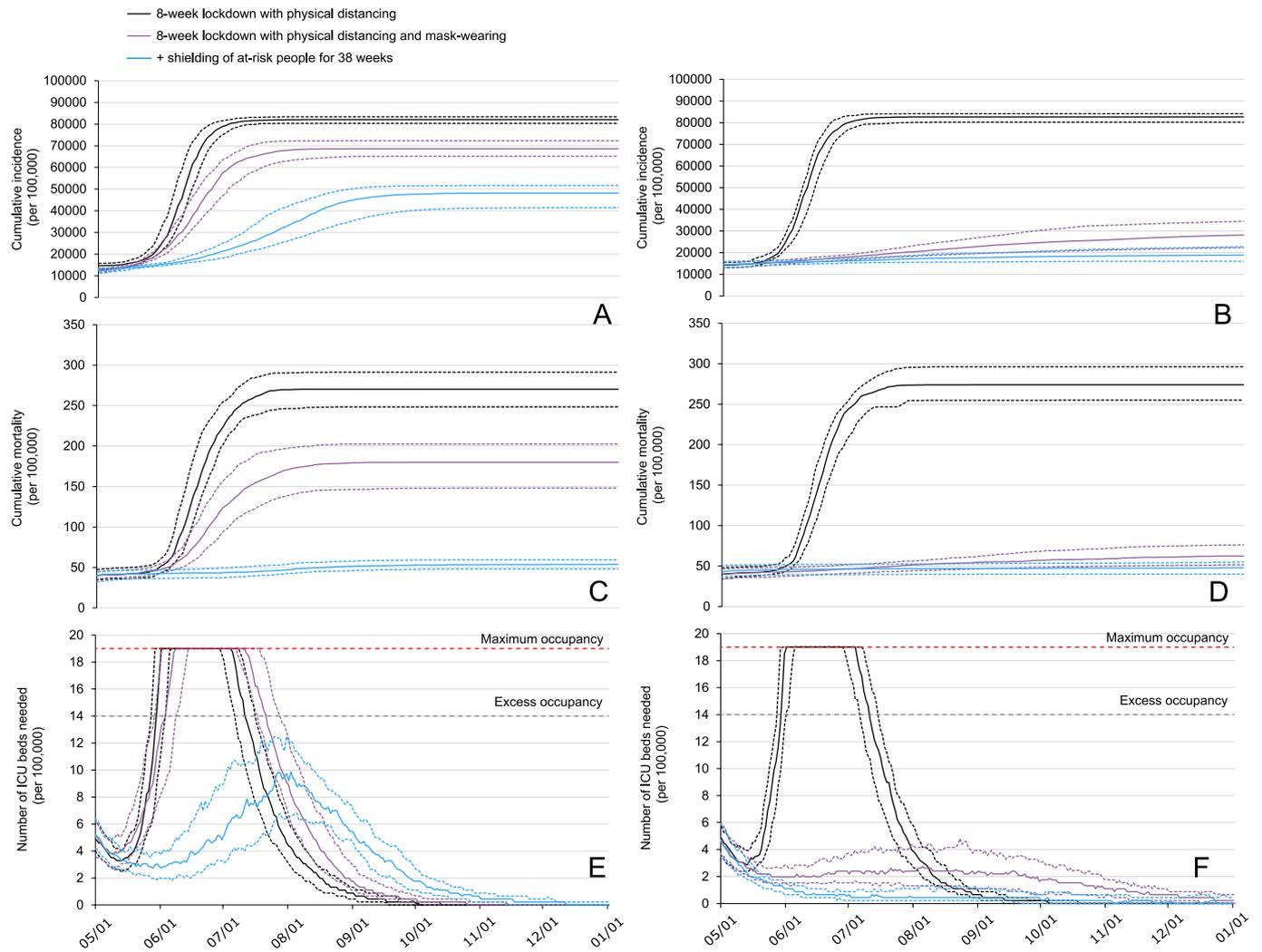
Extended Data Fig. 2 | Impact of varying by +/-20% individual model parameter value for the effect of post-lockdown physical distancing and mask-wearing for the general population on cumulative incidence **a**, cumulative mortality **b**, and ICU-bed occupancy **c**. Only the 10 parameters having the highest impact are presented.



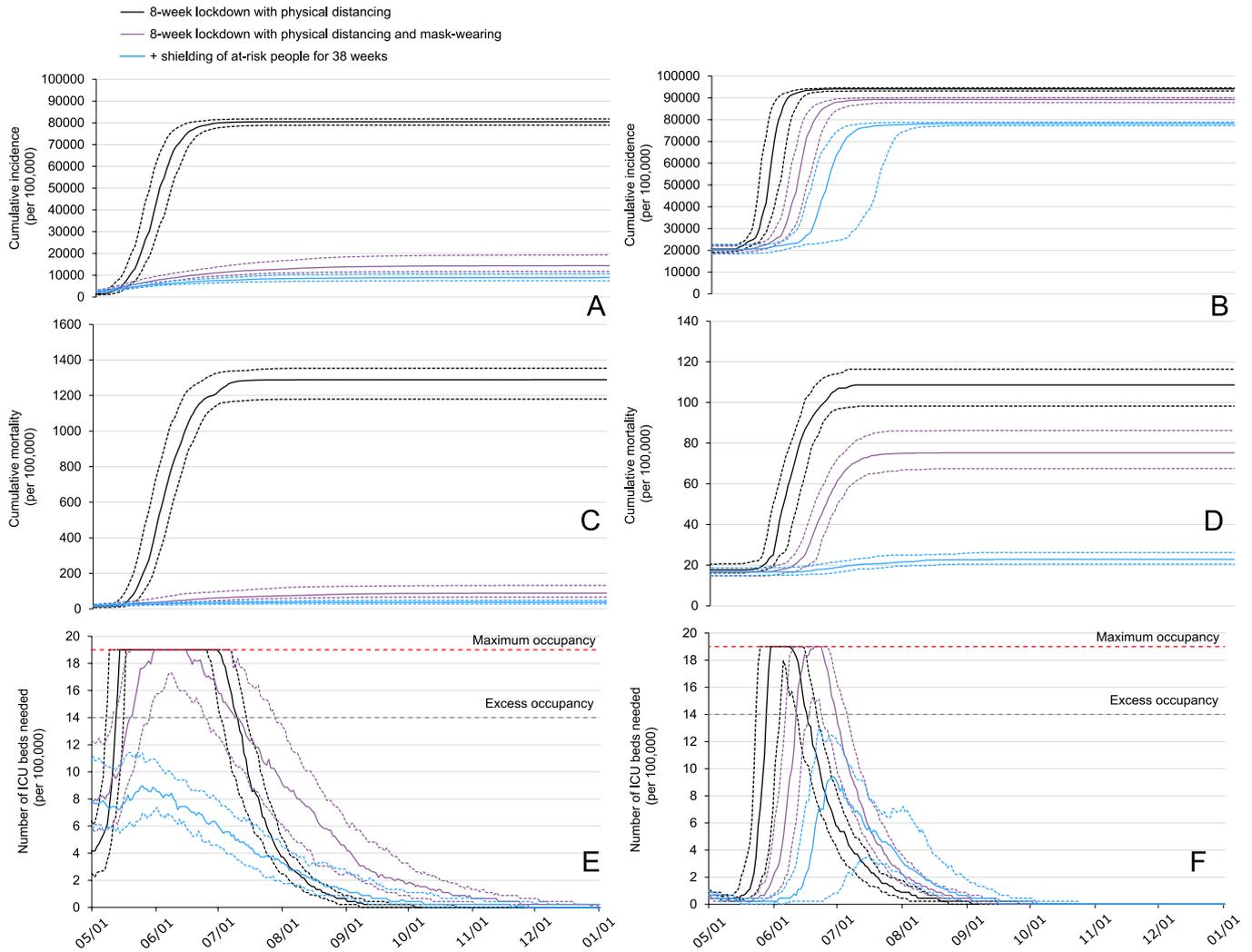
Extended Data Fig. 3 | Effect of varying by +/-20% individual model parameter value on cumulative incidence **a**, cumulative mortality **b**, and ICU-bed occupancy **c**, for the difference between the strategies 'post-lockdown physical distancing and mask-wearing for the general population, and shielding of at-risk individuals' and 'post-lockdown physical distancing and mask-wearing for the general population'. Only the 10 parameters having the highest impact are presented.



Extended Data Fig. 4 | Impact of fixing the parameter value of masks' efficacy at 21% **a, c, e**, and 64% **b, d, f**, instead of 47% as in the main analysis, on cumulative incidence (**a, b**), cumulative mortality (**c, d**), and number of ICU beds needed (**e, f**). The dotted lines represent the uncertainty range (first to last quartile) stemming from the uncertainty in the parameter values.



Extended Data Fig. 5 | Impact of assuming that the risk of contamination is reduced by 50% **a, c, e**, and 80% **b, d, f**, on cumulative incidence (**a, b**), cumulative mortality (**c, d**), and number of ICU beds needed (**e, f**). The dotted lines represent the uncertainty range (first to last quartile) stemming from the uncertainty in the parameter values.



Extended Data Fig. 6 | Impact of assuming a diagnosis rate (that is the number of undiagnosed cases for each diagnosed case) of 1 in 50 **a, c, e**, and 1 in 200 **b, d, f**, instead of 1 in 100 as in the main analysis, on cumulative incidence (**a, b**), cumulative mortality (**c, d**), and number of ICU beds needed (**e, f**). The dotted lines represent the uncertainty range (first to last quartile) stemming from the uncertainty in the parameter values.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data were collected using Microsoft Excel. The agent-based model has been performed using C++. The C++ source code has been deposited in a recognized public source code repository (GitHub, <https://github.com/henrileu/covid19>). All parameters used in our stochastic agent-based model of SARS-CoV-2 epidemic in France are described in the method section and summarized in Supplementary Table 3.

Data analysis

Statistical analyses were conducted using SAS software version 9.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Population demographics, data on social contacts and on SARS-CoV-2 characteristics, and observed data for COVID-19 incidence, mortality, and ICU-bed occupancy in France are available from the publicly available data of French National Statistical Institute (INSEE) (<https://insee.fr/fr/accueil>), Santé Publique France (<https://www.santepubliquefrance.fr/>), Institut Pasteur (<https://www.pasteur.fr/fr/>), London Imperial College (<https://www.imperial.ac.uk/>), and Ministère des Solidarités et de la Santé (<https://drees.solidarites-sante.gouv.fr/etudes-et-statistiques/>) for the year 2020, Direction Générales des Entreprises (DGE) (<https://www.entreprises.gouv.fr/>) for the year 2018, and Ministère de la Culture for the year 2015 (<https://www.culture.gouv.fr/>).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The stochastic agent-based microsimulation model (ABM) was run for 360 days on 500,000 individuals. The results were based on an average of 200 simulations. Analyses were performed on May 17th using data for model parameters until April 15th. Results were extrapolated to the French population of 67 million people.
Data exclusions	None
Replication	We provided uncertainty measures by using 200 bootstrap samples based on the random variation of all non-calibrated parameters simultaneously, either within a 95% confidence interval for parameters estimated from the literature or within a +/-20% interval if the parameter was assumed. We examined the robustness of our results by evaluating the impact on outcomes of varying successively individual parameter values by +/-20%, without recalibrating the model. Given the uncertainty of the two calibrated parameters and of the efficacy of masks, we performed sensitivity analyses and evaluated the impact of varying values of these parameters on the predicted epidemic course. These analyses are detailed for the intervention 'post-lockdown physical distancing and mask-wearing for the general population' and 'shielding of at-risk individuals in addition of physical distancing and mask-wearing'.
Randomization	Not applicable
Blinding	Not applicable

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | n/a | Involvement in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | n/a | Involvement in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |