Physical Biology

PAPER

Why case fatality ratios can be misleading: individual- and population-based mortality estimates and factors influencing them

RECEIVED
26 April 2020

REVISED
3 June 2020

ACCEPTED FOR PUBLICATION
18 June 2020

PUBLISHED
17 September 2020

Lucas Böttcher[1], Mingtao Xia[2] and Tom Chou[1,2,3]

[1] Dept. of Computational Medicine, UCLA, Los Angeles, CA 90095-1766, United States of America
[2] Dept. of Mathematics, UCLA, Los Angeles, CA 90095-1555, United States of America
[3] Author to whom any correspondence should be addressed.

E-mail: tomchou@ucla.edu

Keywords: COVID-19, case fatality ratio, mortality, epidemic, survival probability, SIR model

Abstract

Different ways of calculating mortality during epidemics have yielded very different results, particularly during the current COVID-19 pandemic. For example, the 'CFR' has been interchangeably called the case fatality ratio, case fatality rate, and case fatality risk, often without standard mathematical definitions. The most commonly used CFR is the *case fatality ratio*, typically constructed using the estimated number of deaths to date divided by the estimated total number of confirmed infected cases to date. How does this CFR relate to an infected individual's probability of death? To explore such issues, we formulate both a survival probability model and an associated infection duration-dependent SIR model to define individual- and population-based estimates of dynamic mortality measures to show that neither of these are directly represented by the case fatality ratio. The key parameters that affect the dynamics of different mortality estimates are the incubation period and the time individuals were infected before confirmation of infection. Using data on the recent SARS-CoV-2 outbreaks, we estimate and compare the different dynamic mortality estimates and highlight their differences. Informed by our modeling, we propose more systematic methods to determine mortality during epidemic outbreaks and discuss sensitivity to confounding effects and uncertainties in the data arising from, e.g., undertesting and heterogeneous populations.

1. Introduction

Mortality metrics are key quantities describing the severity of a viral disease [1]. During an outbreak, these metrics typically evolve in time before converging to a constant value and can be defined in a number of ways. Commonly used metrics are the case fatality ratio, case fatality rate, and case fatality risk, which are all confusingly denoted 'CFR' [2, 3]. Fatality *rate* implies a change in deaths per unit time, *risk* implies an individual probability, while *ratio* implies a fraction of two numbers, typically populations. CFR is most often defined as the ratio of the total estimated number of deaths to date, $D(t)$, to the estimated number of all confirmed cases to date $N_c(t)$ [1, 4–6]. These numbers are key to estimating disease severity. Usually, antibody [7] and reverse transcription-polymerase chain reaction (RT-PCR) testing [8] is used to confirm SARS-CoV-2-positive patients. To find $D(t)$, the number of patients who actually die of COVID-19 must also be quantified. In Italy, deaths of patients with positive RT-PCR testing for SARS-CoV-2 are reported as COVID-19 deaths, but the criteria for COVID-19-related deaths are currently not clearly defined and may vary from region to region [9].

Studies that define CFR as the 'case fatality risk' associate it with the probability of death of an individual confirmed case within 'a period of time' [10]. Yet others define case fatality ratio as simply 'case fatality' and reserve the term case fatality ratio to mean the ratio of the case fatalities of two different diseases [3]. Infection fatality ratios (IFR), the number of deaths to date divided by the number of all infected individuals, have also been used [11–13] although the IFR $= D(t)/N(t)$ requires an estimate of $N(t)$, the number of total (including unconfirmed) infected individuals. Similar to the

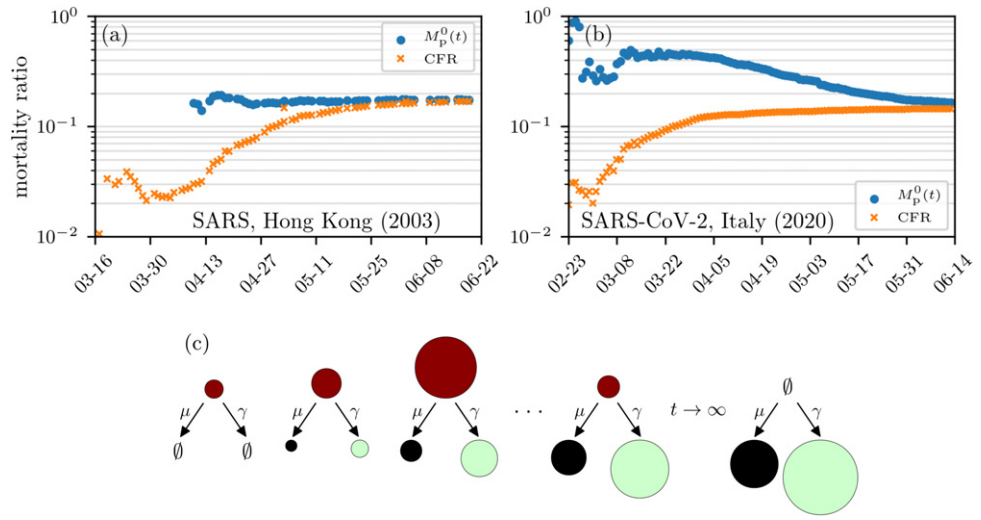

Figure 1. Mortality estimates. (a) and (b) estimates of mortality ratios (see equations (9) and (14)) of SARS-CoV infections in Hong Kong (2003) [20] and SARS-CoV-2 infections in Italy. (c) Evolution of the cumulative number of infected (red), death (black), and recovered (green) cases. The size of the circles indicates the number of cases in the respective compartments on a certain day. Note that CFR and $M_p^0(t)$ have exhibited qualitatively similar behavior across different epidemics. The data are based on reference [21].

ambiguity in the definition of CFR, the IFR has also been called the 'infection fatality risk,' the probability of an individual dying conditioned on being infected. This individual-based definition of IFR is thus equivalent to the individual-based case fatality risk. However, in nearly all practical cases, both the CFR and IFR are estimated from aggregated population data from past outbreaks [4] as well as from those of the recent SARS-CoV-2 outbreaks [1, 11, 13–17].

Since case fatality ratio is the most commonly used, we henceforth define CFR $= D(t)/N_c(t)$. We show examples of CFR curves (orange), which typically vary significantly both by region and in time, in figure 1 and in the results and discussion section. During the severe acute respiratory syndrome (SARS) outbreak in Hong Kong in 2003, the World Health Organization (WHO) also used the aforementioned estimate to obtain an initial CFR $\sim$3% while the final values, after resolution of infections, approached 17.0% [18, 19] (see figure 1(a)).

Another population-based mortality ratio is $M_p(t) = D(t)/(D(t) + R(t))$, the number of deaths divided by the sum of death and recovered cases (the number of *resolved* cases), up to time t is shown in blue in figures 1(a) and (b). In principle, $M_p(t)$ should be a better measure of the likelihood of death, but it is underestimated by the CFR $= D(t)/N_c(t)$. For example, on April 25, 2020, the worldwide $M_p(t)$ was $203164/(203164+836612) \approx 20\%$, significantly higher than the April 25, 2020 CFR$(t) = D(t)/N_c(t) = 203\,164/2919\,404 \approx 7\%$ [22]. Despite this underestimation, the CFR is still commonly used by the WHO and other health officials, such as in the ongoing SARS-CoV-2 outbreaks [1, 12, 13, 15, 17] (see table 1). As shown in figure 1(c), the CFR would correspond to the mortality ratio only if all tested infected individuals recover. Such underestimations by CFRs may lead to insufficient countermeasures and a more severe epidemic [23, 24].

Table 1. Different CFR estimates of COVID-19.

Reference	CFR
Xu *et al* [5, 25] and Mahase [26]	2%
Wu and McGoogan [6]	0.1–1% (outside Wuhan)
World Health Organization [27, 28]	2–4%
Porcheddu *et al* [29]	2.3% (Italy and China)
Peeri *et al* [30]	2%

Since meaningful and accurate mortality metrics are critical for assessing the risks associated with epidemic outbreaks, we first unambiguously define the probability $M_1(t)$ that a single, newly infected individual will die of the disease by a given time. This probability has also been called the case fatality risk, but without specifying its dependence on time after infection [10]. This intrinsic mortality or probability of death, can be identified as one minus the survival probability of a single infected individual. It should be an *intrinsic* property of the virus and the infected individual, depending on age, health, access to health care, etc, and not *directly* on the population-level dynamics of infected and recovered individuals. Whether this individual infects others does not directly affect his probability of eventually dying [31].

In the next section, we derive a survival probability model for $M_1(t)$ similar to that in Ghani *et al* [32]. Importantly, our individual survival model incorporates the duration of infection (including an incubation period) before a patient tests positive at time $t = 0$. However, the CFR and other mortality measures are typically reported based on population data. Do these population-based measures,

including CFR, provide reasonable measures of the probability of death of an individual? To address these and related issues, we develop an analogous population-based mortality metric based on a disease duration-structured SIR model. While population-based estimates of CFR are typically not a meaningful measure of individual mortality, under simplifying assumptions, the population-based mortality ratio $M_p(t)$ is more closely related to the true probability of death $M_1(t)$ [32].

We will use the same rate parameters in our individual and population models to compute and compare the different mortality measures. By critically analyzing and comparing these estimates, the CFR, and a 'delayed' case fatality ratio CFR_d, we illustrate and interpret the differences among these measures and discuss how changes or uncertainty in the data affect them. In the results and discussion section, we identify a correction factor to transform population-level mortality estimates into individual mortality probabilities, and discuss the effects of other possible confounding factors such as heterogeneous populations and undertesting (unconfirmed cases).

2. Mortality measures

In this section, we present different mortality measures for *confirmed* cases and outline their underlying mathematical models.

2.1. Intrinsic individual mortality rate

Consider an individual that, at the time of positive testing ($t = 0$), had been infected for a duration τ_1. A 'survival' probability density can be defined such that $P(\tau, t|\tau_1)\mathrm{d}\tau$ is the probability that the patient is still alive and infected (not recovered) at time $t > 0$ and has been infected for a duration between τ and $\tau + \mathrm{d}\tau$. Since τ_1 is unknown, it must be estimated or averaged over some distribution. The individual survival probability evolves according to [33].

$$\frac{\partial P(\tau, t|\tau_1)}{\partial t} + \frac{\partial P(\tau, t|\tau_1)}{\partial \tau} = -(\mu(\tau, t|\tau_1) + \gamma(\tau, t|\tau_1))P(\tau, t|\tau_1), \quad (1)$$

where the death and recovery rates, $\mu(\tau, t|\tau_1)$ and $\gamma(\tau, t|\tau_1)$, depend explicitly on the duration of infection at time t and can be further implicitly stratified according to patient age, gender, health condition, etc [1, 34]. They may also depend explicitly on time t to reflect changes in clinical policy or available health care. For example, enhanced medical care may decrease the death rate μ, giving the individual's intrinsic physiological processes a chance to cure the patient.

If we assume an initial condition of one individual having been infected for time τ_1 at the time of confirmation, equation (1) can be solved using the method of characteristics shown in the appendix A. From the solution $P(\tau = t + \tau_1, t|\tau_1)$, one can derive the probabilities of death and recovery by time t as

$$P_d(t|\tau_1) = \int_0^t \mathrm{d}s\, \mu(\tau_1 + s, s)P(\tau_1 + s, t|\tau_1),$$
$$P_r(t|\tau_1) = \int_0^t \mathrm{d}s\, \gamma(\tau_1 + s, s)P(\tau_1 + s, t|\tau_1). \quad (2)$$

The probability that an individual died before time t, conditioned on resolution (either death or recovery), is then defined as

$$M_1(t|\tau_1) = \frac{P_d(t|\tau_1)}{P_d(t|\tau_1) + P_r(t|\tau_1)}. \quad (3)$$

Equations (2) and (3) also depend on all other relevant patient attributes such as age, accessibility to health care, etc. In the long-time limit, when resolution has occurred ($P_d(\infty|\tau_1) + P_r(\infty|\tau_1) = 1$), the individual mortality ratio is simply $M_1(\infty|\tau_1) = P_d(\infty|\tau_1)$. In order to capture the dependence of death and recovery rates on the time an individual has been infected, we propose a constant recovery rate γ and a piecewise constant death rate $\mu(\tau|\tau_1)$ that is not explicitly a function of time t:

$$\gamma(\tau, t|\tau_1) = \gamma, \quad \mu(\tau|\tau_1) = \begin{cases} 0 & \tau \leqslant \tau_{\mathrm{inc}} \\ \mu_1 & \tau > \tau_{\mathrm{inc}} \end{cases}. \quad (4)$$

The parameter τ_{inc} is the incubation time during which the patient is asymptomatic, has negligible chance of dying, but can recover by clearing the virus. In other words, some patients fully recover without ever developing serious symptoms.

For coronavirus infections, the incubation period appears to be highly variable with a mean of $\tau_{\mathrm{inc}} \approx 6.4$ days [36]. We can estimate μ_1 and γ using recent individual patient data from Singapore where 178 patients (mean age: 46 years) had been tracked from the date on which their first symptoms occurred until they recovered [35], on average, after 13.7 days. We show the recovery-time distribution in figure 2(a). Compared to other existing datasets, the Singapore COVID-19 dataset provides complete line lists for a large number of patients and is being updated regularly.

We then use the global mortality of all resolved cases (≈20% [22]) to determine the dependence between μ_1 and γ via $\mu_1/(\mu_1 + \gamma) \approx 1/5$ (or $\gamma/\mu_1 \approx 4$). The constant recovery and post incubation death rates [37] are thus

$$\gamma \approx \frac{1}{13.7}/\mathrm{day} = 0.073/\mathrm{day} \quad \text{and}$$
$$\mu_1 \approx \gamma/4 = 0.018/\mathrm{day}. \quad (5)$$

Using these numbers, the recovery and death rate functions $\gamma(\tau, t|\tau_1)$ and $\mu(\tau|\tau_1)$ are plotted as functions of τ in figure 2(b). We show the evolution of

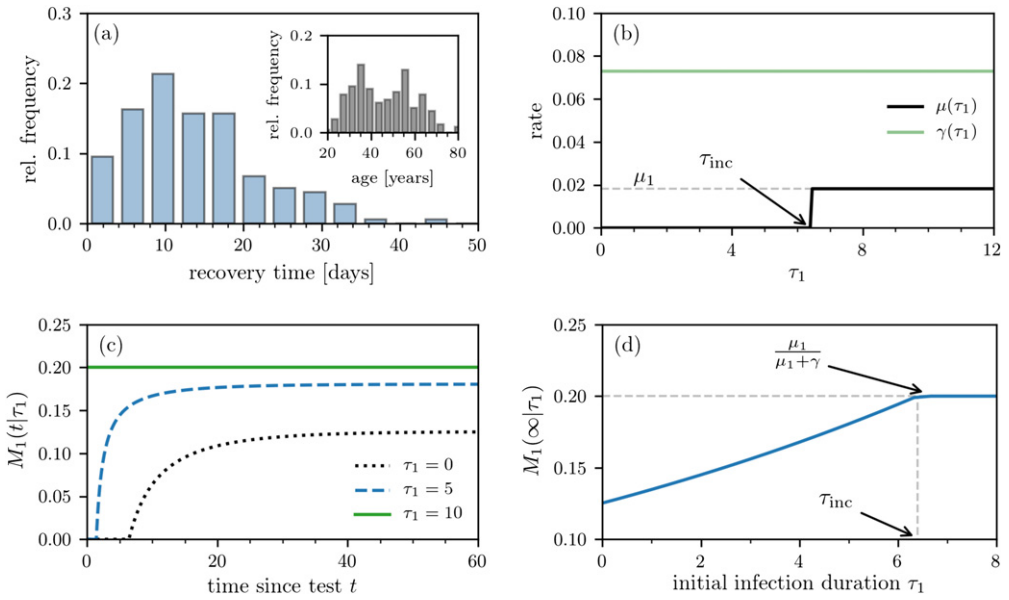

Figure 2. Individual mortality. (a) Recovery time after first symptoms occurred based on individual data of 178 patients [35]. The inset shows the age distribution of these patients. (b) Death- and recovery rates as defined in equation (4). The death rate $\mu(\tau_1)$ approaches μ_1 for $\tau_1 > \tau_{\text{inc}}$, where τ_{inc} is the incubation period and τ_1 is the time the patient has been infected before first being tested positive. (c) The individual mortality ratio $M_1(t|\tau_1)$ for $\tau_{\text{inc}} = 6.4$ days at different values of τ_1. Note that the individual death probability $P_d(t|\tau_1)$ and $M_1(t|\tau_1)$ are nonzero only after $t > \tau_{\text{inc}} - \tau_1$. (d) The asymptotic individual mortality ratio $M_1(\infty)$ (see equation (3)) as a function of τ_1.

$M_1(t|\tau_1)$ at different values of τ_1 in figure 2(c). The corresponding long-time limit $M_1(\infty|\tau_1)$ is readily apparent in figure 2(d): for $\tau_1 \geqslant \tau_{\text{inc}}$, $M_1(\infty|\tau_1) = \mu_1/(\mu_1 + \gamma) \approx 0.2$, while $M_1(\infty|\tau_1) < \mu_1/(\mu_1 + \gamma)$ when $\tau_1 < \tau_{\text{inc}}$. The smaller expected mortality associated with early identification of infection arises from the remaining incubation time during which the patient has a chance to recover without possibility of death. When conditioned on testing positive at or after the incubation period, the patient immediately experiences a positive death rate, increasing his $M_1(\infty|\tau_1)$.

In order to infer M_1 (and also indirectly μ and γ) during an outbreak, a number of statistical issues must be considered. First, if the outbreak is ongoing, there may not be sufficient long-time cohort data. Second, τ_1 is unknown. Since testing typically occurs at the onset of symptoms, most positive patients will have been infected a few days earlier. The uncertainty in τ_1 can be represented by a probability density $\rho(\tau_1)$ for the individual. The expected mortality can then be constructed as an average over $\rho(\tau_1)$:

$$\bar{M}_1(t) = \frac{\bar{P}_d(t)}{\bar{P}_d(t) + \bar{P}_r(t)}, \tag{6}$$

where $\bar{P}_d(t)$ and $\bar{P}_r(t)$ are the τ_1-averaged probabilities death and cure probabilities.

Some properties of the distribution $\rho(\tau_1)$ can be inferred from the behavior of patients. Before symptoms arise, only very few patients will know they have been infected, seek medical care, and get their case confirmed (i.e., $\rho(\tau_1) \approx 0$ for $\tau_1 \approx 0$). The majority of patients will seek care when they have been infected for approximately τ_{inc}. We choose the gamma distribution

$$\rho(\tau_1; n, \lambda) = \frac{\lambda^n}{\Gamma(n)} \tau_1^{n-1} \, \mathrm{e}^{-\lambda\tau_1} \tag{7}$$

with shape parameter $n = 8$ and rate parameter $\lambda = 1.25/\text{day}$ so that the mean n/λ is equal to $\tau_{\text{inc}} = 6.4$. Note that, independent of the distribution ρ, the average $\bar{M}_1(t)$ is bounded from above by $M_1(\infty) = \mu_1/(\mu_1 + \gamma)$ for all times t.

Upon using the rates in equation (4) and averaging over $\rho(\tau_1)$, we derived expressions for $\bar{P}(t)$, $\bar{P}_d(t)$, and $\bar{P}_r(t)$ which are explicitly given in the appendix A. Using the values in equation (5) we find an expected individual mortality ratio $\bar{M}_1(t)$ (which are subsequently plotted in figure 3) and its asymptotic value $\bar{M}_1(\infty) = \bar{P}_d(\infty) \approx 0.19$ (slightly less than $M_1(\infty|\tau_1)$ due to averaging over $\rho(\tau_1)$). Of course, it is also possible to account for more complex time-dependent forms of γ and μ_1 [38], but we will primarily use equation (4) in our subsequent analyses. We stress that $M_1(t)$ tracks mortality of a cohort of individuals infected at about the same time, and does not include mortality of newly infected individuals. Thus, it can be trivially stratified according to different age groups and defined as the mortality $M_1(t|\mu)$ of each age subpopulation with death rate μ.

In the next subsection, we define population-based estimates for mortality ratios, $M_p(t)$, and explore how they can be computed using SIR-type models. By comparing $\bar{M}_1(t)$ to $M_p(t)$, we gain

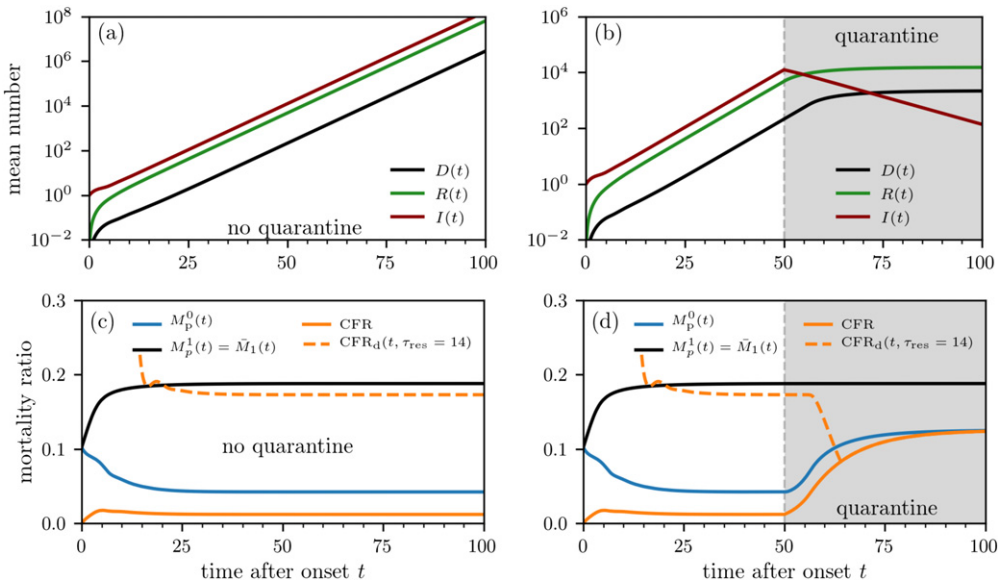

Figure 3. Population-level mortality estimates. Outbreak evolution and mortality ratios without containment measures (a) and (c) and with quarantine (b) and (d). The curves are based on numerical solutions of equation (10) using the initial condition $I(\tau, 0) = \rho(\tau; 8, 1.25)$ (see equation (7)). The death and recovery rates are defined in equations (4) and (5). We use an infection rate (equation (16)) defined by $\beta_0 S_0 = 4.64$/day, which we estimated from the basic reproduction number of SARS-CoV-2 [36]. To model quarantine effects, we set $\beta_0 S_0 = 0$ for $t > 50$ days. We show the mortality-ratio estimates $M_p^0(t)$ and $M_p^1(t)$ (see equation (14)) and $\mathrm{CFR}_d(t, \tau_{res})$ (see equations (8), (12), and (14)). $\mathrm{CFR}_d(t, \tau_{res} = 14$ days) behaves very differently from CFR, initially decreasing for $\tau_{res} > 0$ and significantly *overestimating* $M_p^0(t)$ but providing a reasonable estimate of $\bar{M}_1(t) = M_p^1(t)$ without quarantine. Note that under quarantine, $\mathrm{CFR}(\infty)$, $\mathrm{CFR}_d(\infty)$, and $M_p^0(\infty)$ approach the same value since they reflect the mortality ratio of the total cohort at the time of quarantine. On the other hand, $\bar{M}_1(t) = M_p^1(t)$ reflects the ratio of the initial cohort at the start of the outbreak and remains unchanged from the no-quarantine case.

insight into whether population-based metrics are good proxies for individual mortality ratios.

2.2. Relation to infection duration-dependent SIR model

While individual mortalities can be estimated by tracking many individuals from infection to recovery or death, often, the available data are not resolved at the individual level and only total populations are given. Typically, one only has the total number of confirmed cases accumulated up to time t, $N_c(t)$, the number of deaths to date $D(t)$, and the number of cured/recovered patients to date $R(t)$ (see figure 1). Note that $N_c(t)$ includes unresolved cases and that $N_c(t) \geqslant R(t) + D(t)$. Resolution (death or recovery) of all patients, $N_c(\infty) = R(\infty) + D(\infty)$, occurs only well after the epidemic completely passes.

A variant of the CFR commonly used in the literature is the delayed CFR [5, 6]

$$\mathrm{CFR}_d(t, \tau_{res}) = \frac{D(t)}{N_c(t - \tau_{res})}, \tag{8}$$

which uses an earlier and smaller case number to compensate for underestimation by the standard CFR

$$\mathrm{CFR}(t) = \frac{D(t)}{N_c(t)} \equiv \mathrm{CFR}_d(t, \tau_{res} = 0). \tag{9}$$

The delay τ_{res} used is typically the time between the day symptoms first occurred and the day of death or recovery. To determine a realistic value of the delay time τ_{res} (which can be qualitatively interpreted as a resolution time), we use data on death/recovery periods of 36 tracked COVID-19 patients [39] and find that patients recover/die, on average, $\tau_{res} \approx$ 2 weeks after first symptoms occurred. The delayed $\mathrm{CFR}_d(t, \tau_{res} > 0)$ also underestimates the individual mortality in previous epidemic outbreaks of SARS [18, 32] and Ebola [40], but is highly sensitive to τ_{res}. If the delay between the time of infection and time of resolution were vanishingly small, we can set $\tau_{res} = 0$ and find that the CFR_d and CFR are equivalent (see equation (9)).

Alternatively, a simple and interpretable population-level mortality is $M_p(t) = D(t)/(R(t) + D(t))$, the ratio of infected deaths to all resolved cases of confirmed infections. To provide a concrete model for $D(t)$ and $R(t)$, and hence $M_p(t)$, we will use a variant of the standard infection duration-dependent susceptible-infected-recovered (SIR)-type model described by [41, 42]

$$\frac{dS(t)}{dt} = -S(t) \int_0^\infty d\tau' \, \beta(\tau', t) I(\tau', t),$$
$$\frac{\partial I(\tau, t)}{\partial t} + \frac{\partial I(\tau, t)}{\partial \tau} = -(\mu(\tau, t) + \gamma(\tau, t)) I(\tau, t), \tag{10}$$

and $dR(t)/dt = \int_0^\infty d\tau \gamma(\tau, t) I(\tau, t)$, where $S(t)$ is the number of susceptibles, $I(\tau, t)$ is density of individuals at time t who have been infected for time τ, and

$R(t)$ is the number of recovered individuals. The rate at which an individual infected for time τ at time t infects susceptibles is denoted by $\beta(\tau, t)S(t)$. For simplicity, we assume only community spread and neglect immigration of infected individuals, which could be straightforwardly included [42].

Note that the equation for $I(\tau, t)$ is identical to the equation for the survival probability described by equation (1). It is also equivalent to McKendrick age-structured models [43, 44]. In both the individual model (equation (1)) and population model (equation (10)), the death and recovery rates are insensitive to changes in age a over the $\lesssim 1$ year epidemic timescale. In this limit, we consider only infection-duration dependence in the population dynamics. However, in contrast to the individual survival probability, new infections of susceptibles are described by the boundary condition (or renewal equation)

$$I(\tau = 0, t) = S(t)\int_0^\infty \mathrm{d}\tau'\, \beta(\tau', t)I(\tau', t), \qquad (11)$$

which is similar to that used in age-structured models to represent birth [43]. The initial time $t = 0$ is arbitrary as long as the initial condition $I(\tau, 0)$ is defined. We use an initial condition corresponding to a single infected with the infection duration density given by equation (7): $I(\tau, 0) = \rho(\tau; n = 8, \lambda = 1.25)$. Note that equation (11) assumes that all newly infected individuals are immediately identified; i.e., these newly infected individuals start with $\tau_1 = 0$. After solving for the infected population density, we find the total number of deaths, recoveries, and total cases to date,

$$\begin{aligned} D^0(t) &= \int_0^t \mathrm{d}t' \int_0^\infty \mathrm{d}\tau\, \mu(\tau, t')I(\tau, t'),\\ R^0(t) &= \int_0^t \mathrm{d}t' \int_0^\infty \mathrm{d}\tau\, \gamma(\tau, t')I(\tau, t'),\\ N^0(t) &= R^0(t) + D^0(t) + \int_0^\infty \mathrm{d}\tau\, I(\tau, t), \end{aligned} \qquad (12)$$

and use $D^0(t)$ and $N^0(t)$ for $D(t)$ and $N_c(t)$ in definitions of $\mathrm{CFR}(t)$ and $\mathrm{CFR}_d(t, \tau_{res})$ (equation (8)). In the definitions of $D^0(t)$, $R^0(t)$, and $N^0(t)$, we account for all possible death and recovery cases to date (see appendix A) and that newly infected individuals are immediately identified. We use these case numbers as approximations of the reported case numbers to study the evolution of mortality ratio estimates. Mortalities based on these numbers underestimate the actual individual mortality M_1 (see the previous 'intrinsic individual mortality rate' subsection) since they involve individuals that have been infected for different durations τ, particularly recently infected individuals who have not yet died.

An alternative way to compute populations is to exclude new infections and consider only an initial cohort. The corresponding populations in this case are defined as

$$\begin{aligned} D^1(t) &= \int_0^t \mathrm{d}t' \int_{t'}^\infty \mathrm{d}\tau\, \mu(\tau, t')I(\tau, t'),\\ R^1(t) &= \int_0^t \mathrm{d}t' \int_{t'}^\infty \mathrm{d}\tau\, \gamma(\tau, t')I(\tau, t'). \end{aligned} \qquad (13)$$

Since $D^1(t)$ and $R^1(t)$ do not include infected individuals with $\tau < t$, they exclude the effect of newly infected individuals and may yield more meaningful mortalities as they would be based on an initial cohort of individuals in the distant past. It is superfluous to define CFR using $D^1(t)/N_c$ because the corresponding N_c of a cohort is a constant. The infections that occur after $t = 0$ contribute only to $I(\tau < t, t)$; thus, $D^1(t)$ and $R^1(t)$ do not depend on the transmission rate β, possible immigration of infected individuals, or the number of susceptibles $S(t)$. Note that all the populations derived above implicitly average over $\rho(\tau_1; n, \gamma)$ for the first cohort of identified infected individuals (but not subsequent infections). Moreover, the population density $I(\tau \geqslant t, t)$ follows the same equation as $\bar{P}(t|\tau_1)$ provided the same $\rho(\tau_1; n, \lambda)$ is used in their respective calculations.

The two different ways of partitioning populations (equations (12) and (13)) lead to two different population-level mortality ratios

$$\begin{aligned} M_p^0(t) &\equiv \frac{D^0(t)}{D^0(t) + R^0(t)} \quad \text{and}\\ M_p^1(t) &\equiv \frac{D^1(t)}{D^1(t) + R^1(t)}. \end{aligned} \qquad (14)$$

Since the populations $D^0(t)$ and $R^0(t)$, and hence $M_p^0(t)$, depend on disease transmission through $\beta(\tau, t)$ and $S(t)$, we expect $M_p^0(t)$ to carry a different interpretation from $M_1(t)$ and $M_p^1(t)$.

In the special case in which μ and γ are constants, the time-integrated populations $\int_0^t \mathrm{d}t' \int_0^\infty \mathrm{d}\tau\, I(\tau, t')$ and $\int_0^t \mathrm{d}t' \int_{t'}^\infty \mathrm{d}\tau\, I(\tau, t')$ factor out of $M_p^0(t)$ and $M_p^1(t)$, rendering them time-independent and

$$M_p^{0,1} = \frac{\mu_1}{\mu_1 + \gamma} = M_1. \qquad (15)$$

Thus, only in the special time-homogeneous case do both population-based mortality ratios become *independent* of the population (and transmission β) and coincide with the individual death probability.

To illustrate, in more general cases, the differences between $M_1(t), M_p^{0,1}(t)$ and $\mathrm{CFR}_d(t, \tau_{res})$, we use the simple death and recovery rate functions given by equation (4) in solving equations (1) and (10). For $\beta(\tau, t)$ in equation (11), we use a recently inferred SARS-CoV-2 infectiousness profile [45] which is described by a gamma distribution

$$\beta(\tau) = \beta_0 \rho(\tau; n, \lambda) \qquad (16)$$

with a peak that occurs shortly before the onset of symptoms at time τ_{inc} and coincidentally has $n = 8$ and $\lambda = 1.25$/day as in the testing time distribution $\rho(\tau_1)$ from a single infected (equation (7)). The constant dimensionless prefactor β_0 sets the amplitude of the transmission rate. For the chosen parameters n and λ, the gamma distribution $\rho(\tau; n, \lambda)$ reaches a maximum at $\tau \approx 5.6$ days, about one day before $\tau_{\text{inc}} = 6.4$ days [45]. Assuming that the susceptible pool is not appreciably depleted, $S(t) \approx S_0$ and equation (11) becomes $I(\tau = 0, t) = \beta_0 S_0 \int_0^\infty \mathrm{d}\tau'\, \rho(\tau; n, \lambda) I(\tau', t)$. The amplitude $\beta_0 S_0$ can be found by assuming a single infected for $I(\tau, t)$ in the renewal equation and using the estimated basic reproduction number. The basic reproduction number $\mathcal{R}_0$ is the average number of secondary infections that result from any single infected individual before he dies or recovers [37]. There are two terms to consider when determining $\mathcal{R}_0$: (i) $\beta(\tau)\mathrm{d}\tau$ is the probability that an infection occurs in $[\tau, \tau + \mathrm{d}\tau]$ and (ii) $\exp\left[-\int_0^\tau (\mu(\tau') + \gamma)\mathrm{d}\tau'\right]$ is the probability that a single infected individual has not died or recovered prior to time τ. If we integrate over the product of these quantities and multiply by the total susceptible population S_0 (which is equivalent to the boundary condition (11) applied to a single infected individual), we obtain the average number of susceptibles infected by one infected individual, i.e., $\mathcal{R}_0$. Thus, upon using equation (16), $\beta_0 S_0$ can be found by solving

$$\begin{aligned} S_0 \int_0^\infty \beta(\tau) \exp\left[-\int_0^\tau (\mu(\tau') + \gamma)\mathrm{d}\tau'\right] \mathrm{d}\tau \\ = \beta_0 S_0 \int_0^\infty \rho(\tau; n, \lambda) \exp\left[-\int_0^\tau (\mu(\tau') + \gamma)\mathrm{d}\tau'\right] \mathrm{d}\tau \\ = \mathcal{R}_0 \approx 2.91. \end{aligned} \tag{17}$$

Using the death and recovery rate functions given by equations (4) and (5), we find $\beta_0 S_0 \approx 4.64$/day. Using this value, we numerically solve equations (10) and (11) (see appendix for further details) and use these solutions to compute $D^{0,1}(t)$, $R^{0,1}(t)$, and $N_{0,1}(t)$, which are then used in equation (14) and $\mathrm{CFR}_{\mathrm{d}}(t, \tau_{\text{res}})$.

3. Results and discussion

3.1. Comparison of mortalities

Here, we evaluate and compare the different mortality metrics and show how some of them qualitatively resemble the measured mortality estimates shown in figure 1. In figure 3(a), we show the unbounded subpopulations $I^0(t), D^0(t)$, and $R^0(t)$ computed using equations (10)–(12) when the susceptible population is assumed constant. Figure 3(b) shows the populations when a strict quarantine ($S(t > t_{\mathrm{q}}) = 0$) is applied after $t_{\mathrm{q}} = 50$ days. The mortalities plotted in figure 3(c) show that $M_{\mathrm{p}}^1(t)$ approaches the individual mortality ratio $\bar{M}_1(\infty) \approx 0.19$ given in the 'intrinsic individual mortality rate' subsection above. This occurs because the model for $P(\tau, t)$ and $I(\tau, t)$ are equivalent and we assumed the same initial distribution $\rho(\tau; 8, 1.25)$ for both quantities. However, the population-level mortality ratios $\mathrm{CFR}_{\mathrm{d}}(t, \tau_{\text{res}})$ and $M_{\mathrm{p}}^0(t)$ also take into account recently infected individuals who may recover before symptoms. This difference yields different mortality ratios because newly infected individuals are implicitly assumed to be detected immediately and all have $\tau_1 = 0$. Thus, the underlying infection-time distribution is not the same as that used to compute $\bar{M}_{\mathrm{p}}^1(t)$ (see appendix A for further details). The mortality ratio $M_{\mathrm{p}}^0(t)$ should not be used to quantify the individual mortality probability $\bar{M}_1(t)$ of individuals who tested positive, while the accuracy of $\mathrm{CFR}_{\mathrm{d}}(t, \tau_{\text{res}})$ is sensitive to τ_{res} and quarantining. Moreover, due to evolution of the disease, $D(t)$, $R(t)$, and $N(t)$ do not change with the same rates during an outbreak, the population-level mortality measures $\mathrm{CFR}_{\mathrm{d}}(t, \tau_{\text{res}})$ and $M_{\mathrm{p}}^0(t)$ reach their final steady state values only after sufficiently long times. Figure 3(d) shows the corresponding mortalities with quarantining after $t_{\mathrm{q}} = 50$ days.

The population-level ratios $M_{\mathrm{p}}^0(t)$ and $\mathrm{CFR}(t)$ implicitly depend on new infections and the transmission rate β. Despite this confounding factor, $M_{\mathrm{p}}^0(t)$ and $\mathrm{CFR}_{\mathrm{d}}(t, \tau_{\text{res}})$ approach $\mathrm{e}^{-\gamma\tau_{\text{inc}}}\mu_1/(\mu_1 + \gamma)$ as $t \to \infty$, where $\mathrm{e}^{-\gamma\tau_{\text{inc}}}$ is the probability that no recovery occurred during the incubation time τ_{inc}. Based on these results, we can establish the following connection between the different mortality ratios for initial infection times with distribution $\rho(\tau_1; n, \lambda)$ and mean $\bar{\tau} = n/\lambda$:

$$\begin{aligned} \mathrm{CFR}_{\mathrm{d}}(\infty) = M_{\mathrm{p}}^0(\infty) \approx \mathrm{e}^{-\gamma\bar{\tau}} M_{\mathrm{p}}^1(\infty) \\ = \mathrm{e}^{-\gamma\bar{\tau}} \bar{M}_1(\infty). \end{aligned} \tag{18}$$

According to equation (18), population-level mortality estimates (e.g., CFR and M_{p}^0), can be transformed, at least approximately, into individual mortality probabilities using the correction factor $\mathrm{e}^{-\gamma\bar{\tau}}$ with $\bar{\tau} \approx \tau_{\text{inc}}$. Although population-level quarantining does not directly affect the individual mortality $M_1(t|\tau_1)$ or $\bar{M}_1(t)$, it can be easily incorporated into the SIR-type population dynamics equations through changes in $\beta(\tau, t)S(t)$. For example, we have set $S(t > t_{\mathrm{q}}) = 0$ to represent implementation of a perfect quarantine after $t_{\mathrm{q}} = 50$ days of the outbreak. After $t_{\mathrm{q}} = 50$ days, no new infections occur and the estimates $\mathrm{CFR}(t)$ and $M_{\mathrm{p}}^0(t)$ start to converge toward their common larger value (see figure 3(d)). In other words, without quarantining, the infected and recovered populations are continuously increasing, keeping CFR and $M_{\mathrm{p}}^0(t)$ low. Since the number of deaths decreases after the implementation of quarantine measures, the delayed $\mathrm{CFR}_{\mathrm{d}}(t, \tau_{\text{res}} = 14\,\text{days})$ is first decreasing until $t = t_{\mathrm{q}} + \tau_{\text{res}} = 64$ days. For $t > 64$ days, the $\mathrm{CFR}_{\mathrm{d}}(t, \tau_{\text{res}} = 14\,\text{days})$ measures no new cases and is thus equal to the CFR.

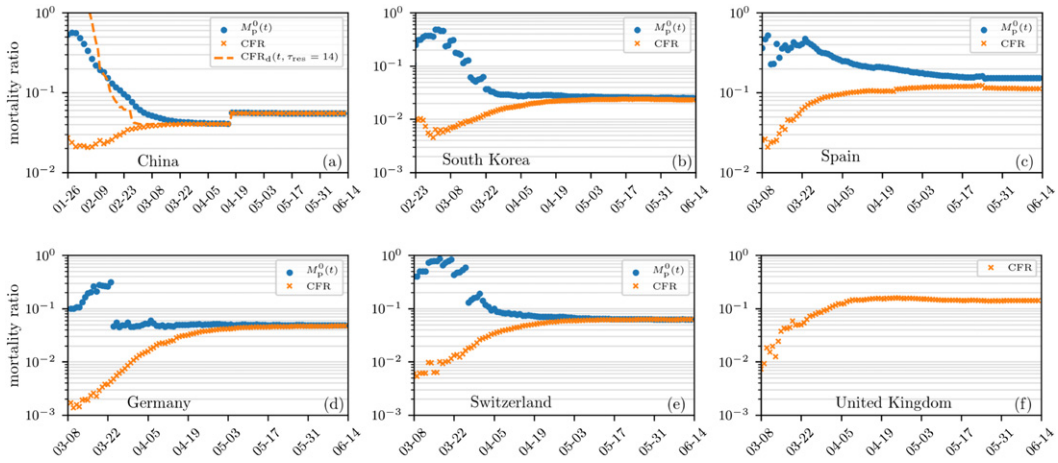

Figure 4. Mortality estimates in different countries. Estimates of mortality ratios (see equations (8) and (14)) of SARS-CoV-2 infections in different countries. The data are derived from reference [21]. The case fatality rate, CFR, corresponds to the number of deaths to date divided by the total number of cases to date. The 'delayed' mortality-ratio estimate $\mathrm{CFR_d}$ corresponds to the number of deaths to date divided by total number of cases at time $t - \tau_{\mathrm{res}}$ is also shown for China. The population-based mortality ratios $M_{\mathrm{p}}(t)$ are also shown, except for the UK which has reported an inexplicable $M_{\mathrm{p}}^0(t) \sim 1$.

The overall time-evolution of some of the mortalities in figure 3 qualitatively resembles the behavior of the mortality estimates in figure 1. As shown in figure 1, the CFR is increasing over time whereas M_{p}^0 provides a more stable mortality estimate for the SARS-CoV outbreak in Hong Kong (2003) and seems to follow a similar behavior in the current SARS-CoV-2 outbreak in Italy. In figure 4, we show additional examples of mortality-ratio estimates for China, South Korea, Spain, Germany, Switzerland, and the United Kingdom. After an initial transient, the CFR, in most cases, increases to a new asymptote after the epidemic passes. As in figure 1, we observe, consistent with their definitions, that the population-based mortality ratio $M_{\mathrm{p}}^0(t)$ is larger than the corresponding CFR in all cases. $M_{\mathrm{p}}^0(t)$ also appears to be a temporally more stable metric. Differences in the evolution of mortality ratios in different regions could result from changing practices in data collection or from explicitly time-inhomogeneous parameters $\mu(\tau, t)$, $\gamma(\tau, t)$, and/or $\beta(\tau, t)$.

In addition to the mathematical differences between $M_1(t)$, and $M_{\mathrm{p}}^0(t)$, and CFR, estimating $M_{\mathrm{p}}^0(t)$ and CFR(t) from aggregate populations implicitly incorporate a number of confounding factors that contribute to their variability. On April 25, 2020, the value of $M_{\mathrm{p}}^0(t)$ in Sweden, Belgium, France, the US, and Italy are $2194/(2194 + 1005) \approx 69\%$, $7094/(7094 + 10\,785) \approx 40\%$, $22\,856/(22\,856 + 44\,903) \approx 34\%$, $54\,941/(54\,941 + 118\,633) \approx 32\%$, and $26\,384/(26\,384 + 63\,120) \approx 29\%$, respectively. These $M_{\mathrm{p}}^0(t)$ have slowly decreased (see figure 4) as patients resolve even if current conditions (e.g., treatment methods, age group proportion of infected individuals, etc) have not changed. By comparison, on April 25, 2020, $M_{\mathrm{p}}^0 \approx 5\%$ and 6% in Germany and China, respectively. These differences result from delays and inaccuracies in reporting, varied guidelines for assigning cause of death, differing medical treatment strategies, and demographic heterogeneity among different countries.

3.2. Heterogeneous populations

Differences in demographics can easily be a source of variability in mortality rates measured across different regions. Older patients and those with underlying medical conditions typically have a higher death rate $\mu(\tau, t)$ and/or lower recovery rate γ. Since we focus on mortality, the different subpopulations within the confirmed population matter only through their differences in μ and/or γ. For the $\bar{M}_1(t)$ and $M_{\mathrm{p}}^1(t)$ metrics, no new infections are used in their determination. Thus, these metrics are associated with the mean death and recovery rates of the original group of infected individuals, i.e., the ratios $\bar{M}_1(t|\mu, \gamma)$ and $M_{\mathrm{p}}^1(t|\mu, \gamma)$ refer to the mortality ratios of each subpopulation or individual described by μ and γ. The effective $M_{\mathrm{p}}^1(t)$ over the entire confirmed population can be trivially constructed by population-averaging $D^1(t|\mu, \gamma)$ and $R^1(t|\mu, \gamma)$ over μ and γ before constructing $M_{\mathrm{p}}^1(t)$.

For the other confirmed mortalities $M_{\mathrm{p}}^0(t)$ and CFR(t), new infections are taken into account and subpopulations with different death and recovery rates can infect each other. Suppose there are two subpopulations 'a' and 'b' (e.g., young and old) with associated death and recovery rates $\mu_{\mathrm{a,b}}$ and $\gamma_{\mathrm{a,b}}$, respectively. The equations for each subpopulation are

$$\begin{aligned} \frac{\partial I_{\mathrm{a}}(\tau, t)}{\partial t} + \frac{\partial I_{\mathrm{a}}(\tau, t)}{\partial \tau} &= -(\mu_{\mathrm{a}}(\tau, t) + \gamma_{\mathrm{a}}(\tau, t)) I_{\mathrm{a}}(\tau, t), \\ \frac{\partial I_{\mathrm{b}}(\tau, t)}{\partial t} + \frac{\partial I_{\mathrm{b}}(\tau, t)}{\partial \tau} &= -(\mu_{\mathrm{b}}(\tau, t) + \gamma_{\mathrm{b}}(\tau, t)) I_{\mathrm{b}}(\tau, t), \end{aligned} \tag{19}$$

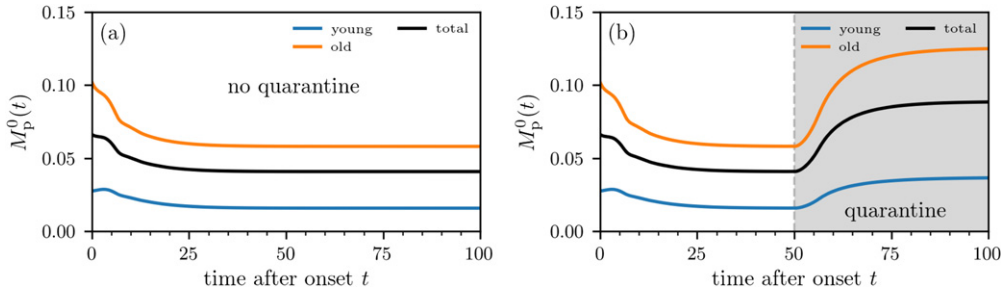

Figure 5. Population-level mortality estimate for two age groups. The mortality ratio $M_p^0(t)$ without containment measures (a) and under quarantining (b). The curves are based on numerical solutions of equations (19) and (20) assuming constant $S(t) \approx S_0$ and using the initial condition $I_a(\tau, 0) = I_b(\tau, 0) = \rho(\tau; 8, 1.25)/2$ (see equation (7)), where the subscripts 'a' and 'b' denote the young and old age group, respectively. The death and recovery rates for the younger age group are defined in equations (4) and (5). For the older age group, we set $\mu_b = 4\mu_a$ and $\gamma_b = \gamma_a$. We use an infection rate (equation (16)) defined by $\beta_{aa}S_0 = 4.64$/day, which we estimated from the basic reproduction number of SARS-CoV-2 [36]. The remaining infection rates are defined via $\beta_{aa} = \sqrt{2}\beta_{ba} = \sqrt{2}\beta_{ab} = 2\beta_{bb}$. To model quarantine effects, we set $\beta_0 S_0 = 0$ for $t > 50$ days in (b).

indicating that each subpopulation follows their own dynamics for $\tau > 0$. However, the subpopulations interact with each other through the coupled boundary conditions

$$\begin{aligned} I_a(0,t) &= S(t)\int_0^\infty d\tau' \left[\beta_{aa}(\tau',t)I_a(\tau',t)\right. \\ &\quad \left. + \beta_{ab}(\tau',t)I_b(\tau',t)\right] \\ I_b(0,t) &= S(t)\int_0^\infty d\tau' \left[\beta_{ab}(\tau',t)I_a(\tau',t)\right. \\ &\quad \left. + \beta_{bb}(\tau',t)I_b(\tau',t)\right] \end{aligned} \tag{20}$$

that describe cross-infections between the 'a' and 'b' subpopulations. Thus, the infection levels in each subpopulation also depend on the transmission rates β_{aa}, β_{ab}, and β_{bb}. To compute the overall confirmed mortality $M_p^0(t)$ or CFR(t) of the entire population, we must solve equations (19) and (20) for I_a and I_b, and hence $D_a(t)$, $D_b(t)$, and $D(t) = D_a(t) + D_b(t)$.

In figure 5, we show the evolution of $M_p^0(t)$ for two age groups representing young and old individuals with different mortality and infection rates. The behavior of $M_p^0(t)$ for the entire population is qualitatively similar to the behavior of the corresponding mortality ratios of both age groups (see figure 3). Whether the overall mortality is closer to that of the young or old population depends on the relative populations of young and old infecteds, their death and recovery rates, and their cross transmission rates β_{ab}. For age-stratified case data, the subpopulation model outlined above, or other approaches such as scaling approximations [46] may be useful for capturing age-dependent variations in $M_p^0(t)$.

3.3. Undertesting and unconfirmed cases

Another important confounding factor is the large number of untested and often asymptomatic infected individuals. The mortality rate often quoted in the literature ranges from < 1–3%, which is much smaller than the resolved mortality ratios we have used for illustration. Our estimates of $M_p^{0,1}(t)$ and CFR$_d(t, \tau_{res})$ using $I(\tau, t)$ actually describe the mortality of the population *conditioned* on being tested positive. Since we used equation (10) to compute infected populations, we implicitly assumed that all infected individuals have been tested/confirmed. However, the total infected population is comprised of tested and untested individuals, which may or may not carry different death and recovery rates. Typically, only a small fraction f of the total number of infected individuals might be tested and confirmed positive.

Our confirmed mortalities (derived from only the positively tested population) can be extended to the entire population, tested or untested. The 'true' $\mathcal{M}_p^0$ and the fatality ratio *conditioned on having been infected* (the IFR) would typically be much smaller than the M_p^0 and CFR calculated using only confirmed cases. How the testing fraction $f < 1$ might qualitatively affect the 'true' underlying mortality measures (the mortality conditioned on simply being infected) is illustrated in figure 6.

Estimates for SARS-CoV-2 show that f is small (e.g., $f \approx 14\%$ in China before January 23, 2020) [47]. At early times (figure 6(a)) most patients, tested or untested, have not yet resolved. A reported/tested fraction $f < 1$ *would not* directly affect or alter the CFRs or mortality ratios if the unreported/untested population dies and recovers in the same proportion as those tested, as depicted in figure 6(b). That is, undertesting would still provide a good estimate of the true mortality if the entire population were homogeneous in death and recovery rates. However, if the untested (presumably mildly or asymptomatic infected) are less likely to die than the tested infected individuals, undertesting would give rise to $M_p^0(t)$ and CFR(t) that overestimate the true mortality $\mathcal{M}_p^0(t)$ and the IFR. If untested infected individuals do not die at all, as depicted in figure 6(c), the true long-time mortality $\mathcal{M}_p^{0,1}(\infty) \approx f M_p^{0,1}(\infty)$. In the unlikely scenario in which untested individuals do not receive medical care and hence die at a faster rate (figure 6(d)), $M_p^{0,1}(\infty)$ and CFR based

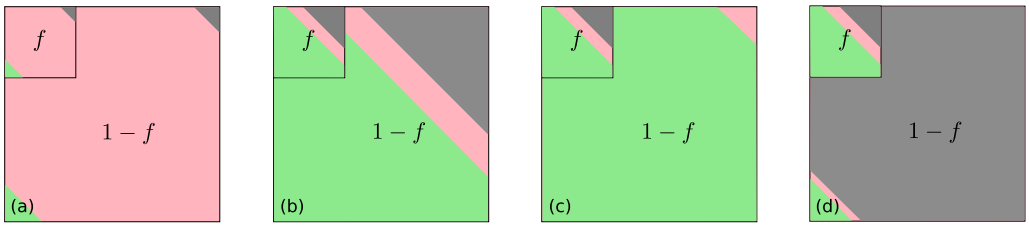

Figure 6. Fractional testing. An example of fractional testing in which a fixed fraction f of the real total infected population is assumed to be tested. The remaining fraction of infected individuals, $1-f$, is untested. Equivalently, if the total tested fraction has unit population, then the fraction of the population that remains untested is $1/f-1$. (a) At short times after an outbreak, most of the infected patients, tested and untested, have not yet resolved (red). Only a small number have died (gray) or have recovered (green). (b) At later times, if the untested population dies at the same rate as the tested population, $M_{\rm p}(t)$ and CFR remain accurate estimates for the entire infected population. (c) If the untested population is, say, asymptomatic and rarely dies, the true mortality $\mathcal{M}_{\rm p}^{0,1}(\infty)\approx fM_{\rm p}^{0,1}(\infty)$ can be significantly overestimated by the tested mortality $M_{\rm p}^{0,1}(t)$. (d) Finally, in a scenario in which untested infected individuals die at a higher rate than tested ones, $M_{\rm p}^{0,1}(t)$ and CFR based on the tested fraction *underestimate* the true mortality $\mathcal{M}_{\rm p}^{0,1}$.

Table 2. Definitions of the main metrics. The superscript '0' and '1' denote quantities that are based on the total population (including new infections) and a cohort (excluding new infections), respectively. Quantities with subscript 'c' and 'u' denote confirmed and untested pools (for example, $N_{\rm u}^0(t)$ is the total number of untested individuals at time t) that must be inferred using other measurements such as random testing. For the resolved mortalities, 'w/inf' indicates quantities that include new infections, while 'w/o inf' indicates quantities that exclude new infections. We have suppressed the time dependences for notational simplicity.

Subpopulation	Metric			
	Fatality ratios	Resolved mortality w/inf	Resolved mortality w/o inf	Individual risk
Confirmed (tested)	$\mathrm{CFR}=\frac{D_{\rm c}^0}{N_{\rm c}^0}$	$M_{\rm p}^0=\frac{D_{\rm c}^0}{D_{\rm c}^0+R_{\rm c}^0}$	$M_{\rm p}^1=\frac{D_{\rm c}^1}{D_{\rm c}^1+R_{\rm c}^1}$	$\bar{M}_1=\frac{\bar{P}_{\rm d}}{\bar{P}_{\rm d}+P_{\rm r}}$
Total (tested + untested)	$\mathrm{IFR}=\frac{D_{\rm c}^0+D_{\rm u}^0}{N_{\rm c}^0+N_{\rm u}^0}$	$\mathcal{M}_{\rm p}^0=\frac{D_{\rm c}^0+D_{\rm u}^0}{D_{\rm c}^0+D_{\rm u}^0+R_{\rm c}^0+R_{\rm u}^0}$	$\mathcal{M}_{\rm p}^1=\frac{D_{\rm c}^1+D_{\rm u}^1}{D_{\rm c}^1+D_{\rm u}^1+R_{\rm c}^1+R_{\rm u}^1}$	Not defined

on the tested fraction would underestimate the true long-time mortality $\mathcal{M}_{\rm p}^{0,1}(\infty)$ and IFR, respectively.

To quantitatively estimate the underlying mortality of the population conditioned simply on being infected, we have to quantify the number of confirmed and untested infected individuals, $I_{\rm c}(\tau,t)$ and $I_{\rm u}(\tau,t)$, which can be further divided into subpopulations with intrinsically different transmission, death, and recovery rates. The act of confirmation itself may change behavior and/or treatment, further changing transmission, death, and recovery parameters.

By constructing the accumulated deaths and recoveries associated with $I_{\rm c}(\tau,t)$ and $I_{\rm u}(\tau,t)$, $D_{\rm c,u}^{0,1}(t)$ and $R_{\rm c,u}^{0,1}(t)$, respectively, we can define true, whole population mortality ratios as listed in table 2. For example,

$$D_{\rm c,u}^0(t)=\int_0^t{\rm d}t'\int_0^\infty{\rm d}\tau\,\mu_{\rm c,u}(\tau,t')I_{\rm c,u}(\tau,t'),$$
$$R_{\rm c,u}^0(t)=\int_0^t{\rm d}t'\int_0^\infty{\rm d}\tau\,\gamma_{\rm c,u}(\tau,t')I_{\rm c,u}(\tau,t'),\qquad(21)$$

where $\mu_{\rm c,u}$ and $\gamma_{\rm c,u}$ are the death and recovery rates associated with infected individuals who are confirmed and untested, respectively. Analogous expressions arise for $D_{\rm c,u}^1(t)$ and $R_{\rm c,u}^1(t)$. If the confirmed and untested populations are further subdivided, the $\mu_{\rm c,u}I_{\rm c,u}$ and $\gamma_{\rm c,u}I_{\rm c,u}$ integrands would be replaced by a Hadamard (i.e., element-wise) product of two vectors representing subpopulations and their corresponding rates. The populations $I_{\rm c,u}$ themselves can be found from a specific disease transmission model that also includes a testing process that converts $I_{\rm u}$ to $I_{\rm c}$.

4. Summary and conclusions

The CFR has been predominantly used but appears to evolve in qualitatively similar ways as epidemics evolve. Although $\mathrm{CFR}_{\rm d}(t,\tau_{\rm res})$ is based on a delay reflecting the timescales for recovery, in general, there is no clear mechanistic interpretation for using the CFR or IFR as mortality ratios.

Here, we stress that more mechanistically meaningful and interpretable metrics can be readily defined and just as easily estimated from population data as CFRs are. Our proposed mortality ratios for viral epidemics are defined in terms of (i) individual survival probabilities and (ii) population ratios using numbers of deaths and recovered individuals. Both of these measures are based on the within-host evolution of the disease, and in the case of $M_{\rm p}^{0,1}(t)$, the population-level transmission dynamics. On a single patient level, $\bar{M}_1(t)$ is the metric of interest. However, to estimate this, one needs accurate cohort data, for which few exist for coronavirus. Nonetheless, cumulative population-based mortalities can provide insight.

Among the metrics we describe, $M_{\rm p}^1(t)$ is structurally closest to the individual mortality $\bar{M}_1(t)$ in

that both are independent of disease transmission since new infections are not counted. Both of these mortality ratios converge after an incubation time τ_{inc} to a value smaller than or equal to $\mu_1/(\mu_1+\gamma)$ and are best interpreted as approximately the mortality probability *conditioned on being tested positive.* The most accurate estimates of $\bar{M}_1$ can be obtained if we keep track of the fate of cohorts who were confirmed within a small time window in the past. By following only these individuals, one can track how many of them die as a function of time. As more cases arise, one should stratify them according to their estimated times since infection to obtain better statistics for $M_1(\infty)$. With the further spread of SARS-CoV-2 in different countries, data on more individual cases of death and recovery can also be more easily stratified according to other central factors in COVID-19 mortality: age, sex, health condition. Population heterogeneity and uncertainty in intrinsic disease parameters such as the incubation period and the time τ_1 a patient had been infected before confirmation can affect the mortality measures.

Besides demographic heterogeneity and the highly variable estimates of COVID-19 mortality due to different clinical protocols for assigning cause of death, undertesting also confounds accurate estimation of the true underlying mortality. Infected individuals in the population at large who are untested comprise an unknown population I_{u} which contributes to deaths and recovery, and need to be factored into the 'true' mortalities $\mathcal{M}_{\text{p}}^{0,1}$ or the IFR.

These untested/unconfirmed populations can, in principle, be computed from a multicompartment mathematical model for disease transmission and testing. The relevant expressions for $\mathcal{M}_{\text{p}}^{0,1}$ are listed in table 2. Even though $M_{\text{p}}(t)$ typically overestimates the true mortality, tracking $\bar{M}_1(t)$ or $M_{\text{p}}^1(t)$ of an initially confirmed cohort can still provide a reasonable estimate of the mortality ratio, especially if untested infected individuals die at the same rate as confirmed individuals.

Acknowledgments

LB acknowledges financial support from the SNF Early Postdoc.Mobility fellowship on 'multispecies interacting stochastic systems in biology'. The authors also acknowledge financial support from the Army Research Office (W911NF-18-1-0345), the NIH (R01HL146552), and the National Science Foundation (DMS-1814364). TC also acknowledges the hospitality of the Beijing Computational Science Research Center.

Appendix A

A.1. Numerical scheme

To numerically solve equations (10) and (11), we use a uniform discretization $\tau_k = k\Delta\tau, k = 0, 1, \ldots, K$. A backward difference operator $[I(\tau_k,t) - I(\tau_{k-1},t)]/(\Delta\tau)$ is used to approximate $\partial_\tau I(\tau,t)$ and a predictor-corrector Euler scheme is used to advance time [48]. Setting the cut-offs $I(-\Delta\tau,t) \equiv 0$ and $I(K\Delta\tau,t) \equiv 0$, the resulting discretized equations for the full SIR model are

$$
\begin{aligned}
S(t+\Delta t) &= S(t) - \Delta t S(t) \sum_{k=0}^{K} \beta(\tau_k,t) I(\tau_k,t) \Delta\tau,\\
\tilde{I}(\tau_k,t) &= I(\tau_k,t) - \Delta t \frac{I(\tau_k,t) - I(\tau_{k-1},t)}{\Delta\tau}\\
&\quad - \Delta t(\gamma(\tau_k,t) + \mu(\tau_k,t)) I(\tau_k,t),\\
I(\tau_k,t+\Delta t) &= I(\tau_k,t) - \frac{\Delta t}{2}\left[\frac{I(\tau_k,t) - I(\tau_{k-1},t)}{\Delta\tau}\right.\\
&\quad + (\gamma(\tau_k,t) + \mu(\tau_k,t)) I(\tau_k,t)\\
&\quad + \frac{\tilde{I}(\tau_k,t) - \tilde{I}(\tau_{k-1},t)}{\Delta\tau} + (\gamma(\tau_k,t+\Delta t)\\
&\quad \left. + \mu(\tau_k,t+\Delta t))\,\tilde{I}(\tau_k,t)\right]\\
&\quad + \delta_{k,0} \frac{\Delta t}{\Delta\tau} S(t) \sum_{j=0}^{K} \beta(\tau_j,t) I(\tau_j,t) \Delta\tau,
\end{aligned}
\tag{A1}
$$

where $\tilde{I}$ is the initial predicted guess, and the last term proportional to $\delta_{k,0}$ encodes the boundary condition equation (11). Note that we use $\sum_{k=0}^{K} \beta(\tau_k,t) I(\tau_k,t) \Delta\tau$ to indicate the numerical evaluation of $\int_0^\infty \mathrm{d}\tau' \beta(\tau',t) I(\tau',t)$. Quadrature methods such as Simpson's rule and the trapezoidal rule can be used to approximate the integral more efficiently.

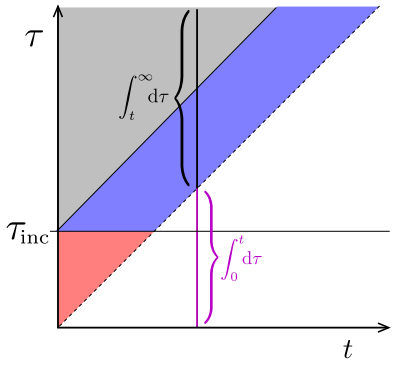

Figure A1. Phase plot for $\bar{P}(\tau > t, t)$ and $I(\tau > t, t)$. The regions delineate the different forms for the solution (equation (A6)). Here, we have included an incubation time τ_{inc} before which no death occurs. The solution for $\bar{P}(\tau,t)$ or $I(\tau,t)$ in the $\tau < t$ region must be self-consistently solved using the boundary condition equation (11). At any fixed time, the integral of $I(\tau,t)$ over $t < \tau \leqslant \infty$ captures only the initial population, excludes newly infected individuals, and is used to compute $D^1(t), R^1(t)$, and $M_{\text{p}}^1(t)$. To compute $D^0(t), R^0(t)$, and $M_{\text{p}}^0(t)$, we integrate across all infected individuals (also including the integral over $t > \tau \geqslant 0$ shown in magenta).

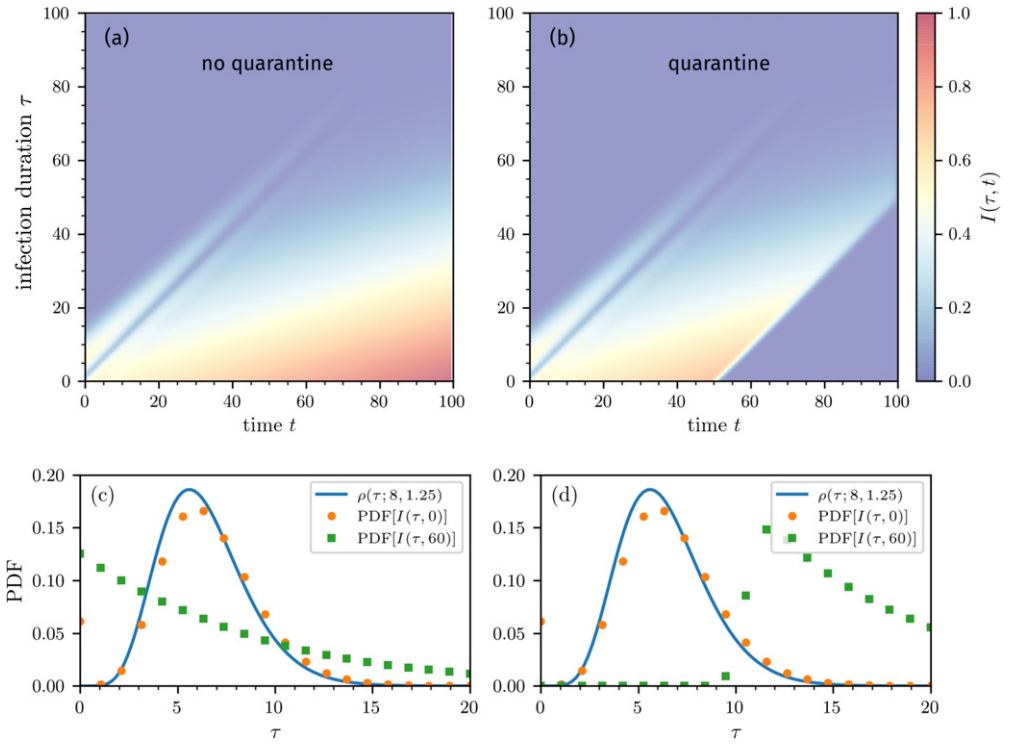

Figure A2. Density plots of $I(\tau, t)$ in the t–τ plane. Numerical solutions of the equation for $I(\tau, t)$ in equation (10) under the assumption of a fixed susceptible size and $\beta_0 S_0 = 4.64$/day. (a) The density without quarantine monotonically grows with time t in the region $\tau < t$ as an unlimited number of susceptibles continually produces infections. (b) With quarantining after $t_q = 50$ days, we set $\beta_0 S_0 = 0$ for $t > t_q$, which shuts off new infections. Both plots were generated using the same initial density $\rho(\tau_1)$ defined in equation (7). In both cases, the density $I(\tau > t)$ is identical to $P(\tau > t)$ if the same $\rho(\tau_1)$ is used and is independent of disease transmission, susceptible dynamics, etc. (c) and (d) probability-density functions (PDFs) of the number of infected individuals $I(\tau, t)$ for $t = 0, 60$ days (b) without and (c) with quarantine. The blue solid line corresponds to the initial distribution $\rho(\tau; n = 8, \lambda = 1.25)$ (see equation (7)).

The total number of dead, recovered, and infected individuals at time t are found by

$$D^0(m\Delta t) = \frac{1}{2}\sum_{j=0}^{m}\sum_{k=0}^{K} c(k\Delta\tau, j\Delta t) \times \left[I(k\Delta\tau, j\Delta t) + \tilde{I}(k\Delta\tau, j\Delta t)\right]\Delta\tau\Delta t,$$

$$R^0(t) = \frac{1}{2}\sum_{j=0}^{m}\sum_{k=0}^{K} \mu(k\Delta\tau, j\Delta t) \times \left[I(j\Delta\tau, j\Delta t) + \tilde{I}(k\Delta\tau, j\Delta t)\right]\Delta\tau\Delta t,$$

$$I(m\Delta t) = \sum_{k=0}^{K} I(k\Delta\tau, m\Delta t)\Delta\tau,$$

with analogous expressions for $D^1(m\Delta t)$ and $R^1(m\Delta t)$. To obtain a stable integration scheme, the time steps Δt and $\Delta\tau$ have to satisfy $\Delta t/(2\Delta\tau) < 1$. In all of our numerical computations, we thus set $\Delta t = 0.002, \Delta\tau = 0.02$, and $K = 10^4$. In the next section, we show additional plots of the magnitude of $I(\tau, t)$ in the t–τ plane.

A.2. Solutions for τ_1-averaged probabilities

Using the method of characteristics, we find the formal solution to equation (1):

$$P(\tau, t|\tau_1) = \delta(\tau - t - \tau_1)\mathrm{e}^{-\int_0^t (\mu(\tau - t + s, s|\tau_1) + \gamma(\tau - t + s, s|\tau_1))\mathrm{d}s}, \tag{A2}$$

which can be used to construct the death and cure probabilities

$$P_\mathrm{d}(t|\tau_1) = \int_0^t \mathrm{d}t'\, \mu(\tau_1 + t', t')\mathrm{e}^{-\int_0^{t'} (\mu(\tau_1 + s, s) + \gamma(\tau_1 + s, s))\mathrm{d}s}$$

$$P_\mathrm{r}(t|\tau_1) = \int_0^t \mathrm{d}t'\, \gamma(\tau_1 + t', t')\mathrm{e}^{-\int_0^{t'} (\mu(\tau_1 + s, s) + \gamma(\tau_1 + s, s))\mathrm{d}s}. \tag{A3}$$

If we now invoke the functional forms of μ and γ given in equation (4), we find explicitly

$$P_{\rm d}(\tau,t|\tau_1)=\begin{cases}\dfrac{\mu_1}{\mu_1+\gamma}\left(1-{\rm e}^{-(\mu_1+\gamma)t}\right) & \tau>t+\tau_{\rm inc}\\ 0 & \tau_{\rm inc}\geqslant\tau>\tau_1\\ \dfrac{\mu_1\ {\rm e}^{-\gamma(\tau_{\rm inc}-\tau_1)}}{\mu_1+\gamma}\left(1-{\rm e}^{-(\mu_1+\gamma)(\tau-\tau_{\rm inc})}\right) & \tau>\tau_{\rm inc}\geqslant\tau_1\end{cases}\tag{A4}$$

and

$$P_{\rm r}(\tau,t|\tau_1)=\begin{cases}\dfrac{\gamma}{\mu_1+\gamma}\left(1-{\rm e}^{-(\mu_1+\gamma)t}\right) & \tau>t+\tau_{\rm inc}\\ 1-{\rm e}^{-\gamma t} & \tau_{\rm inc}\geqslant\tau>\tau_1\\ 1-{\rm e}^{-\gamma(\tau_{\rm inc}-\tau_1)}+\dfrac{\gamma\ {\rm e}^{-\gamma(\tau_{\rm inc}-\tau_1)}}{\mu_1+\gamma}\left(1-{\rm e}^{-(\mu_1+\gamma)(\tau-\tau_{\rm inc})}\right) & \tau>\tau_{\rm inc}\geqslant\tau_1.\end{cases}\tag{A5}$$

Finally, we can also find the τ_1-averaged probabilities for $\tau\geqslant t$ by weighting over $\rho(\tau_1;n,\lambda)$. For example,

$$\bar{P}(\tau,t)=\begin{cases}\rho(\tau-t;n,\lambda){\rm e}^{-(\mu_1+\gamma)t} & \tau\geqslant t+\tau_{\rm inc}\\ \rho(\tau-t;n,\lambda){\rm e}^{-\gamma t} & \tau_{\rm inc}\geqslant\tau>t\\ \rho(\tau-t;n,\lambda){\rm e}^{-\gamma t}\ {\rm e}^{-\mu_1(\tau-\tau_{\rm inc})} & t+\tau_{\rm inc}\geqslant\tau>\tau_{\rm inc}\end{cases}.\tag{A6}$$

These solutions hold for the different regions shown in the phase plot of figure A1 and are equivalent to those for $I(\tau>t,t)$. Corresponding expressions for $\bar{P}_{\rm d}(t)$ and $\bar{P}_{\rm r}(t)$ can be found and used to construct $M_{\rm p}^1(t)$. Figure A2(a) shows the magnitude of $I(\tau,t)$ in the t–τ plane when we use equation (16), set $S(t)=S_0$ constant (so that the first equation in equation (A1) does not apply) and assign $\beta_0 S_0=4.64/\text{day}$. In this case, the epidemic continues to grow in time, but the mortality rates $M_{\rm p}^{0,1}(t)$ nonetheless converge as $t\to\infty$. In figure A2(b), we set $\beta_0 S_0=0$ for $t>t_{\rm q}$ to model strict quarantining after $t_{\rm q}=50$ days. We observe no new infection after the onset of strict quarantine measures. In both cases (quarantine and no quarantine), we use $\rho(\tau;n=8,\lambda=1.25)$ (see equation (7) in the main text) to describe the initial distribution of infection times τ. As time progresses, more of the distribution of τ moves toward smaller values until quarantine measures take effect (see figures A2(c) and (d)).

ORCID iDs

Lucas Böttcher https://orcid.org/0000-0003-1700-1897
Tom Chou https://orcid.org/0000-0003-0785-6349

References

[1] Verity R *et al* 2020 Estimates of the severity of coronavirus disease 2019: a model-based analysis *Lancet Infect. Dis.* **20** 669–77
[2] Kelly H and Cowling B J 2013 Case fatality *Epidemiology* **24** 622–3
[3] European Centre for Disease Prevention and Control 2020 Field epidemiology manual wiki https://wiki.ecdc.europa.eu/fem/Pages/Attack;20rates;20an;20case;20fatality.aspx (Accessed: 21 April 2020)
[4] Garske T, Legrand J, Donnelly C A, Ward H, Simon C, Fraser C, Ferguson N M and Ghani A C 2009 Assessing the severity of the novel influenza A/H1N1 pandemic *BMJ* **339** b2840
[5] Xu Z *et al* 2020 Pathological findings of COVID-19 associated with acute respiratory distress syndrome *Lancet Respir. Med.* **8** 420–2
[6] Wu Z and McGoogan J M 2020 Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention *JAMA* **323** 1239–42
[7] Centers for Disease Control and Prevention 2020 CDC Diagnostic Tests for COVID-19 www.cdc.gov/coronavirus/2019-ncov/lab/testing.html (Accessed: 13 May 2020)
[8] Centers for Disease Control and Prevention 2020 Research use only 2019-novel coronavirus (2019-nCoV) real-time RT-PCR primer and probe information https://cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html (Accessed: 13 May 2020)
[9] Onder G, Rezza G and Brusaferro S 2020 Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy *JAMA* **323** 1775–6
[10] Lipsitch M *et al* 2015 Potential biases in estimating absolute and relative case-fatality risks during outbreaks *PLoS Neglected Trop. Dis.* **9** 1–16
[11] Jung S-m, Akhmetzhanov A R, Hayashi K, Linton N M, Yang Y, Yuan B, Kobayashi T, Kinoshita R and Nishiura H 2020 Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases *J. Clin. Med.* **9** 523
[12] Famulare M 2020 2019-nCoV: preliminary estimates of the confirmed-case-fatality-ratio and infection-fatality-ratio, and initial pandemic risk assessment https://institutefordiseasemodeling.github.io/nCoV-public/analyses/first_adjusted_mortality estimates_and_risk_assessment/2019-nCoV-preliminary_age_and_time_adjusted_mortality_rates_and_pandemic_risk_assessment.html (Accessed: 31 March 2020)
[13] Oke J and Heneghan C 2020 Global covid-19 case fatality rates: Oxford COVID-19 evidence service www.cebm.net/covid-19/global-covid-19-case-fatality-rates/ (Accessed: 27 March 2020)
[14] Angelopoulos A N, Pathak R, Varma R and Jordan M I 2020 On identifying and mitigating bias in the estimation of the

COVID-19 case fatality rate *Harvard Data Sci. Rev.* (https://doi.org/10.1162/99608f92.f01ee285)
[15] Mizumoto K and Chowell G 2020 Estimating risk for death from 2019 novel coronavirus disease, China, January-February 2020 *Emerg. Infect. Dis.* **26** 1251–6
[16] Ruan S 2020 Likelihood of survival of coronavirus disease 2019 *Lancet Infect. Dis.* **20** 630–1
[17] Spychalski P, Błażyńska-Spychalska A and Kobiela J 2020 Estimating case fatality rates of COVID-19 *Lancet Infect. Dis.* **20** 774–5
[18] Yip P S F, Lam K F, Lau E H Y, Chau P-H, Tsang K W and Chao A 2005 A comparison study of realtime fatality rates: severe acute respiratory syndrome in Hong Kong, Singapore, Taiwan, Toronto and Beijing, China *J. R. Stat. Soc.* A **168** 233–43
[19] Yip P S F, Lau E H Y, Lam K F and Huggins R M 2005 A chain multinomial model for estimating the real-time fatality rate of a disease, with an application to severe acute respiratory syndrome *Am. J. Epidemiol.* **161** 700–6
[20] World Health Organization 2020 Cumulative number of reported probable cases of severe acute respiratory syndrome (SARS) https://who.int/csr/sars/country/en/ (Accessed: 30 March 2020)
[21] Dong E, Du H and Gardner L 2020 An interactive web-based dashboard to track COVID-19 in real time *Lancet Infect. Dis.* **20** 533–4
[22] Worldometers.info 2020 COVID-19 statistics https://worldometers.info/coronavirus/ (Accessed: 26 February 2020)
[23] Böttcher L, Woolley-Meza O, Araújo N A M, Herrmann H J and Helbing D 2015 Disease-induced resource constraints can trigger explosive epidemics *Sci. Rep.* **5** 1–11
[24] Böttcher L, Woolley-Meza O, Goles E, Helbing D and Herrmann H J 2016 Connectivity disruption sparks explosive epidemic spreading *Phys. Rev.* E **93** 042315
[25] Zhou X, Li S, Shen T, Li H and Kong L-q 2020 Full spectrum of COVID-19 severity still being depicted *Lancet* **395** 947–8
[26] Mahase E 2020 Coronavirus: COVID-19 has killed more people than SARS and MERS combined, despite lower case fatality rate *BMJ* **368** m641
[27] World Health Organization 2020 WHO Director-General's opening remarks at the media briefing on COVID-19—24 February 2020 https://who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---24-february-2020 (Accessed: 28 February 2020)
[28] World Health Organization 2020 WHO Director-General's opening remarks at the media briefing on COVID-19 - 3 March 2020 https://who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---3-march-2020\ignorespaces (Accessed: 5 March 2020)
[29] Porcheddu R, Serra C, Kelvin D, Kelvin N and Rubino S 2020 Similarity in case fatality rates (CFR) of COVID-19/SARS-COV-2 in Italy and China *J. Infect. Dev. Ctries.* **14** 125–8
[30] Peeri N C *et al* 2020 The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *Int. J. Epidemiol.* **49** 717–26
[31] Of course, at the population level, if there are many deaths, medical facilities may be stressed, which can indirectly lead to an increase in death rates.
[32] Ghani A C *et al* 2005 Methods for estimating the case fatality ratio for a novel, emerging infectious disease *Am. J. Epidemiol.* **162** 479–86
[33] Since the disease timescale is much shorter than the timescale over which aging appreciably affects death or recovery, the age-dependent transport terms $\partial P/\partial a$ can be neglected.
[34] Hauser A, Counotte M J, Margossian C C, Konstantinoudis G, Low N, Althaus C L and Riou J 2020 Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: a modeling study in Hubei, China, and six regions in Europe *PLoS Med* **17** e1003189
[35] covid19 SG 2020 Dashboard of the COVID-19 virus outbreak in Singapore https://co.vid19.sg/cases (Accessed: 4 April 2020)
[36] Lai C-C, Shih T-P, Ko W-C, Tang H-J and Po-Ren H 2020 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges *Int. J. Antimicrob. Agents* **55** 105924
[37] Keeling M J and Rohani P 2011 *Modeling Infectious Diseases in Humans and Animals* (Princeton, NJ: Princeton University Press)
[38] Böttcher L and Antulov-Fantulin N 2020 Unifying continuous, discrete, and hybrid susceptible-infected-recovered processes on networks *Phys. Rev. Res.* **2** 033121
[39] nCoV 2019 Data Working Group 2020 Epidemiological data from the nCoV-2019 outbreak: early descriptions from publicly available data http://virological.org/t/epidemiological-data-from-the-ncov-2019-outbreak-early-descriptions-from-publicly-available--data/337 (Accessed: 26 February 2020)
[40] Atkins K E, Wenzel N S, Ndeffo-Mbah M, Altice F L, Townsend J P and Galvani A P 2015 Under-reporting and case fatality estimates for emerging epidemics *BMJ* **350** h1115
[41] Webb G F 2008 Population models structured by age, size, and spatial position *Structured Population Models in Biology and Epidemiology* ed P Magal and S Ruan (Berlin: Springer) pp 1–49
[42] Wu J T, Leung K, Bushman M, Kishore N, Niehus R, de Salazar P M, Cowling B J, Lipsitch M and Leung G M 2020 Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China *Nat. Med.* **26** 506–10
[43] McKendrick A G 1926 Applications of mathematics to medical problems *Proc. Edinburgh Math. Soc.* **44** 98–130
[44] Chou T and Greenman C D 2016 A hierarchical kinetic theory of birth, death and fission in age-structured interacting populations *J. Stat. Phys.* **164** 49–76
[45] He X *et al* 2020 Temporal dynamics in viral shedding and transmissibility of COVID-19 *Nat. Med.* **26** 672–5
[46] Seoane B 2020 A scaling approach to estimate the covid-19 infection fatality ratio from incomplete data (arXiv:2006.02757 [q-bio.PE])
[47] Li R, Pei S, Chen B, Song Y, Zhang T, Yang W and Shaman J 2020 Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2) *Science* **368** 489–93
[48] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 2007 *Numerical Recipes: The Art of Scientific Computing* (Cambridge: Cambridge University Press)