Epidemiology and Infection

cambridge.org/hyg

Estimating the case fatality ratio for COVID-19 using a time-shifted distribution analysis

B. S. Thomas and N. A. Marks

Curtin University, School of Electrical Engineering, Computing and Mathematical Sciences, Perth, Australia

Original Paper

Cite this article: Thomas BS, Marks NA (2021). Estimating the case fatality ratio for COVID-19 using a time-shifted distribution analysis. *Epidemiology and Infection* **149**, e197, 1–12. https://doi.org/10.1017/S0950268821001436

Received: 21 October 2020
Revised: 19 June 2021
Accepted: 25 June 2021

Key words:
COVID-19; epidemics; infectious disease epidemiology; mathematical modelling; SARS

Author for correspondence:
N. A. Marks, E-mail: n.marks@curtin.edu.au

Abstract

Estimating the case fatality ratio (CFR) for COVID-19 is an important aspect of public health. However, calculating CFR accurately is problematic early in a novel disease outbreak, due to uncertainties regarding the time course of disease and difficulties in diagnosis and reporting of cases. In this work, we present a simple method for calculating the CFR using only public case and death data over time by exploiting the correspondence between the time distributions of cases and deaths. The time-shifted distribution (TSD) analysis generates two parameters of interest: the delay time between reporting of cases and deaths and the CFR. These parameters converge reliably over time once the exponential growth phase has finished. Analysis is performed for early COVID-19 outbreaks in many countries, and we discuss corrections to CFR values using excess-death and seroprevalence data to estimate the infection fatality ratio (IFR). While CFR values range from 0.2% to 20% in different countries, estimates for IFR are mostly around 0.5–0.8% for countries that experienced moderate outbreaks and 1–3% for severe outbreaks. The simplicity and transparency of TSD analysis enhance its usefulness in characterizing a new disease as well as the state of the health and reporting systems.

Introduction

The novel coronavirus SARS-CoV-2, and its attendant disease, COVID-19, first appeared in late 2019 in Wuhan, China. Since then, studies and estimates of the transmissibility and virulence of COVID-19 have abounded, with widely varying results [1–6]. Virulence is often measured using the case fatality ratio (also called case fatality rate or case fatality risk, CFR), which is the number of deaths due to a disease as a proportion of the number of people diagnosed with the disease. The CFR is dependent on the particular pathogen (and its mechanism of action) and the immune response of the host, which can depend on age, sex, genetic factors and pre-existing medical conditions. Environmental factors such as climate and health system may also affect the CFR. Collectively, these effects can be understood within the framework of the One Health concept [7], which integrates the full spectrum of interactions between the pathogen, the host and the biological and social environment. In this complex adaptive system [8], it is important to accurately quantify the CFR of a new disease to inform policy, communication and public health measures.

Calculating the CFR requires data on cases and deaths over time, either for individuals or populations. In general, the CFR is based on diagnosed cases of disease rather than the number of actual infections (which is difficult to measure); there may be many more infections than reported cases, depending on the expression of symptoms and the degree of testing. The simplest estimate of CFR is to divide the cumulative number of deaths by the cumulative number of cases at a given time, known as the crude (or naïve) CFR. However, the crude CFR tends to underestimate the CFR during an outbreak because at any given time, some of the existing known cases will prove fatal and need to be included in the death count. This bias is known as right-censoring and obscures the CFR of a new disease early in the course of the outbreak, particularly before the time course of the disease is characterised. Further, even once the distribution of times from the onset of disease to death is known, it can be difficult to use this information to accurately correct the crude CFR. An alternative method is to use data for closed cases only, once patients have recovered or died (e.g. [9, 10]), yet this information is also difficult to obtain during an outbreak and may be biased towards a particular demographic or skewed by delays in reporting of recoveries. Other biases in calculating CFR include under-ascertainment of mild or asymptomatic cases, time lags in testing and reporting, and the effects of intervention approach, reporting schemes, demographics and increased mortality due to pre-existing conditions (co-morbidities) [6]. The complexity of the CFR is well-summarised by Angelopoulos *et al.* [11] who write, 'Current estimates of the COVID-19 case fatality rate are biased for dozens of reasons, from under-testing of asymptomatic cases to government misreporting'.

There are many published calculations of CFRs for COVID-19 using various datasets from different countries and using a range of methods. In some places, initial outbreaks have now

CAMBRIDGE UNIVERSITY PRESS

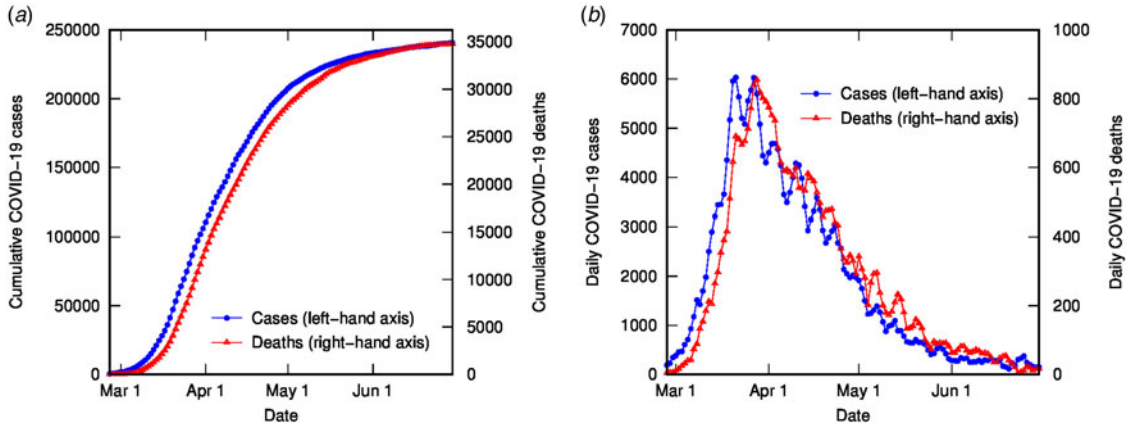

Fig. 1. COVID-19 cases and deaths in Italy to end of June (2020), using 3-day averaged data: (a) cumulative cases (left-hand axis) and deaths (right-hand axis); (b) daily cases (left-hand axis) and deaths (right-hand axis).

concluded and the final crude CFR accurately reflects the overall ratio of reported deaths to cases. In many other places, outbreaks are continuing. Questions remain regarding the quality of data (including the capacity of data collection, data governance and misclassification errors relate to diagnostic methods), methods of calculation and even the possibility of changes in the CFR over time [12]. These continuing uncertainties make it necessary to improve the estimates of the CFR by refining the methods used to calculate it. In essence, this means finding the best way to correct the crude CFR for biases due to time lags and other factors. Most previously published studies make use of a parametrised distribution of times from onset (or hospitalisation) to death, determined from individual case data from early in the outbreak (largely from China) [5, 13–15], which is then used in combination with statistical methods to estimate the CFR using population-level data on cases and deaths [5, 13, 14, 16]. Various assumptions are made in these analyses, including the form (and transferability) of the time course of cases, time lags in reporting or testing or hospitalisation, and estimates of the proportion of cases being detected. Early values of CFR obtained using these methods range from 1% to 18%, with the highest values obtained for China: 4–18% early in the outbreak [5, 13], 12% in Wuhan and as low as 1% outside Hubei province [14]. Values reported outside China include 1–5% for early cases in travellers [5, 13], and 1–4% in Korea [16]. CFRs have also been shown to vary greatly with the age of the patient [5]; Goldstein and Lee [17] found that COVID-19 mortality increases by about 11% per year of age. This obviously limits the transferability of parameters based on case studies, which will depend on demographic distributions. The specific data requirements and the range of approximations and assumptions required by statistical methods can make it difficult to interpret or rely on the results of such analyses since biases can be obscured.

Time-shifted distribution analysis for COVID-19 data

The time-shifted distribution (TSD) analysis method began with an observation that the shape of the evolving time distribution of COVID-19 cases in a given country often closely matches the shape of the corresponding distribution of COVID-19 deaths – simply shifted by a number of days and linearly scaled in magnitude. This is illustrated in Figure 1 for COVID-19 cases and deaths in Italy (data from [18], 3-day averaged data shown); the time-shifted relationship between case and death distributions can be seen in both cumulative and daily tallies. We can understand this shift from the perspective of the time delay between diagnosis and death or recovery. However, the closeness of the match reflects a much simpler apparent relationship than that suggested or assumed by conventional analyses, which relate deaths and cases using statistical parametric models that incorporate a broad distribution of expected times between diagnosis (or onset) and death, usually generated from case study data (e.g. [19]).

This observation suggests that there are two parameters of interest: the number of days separating the case and death distributions (called the delay time or t_d), and the scaling factor between the time-shifted case data and the death data, λ. For the optimal value of t_d, there is a simple linear relationship between cumulative number of deaths at time t, $D(t)$, and cumulative number of cases at time $t - t_d$, $C(t - t_d)$, with gradient λ:

$$D(t) = \lambda C(t - t_d)$$

To find the optimal value for t_d, we test integer values from zero to 25 days. For each value of t_d, we plot $D(t)$ as a function of $C(t - t_d)$ (for all t) and perform a linear regression using Matlab. The value of t_d is chosen on the basis of the lowest root-mean-squared error in the linear regression analysis and the value of λ is the gradient of the corresponding line.

Figure 2 contains the results of this analysis for Italy. Figure 2(a) shows the error from the linear regression of $D(t)$ *vs.* $C(t - t_d)$ as a function of delay time, with a clear minimum at 4 days. Figure 2 (b) shows $D(t)$ *vs.* $C(t - t_d)$ for different delay times: the optimal value of 4 days (with linear fit shown) as well as some other representative values, displaying the convergence of non-linear to linear relationship with optimised t_d. Figures 2(c) and (d) show the excellent correlation of time-shifted and scaled case data and death data (cumulative and daily, respectively), using a delay time of 4 days and a linear scaling factor of 0.144. What do these parameters represent? The delay time is presumably a measure of the delay between reporting of confirmed cases and reporting of COVID-19-related deaths. While 4 days seem very short compared to current estimates of the mean delay between the onset of COVID-19 symptoms and death (or even between hospitalisation and death), which is around 12–22 days with a large

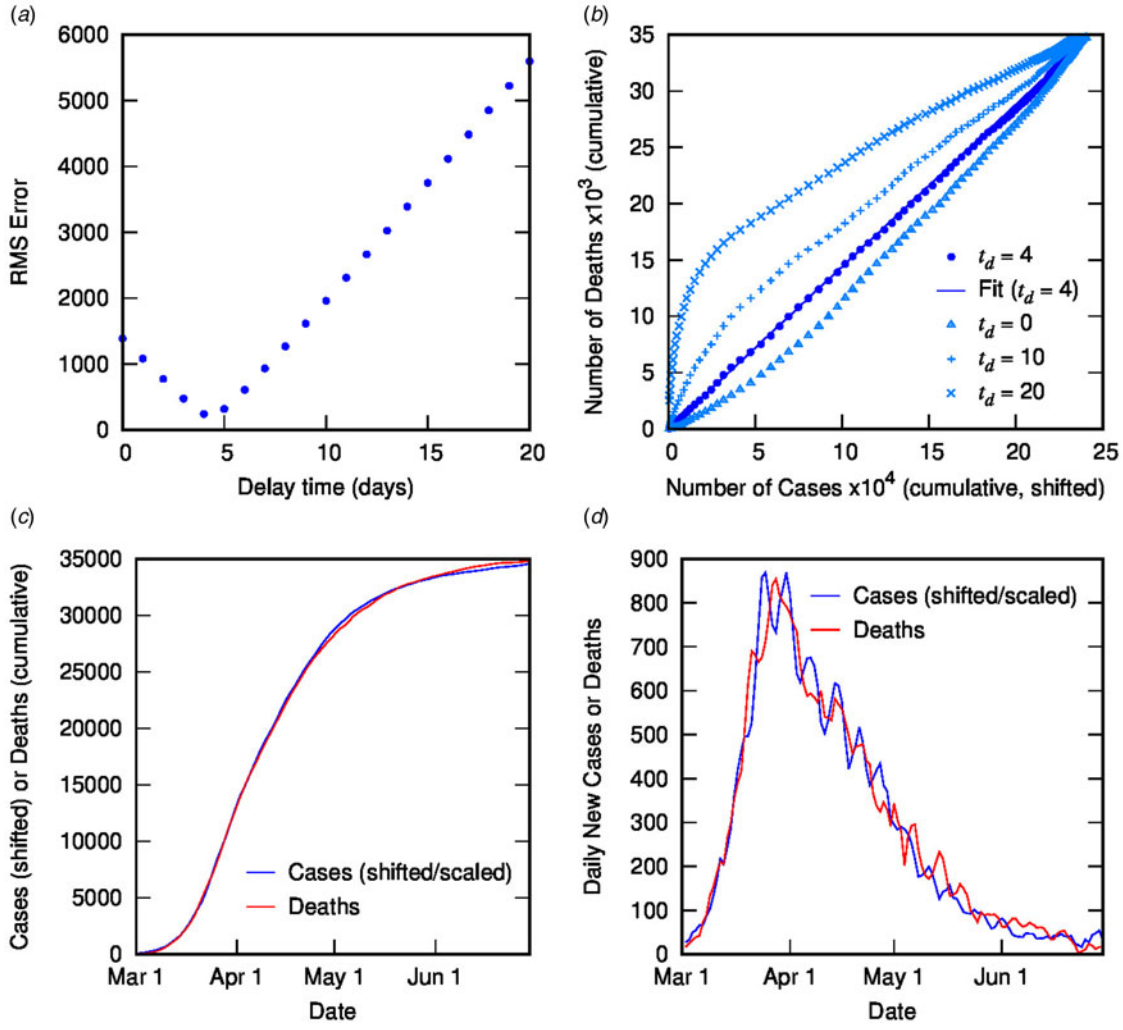

Fig. 2. Time-shifted distribution analysis for Italy: (a) root-mean-squared error in linear regression as a function of delay time, t_d; (b) cumulative deaths as a function of cumulative cases, time-shifted by various t_d values, including the optimal value of 4 days with linear regression shown; (c) overlay of cumulative deaths and time-shifted (and scaled) cases as a function of time, using optimal t_d; (d) overlay of daily deaths and time-shifted (and scaled) cases as a function of time using optimal t_d.

variance [2, 5, 9, 13, 20], it is possible to rationalise the shorter apparent delay on the basis of delays in testing, diagnosis and reporting of the disease, particularly in countries where the outbreak is severe. For example, in Italy from late February, testing was prioritised for 'patients with more severe clinical symptoms who were suspected of having COVID-19 and required hospitalisation' [21]; a subsequent delay in test results could account for the rather short delay between reported diagnosis and death. This shows the inherent danger in analysing such datasets using time-delay distributions from specific case data (which presumes a much longer delay time). Moreover, the delay time may provide some useful information about relative conditions in various countries.

Using a time delay of 4 days in Italy, the scaling factor of 0.144 represents the ratio of deaths to cases, or in other words, an estimate for the CFR, converging towards the crude CFR with time. The calculated CFR of 14.4% is almost identical to the crude CFR of 14.5% at the end of June, which is a good estimate for the 'true' CFR at the end of the outbreak. An interesting question is, at what point in the outbreak does the CFR calculated using the TSD analysis give a good approximation to the final value? This is important because early estimates of CFR are vital for informing public health decisions. Figure 3 shows the CFR calculated at various stages of the outbreak using data available to that point. Errors represent uncertainty in the linear regression as well as in t_d. Once the value of t_d has stabilised (from 26 March), the predicted value of CFR is very stable, and also remarkably accurate (14.4%), compared to the crude value of 10.3% at that time. Even a week earlier, the calculated CFR of 16.6% is a much better estimate than the crude estimate of 8.4%.

It appears that this simple analysis generates two parameters of significant interest: the apparent delay between reporting of related cases and deaths, and the CFR. The estimates of these parameters (which can be determined unequivocally once an outbreak is concluded) can be calculated during the course of an outbreak and give a better approximation than the crude CFR. It should be noted that such an analysis cannot be applied during purely exponential growth, because time-shifting (horizontally) and scaling (vertically) an exponential function are equivalent operations, as: $Ae^{b(t-t_0)} = [Ae^{-bt_0}]e^{bt} = Ce^{bt}$, which means that any value of t_d will give an equivalent relationship between $C(t - t_d)$ and $D(t)$ with gradient depending on t_d. Therefore, the

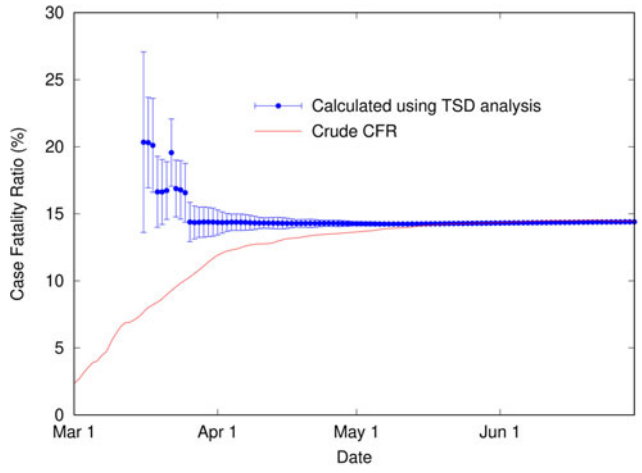

Fig. 3. Calculated case fatality ratio (using TSD analysis) for COVID-19 in Italy (2020) as a function of time during an outbreak, alongside the crude CFR.

TSD analysis is only valid once exponential growth ends and the daily case rate is approaching (or past) its peak. Alternatively, an estimate for t_d could be used, but this reduces the simplicity and transparency of the model. We note that others have calculated an 'adjusted' CFR early in the COVID-19 outbreak using a related method with an assumed value of the time delay between onset and death (because the true value was not known): Yuan *et al.* [22] chose sample values of 1, 3 and 5 days to give estimates from 3% to 13% for Italy in early March, while Wilson *et al.* [23] used 13 days to give 0.8–3.5% for China in early March.

Application of TSD analysis to SARS 2003 outbreak in Hong Kong

To test the TSD analysis method in determining CFR in the middle of an outbreak, and compare it to alternative methods, we analyse data from the SARS 2003 outbreak in Hong Kong (17 March to 11 July), obtained from the World Health Organization [24] and 3-day averaged. Figure 4 shows the cumulative and daily number of SARS cases and deaths in Hong Kong as a function of time.

It is apparent in Figure 4 that, as for COVID-19 data in Italy, the shapes of the distributions of cases and deaths are analogous. TSD analysis gives the following at the end of the outbreak: delay time is 22 days, and calculated CFR is 16.7%, close to the final crude CFR of 17.0%. The linear fit is reasonable given the noise in the data, as shown in Figure 5. If we perform TSD analysis serially over the course of the outbreak, reasonable estimates can be obtained from 17 April, giving values of 12–17% (with delay times of 17–22 days) converging on 16.7%, as shown in Figure 6. To compare, on 17 April the crude CFR is 5.3%, which is a significant underestimate of the true value. The delay time of 22 days is consistent with observations that the delay between onset and death for SARS is approximately 3 weeks [25]. We also applied TSD analysis to SARS data for other countries, giving a calculated CFR of 15% for Singapore and Canada (although data are noisy), and 13% for Taiwan.

We can compare these estimates of the CFR with the more complex mathematical models of Nishiura *et al.* [26] and Ghani *et al.* [19] for the same SARS outbreak. The simple TSD analysis gives better predictions than both the parametric mixture model and modified Kaplan–Meier method described by Ghani *et al.* [19], which use individual case data (dates of hospitalisation and death or discharge from hospital) to estimate CFR using statistical methods. Such methods can provide earlier estimates (from 1 April, giving around 7–8% CFR) but are less accurate at this early stage than a simple estimate of CFR from data on closed cases (recoveries and deaths) at the same dates [19], and are later outperformed by our simple TSD method once sufficient data to perform the analysis are available. Further, TSD analysis requires only publicly reported case and death data (over time), which are easier to obtain than individual case data including onset dates.

Similarly, the model of Nishiura *et al.* [26] can provide much earlier estimates of CFR than our analysis but the accuracy of these estimates is uncertain and depends on the assumptions made. Their analysis requires data on the dates of onset of confirmed cases and the distribution of times from onset to death; the latter, in particular, is poorly known at the start of an outbreak of a new disease. Nishiura *et al.* [26] analyse the Hong Kong SARS data by assuming a simple exponential distribution for the time between onset and death, with a mean of 36 days (from Donnelly *et al.* [25] for SARS cases in Hong Kong up to 28 April, although Donnelly used a γ distribution), and using statistical sampling to predict the CFR. The fact that this model provides a reasonable prediction of CFR at a specific time (around the end of March) is likely fortuitous, given that it involves scaling the crude CFR by a constant factor and will therefore overestimate the CFR at later times (as well as very early times). Further, this method requires the use of parametrised data (the time distribution from onset to death) that are not available at the time that the predictions are purported to be made. In fact, when Nishiura *et al.* [26] apply the method to early H1N1 (swine flu) data in 2009, they are forced to use a time distribution calculated from historical data for H1N1 (Spanish) influenza from 1918 to 1919, which is problematic; a sensitivity analysis shows that the predicted CFR is sensitive to the choice of distribution parameters, making this method somewhat difficult to apply in the circumstances for which it is proposed.

In comparison, the TSD analysis is both transparent and straightforward to implement, using only publicly available data and no assumptions, and can provide a reasonably early estimate (once exponential growth has sufficiently slowed) of CFR that converges to the 'true' value. If the value of the time delay is approximately known early in the outbreak, this could be used to constrain the fitting procedure, but as observed already, it is difficult either to know the time delay between onset and death or to apply it to the time delay between reporting of cases and deaths.

Time-shifted distribution analysis of international COVID-19 data

TSD analysis was performed on COVID-19 data from an extensive range of countries, using datasets from Johns Hopkins Center for Systems Science and Engineering [18], cross-checked and supplemented with data from Worldometers.com and 3-day averaged. For most countries (as for Italy), the analysis results in a robust linear fit and provides a stable estimate for CFR and delay time. These data are shown in Table 1, organised by region (Europe, Middle East, Asia, Oceania, North/Central America, South America, Africa) and then by CFR (decreasing). The corresponding plots of cases and deaths for each country

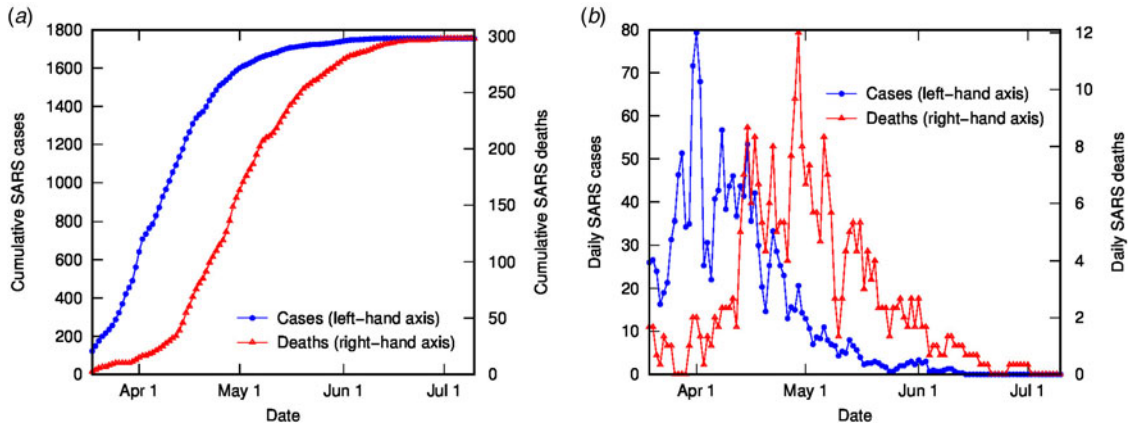

Fig. 4. SARS cases and deaths in Hong Kong (2003), using 3-day averaged data: (a) cumulative cases (left-hand axis) and deaths (right-hand axis); (b) daily cases (left-hand axis) and deaths (right-hand axis).

are included in Appendix 1, to demonstrate the astonishing correlation between case and death time profiles over a huge range of locations and outbreak characteristics. For some countries, either the data are insufficiently resolved (e.g. still in exponential growth or very low numbers) or too noisy or unreliable for rigorous analysis. For other countries, the linear correlation is not robust and varies over time; notable examples are Sweden, Brazil and the USA, which are discussed in the following section. For countries listed in Table 1, most analyses use data up until the end of May, which is generally representative of the initial outbreak; for some countries with later outbreaks, later end dates are used. In many countries, more recent outbreaks have had dramatically different CFR values to initial outbreaks (due largely to improved testing rates); these can be analysed independently by selecting the time frame studied, but values presented here are for the initial outbreak in each country.

The most notable result is the huge range in both delay times and calculated CFR estimates over different countries: from 0 to 24 days' delay and from <1% to 20% CFR. The highest ratios are calculated in Western Europe (up to 20%), followed by North America (up to 15%), South America (up to 10%), Africa (up to 7%), and lowest in the Middle East, Asia and Oceania (up to 5%). It is problematic to draw conclusions about relative COVID-19 virulence by comparing CFR values between countries, because of vast differences in testing and reporting regimes – in particular, the under-reporting of cases (including mild or asymptomatic cases) due to inadequate testing, but also differences in the classification or recognition of COVID-19-related deaths. However, it is instructive to calculate in this way, for any given country, the proportion of detected cases that are currently proving fatal, for the purposes of public health management and planning. For comparison, Mazumder *et al.* [10] calculated CFRs for a range of countries using recovery and death data from closed cases. They analysed 11 countries with high outcome rates and sufficient progression in the outbreak for analysis (at the end of April), but many of their calculated CFR values are much higher than our estimates – for example, estimated CFR above 30% for Italy, France and the USA at the end of April – probably due to delays in recovery reporting, whereas estimated CFR values for Germany, China and South Korea match ours. The TSD analysis provides more reliable estimates for a broader range of countries, due to the greater availability of death and case data over recovery data.

The differences in delay times are also startling, ranging from 0 to 24 days with no clear pattern. This delay between reported cases and deaths may be informative regarding the state of reporting or testing in a country but it is difficult to interpret. The mean delay between the onset of symptoms and death has been estimated at 12–22 days using case data [2, 5, 9, 13, 15, 20], but there are also delays between the onset of symptoms and testing, between testing and reporting of results, and in reporting of deaths. For example, in Sweden, a mean delay of 5 days between the onset of symptoms and the 'statistical date' of a reported case (including 1 day from test to statistic) was reported [9]. In some countries, tests are only administered to the sickest patients (many days after onset), and in others, test results can take up to a few weeks. We note that for Australia and New Zealand, where case numbers have been low and testing extensive and rapid, the calculated time delay is more than 10 days, whereas many of the harder-hit countries in western Europe and North America have much shorter calculated time delays.

Spain is an interesting case. Until 12 August, TSD analysis using Spanish data from the Worldometer website [27] gave a stable CFR of 10% with a delay time of 1 day. On that day, data were 'adjusted retrospectively by national authorities: case counts adjusted from 2 February to 11 August and death counts adjusted from 26 April to 11 August' according to the World Health Organisation (WHO) [28]. Using the revised data, the TSD analysis provided an even more robust fit; the CFR was almost unchanged at 11% but the delay time was increased to 14 days. This means that early data from Spain, which were erratic, reflected a much shorter delay between reported cases and deaths. In fact, the death data were largely unaffected by the August revision, but the dates of reported cases had shifted nearly 2 weeks earlier, presumably to better capture the onset time. This shows that a short delay time can reflect late reporting of cases, due either to testing late in the progress of the disease (well after onset) or delays in providing test results (or both). This may explain the short delay times for the UK, Italy, the Netherlands and the USA, as well as many other countries (e.g. 0 days' delay in Mexico). For countries that demonstrated reliable contact-tracing and testing regimes, such as Australia, New Zealand and Germany, delay times are close to 14 days, similar to the revised Spain data and consistent with the estimated time course of fatal disease.

An important conclusion from this analysis concerns the perils in calculating the CFR using established time distributions from

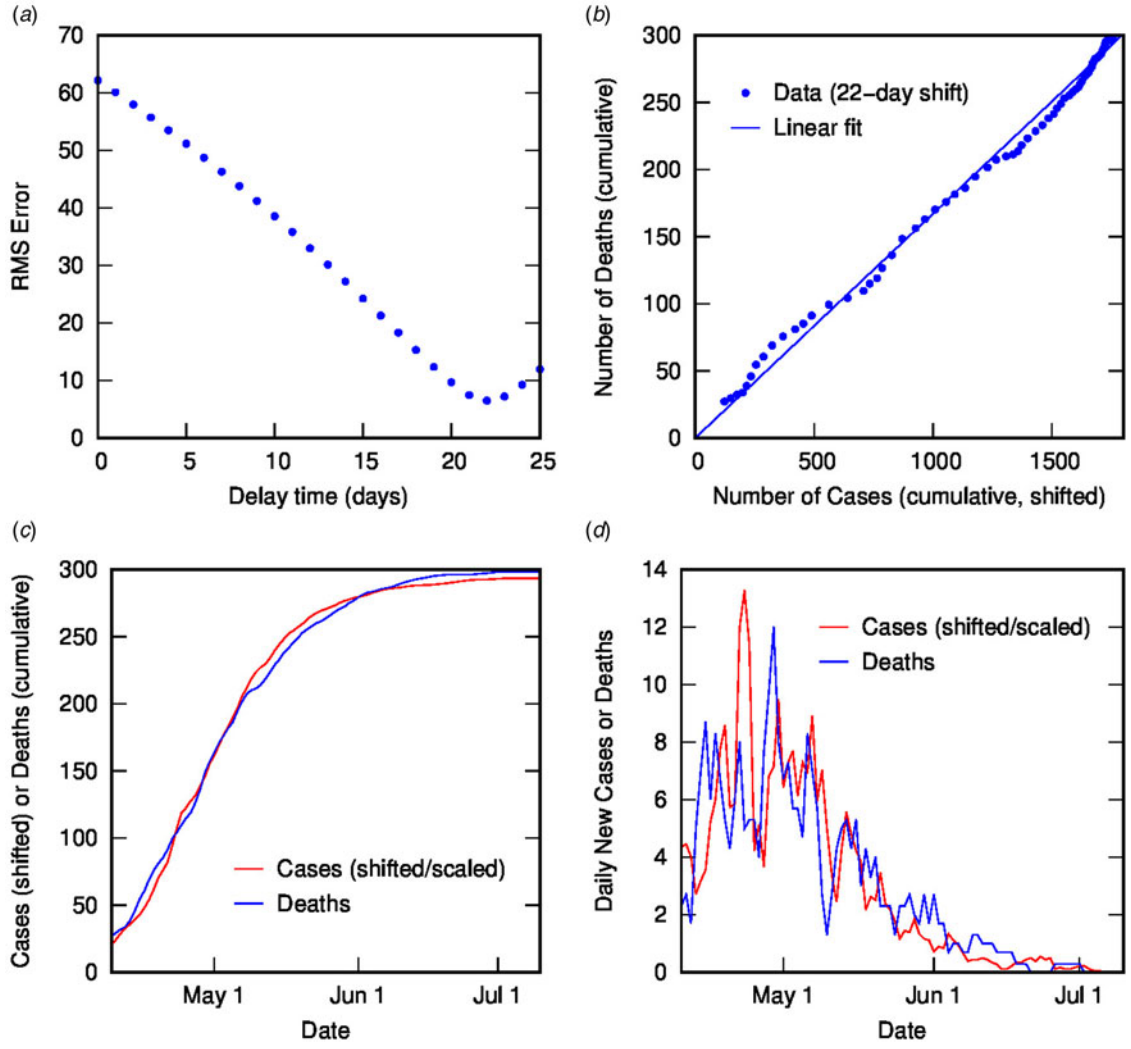

Fig. 5. Time-shifted distribution analysis of SARS (2003) data for Hong Kong: (a) root-mean-squared error in linear regression as a function of delay time, t_d; (b) linear regression for cumulative number of deaths as a function of cumulative number of cases (time-shifted by optimal t_d); (c) overlay of cumulative deaths and time-shifted (and scaled) cases as a function of time, using optimal t_d; (d) overlay of daily deaths and time-shifted (and scaled) cases as a function of time, using optimal t_d.

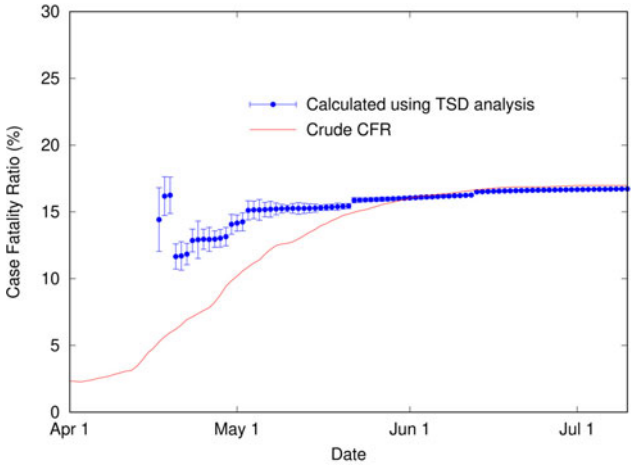

Fig. 6. Calculated case fatality ratio (using TSD analysis) for SARS in Hong Kong (2003) as a function of time during an outbreak, alongside the crude CFR.

onset to death obtained from case studies, as is common. The beauty of this simple method is its transparency – nothing is assumed and the data are enabled to speak for themselves, and can therefore give us information that we might not expect, rather than merely reflecting our assumptions.

Another benefit of TSD analysis is that it provides near-term predictive capacity for numbers of deaths, using the linear relationship between deaths and time-shifted cases. This predictive capacity is intrinsically linked to the delay time, and hence has the greatest utility when the delay time is significant. Figure 7 shows an example of this capacity for the second phase of the COVID-19 outbreak in France from August. Using parameters calculated from TSD analysis for August to mid-October, reported case data can be time-shifted and linearly scaled to predict daily deaths for France for the next 3 weeks. This is useful for public health planning and managing public expectations, as well as decision-making regarding the implementation of restrictions. Figure 7 also shows a sensitivity analysis for the same dataset, using a fixed delay time of 15 days (dashed line; CFR = 0.7%)

Table 1. Case fatality ratio values and delay times calculated using time-shifted distribution analysis for a range of countries (initial outbreak), ordered by region and by CFR

Country	CFR (%)	Delay time (days)	End date
Europe			
France	20	7	31 May
Belgium	17	6	31 May
UK	16	3	31 May
Italy	14	4	31 May
Hungary	14	8	31 May
Netherlands	13	4	31 May
Sweden	13	5	31 May
Spain	11	14	31 May
Romania	7	6	31 May
Ireland	7	7	31 May
Slovenia	7	15	31 May
Bulgaria	6	7	31 May
North Macedonia	6	8	31 May
Greece	6	8	31 May
Poland	6	8	16 May
Switzerland	6	11	31 May
Denmark	5	4	31 May
Finland	5	9	30 June
Germany	5	13	31 May
Croatia	5	18	31 May
Portugal	4	7	31 May
Czechia	4	11	31 May
Austria	4	13	31 May
Estonia	4	13	31 May
Moldova	3	0	31 July
Ukraine	3	4	31 May
Norway	3	13	31 May
Luxembourg	2.6	10	31 May
Latvia	2.6	23	31 May
Serbia	2.1	1	31 May
Armenia	2.0	7	31 July
Russia	1.8	14	31 July
Azerbaijan	1.5	6	31 July
Middle East			
Iraq	5	5	16 July
Egypt	5	8	31 July
Afghanistan	3	10-17	31 July
Turkey	2.8	3	31 May
Israel	1.6	10	31 May
Saudi Arabia	1.2	14	31 July
Kuwait	0.8	2	30 June
Oman	0.6	6	31 July
Qatar	0.17	21	31 July
Asia			
Japan	5	14	30 June
China	4	6	31 March
India	3	0	31 May
South Korea	2.4	18	30 June
Pakistan	2.1	2	31 July
Thailand	1.9	9	31 May
Malaysia	1.7	2	31 May
Taiwan	1.5	4-5	31 May
Bangladesh	1.3	1	31 July
Oceania			
New Zealand	1.5	17	30 June
Australia	1.4	12	31 May
North/Central America			
Mexico	11	0	31 August
Canada	9	10	30 June
USA	7	4	31 May
Guatemala	4	0	31 August
Cuba	4	4	31 May
Panama	3	2	31 May
South America			
Ecuador	10	12	31 May
Bolivia	5	7	31 August
Brazil	3	0	31 August
Colombia	3	0	31 August
Peru	3	5	30 June
Chile	3	18	31 August
Argentina	2.4	7	31 August
Venezuela	0.8	0	31 August
French Guiana	0.7	12	30 September
Africa			
Sudan	7	4	31 July
Tunisia	4	3	31 May
Senegal	2.3	14	30 September
Nigeria	2.3	0	31 July
South Africa	2.4	14	30 September
Ethiopia	1.6	0	30 September
Mayotte	1.3	1-2	31 May
Gabon	0.7	0	31 August
Guinea	0.6	3	31 August

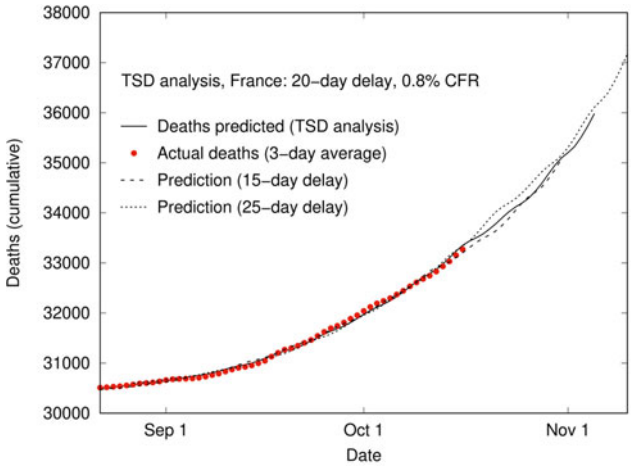

Fig. 7. Application of TSD analysis to predict deaths over time based on case data, delay time and CFR in France from August. Solid line shows the prediction using 3-day averaged case data up to 16 October, shifted (20 days) and linearly scaled using the CFR (0.8%). Dashed and dotted lines show a sensitivity analysis assuming fixed delay times of 15 and 25 days, respectively.

and 25 days (dotted line; CFR = 1.0%). The similarity between the three lines shows that predictions are not very sensitive to the delay time, even though the predicted CFR changes with the delay time.

Estimating the infection fatality ratio from the CFR

We are interested not only in the versatility and simplicity of this method, but also in what conclusions may be drawn from the parameters calculated – namely, the CFR and the delay time. Reported CFR values for COVID-19 vary widely, but the best current estimates of the true infection fatality ratio (IFR; taking into account all infections including undiagnosed and asymptomatic) are around 0.6–0.7% [4, 6] based on cruise ship and population serology data. The very high CFR values calculated for many European countries, in particular, are probably vastly inflated due to the inadequate testing and overwhelmed health systems in these countries, which result in the underestimation of case numbers. However, it is an oversimplification to assume that this is the only relevant factor that differs between countries, since we know that demographics and health systems (among other things) can also affect survival probability. Such an assumption has been used in various studies, in order to compare the effectiveness of different countries' reporting systems and to correct case numbers [4, 29]. However, by assuming that the IFR is identical everywhere at all times, valuable information is lost and conclusions may be misleading.

In this study, we estimate the IFR from the CFR for a subset of countries using seroprevalence data (to correct case numbers) and excess death data (to correct death numbers). Along similar lines, Ioannidis [30] previously estimated the IFR for various countries using seroprevalence data and cumulative reported deaths at a corresponding date, although this does not account for either excess deaths or the relationship between cases and deaths over time; in fact, using seroprevalence and death data alone reintroduces the issue of the unknown time delay between cases and deaths, which must be approximated. We note that studies from very early in the pandemic provided initial estimates for the true prevalence of COVID-19 in specific places; a spatiotemporal transmission model applied to Wuhan [31] gave a prevalence factor of seven in January (published mid-March), and a statistical analysis study of testing data in the USA [32] estimated a prevalence factor of nine in April (published in May). These early prevalence studies can be useful in roughly correcting the CFR to estimate the IFR before rigorous seroprevalence data are available. We also note that, while excess mortality may not exclusively represent COVID-19 deaths, it is a more comprehensive and reliable measure than reported deaths alone [33], especially for comparative purposes.

In Australia, case numbers have been generally low (especially before June) and testing rates high. It is unlikely that there have been appreciable unreported COVID-related deaths [34]. However, even with robust testing, many cases will be undiagnosed, especially asymptomatic cases, which could constitute half of all infections [35]. A recent seroprevalence study of elective surgery patients in four states [36] estimated that the number of true infections was around 5–10 times the number of reported cases, although the authors state that the study cohort may not reflect the general population (older individuals overrepresented). This prevalence ratio gives an approximate IFR for Australia of 0.1–0.3%. Note that before June, most of Australia's COVID-19 cases were returned travellers, which may affect the age distribution and baseline health of cases compared to the general population. New Zealand, Taiwan and Thailand are similarly circumstanced and have very similar CFR values, which are expected to reflect similar IFR values to Australia. Singapore, with its extremely low fatalities and extensive testing, did not return a robust result from TSD analysis; nonetheless, the crude CFR of 0.07% at the end of May is likely a lower bound for the IFR.

The USA is an interesting case study. The TSD analysis is problematic because the relationship between cases and deaths changes over time, causing a mismatch between case and death distributions and a downward drift in both CFR and delay time. This may be due to incomplete data, or changes in testing or reporting over time, which can affect both delay time and case numbers. Alternatively, the CFR may be truly changing over time, due to changes in treatment approach or in the demographics (or location) of COVID-19 cases [12]. In the USA, there is also heterogeneity between states. To demonstrate, we present the TSD analysis for the USA in Figure 8, and for the state of New Jersey (which has the highest mortality rate in the USA) in Figure 9. For the USA as a whole, there is clear variation over time in the relationship between time-shifted case and death data, demonstrated in both the poor linear fit and the mismatch in distribution profiles. If we scrutinise individual state data, some US states (including New Jersey, Illinois, Massachusetts, New Mexico, Ohio and Pennsylvania) manifest a very reliable TSD analysis, but others do not (e.g. California, North Carolina, Oklahoma and Texas). Data from New Jersey (Fig. 9) give a stable CFR around 8–9%, comparable to New York, Massachusetts and Pennsylvania, while Ohio gives 7% and Illinois and New Mexico give 5%.

One potential reason for the mismatch of case and death data in the USA as a whole (and many of its states) is the underreporting of cases due to the low level of testing, which varies over time. One measure of the adequacy of testing is the share of daily COVID-19 tests that return a positive result, known as the positive test rate (PTR). The WHO has suggested a PTR of around 3–12% (or less) as a benchmark of adequate testing [37]. In the USA, the PTR reached maximum levels in April, with values between 18% and 22% from 1 to 21 April [37],

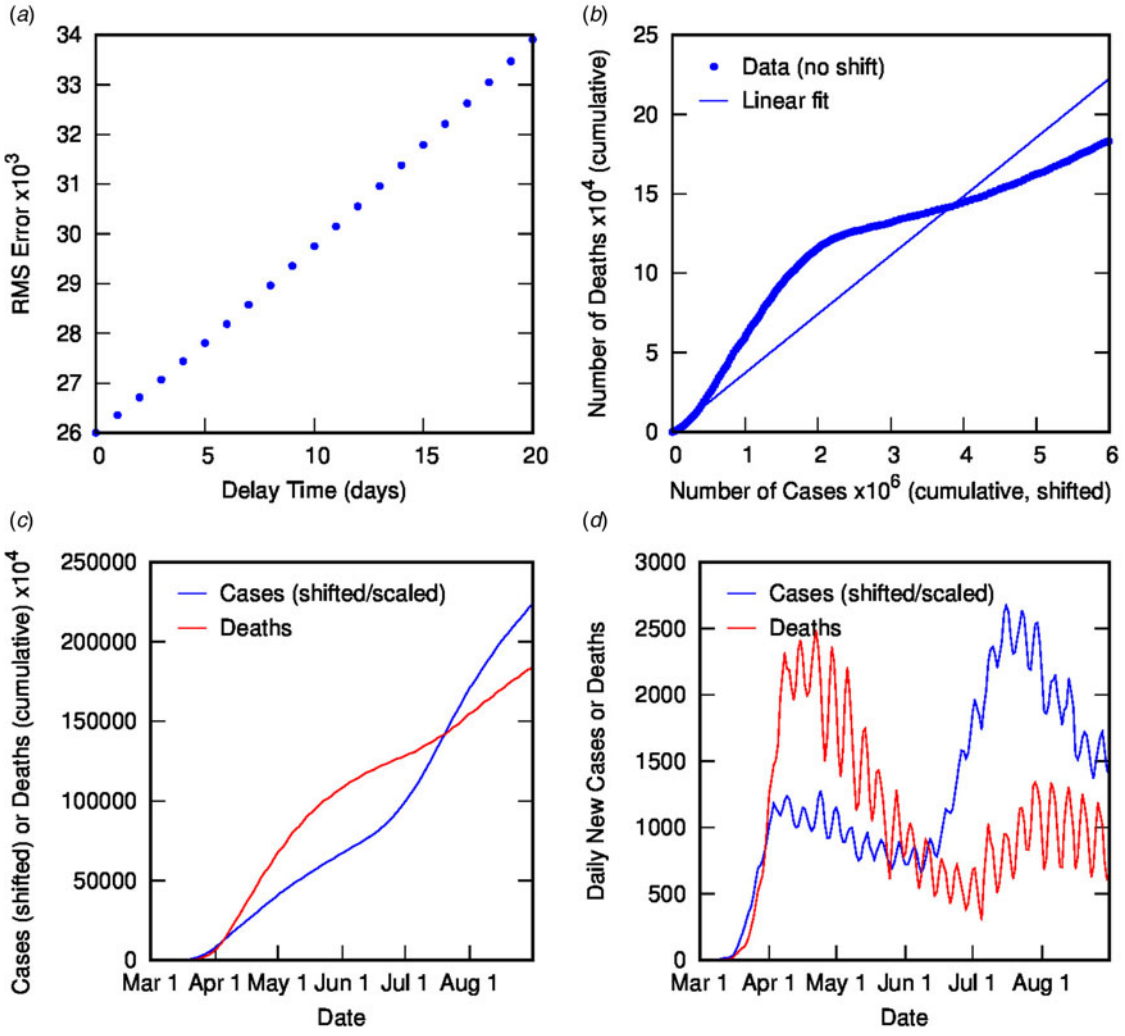

Fig. 8. Time-shifted distribution analysis for the USA at the end of August: (a) root-mean-squared error in linear regression as a function of delay time, t_d; (b) linear regression for cumulative number of deaths as a function of cumulative number of cases (time-shifted by optimal t_d); (c) overlay of cumulative deaths and time-shifted (and scaled) cases as a function of time, using optimal t_d; (d) overlay of daily deaths and time-shifted (and scaled) cases as a function of time, using optimal t_d. Note the mismatch between distributions of death and cases.

which is the region of greatest discrepancy between case and death profiles in the initial outbreak, as seen in Figure 8. We would expect that such a high PTR indicates that case numbers during this time are greatly underestimated, which may explain the poor fit from TSD analysis and the high CFR. Similar effects are seen in data from Sweden and Brazil, which also had low and variable testing rates and high PTR. Recent seroprevalence studies in many states of the USA from March to May [38] suggest that there were at least 11 times as many infections as reported cases before the end of May. Excess death data indicate that COVID-related deaths may be higher than reported by a factor of 1.4 for the same period [39]. Using these correction factors for the CFR, the estimated IFR for the USA is 1.0% or below. For comparison, a Worldometers calculation estimated an IFR of 1.4% in New York City in May [27], using a prevalence ratio of 10 from an early antibody study [40, 41].

In Europe, many of the most affected countries have very high CFRs, often combined with relatively short delay times. For some of these countries, seroprevalence studies provide the estimates of the degree of undercounting of cases during the initial outbreak [42–50], which can be utilised along with excess death data [39] to estimate the IFR. These IFR values are shown in Table 2 along with the correction factors used. Some of the seroprevalence data are preliminary, including studies of Germany, Sweden and Italy, and others are for specific regions of the country and may not be representative. Nonetheless, the estimated IFR values are reasonable: Switzerland and Germany are around 0.6%, above Australia and below Sweden and the USA at around 0.8%; Belgium, UK and Spain are between 1% and 2%; and Italy higher at around 3%. Ioannidis [30] also calculated the IFR for many of these countries using seroprevalence studies, but using only single-time seroprevalence and death data with an assumed delay time (generally a week after the midpoint of the seroprevalence survey); these are also shown in Table 2 and are broadly consistent with our values except where excess deaths are significant (e.g. Spain). Our value for Germany is somewhat higher but we expect that it is more reliable, using the scaling factor for cases [42] with our calculated CFR rather than the absolute number of deaths at a certain date in the German town of Gangelt [30], which is very low and reflects a date early in the German outbreak.

Although the calculated IFR values are only approximate and subject to revision, it is conceivable that higher IFR values may

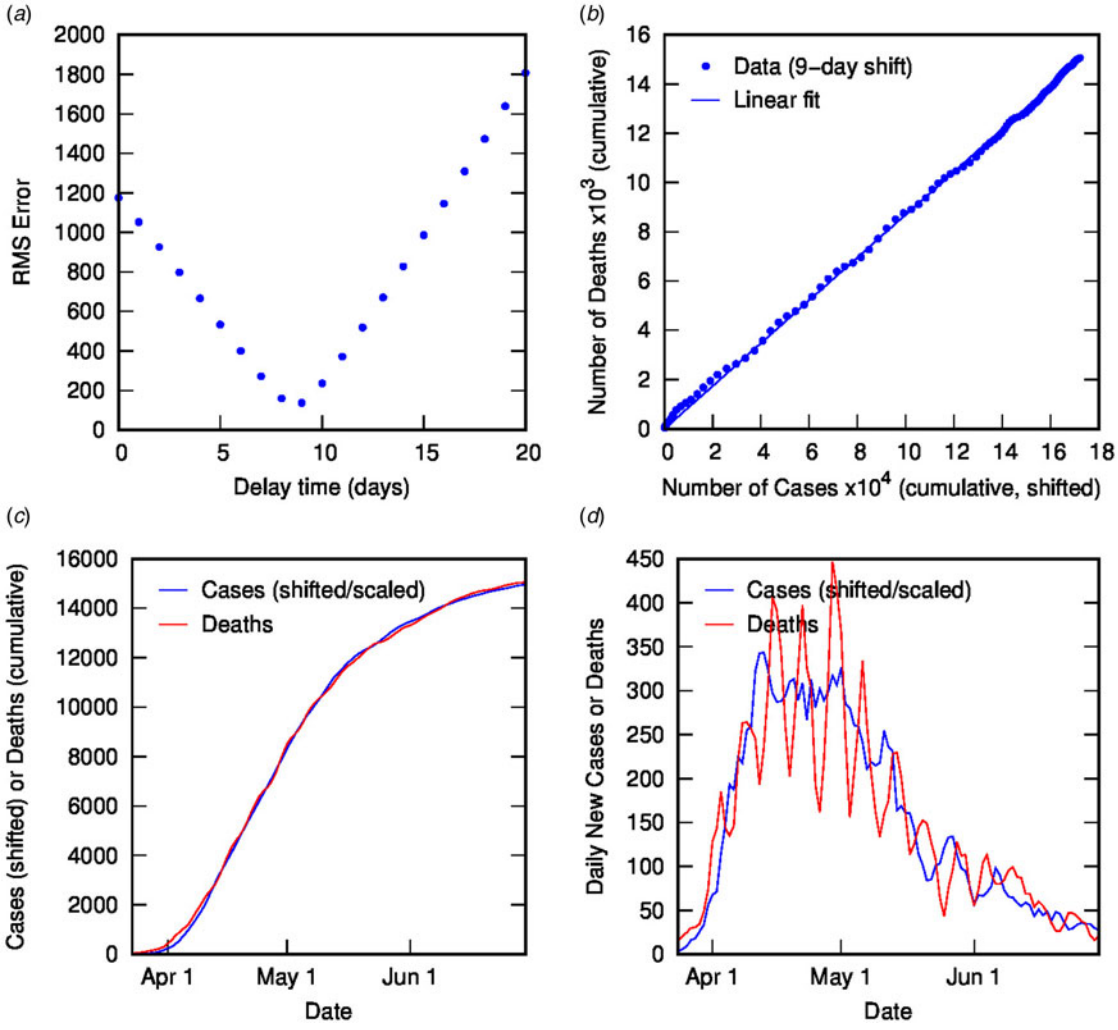

Fig. 9. Time-shifted distribution analysis for New Jersey: (a) root-mean-squared error in linear regression as a function of delay time, t_d; (b) linear regression for cumulative number of deaths as a function of cumulative number of cases (time-shifted by optimal t_d); (c) overlay of cumulative deaths and time-shifted (and scaled) cases as a function of time, using optimal t_d; (d) overlay of daily deaths and time-shifted (and scaled) cases as a function of time, using optimal t_d.

Table 2. Estimated IFR from CFR (calculated in this work), using scaling factors from seroprevalence and excess death data

Country	CFR [this work]	Scaling factors		IFR estimate [this work]	IFR from Ioannidis[a] [30]
		Deaths (excess) [39]	Cases (prevalence)		
Australia	1.4	1.0	5–10 [36]	0.1–0.3	
Switzerland	6	1.0	10–12 [43]	0.5–0.6	0.45
Germany	5	0.8	5–7 [42]	0.5–0.8	0.28
Sweden	13	1.2	17–21 [44]	0.7–0.9	0.71
USA[b]	7	1.4	9–13 [38]	0.7–1.0	0.65
Belgium	17	1.0	13–15 [46]	1.1–1.3	1.09
UK	15	1.3	14–15 [45]	1.3–1.4	
Spain	11	1.5	9–12 [47, 48]	1.3–1.8	1.15
Italy	14	1.5	6–7 [49, 50]	3.0–3.5	

[a]IFR from Ioannidis [30] uses seroprevalence and concomitant deaths at a single time point.
[b]USA value is a weighted mean of six states.

reflect higher fatality ratios in particular places at particular times, due to overwhelmed health systems in hard-hit areas or specific demographics or baseline health of affected populations. For example, it is reasonable to conclude that in Lombardy, Italy, the older population and overwhelmed health system caused a higher fatality ratio compared to other places. In fact, the difference in age distribution of cases between Italy and Australia up to the end of May (using data from [51] and [52]) can alone account for a factor of three in the IFR. Therefore, while differences in testing and reporting between different countries undoubtedly account for much of the variation in IFR between countries, we neither expect nor find that IFR is the same for all COVID-19 outbreaks. Country-specific factors that influence IFR and differ between countries include testing and reporting, age demographics [53], health-care systems and treatments [12], mask-wearing and other behaviours, climate and culture, transport infrastructure and community mobility [54], genetic factors or prevalence of particular antibodies that affect immune response [55].

There is some evidence that the IFR might be decreasing over time in some countries, especially those experiencing a 'second wave'. This is observed, for example, in the data for the USA in Figure 8, demonstrated in the increasing mismatch in case and death distributions later in the outbreak. We can use the TSD analysis to analyse the latter part of the outbreak (from July to September), giving a CFR of 1.5–1.7% and a delay time of 2–3 weeks. Similar analyses for individual states of the USA give stable CFR values from 1.1% to 2.3% with delay times between 4 and 24 days, with a mean of 1.6% CFR and 17 days' delay over states with robust fits. This later CFR is far lower than the value of 7% calculated early in the outbreak. We observe similar effects in various other countries post-July including Japan (reduced to 1.1% and 22 days' delay) and Spain, France and Portugal (all reduced to 0.8–1.3%, 12–29 days' delay). These values are all similar and may reflect a reasonable estimate for CFR when testing is adequate; we would still expect the IFR to be lower by a factor of at least two due to undiagnosed and asymptomatic cases. A decrease in CFR over time may also indicate a change in the demographics of the case load or improvements in treatment or even an increasing time delay between reported cases and deaths, perhaps due to earlier diagnosis.

Conclusion

The TSD analysis is a straightforward way to predict CFR over time, using only publicly available data on cases and deaths and requiring no assumptions or parametrisations regarding the progress of the illness. The beauty of this method is in its transparency and simplicity; the lack of assumptions allows more to be gained from the data, including trends that may be unexpected or changing over time. This analysis method has particular utility early in an outbreak, once sufficient data are available for a robust fit (beyond the exponential growth phase). Without the benefit of hindsight, the TSD-calculated values for CFR and time delay between cases and deaths can shed light on the virulence of a disease and on the conditions that a particular country may be facing. Excess death data (where available) may be used to correct death data, while PTRs and other indicators or models of testing adequacy can often give an early rough idea of the true prevalence relative to reported case numbers. These data can be used to interpret the CFR calculated using TSD analysis early in an outbreak, and to approximate the IFR.

Our estimates of IFR range from 0.3% to 3%, with higher values observed for countries that experienced more severe outbreaks, perhaps reflecting the negative influence of overwhelmed health systems and the spread of disease to more vulnerable populations. The calculated time delay is also potentially informative; for example, the 1-day delay calculated from early data in Spain reflects the breakdown of testing and reporting systems at that time, whereas the revised delay time of 14 days shows the recovery of the system and the likely delay between case diagnosis and death. In this way, TSD analysis of data from a particular place at a particular time can give useful local information on the progression of an outbreak to inform public health planning and policy.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0950268821001436

Acknowledgements. We thank Dr Nick Golding (Curtin University) for many helpful conversations and comments on the manuscript.

Conflict of interest. None.

Data availability statement. The data used in this study are publicly available. COVID-19 data are from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University at https://github.com/CSSEGISandData/COVID-19 (via https://github.com/pomber/covid19), and from Worldometer at https://www.worldometers.info/coronavirus/. SARS data are from the World Health Organization at https://www.who.int/csr/sars/country/en/ (via https://www.kaggle.com/imdevskp/sars-outbreak-2003-complete-dataset). Excess death data are from the Economist's COVID-19 excess deaths tracker repository at https://github.com/TheEconomist/covid-19-excess-deaths-tracker, and positive COVID-19 test rates are from Our World in Data at https://ourworldindata.org/coronavirus-testing.

References

1. **Li Q *et al.*** (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine* **382**, 1199–1207.
2. **Linton NM *et al.*** (2020) Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of Clinical Medicine* **9**, 538.
3. **Rajgor DD *et al.*** (2020) The many estimates of the COVID-19 case fatality rate. *The Lancet Infectious Diseases* **20**, 776–777.
4. **Russell TW *et al.*** (2020) Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Eurosurveillance* **25**, 2000256.
5. **Verity R *et al.*** (2020) Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases* **20**, 669–677.
6. **Meyerowitz-Katz G and Merone L** (2020) A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *International Journal of Infectious Diseases* **101**, 138–148.
7. **Mackenzie J and Jeggo M** (2014) One Health: from concept to practice. In Yamada A *et al.* (ed.) *Confronting Emerging Zoonoses*. Tokyo: Springer, pp. 163–189.
8. **de Garine-Wichatitsky M *et al.*** (2020) Will the COVID-19 crisis trigger a One Health coming-of-age? *The Lancet Planetary Health* **4**, e377–e378.
9. **Public Health Agency of Sweden**. The Infection Fatality Rate of COVID-19 in Stockholm – Technical Report: Public Health Agency of Sweden; 202020094-2. Available at www.folkhalsomyndigheten.se/publicerat-material/.
10. **Mazumder A *et al.*** (2020) Geographical variation in case fatality rate and doubling time during the COVID-19 pandemic. *Epidemiology & Infection* **148**, 1–12.
11. **Angelopoulos AN *et al.*** (2020) On identifying and mitigating bias in the estimation of the COVID-19 case fatality rate. *Harvard Data Science Review*. doi: 10.1162/99608f92.f01ee285

12. **von Kügelgen J, Gresele L and Schölkopf B** (2021) Simpson's paradox in Covid-19 case fatality rates: a mediation analysis of age-related causal effects. *IEEE Transactions on Artificial Intelligence* **2**, 18–27.
13. **Dorigatti I *et al.*** (2020) Report 4: Severity of 2019-Novel Coronavirus (nCoV). Imperial College London, London 2020.
14. **Mizumoto K and Chowell G** (2020) Estimating risk for death from coronavirus disease, China, January-February 2020. *Emerging Infectious Diseases* **26**, 1251–1256.
15. **Wang W, Tang J and Wei F** (2020) Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. *Journal of Medical Virology* **92**, 441–447.
16. **Shim E *et al.*** (2020) Estimating the risk of COVID-19 death during the course of the outbreak in Korea, February–May 2020. *Journal of Clinical Medicine* **9**, 1641.
17. **Goldstein JR and Lee RD** (2020) Demographic perspectives on the mortality of COVID-19 and other epidemics. *Proceedings of the National Academy of Sciences* **117**, 22035–22041.
18. **Dong E, Du H and Gardner L** (2020) An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* **20**, 533–534.
19. **Ghani AC *et al.*** (2005) Methods for estimating the case fatality ratio for a novel, emerging infectious disease. *American Journal of Epidemiology* **162**, 479–486.
20. **Yang X *et al.*** (2020) Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *The Lancet Respiratory Medicine* **8**, 475–481.
21. **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**, 1775–1776.
22. **Yuan J *et al.*** (2020) Monitoring transmissibility and mortality of COVID-19 in Europe. *International Journal of Infectious Diseases* **95**, 311–315.
23. **Wilson N *et al.*** (2020) Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality. *Emerging Infectious Diseases* **26**, 1339–1441.
24. **World Health Organization**. Cumulative number of reported probable cases of severe acute respiratory syndrome (SARS). Available at https://www.who.int/csr/sars/country/en/; https://www.kaggle.com/imdevskp/sars-outbreak-2003-complete-dataset, 30 June 2020.
25. **Donnelly CA *et al.*** (2003) Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *The Lancet* **361**, 1761–1766.
26. **Nishiura H *et al.*** (2009) Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic. *PLoS ONE* **4**, e6852.
27. **Worldometer**. Coronavirus (COVID-19) mortality rate. Available at https://www.worldometers.info/coronavirus/coronavirus-death-rate/, 14 May 2020.
28. **World Health Organization** (2020) Coronavirus disease (COVID-19): log of major changes and errata in WHO daily aggregate case and death count data. 23 August 2020. Available at https://www.who.int/publications/m/item/log-of-major-changes-and-errata-in-who-daily-aggregate-case-and-death-count-data.
29. **Kuster AC and Overgaard HJ** (2021) A novel comprehensive metric to assess effectiveness of COVID-19 testing: inter-country comparison and association with geography, government, and policy response. *PLoS ONE* **16**, e0248176.
30. **Ioannidis JPA** (2021) Infection fatality rate of COVID-19 inferred from seroprevalence data. *Bulletin of the World Health Organization* **99**, 19–33.
31. **Li R *et al.*** (2020) Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science (New York, N.Y.)* **368**, 489–493.
32. **Wu SL *et al.*** (2020) Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications* **11**, 1–10.
33. **Our World in Data**. Excess mortality during the Coronavirus pandemic (COVID-19). Available at https://ourworldindata.org/excess-mortality-covid (Accessed 16 March 2021).
34. **Martino M.** How accurate are Australia's coronavirus numbers? The answer lies in our death data. Available at https://www.abc.net.au/news/2020-06-23/coronavirus-australia-excess-deaths-data-analysis/12321162, 23 June 2020.
35. **Lavezzo E *et al.*** (2020) Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* **584**, 425–429.
36. **Hicks SM *et al.*** (2021) A dual-antigen enzyme-linked immunosorbent assay allows the assessment of severe acute respiratory syndrome coronavirus 2 antibody seroprevalence in a low-transmission setting. *Journal of Infectious Diseases* **223**, 10–14.
37. **Our World in Data**. Coronavirus (COVID-19) testing. Available at https://ourworldindata.org/coronavirus-testing (Accessed 16 July 2020).
38. **Havers FP *et al.*** (2020) Seroprevalence of antibodies to SARS-CoV-2 in 10 sites in the United States, March 23–May 12, 2020. *JAMA Internal Medicine* **180**, 1576–1586.
39. **The Economist**. Tracking covid-19 excess deaths across countries. Available at https://www.economist.com/graphic-detail/2020/07/15/tracking-covid-19-excess-deaths-across-countries. (Accessed 15 October 2020).
40. **Rosenberg ES *et al.*** (2020) Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York. *Annals of Epidemiology* **48**, 23–29.
41. **New York State Government**. Amid ongoing COVID-19 pandemic, Governor Cuomo announces results of completed antibody testing study of 15,000 people showing 12.3 percent of population has COVID-19 antibodies; 2 May 2020. Available at https://www.governor.ny.gov/news/amid-ongoing-covid-19-pandemic-governor-cuomo-announces-results-completed-antibody-testing.
42. **Streeck H *et al.*** (2020) Infection fatality rate of SARS-CoV2 in a superspreading event in Germany. *Nature Communications* **11**, 5829.
43. **Stringhini S *et al.*** (2020) Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *The Lancet* **396**, 313–319.
44. **Folkhälsomyndigheten (Public Health Agency of Sweden)** (2020) Första resultaten om antikroppar efter genomgången covid-19 hos blodgivare; 18 June 2020. Available at https://www.folkhalsomyndigheten.se/nyheter-och-press/nyhetsarkiv/2020/juni/forsta-resultaten-om-antikroppar-efter-genomgangen-covid-19-hos-blodgivare/.
45. **Ward H *et al.*** (2021) SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. *Nature Communications* **12**, 905.
46. **Herzog S *et al.*** (2020) Seroprevalence of IgG antibodies against SARS coronavirus 2 in Belgium: a prospective cross-sectional study of residual samples. *Medrxiv*. doi: 10.1101/2020.06.08.20125179
47. **Pollán M *et al.*** (2020) Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet* **396**, 535–544.
48. **Pastor-Barriuso R *et al.*** (2020) SARS-CoV-2 infection fatality risk in a nationwide seroepidemiological study. *BMJ* **371**, m4509.
49. **Italian National Institute of Statistics** (2020) Primi risultati dell'indagine di sieroprevalenza SARS-CoV-2; 3 August 2020. Available at http://www.salute.gov.it/imgs/C_17_notizie_4998_0_file.pdf.
50. **Pagani G *et al.*** (2020) Seroprevalence of SARS-CoV-2 significantly varies with age: preliminary results from a mass population screening. *Journal of Infection* **81**, e10–e12.
51. **Instituto Superiore di Sanita (ISS)** (2020) Epidemia COVID-19: Aggiornamento nazionale. Roma; 30 giugno 2020. Available at https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_30-giugno-2020.pdf.
52. **Australian Government: Department of Health**. Coronavirus (COVID-19) current situation and case numbers. Available at https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/coronavirus-covid-19-current-situation-and-case-numbers (Accessed 3 August 2020).
53. **Levin AT *et al.*** Assessing the age specificity of infection fatality rates for COVID-19: meta-analysis & public policy implications. Cambridge MA: National Bureau of Economic Research; July 2020, revised October 2020; Working Paper 27597. Available at http://www.nber.org/papers/w27597.
54. **Valero M and Valero-Gil JN** (2021) Determinants of the number of deaths from COVID-19: differences between low-income and high-income countries in the initial stages of the pandemic. *International Journal of Social Economics* **48**, 1229–1244.
55. **Grifoni A *et al.*** (2020) Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501.