Appendix

Table of Contents

S1 Mathematical notations

Summary table with all the notations used in the study.

notation	meaning
tracked densities of individuals	
S	susceptible
J	non-critical infectious
Y	critical infectious
H	long-stay ICU hospitalised
W	other critical hospitalised patients
R	recovered immunised
D	dead (cumulative mortality)
E	latent$^\sharp$
derived densities or flows	
C	cumulative incidence
ΔC	daily incidence
I	community infectious
A	daily ICU admissions
L	ICU discharges
M	daily mortality

Table S-1: **Density related notations.**
$^\sharp$: only applies to the Markovian continuous-time model.

notation	meaning	value origin
infection related quantities		
Λ_i	force of infection	calculated
Z, ζ_k	generation time distribution	[3]
$\overline{I}$	effective community infectious density	calculated
$\mathcal{R}_0$	basic reproduction number	estimated
S_0	population size	[28, 29]
$k_{\cdot}$	current *per capita* contact rate	unknown$^\flat$
$k_{\cdot,0}$	pre-epidemic *per capita* contact rate	unknown$^\flat$
$c_{\cdot}$	*per capita* contact ratio	unknown$^\flat$
κ	lock-down effect	estimated
illness related quantities		
$\mathrm{IFR}_{\cdot}$	infection fatality ratio	[30]
$\mathrm{H}, \eta_{\cdot}$	contamination to hospitalisation interval distribution	[31, 32]
$\mathrm{P}, \rho_{\cdot}$	long stay ICU length distribution	[32] and fitted
$\Upsilon, \upsilon_{\cdot}$	hospitalisation to death outside ICU interval distribution	[32] and fitted
θ	critical illness frequency	calculated and fitted
ψ	long-stay ICU admission frequency	calculated and fitted
μ	long-stay ICU fatality ratio	[32]
b	long-stay ICU admission frequency among hospitalised patients	[20]
d	fatality ratio among hospitalised patients	[20]
ω^{-1}	average latency period$^\sharp$	fitted
γ^{-1}	average infectious period$^\sharp$	fitted

Table S-2: **Main parameter notations.**
$^\sharp$: only applies to the Markovian continuous-time model. $^\flat$: the calculation of these values are bypassed by the estimation of κ, as shown in S2.3.

notation	**meaning**
derived output quantities	
$\mathcal{R}_t$	temporal reproduction number
ι	population immunisation
ι_{h}	herd immunity threshold
q_{f}	final size proportion
π	community prevalence
other estimated/fitted parameters	
t_0	epidemic initiation date
$\mathbb{E}[\mathrm{H}]$	critical case contamination to hospitalisation interval expectation
$\mathbb{V}[\mathrm{H}]$	critical case contamination to hospitalisation interval variance
$\mathbb{E}[\mathrm{P}]$	long ICU stay length expectation
$\mathbb{E}[\Upsilon]$	critical case hospitalisation to death interval expectation (non long-stay ICU patients)
$\mathfrak{C}_{\mathrm{F}}$	infection fatality ratio correction factor
$\mathfrak{C}_{\mathrm{M}}$	long-stay ICU fatality ratio correction factor
$\mathfrak{C}_{\Psi}$	long-stay ICU frequency correction factor

Table S-3: **Output-related quantities and fitted accessory parameters.**

notation	meaning	**value origin**
other notations		
$\cdot_i$	age group (average, density...)	calculated
$\tilde{\cdot}$	moving-averaged data	[15]
$\hat{\cdot}$	adjusted parameter	fitted
$\mathfrak{f}_{\cdot}$	demographic frequency	[28, 29]

Table S-4: **Other generic notations.**

S2 Model details

S2.1 Recurrence relation system

Instead of classical ordinary differential equations (ODE), the dynamics of the model satisfy a system of recurrence relations (one could as well write as finite difference equations (FDE)), which is detailed below. For the sake of simplicity, we omit time dependence in the notations. Instead, X_i denotes a density at a given time t and X_i' the density at time $t+1$. The equations are formally identical for all age groups i. Between-group dynamics are coupled through the forces of infection $\Lambda_\cdot$ defined in the next subsection.

$$
\begin{cases}
S_i' = (1-\Lambda_i)\, S_i, & \\
J_{i,1}' = (1-\theta_i)\, \Lambda_i S_i, & J_{i,k}' = J_{i,k-1}, \qquad 1 < k \leq g, \\
Y_{i,1}' = \theta_i \Lambda_i S_i, & Y_{i,k}' = (1-\eta_{k-1})\, Y_{i,k-1}, \qquad 1 < k \leq h, \\
H_{i,1}' = \psi_i \sum_{k=1}^{h} \eta_k Y_{i,k}, & H_{i,k}' = (1-\rho_{k-1})\, H_{i,k-1}, \qquad 1 < k \leq r, \\
W_{i,1}' = (1-\psi_i) \sum_{k=1}^{h} \eta_k Y_{i,k}, & W_{i,k}' = (1-\upsilon_{k-1})\, W_{i,k-1}, \qquad 1 < k \leq u, \\
D_i' = D_i + \sum_{k=1}^{u} \upsilon_k W_{i,k} + \mu_i \sum_{k=1}^{r} \rho_k H_{i,k}, & \\
R_i' = R_i + J_{i,g} + (1-\mu_i) \sum_{k=1}^{r} \rho_k H_{i,k}. &
\end{cases}
\tag{S-1}
$$

S2.2 Force of infection

In the SIR-like continuous-time modelling framework, the force of infection refers to the infection rate per capita of susceptibles, often expressed as $\lambda := \beta I$ [13]. Equivalently, the instantaneous incidence is $\beta IS = \lambda S$, which is the translation of the mass action law implied by the mean-field approximation made by such spatially unstructured models.

In our discrete-time model, the force of infection Λ_i is not a rate but a daily probability of infection (*per capita* of susceptibles from group i) that saturates with the prevalence. Individual contributions of infected individuals are not additive when prevalence is high because a susceptible host surrounded by contagious individuals can be infected by several

of them the same day. When prevalence is low, the probability of contamination by multiple infectors the same a day is low and the force of infection is well approximated by the sum of contributions of each infected individual. Λ_i is therefore a monotonically increasing function of prevalence, bounded by 1 and with a positive initial slope recovering the continuous-time mass action law.

The most parsimonious expression compatible with such constraints is the well-known Michaelis-Menten (or Holling type II physiological response) function of the form $x \mapsto \frac{x}{a+x}$. However, we cannot use the prevalence as the argument of Λ_i here because all infected individuals, whether they are critically ill (Y) or not (J), do not contribute equally to transmission events. This heterogeneity in contagiousness originates from differences in infection ages (individuals contaminated 6 days earlier are more contagious than those 10 days earlier) and in contact rates.

To address this issue, we introduce the effective infectious density $\overline{I}(t)$,

$$\overline{I}(t) := \sum_j c_j(t) \sum_k \zeta_k \left(J_{j,k}(t) + Y_{j,k}(t) \right), \tag{S-2}$$

which is the sum over all infected community compartments weighted by both the generation time distribution ζ_k, which is the time between the infection of an 'infector' and the infection of his or her 'infectee', and the per-capita contact ratio $c_i(t)$. The latter is defined as the current contact rate per-capita of individuals of age group i ($k_i(t)$) relative to their pre-epidemic baseline contact rate ($k_{i,0}$),

$$c_i(t) := \frac{k_i(t)}{k_{i,0}}. \tag{S-3}$$

With $\overline{I}$ kept constant, Λ_i is expected to display a Michaelis-Menten behavior (i.e. positive initial slope, increasing and upper-bounded) with respect to the current per-capita rate as well. Consequently, the force of infection should satisfy

$$\Lambda_i(t) = \frac{k_i(t)\,\overline{I}(t)}{a + k_i(t)\,\overline{I}(t)}, \tag{S-4}$$

as shown in Fig.S-1.

We then derive the expression of a using known parameters. To do so, we consider the probability for a given susceptible individual to be part of the first generation of cases, that is, to have been infected by the index case. Let us denote $\Lambda_{i,\text{index}}(t)$ the probability of being infected by the index case on day t. By definition of the basic reproduction number, the index case infects $\mathcal{R}_0$ secondary cases, on average, by the end of its contagious period. Under the mean-field approximation, the probability of being part of these secondary cases is simply $\mathcal{R}_0/S_0$ (which is an extremely rare event). Summing over all possible days, we therefore have

$$\sum_{t\geq 1}\Lambda_{i,\text{index}}(t) = \mathcal{R}_0/S_0 \lll 1. \tag{S-5}$$

For simplicity, we make the following three assumptions:

- the index case is not critically ill (less than 5% of cases are),
- the index case has infected all his or her secondary cases before public health measures are implemented.

It follows from these assumptions that if we set the contamination day of the index case to $t = 0$, the force of infection generated by the index case is proportional to

$$\overline{I}_{\text{index}}(t) = \zeta_t,$$

(note that, for all j, $k_j(t) = k_{j,0}$ over the considered period of time, hence $c_j(t) = 1$ and all densities are equal to 0 except $J_{j,t} = 1$).

Applying these results to equation S-4, the daily probability of infection by the index case therefore becomes

$$\Lambda_{i,\text{index}}(t) = \frac{k_{i,0}\zeta_t}{a + k_{i,0}\zeta_t}.$$

Now, from the magnitude comparison (S-5), we have $a \ggg k_i(t)\,\zeta_t$, and hence

$$\sum_{t\geq 1}\Lambda_{i,\text{index}}(t) \approx \frac{k_{i,0}}{a}\sum_{t\geq 1}\zeta_t,$$

In the limit of low prevalence (or low contact rates), the mass action law is recovered.

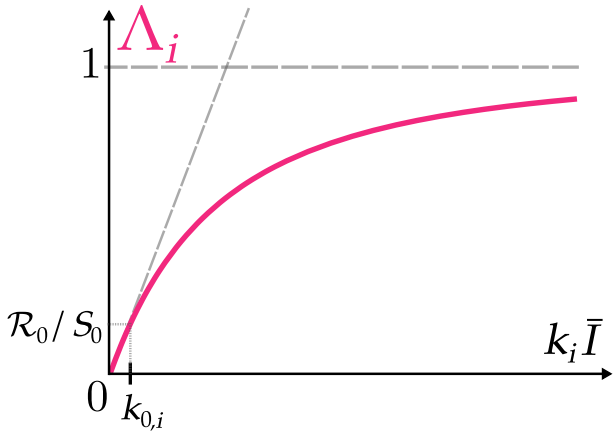

Figure S-1: **Daily force of infection as a function of the product of daily contact rate and effective infectious density**
Λ_i is the probability for one susceptible individual from group i to be infected a given day. The initial non-zero slope comes from the law of mass action implied by the mean-field approximation. The derivation of Λ_i is based on the remarkable coordinates shown near the origin of the graph (not to scale). When the effective infectious density is equal to 1 (i.e. as if the generation time distribution were concentrated in a single day) and the contact rate is that in absence of any health measure (denoted by $k_{0,i}$), then, by definition of the basic reproduction number, Λ_i equals $\mathcal{R}_0/S_0$.

Using equation (S-5) and the fact that the sequence $(\zeta_t)_{t\geq 1}$ is a mass function, it sums up to 1, we get

$$a = k_{i,0} \frac{S_0}{\mathcal{R}_0}. \tag{S-6}$$

Combining equations (S-3), (S-4), and (S-6) finally leads to a generic expression of the force of infection:

$$\Lambda_i(t) = \frac{c_i(t)\,\bar{I}(t)}{\frac{S_0}{\mathcal{R}_0} + c_i(t)\,\bar{I}(t)}. \tag{S-7}$$

The equation indicated in the main text is finally obtained using (S-2), (S-3), elementary algebra and assuming that the total population size is almost constant over the investigated time period.

S2.3 Lock-down effect

In the special case where strong public health control measures such as lock-down are being implemented, all individuals may exhibit similar per capita contact rates, $k_i(t) = k_{\text{lock}}$.

From (S-4), it is then possible to express the force of infection in the following way:

$$\begin{aligned}\Lambda_i &= \frac{\overline{I}}{\frac{a}{k_{\text{lock}}} + \overline{I}}, \\ &= \frac{\sum_j \frac{k_{\text{lock}}}{k_{i,0}} \sum_k \zeta_k \left(J_{j,k} + Y_{j,k}\right)}{\frac{k_{i,0}}{k_{\text{lock}}} \frac{S_0}{\mathcal{R}_0} + \sum_j \frac{k_{\text{lock}}}{k_{i,0}} \sum_k \zeta_k \left(J_{j,k} + Y_{j,k}\right)}, \\ \Lambda_i &= \frac{\overline{I}_+}{\left(\frac{k_{i,0}}{k_{\text{lock}}}\right)^2 \frac{S_0}{\mathcal{R}_0} + \overline{I}_+}, \end{aligned} \tag{S-8}$$

where $\overline{I}_+$ is the effective infectious density as if it were calculated in absence of health measures (i.e. with $c_i = 1$). In equation (S-8), one can interpret the quantity $c_{i,\text{lock}}^2 := \left(\frac{k_{\text{lock}}}{k_{i,0}}\right)^2 < 1$ as a factor lowering the basic reproduction number $\mathcal{R}_0$. The lock-down effect, defined as the reduction of $\mathcal{R}_0$ due to this measure, can be calculated by averaging $c_{i,\text{lock}}$ over age groups (according to demography). Hence,

$$\kappa := 1 - \overline{c_{\text{lock}}}^2. \tag{S-9}$$

S2.4 Times series

The largest and most reliable nationwide data for the COVID-19 epidemic in France is that of daily COVID-related death toll in hospitals, communicated daily since Feb 16 2020 by the national public health agency (Santé Publique France). This is why our model neglects COVID-related deaths occurring outside hospitals. In particular, we removed nursing homes (or EHPAD) from calculations.

Starting from Mar 18 2020, two additional time series are communicated: daily ICU admission and current ICU occupied beds. While the former capture the dynamics from contamination to ICU admission, the latter captures moreover the kinetics of ICU stay.

These time series are altered by week-ends and bank days: e.g. death tolls are notably lower on Sundays than previous days, while it increases the next Mondays. It has even been suggested that reporting delays propagate also to Tuesdays. In order to smooth these artifactual weekly oscillations, a right-shifted 7-days moving average was performed

over all time series prior to analysis. We will refer to these smoothed datasets as $\tilde{M}$ for daily hospital mortality, $\tilde{A}$ for daily ICU admissions and $\tilde{H}$ for current ICU occupied beds. Since $\tilde{H}$ contains part of the cumulative information of $\tilde{A}$, we also considered $\tilde{D}$, the cumulative counterpart of $\tilde{M}$ to equilibrate the first step of the fitting procedure (see below).

S2.5 Criticality-related probabilities

Symptom severity of COVID-19 is increasingly classified into mild, moderate, severe and critical. Because we rely on hospital mortality and ICU flow data, we focus on critical cases, i.e. the ones concerned by intensive care and COVID fatality. In the absence of detailed large-scale hospitalisation data, we made the following assumptions:

- non-critical cases are not admitted into ICU (even though some do need hospitalisation),
- all critical cases need hospitalisation, and only survive if they go through intensive care.

For medical reasons not addressed here, in France not all critical cases are admitted soon enough into ICU. Part of them die in non-intensive care wards, while others die shortly after entering the ICU, therefore not contributing to ICU bed occupancy. We therefore need to estimate two key criticality-related probabilities, namely θ_i, the proportion of critical cases within age group i, and ψ_i, the proportion of critical cases in age group i that contribute to ICU bed occupancy (i.e. their stay in the ward exceeds one day). Because these probabilities differ among age classes and improper averaging could lead to substantial bias (see below), we first need to make calculations focused on the smallest age stratification unit (usually a decade). In the following, $\theta(a)$ and $\psi(a)$ are the age-specific critical case frequency and long-stay admission given critical illness respectively. In addition, $\mathbb{P}_a[\mathtt{X_1}|\mathtt{X_2}]$ reads as the probability of event $\mathtt{X_1}$ given $\mathtt{X_2}$ has occurred for an individual of age a.

Let us consider the four events needed to derive $\theta(a)$ and $\psi(a)$ from available data:

- $\mathtt{I}$, being infected by SARS-CoV-2,
- $\mathtt{U}$, being hospitalised,
- $\mathtt{B}$, occupying an ICU bed for more than a day,
- $\mathtt{D}$, dying at the hospital from COVID-19.

The interplay between these events is formally depicted by the tree diagram in Fig.S-2.

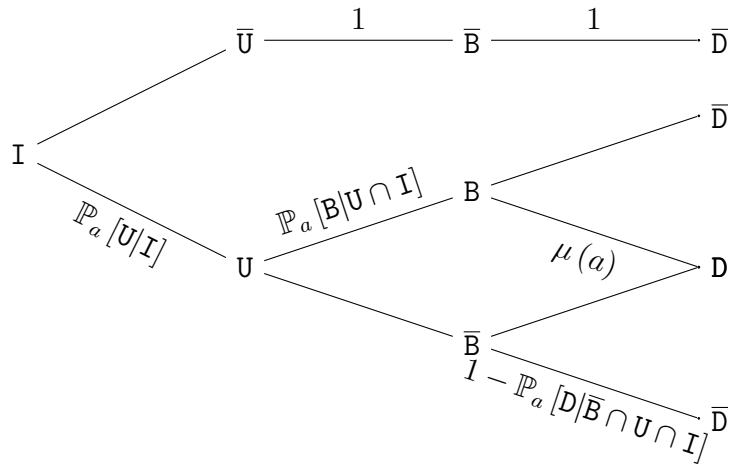

Figure S-2: **Tree diagram of critical COVID-19 related events**. A fraction $\mathbb{P}_a[\mathtt{U}|\mathtt{I}]$ of infected ($\mathtt{I}$) are hospitalised ($\mathtt{U}$). Among these, a proportion $\mathbb{P}_a[\mathtt{B}|\mathtt{U}\cap\mathtt{I}]$ are admitted into ICU for a stay longer than a day $\mathtt{B}$. The fatality ratio ($\mathtt{D}$) equals $\mu(a)$ for these patients and $\mathbb{P}_a[\mathtt{D}|\overline{\mathtt{B}}\cap\mathtt{U}\cap\mathtt{I}]$ for the others.

We need to find two independent equations involving probabilities related to these events for which age-stratified data is known in order to solve $\theta(a)$ and $\psi(a)$. First, we need to identify such data:

- $\mathbb{P}_a[\mathtt{D}|\mathtt{I}]$ is the proportion of deaths among COVID-19 infected individuals, better known as the Infection Fatality Ratio (IFR), which has been calculated using the Diamond Princess data by Verity et al. and corrected for non-uniform attack rate by Ferguson et al. and will be denoted by $\mathrm{IFR}(a)$ hereafter,
- $\mathbb{P}_a[\mathtt{D}|\mathtt{B}\cap\mathtt{U}\cap\mathtt{I}]$ is the proportion of deaths among COVID+ ICU hospitalised patients; this age stratified data has been communicated by Santé Publique France as a weekly epidemic report on May 7 2020 [32], and will be denoted by $\mu(a)$ hereafter

- $\mathbb{P}_a\left[\mathtt{B}|\mathtt{U}\cap\mathtt{I}\right]$ and $\mathbb{P}_a\left[\mathtt{D}|\mathtt{U}\cap\mathtt{I}\right]$ are respectively the proportions of hospitalised patients (whether they are critical or not) admitted in ICU, and those that die (whether in ICU or not), and hereafter denoted by $b\left(a\right)$ and $d\left(a\right)$. These data come from the SI-VIC database and made available by [20]. (Tables S1 and S2).

A first equation comes by noticing that the probability for an infected individual to die from COVID-19 is the probability of developing critical illness if infected (namely $\theta\left(a\right)$) times the proportion of deaths among critical cases. The latter is the sum of critical cases that die in ICU after a stay longer than one day ($\mu\left(a\right)\psi\left(a\right)$) and the critical cases that are not lengthily admitted in ICU and cannot be saved ($1-\psi\left(a\right)$):

$$\mathrm{IFR}\left(a\right)=\left(\mu\left(a\right)\psi\left(a\right)+1-\psi\left(a\right)\right)\theta\left(a\right),$$

hence

$$\theta\left(a\right)=\frac{\mathrm{IFR}\left(a\right)}{1-\left(1-\mu\left(a\right)\right)\psi\left(a\right)}.$$

Now, to find $\psi\left(a\right)$, let us notice that the proportion of deaths not occurring in ICU can be expressed as

$$\frac{1-\psi\left(a\right)}{\mu\left(a\right)\psi\left(a\right)+1-\psi\left(a\right)}=\mathbb{P}_a\left[\overline{\mathtt{B}}|\mathtt{D}\cap\mathtt{U}\cap\mathtt{I}\right],$$

hence

$$\psi\left(a\right)=\frac{1-\mathbb{P}_a\left[\overline{\mathtt{B}}|\mathtt{D}\cap\mathtt{U}\cap\mathtt{I}\right]}{1-\left(1-\mu\left(a\right)\right)\mathbb{P}_a\left[\overline{\mathtt{B}}|\mathtt{D}\cap\mathtt{U}\cap\mathtt{I}\right]}.$$

The unknown probability can be calculated as follows

$$\begin{aligned}\mathbb{P}_a\left[\overline{\mathtt{B}}|\mathtt{D}\cap\mathtt{U}\cap\mathtt{I}\right]&=\frac{\mathbb{P}_a\left[\overline{\mathtt{B}}\cap\mathtt{D}|\mathtt{U}\cap\mathtt{I}\right]}{\mathbb{P}_a\left[\mathtt{D}|\mathtt{U}\cap\mathtt{I}\right]},\\&=\frac{\mathbb{P}_a\left[\mathtt{D}|\mathtt{U}\cap\mathtt{I}\right]-\mathbb{P}_a\left[\mathtt{B}\cap\mathtt{D}|\mathtt{U}\cap\mathtt{I}\right]}{d\left(a\right)},\\&=\frac{d\left(a\right)-\mu\left(a\right)b\left(a\right)}{d\left(a\right)}.\end{aligned}$$

After elementary calculations, we finally get

$$\psi(a) = \frac{1}{1 - \mu(a) + \frac{d(a)}{b(a)}},$$

hence both $\theta(a)$ and $\psi(a)$ can be calculated from age-stratified available data.

The mean proportion of critical cases among a specific age group, θ_i, is simply the demographic-weighted average of $\theta(a)$ over the considered ages, as infection samples uniformly the susceptible compartment (NB: the IFR stratified data used here from [30] is already corrected for non-uniform attack rate). However, as the probability of the next events related to critical illness are not homogeneous with respect to age, the age group averages ψ_i and μ_i cannot be weighted directly with relative demographic age frequencies. Instead, they must be calculated by taking into account that ages with higher $\psi(a)$ and then $\mu(a)$ will be over-sampled in $Y_i \to H_i$ and $H_i \to D_i$ transitions respectively.

To account for this bias and as well to allow adjusting the parameters to both the model (for $\psi(a)$ and $\mu(a)$) and the French epidemic (for the IFR), we introduce a series of corrections detailed hereafter in the calculation of the parameters used to run the model.

First, we account for the fact that the ICU fatality ratio might mix both short and long-stay patients, while our model splits these two flows. The corrected age-specific long-stay ICU fatality ratio will be denoted by $\hat{\mu}(a)$ and calculated as the product of the corresponding data and a correcting factor denoted by $\mathfrak{C}_{\mathrm{M}}$, i.e. $\hat{\mu}(a) := \mathfrak{C}_{\mathrm{M}}\mu(a)$. Likewise, the age-specific IFR (which was not estimated from French data) will be corrected as $\hat{\mathrm{IFR}}(a) := \mathfrak{C}_{\mathrm{F}}\mathrm{IFR}(a)$. Now, equations S2.5 and S2.5 rewrite as

$$\hat{\psi}(a) := \frac{\mathfrak{C}_{\Psi}}{1 - \hat{\mu}(a) + \frac{d(a)}{b(a)}} \text{ and } \hat{\theta}(a) := \frac{\hat{\mathrm{IFR}}(a)}{1 - (1 - \hat{\mu}(a))\,\hat{\psi}(a)}.$$

Note that there is no need for a fourth correction factor as $\mathfrak{C}_{\mathrm{M}}$, $\mathfrak{C}_{\mathrm{F}}$ and $\mathfrak{C}_{\Psi}$ already capture possible corrections for $\hat{\theta}$. These corrections are yet of unknown values but will be estimated by the fitting procedure detailed below. As the corrected values are still probabilities, the correction factors are necessarily upper bounded, respectively by

$(\max_a \mu(a))^{-1}$, $(\max_a \mathrm{IFR}(a))^{-1}$, and $(\max_a \psi(a))^{-1}$.

The last step consists in age-group averaging as mentioned above. Let us denote by $\mathfrak{f}(a)$ the frequency of individuals of age a in the French metropolitan population (after having removed the ca 730,000 individuals living in nursing homes). We call the i-group relative frequency of age a as the standardised age frequency $\mathfrak{f}_i(a) := \mathfrak{f}(a) \Big/ \sum_{j \in \mathcal{A}_i} \mathfrak{f}(j)$, where $\mathcal{A}_i$ is the set of ages belonging to age group i. As previously implied, the frequency of critical cases in age group i is the straightforward demographic weighted average

$$\theta_i = \sum_{a \in \mathcal{A}_i} \mathfrak{f}_i(a) \hat{\theta}(a).$$

The frequency of long-stay ICU patients among hospitalised critical cases in age group i is then weighted by both the relative age frequencies and the ratio of critical illness probability to the group average, i.e.

$$\psi_i = \sum_{a \in \mathcal{A}_i} \mathfrak{f}_i(a) \hat{\psi}(a) \frac{\hat{\theta}(a)}{\theta_i}.$$

Finally, as the last event to occur, the average fatality ratio for long-stay ICU patients belonging to group i must be corrected by the ratio of the product of the two previous frequencies relative to the group average ϕ_i, i.e.

$$\mu_i = \sum_{a \in \mathcal{A}_i} \mathfrak{f}_i(a) \hat{\mu}(a) \frac{\hat{\theta}(a) \hat{\psi}(a)}{\phi_i},$$

where $\phi_i := \sum_{a \in \mathcal{A}_i} \mathfrak{f}_i(a) \hat{\psi}(a) \hat{\theta}(a)$.

S2.6 Waiting times

Four time distributions underlie the dynamics: the generation time (or index-contamination-to-secondary-contamination interval), the contamination-to-hospitalisation interval of critical cases, the ICU length of stay and the hospitalisation-to-outside-ICU death interval of critical cases. Each of these events can be seen as a random waiting time variable, denoted by Z° (capital zeta), H° (capital eta), P° (capital rho) and Υ° (capital upsilon)

respectively.

Initially, all four random variables are assumed to follow Weibull distributions, with shape parameters greater than one. Such distributions are widely used in the biomedical literature, along with Gamma and Lognormal distributions, for fitting the probability density function (PDF) of ageing processes, for which the probability for the focal event to occur increases with lapsed time [33]. Preliminary exploratory fittings of the model to daily mortality data indicated that maximum likelihood estimates of the shape parameter of P° and Υ° were close to unity. Because the computational procedures used for maximum likelihood estimation misbehave in the vicinity of parameter range boundaries, the shape parameter of these two distributions was set to 1, turning them into exponential distributions by fitting (the exponential distribution being a special case of Weibull distributions), though not by assumption.

Weibull distributions have a right-unbounded support $[0, \infty)$, which means that true distributions require truncation for obvious computational reasons. Let us introduce the generic notations $\Xi \equiv \mathrm{Z}, \mathrm{H}, \mathrm{P}, \Upsilon$ and $x \equiv g, h, r, u$. We construct the right-truncated analogous distributions by setting the finite upper boundary of their support $x := \min\{n \in \mathbb{N} : F_{\Xi^{\circ}}(n) \geq 0.99\}$, i.e. the upper-integer-rounded 99%-quantile of the original distribution Ξ°, where $F.$ denotes the cumulative distribution function (CDF). The truncated distributions Ξ are therefore such that their CDF satisfy $F_{\Xi} = F_{\Xi^{\circ}} \Big/ F_{\Xi^{\circ}}(x)$, and having as their supports $F_{\Xi}(\Omega) = [0; x]$.

Dynamics unfold in discrete time in our model, which means these continuous distributions need to be discretised into sequences to be implemetend into the framework. The generation time sequence is straigtforwardly defined as $\zeta_k := F_{\mathrm{Z}}(k) - F_{\mathrm{Z}}(k-1)$ for $1 \leq k \leq g$. Indeed, transmission events do not affect the progression of the infector within its compartment. However, the three other sequences of parameters, $\xi_k \equiv \eta_k, \rho_k, \upsilon_k$, represent the proportion of individuals that leave the compartment k days after having entered it. These parameters need to capture the probability that the corresponding event occurs on day k but they also need to be standardised by the probability of not having left the

compartment by that day. They are therefore calculated as

$$\xi_k := \frac{F_\Xi(k) - F_\Xi(k-1)}{1 - F_\Xi(k-1)} \text{ for } 1 \leq k \leq x.$$

S2.7 Fitting and estimation procedure

The fitting procedure was performed using the mle2 routine from the bbmle package [34] implemented in R [35]. Starting from the initial parameter values $\mathbf{v}_0$, an ordinary least square optimum $\mathbf{v}_1$ was found by minimising the euclidean distance to $\tilde{A}$, $\tilde{M}$, $\tilde{H}$ and $\tilde{D}$ simultaneously, thus accounting for all events, from contamination to death or recovery. Having great confidence in both ICU admissions and current ICU occupancy is especially valuable for forecasting hospital needs, while mortality predictions cannot be ignored by decision makers. This first step is only used to locate the closest parameter region from $\mathbf{v}_0$ where likelihoods further calculated might reach their maximum value.

Maximum likelihood estimates (MLE) and 95%-likelihood intervals (LI) were calculated using the same routine. We assume observed data to be Gaussian-noised realisations of the model prediction, then considering each daily count to be distributed as $X_{\text{obs}}(t) \sim \mathcal{N}(X_{\text{sim}}(t), X_{\text{sim}}(t))$, where X_{sim} is the simulated daily count. The choice of the distribution is supported by the large numbers involved and the Poissonian nature of count processes (NB: pre-lock-down mortality data was ignored for the central limit to apply). Contrary to the first step, only one time series was used for each estimation. MLE and LI of $\mathcal{R}_0$, t_0, $\mathbb{E}[\mathrm{H}]$, $\mathbb{V}[\mathrm{H}]$ and κ were estimated with respect to $\tilde{A}$, while the other parameter values were set as in $\mathbf{v}_0$. The resulting MLE of the free parameters replaced the corresponding values in $\mathbf{v}_0$, providing vector $\mathbf{v}_1$. This new parameter set served as the starting point to estimate the MLE and LI for $\mathbb{E}[\mathrm{P}]$, $\mathfrak{C}_{\mathrm{M}}$ and $\mathfrak{C}_{\Psi}$ with respect to $\tilde{L}$, the derived time series of daily ICU discharges (calculated as the daily ICU admission minus the daily difference in ICU bed occupancy). The resulting parameter set $\mathbf{v}_2$ then initiated the estimation of the remaining parameters, $\mathbb{E}[\Upsilon]$ and $\mathfrak{C}_{\mathrm{F}}$, with respect to $\tilde{M}$, giving vector $\mathbf{v}_3$.

Using the fact that the maximum likelihood estimators are asymptotically normally

distributed, we used the range of the LI as proxies for the standard deviation of the marginal distribution of each MLE. Then, we randomly drew parameter sets in a multivariate Gaussian distribution with mean $\mathbf{v}_3$ and diagonal variance-covariance matrix. For each of them, we calculated their likelihood with respect to $\tilde{M}$ and kept only those whose likelihood was not significantly different from that of $\mathbf{v}_3$, according to Wilk's theorem [36]. Sampling stopped once 10^3 draws have satisfied the condition. Importantly, we considered all the retained parameters equivalent from the likelihood point of view, i.e. they are assumed to represent equally likely versions of adjusted parameter sets. Their diversity thus allows to account for uncertainty in the real parameter values.

For any further analysis of the model, system S-1 was run independently with each of the 10^3 parameter sets. The confidence intervals of simulated tracked densities ($S(t)$, $J(t)$...) as well as any derived quantity ($\mathcal{R}(t)$, $\iota(t)$...) were then simply calculated for each time point as the unweighted 2.5% and 97.5% sample quantiles of the 10^3 outputs at the given time point. The central estimations correspond to the median value of these distributions.

S2.8 Derived outputs

Tracked densities are the number of individuals in each clinical-epidemiological compartment $X_{i,k}$, the dynamics of which satisfy S-1 and are thus directly provided by numerical iteration of the recurrence relations. However, several quantities of interest require additional calculations, hereafter exposed.

Let us first introduce two notations for the sake of concision. Tracked densities without indices will refer to sum over all groups, and, for multiple-days compartments, over all possible days of progression as well: $X \equiv \sum_i \sum_k X_{i,k}$ (e.g. $J(t)$ represents all individuals belonging to $J_{i,k}$ for all groups i and all ages of infection k). The daily difference $\Delta X(t) \equiv X(t) - X(t-1)$ is straightforwardly used to extract the instantaneous dynamics from a cumulative time series.

The following three time series are crucial as they are used for likelihood calculations with respect to their data counterpart:

- $M(t) := \Delta D(t)$ is the daily mortality,
- $A(t) := \sum_i H_{i,1}(t)$ is the daily ICU admissions number,
- $L(t) := A(t) - \Delta H(t)$ is the daily ICU discharge number.

The next times series are intermediate calculations required for further key quantities.

- $I(t) := J + Y$ is the community infectious density and represents all not hospitalised infected individuals, which can be used to estimate the expected proportion of COVID PCR+ in the general population,
- $C(t) := \sum_{\tau \leq t} \sum_i (J_{i,1}(\tau) + Y_{i,1}(\tau)) = S_0 - S(t)$ and $\Delta C(t)$ are the cumulative and instantaneous incidence respectively.

The latter is used for the calculation of the most scrutinised indicator in epidemic monitoring, namely the temporal (or effective) reproduction number, $\mathcal{R}(t)$, which we calculate here following Wallinga & Lipsitch, at the time of the infectees' contamination:

$$\mathcal{R}(t) := \frac{\Delta C(t)}{\sum_{\tau \geq 1} \Delta C(t-\tau)\, \zeta_\tau}.$$

In this work, (population) immunisation ι refers to the proportion of individuals that have been infected by SARS-CoV-2. We assume waning immunity is negligible at this timescale. Its calculation is simply

$$\iota(t) := \frac{R - D}{S_0 - D}.$$

In absence of waning immunity from the host and antigenic drift from the virus, and assuming public health measures are fully relaxed, further epidemic can only be passively prevented if the herd immunity threshold is reached, i.e. $\iota \geq \iota_\mathrm{h}$ where

$$\iota_\mathrm{h} = 1 - \frac{1}{\mathcal{R}_0}.$$

A classical result from Kermack & McKendrick is that if the epidemic has started spreading, the final cumulative relative incidence will not stop at the herd immunity

threshold, but continue to a greater value, known as the final size proportion, that has no close form solution, but can implicitly be defined as

$$q_{\mathrm{f}} := \{q \in (0,1) : \mathcal{R}_0 q + \log(1-q) = 0\}.$$

Finally, current prevalence in the community is simply given by

$$\pi(t) = \frac{I}{S+I+R}.$$

S2.9 Continuous time model

The Markovian continuous-time model analogous to S-1 is illustrated in Figure S-3

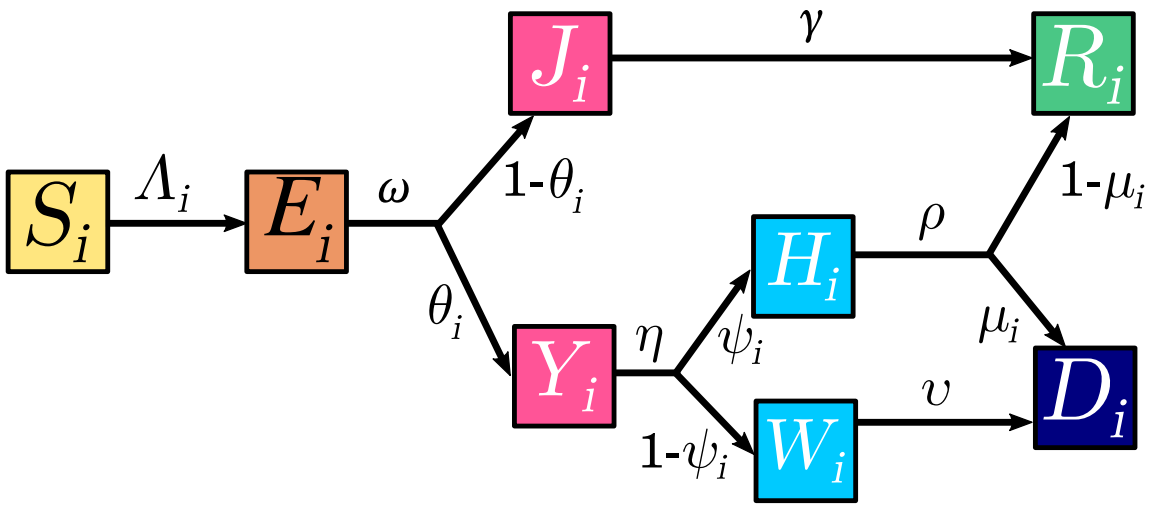

Figure S-3: **COVID-19 epidemic continuous time model structure**
Each square represents a group of individuals who share the same clinical kinetics and who contribute equally to the epidemic dynamics. Pink boxes correspond to infected individuals in the community. Light blue boxes represent critical cases cared for in hospitals. Arrows between boxes correspond to instantaneous flow of individuals.

and corresponds to the following set of ordinary differential equations

$$\begin{cases} \frac{\mathrm{d}S_i}{\mathrm{d}t} &= -\Lambda_i S_i, \\ \frac{\mathrm{d}E_i}{\mathrm{d}t} &= \Lambda_i S_i - \omega E_i, \\ \frac{\mathrm{d}J_i}{\mathrm{d}t} &= (1-\theta_i)\,\omega E_i - \gamma J_i, \\ \frac{\mathrm{d}Y_i}{\mathrm{d}t} &= \theta_i \omega E_i - \eta Y_i, \\ \frac{\mathrm{d}H_i}{\mathrm{d}t} &= \psi_i \eta Y_i - \rho H_i, \\ \frac{\mathrm{d}W_i}{\mathrm{d}t} &= (1-\psi_i)\,\eta Y_i - \upsilon W_i, \\ \frac{\mathrm{d}R_i}{\mathrm{d}t} &= (1-\mu_i)\,\rho H_i + \gamma J_i, \\ \frac{\mathrm{d}D_i}{\mathrm{d}t} &= \mu_i \rho H_i + \upsilon W_i, \end{cases}$$

where the force of infection is given by

$$\Lambda_i(t) = \frac{c_i(t)\,\gamma \mathcal{R}_0}{S_0} \sum_j c_j(t)\,(J_j + Y_j)\,.$$

Note that a latent compartment E has been added to account for a delay between contamination time and the beginning of the infectious period. The average latency and infectious period are equal to ω^{-1} and γ^{-1}. By construction, all transition times are exponentially distributed.

To compare this model to the focal non-Markovian discrete-time model introduced in this work, the output of the numerical integration was sampled at integer-valued time points. Then the fitting procedure used for the focal model was applied, though with a supplementary degree of freedom. Indeed, in the one hand, H is here exponentially distributed, therefore $\mathbb{V}[\mathrm{H}]$ is determined by $\mathbb{E}[\mathrm{H}]$ and, in the other hand, proper parameters ω and γ need to be fitted as there is no one-way relationship from generation time to latency and contagious periods. For information purposes, the maximum likelihood estimates and corresponding 95%-likelihood intervals of $\mathcal{R}_0$, t_0 and κ found by the estimation procedure applied to this Markovian model were respectively $4.3\ [2.9, 5.8]$, $01-22$ $[01-20, 01-23]$ and $26\%\ [5, 47]\%$.

S3 Supplementary Results

S3.1 Maximum likelihood parameter estimates

main input parameter	**notation**	**maximum likelihood estimates**	**95% - likelihood interval**
basic reproduction number	$\mathcal{R}_0$	2.99	$[2.59, 3.39]$
initiation day (YY-MM-DD)	t_0	20-01-20	[20-01-12, 20-01-28]
lock-down control (%)	κ	75.9	$[72.9, 78.7]$
critical case contamination to hospitalisation interval expectation (days)	$\mathbb{E}[\mathrm{H}]$	14.5	$[13.6, 15.4]$
critical case contamination to hospitalisation interval variance (days^2)	$\mathbb{V}[\mathrm{H}]$	20.0	$[11.4, 30.9]$
long ICU stay length expectation (days)	$\mathbb{E}[\mathrm{P}]$	16.7	$[14.9, 18.8]$
critical case hospitalisation to death interval expectation (non long-stay ICU patients (days)	$\mathbb{E}[\Upsilon]$	6.63	$[6.19, 7.10]$
infection fatality ratio correction factor(%)	$\mathfrak{C}_{\mathrm{F}}$	87.2	$[85.8, 88.5]$
long-stay ICU fatality ratio correction factor (%)	$\mathfrak{C}_{\mathrm{M}}$	100.3	$[100.2, 100.5]$
long-stay ICU frequency correction factor (%)	$\mathfrak{C}_{\Psi}$	93.8	$[93.0, 94.5]$

Table S-5: **Maximum likelihood estimates and associated 95% - likelihood intervals for the ten input parameters.** Details about the estimation procedure are provided in section S2.7.

S3.2 Alternative IFR initial values

The age-stratified IFR values used for the initialisation of the fitting procedure were the same as in [30]. The resulting demographic averaged IFR ranging from 1.13 to 1.17%, which is close to the corresponding weighted mean IFR calculated from the estimates found in [20], namely 0.9%. To investigate the impact of the choice of initial age-stratified IFR values on our results, we ran the fitting and estimation procedure replacing the initial the age-stratified IFR values with the estimates provided by [20]. Table S-6 shows the MLE and corresponding likelihood intervals of the parameters following this change.

main input parameter	**notation**	**maximum likelihood estimates**	**95% - likelihood interval**
basic reproduction number	$\mathcal{R}_0$	3.33	$[3.16, 3.50]$
initiation day (YY-MM-DD)	t_0	20-01-22	[20-01-20, 20-01-25]
lock-down control (%)	κ	77.6	$[76.4, 78.7]$
critical case contamination to hospitalisation interval expectation (days)	$\mathbb{E}[\mathrm{H}]$	14.0	$[13.7, 14.3]$
critical case contamination to hospitalisation interval variance (days^2)	$\mathbb{V}[\mathrm{H}]$	19.4	$[18.2, 20.6]$
long ICU stay length expectation (days)	$\mathbb{E}[\mathrm{P}]$	19.2	$[18.4, 19.9]$
critical case hospitalisation to death interval expectation (non long-stay ICU patients (days)	$\mathbb{E}[\Upsilon]$	9.43	$[9.20, 9.66]$
infection fatality ratio correction factor (%)	$\mathfrak{C}_{\mathrm{F}}$	62.0	$[61.0, 63.0]$
long-stay ICU fatality ratio correction factor (%)	$\mathfrak{C}_{\mathrm{M}}$	110.0	$[74.4, 146.0]$
long-stay ICU frequency correction factor (%)	$\mathfrak{C}_{\Psi}$	100.2	$[100.1, 100.2]$

Table S-6: **Maximum likelihood estimates and associated 95% - likelihood intervals for the ten input parameters using alternative IFR values.** Details about the estimation procedure are provided in section S2.7.

Using this alternative parameter setting, the reproduction number value by the end of lock-down was estimated to 0.70 [0.66, 0.73] while the immunisation proportion was equal to 5.05 [4.45, 5.58]%. Additionally, the model thus configured estimates that implementing the lock-down a week earlier could have led to 13,800 [13,400, 14,100] less deaths, while a one-week delay would have cause 54,700 [47,600, 61,600] more. These values are in line with those provided by the main parameter configuration.

S3.3 Forecasting

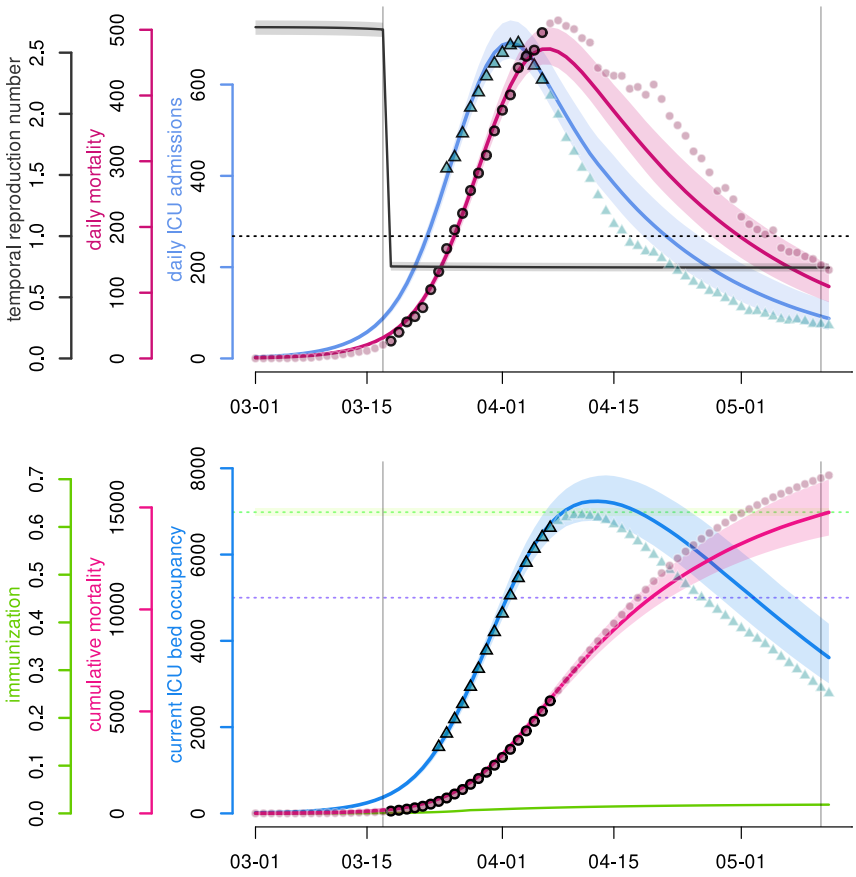

Figure S-4: **The COVID-19 epidemic wave in France as forecasted by early datasets.**
The nationwide hospital time series are here simulated by the model solely based on the publicly released daily figures available between Mar 18 and Apr 7 2020. The corresponding 7-day rolling-averaged data points on which the inference was based are outlined in black. Because of the limited number of points, the correction factors were excluded from the analysis (i.e. set to 1). **Top panel.** The blue and pink curves respectively represent the median daily ICU admissions and the median daily (hospital) mortality as generated by the fitted model. Turquoise triangles and red circles are the (rolling 7-day average) data counterparts. The black curve shows the median daily temporal reproduction number calculated from the simulated epidemic. The dotted horizontal line shows the reproduction number threshold value, i.e. 1. **Bottom panel.** The blue and pink curves respectively represent the median number of occupied beds in ICU nationwide and the median cumulative (hospital) mortality as generated by the fitted model. The turquoise triangles and red circles are the (rolling 7-day average) data counterparts. The purple dotted horizontal line shows the initial French ICU capacity, ca. 5,000 beds. The green curve shows the median proportion of the population that has recovered (and is assumed to be immune). The green dotted horizontal line corresponds to the median herd immunity threshold $1 - \mathcal{R}_0^{-1}$.The two vertical lines show respectively (from left to right) the beginning and the end of the French national lock-down. Shaded areas correspond to 95% confidence intervals.