Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Letter to the Editor

Rethinking case fatality ratios for covid-19 from a data-driven viewpoint

Dear Editor,

While examining the association between the Case Fatality Ratio (CFR) and the cumulative number of COVID-19 infections in this journal, Kenyon\(^1\) recently came across various difficulties in estimating the CFR. One of these was addressed by Baud and colleagues\(^2\), who pointed out that the CFR (number of reported deaths divided by reported cases) ignores the time delay between incubation and death. Various problems\(^3\) with their approach have been identified, but a concrete solution is unclear. While we agree that time lag plays an important role, it is overestimated by Baud and coworkers\(^2\), whereas reported CFR values\(^4\) ignore time lag completely. We find that either of these approaches introduces a spurious time dependence that severely distorts the magnitude and true meaning of the CFR. Instead, a suitably corrected CFR is far more useful as an indicator of COVID-19 fatality, because it turns out to be constant in time for many countries, as we show.

The CFR is unfavorably compared with the Infection Fatality Ratio (IFR)\(^2\) of deaths over total actual infections, often because asymptomatic cases do not contribute to it, unless identified by testing. The IFR is important, but practically impossible to measure, due to lack of data for the denominator, which requires widespread, continuous random testing\(^5\).

The CFR (only including reported cases) may have its uses in estimating fatalities\(^6\). Assuming random testing is very limited\(^4\), the majority of reported cases have developed symptoms severe enough to seek medical assistance; these individuals are far more likely to die from the disease than asymptomatic cases\(^8\), which would go undetected in the absence of testing. In this sense, the CFR is a meaningful measure of fatality risk among symptomatic individuals. This begs the question whether CFR versus time might be roughly constant for each country, at least during a period of fixed social distancing measures. This constant value would be different for each country, because of differing age distributions, mortality being a strongly increasing function of age\(^9\), and possibly other factors.

At first glance, this hypothesis is not supported by the COVID-19 data\(^4\). Most countries have an increasing, and some a decreasing CFR that eventually levels off to a constant.

To test the constant CFR hypothesis, we started with a hard-hit country, Italy, plotted deaths and reported cases versus time (Fig. 1a), and observed that multiplying deaths by roughly a factor of 7, made the two graphs almost the same (Fig. 1b), except for a shift \(\Delta t=4\) days; after compensating for which they became nearly indistinguishable (Fig. 1c), implying a CFR 1/7 = 0.14 that remains virtually constant within 3% of CFR.

Baud and coworkers\(^2\) used a lag of 14 days, representing symptom onset to death. Instead we feel that the time lag \(\Delta t\) should reflect time from reporting to death. Delays from onset to reporting do occur\(^1\)\(^9\). In Singapore these delays had a mean of a week\(^9\) and could exceed two weeks. Moreover, delays in reporting bring delays in critical medical care, hence may accelerate death. By increasing the time from onset to reporting, such factors decrease the time lag \(\Delta t\) from reporting to death, so we might expect \(\Delta t\) to be much less than 14 days, but uncertainty is introduced. Here, instead of arbitrarily picking CFR and \(\Delta t\), or using estimates from a different location\(^3\), we let the data decide. Data-driven predictions\(^11\) of epidemic metrics are promising. We use a simple data-driven approach to find the right constant values, CFR and \(\Delta t\).

Simply put, we choose these values to be the ones that minimize the root mean square deviation between cases and deaths versus time, with deaths multiplied by cCFR and shifted back by c\(\Delta t\); see the Appendix for details. This gives c\(\Delta t\)=4 days for Italy. Shifting deaths back in time by this c\(\Delta t\), then dividing by cases, yields a virtually constant CFR versus time (red curve, Fig. 1d), equal to cCFR=0.14 (black dashed line, Fig. 1d) within a few percent. The statement “14% of reported cases die after four days” remains closer to the truth for much longer than any analogous statement regarding the reported, variable CFR that nearly doubles its value in two months (orange curve Fig. 1d).

This procedure works for many countries (Fig. 2), producing a different cCFR and c\(\Delta t\) for each, but also for the entire world: cCFR=0.08, c\(\Delta t\)=3 days (black dashed line, Fig. 2), but a nearly constant corrected CFR for all cases considered.

The reported\(^2\) CFR (orange curve, Fig. 1d), which ignores time lag, increases with time and underestimates Italy’s cCFR by a time-dependent amount. Baud et al. approach, shifting deaths back by \(\Delta t=14\) days\(^2\) (green curve, Fig. 1d) overestimates Italy’s cCFR and decreases with time.

In summary, by allowing for an initially unknown time lag between case reporting and death, we find that many countries, and the entire world, exhibit a corrected CFR that is essentially constant during a long period of imposed social distancing. This value can be estimated long before the full evolution of the pandemic, hence it is useful for early prediction of fatalities, in situations where extensive random testing is not available.

https://doi.org/10.1016/j.jinf.2020.06.010
0163-4453/© 2020 The British Infection Association. Published by Elsevier Ltd. All rights reserved.
Fig. 1. The case for a constant CFR. (a) Italy deaths (red) and cases (blue). (b) as in (a) but with deaths scaled by a factor of 7. (c) as in (a) but with deaths scaled by 7 and shifted back by 4 days. The result is a constant CFR = 1/7. (d) Various estimates of CFR versus time (March 20–May 20, 2020) for Italy. Orange: reported value ignoring time delay ($\Delta t = 0$). Black dashed line: our prediction of $c\text{CFR} = 0.14$. Red: Corrected CFR with deaths shifted back by our predicted $c\Delta t = 4$ days. Green: Using Baud et al. method with deaths shifted back by $\Delta t = 14$ days. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Authors' contributions

PR and MEM designed the research. PR performed the analysis with input from MEM. Both authors discussed the results and wrote the manuscript.

Funding

None.

Ethical approval

The work involved a secondary analysis of public access data. No ethics approval was necessary.

Informed consent

Not applicable.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jinf.2020.06.010.

References

1. Kenyon C. Flattening-the-curve associated with reduced COVID-19 case fatality rates—an ecological analysis of 65 countries. J Infect Pr 16, 2020.
2. Baad D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. Lancet Infect Dis Mar 12, 2020.
3. Spychalski P, Blazynska-Spychalska A, Kobela J. Estimating case fatality rates of COVID-19. Lancet Infect Dis Mar 31, 2020.
4. Roser M, Ritchie H, Ortiz-Ospina E, Hasell J. Coronavirus pandemic (COVID-19). Our World Data Mar 4, 2020. Published online at OurWorldInData.org. Retrieved from: https://ourworldindata.org/coronavirus.
5. Russell TW, Hellewell J, Jarvis CI, Van Zandvoort K, Abbott S, Ratnayake R, Flasche S, Egoq RM, Edmunds WJ, Kucharski AJ. MMID COVID-19 working group: 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 Mar 26, 2020:26(12):2000256.
6. Lipstich M, Donnelly CA, Fraser C, Blake IM, Gor A, Dorigatti I, Ferguson NM, Garske T, Mills HL, Riley S, Van Kerkhove MD. Potential biases in estimating absolute and relative case-fatality risks during outbreaks. PLoS Negl Trop Dis Jul 2015;9(7).
7. Bendavid E, Mulaney B, Sood N, Shah S, Ling E, Bromley-Duffano R, Lai C, Weissberg Z, Saavedra R, Tedrow J, Tversky D. COVID-19 antibody seroprevalence in Santa Clara County, California: MedRxiv; Jan 1, 2020.
8. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. JAMA Mar 23, 2020.
9. Leung K, Wu JT, Liu D, Leung GM. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a model-elling impact assessment. Lancet Apr 8, 2020.
10. Tariq A, Lee Y, Boska K, Blumberg S, Van P, Ma S, Chowell G. Real-time monitoring the transmission potential of COVID-19 in Singapore. Febra 2020. medRxiv Jan 1, 2020.
11. Barmparis GD, Tzironis GP. Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. Chaos, Solitons Fractals Apr 27, 2020:109842.