

Article

The lag between daily reported Covid-19 cases and deaths and its relationship to age

Raymond Jin

Farragut High School, Knoxville, TN, USA

Abstract

Background: The Coronavirus disease 2019 (Covid-19) has infected millions and killed tens of thousands of people. Public health measures put in place by governments are essential to the success of controlling this disease. However, governments may not feel as incentivized to implement these measures when deaths are not rising along with cases. However, it is known that a delay exists between the time of infection and the time of death. This study attempted to find how long that lag is and how the age of people infected may affect that lag.

Design and Method: A descriptive and correlational study was carried out to investigate the length of the lag and the relationship between lag and age.

Results: The average lag between daily Covid-19 cases and deaths was 8.053 days with a standard deviation of 4.116 days for nineteen regions. After excluding data from three more regions due to unavailable age data, the regression yielded an equation of $\text{lag} = 14.015 - 0.153 (\% \text{ cases above } 60)$ with a p-value of 0.066. Because the p-value of 0.066 is lower than the 0.10 significance level, there is evidence that a relationship exists between the lag and the age of cases.

Conclusion: The results show that regions must remain vigilant when Covid-19 cases rapidly increase without similar increases in deaths since there exists a significant lag between the two. Additionally, a younger demographic of cases may lead to an increased lag, further pushing regions into a false sense of security that should be avoided.

Introduction

On December 31st, 2019, health officials in China reported a group of patients with pneumonia-like symptoms in the city of Wuhan.¹ While the cause of these symptoms was unknown at the time, it was quickly identified to come from a new strand of coronavirus known as severe acute respiratory syndrome coronavirus 2, or Sars-CoV-2.^{2,3} The disease itself was then named Covid-19.³ Initially, the disease spread most prevalently in throughout Wuhan

and other parts of China. However, due to the interconnectedness of modern society, Covid-19 has since spread around the entire world. As of July 30th, it has infected more than 17 million people and caused more than 650,000 deaths.⁴

The United States of America has become the epicenter of Covid-19, having had nearly 4.5 million reported cases and more than 150,000 reported deaths.⁴ However, unlike most other countries in Europe and Asia, the United States has been unable to contain this disease. An initial spike of cases and deaths occurred in April, followed by a slow decline in cases and deaths.⁵ Around mid-June, this slight decrease suddenly became the second spike in cases. However, while this second spike in cases is around twice as large as the initial spike in April, the deaths have not increased accordingly.⁵ Many have pointed to this as a sign that there is little to fear about this virus and that things can simply go back to normal.^{6,7} However, it is well known that it takes time for deaths caused by a surge in cases to be reported. What is still unclear is how long that period might be. The World Health Organization has estimated that the time from onset of symptoms to death ranged from two to eight weeks, and other studies have shown that a median of 13 days transpired between pneumonia onset to death.⁸⁻¹⁰ However, these findings were mainly done with preliminary data from China and focused on individual cases. A working paper from the Harvard Center for Population and Development Studies visualized the lag between Covid-19 cases and deaths and estimated the lag to be four to six weeks but did so only within the context of the United States.¹¹ Additionally, there may be other factors that can affect this lag. It is known that older people are at a higher risk of dying from Covid-19, but it is unknown if this might affect the lag in any way.^{12,13} This research hopes to rectify gaps in the understanding of the lag between reported daily cases and deaths as well as see if age has any relationship with this lag.

Methods

Data collection

Data on daily Covid-19 cases and deaths for various countries

Significance for public health

After the first wave of Covid-19 infections has passed, it is very likely that a second surge in infections will occur sometime afterwards. However, during this second wave of infections, governments, for a variety of reasons, may be reluctant to reimplement certain public measures. They might attempt to justify this decision by pointing out how new deaths from Covid-19 have not been rising accordingly with new cases. But it is known that there exists a certain delay between when a person becomes infected with the Covid-19 disease and when they pass away. As such, attaining a better understanding of the extent of this lag and how differences between the initial wave and any subsequent waves may affect the delay will help governments make decisions regarding public health measures. This information could ensure that proper action is taken so that fewer people will become infected and possibly die.

were collected mostly from a database compiled by Our World in Data.⁵ Data for U.S. states came directly from their state health departments. All data is up to date as of July 15th, which is when this study was carried out. I denotes the set of all regions chosen to be included in the data with i as its index, n_i is the number of dates with data available for region i with t as its index, X_i^t is the number of new cases reported in region i on day t , and Y_i^t is the number of new deaths reported in region i on day t .

The data for these regions, however, were not randomly chosen. While a global perspective is necessary to obtain a more comprehensive view of this disease, many regions either have inadequate Covid-19 data or are attempting to censor the data for various reasons. Additionally, the regions chosen are those that have already seen a spike in Covid-19 cases and deaths and have now brought the numbers down. This clear spike in reported cases and deaths is not evident in all regions analyzed, but it holds true for most of them. These two criteria for data selection ensure that the data is as accurate as it can be in order to represent the actual situation as precisely as possible.

The data for the percentage of cases in people 60 years or older within region I , denoted by A_i , was primarily collected from individual regional databases or reports. For some regions, data were only available for cases above 65 years old. In others, some data regarding the age of cases were missing.

Data analysis

To begin analyzing the data, X_i^t and Y_i^t had to be aligned to begin on the same date t . The date t by which region i is considered to have had its first case is when a date with reported cases is followed by two or fewer consecutive days of zero cases, which is then followed by multiple consecutive days of cases. Y_i^t would then begin on the same date that the first confirmed cases occurred on. This method was employed because some regions reported very limited numbers of Covid-19 cases long before the majority of their cases were reported. If the first day of a confirmed case was established as the actual first day when a country or state detected a case, then information would have been heavily skewed. This would occur because deaths would likely not begin occurring until there were a substantial number of people who had become infected with Covid-19.

Occasionally, there would be anomalies within the data. For some region i , a date t would exist such that there were apparently no new reported cases when the trend at the time would suggest otherwise. This is likely due to either a reporting lapse by the governments of those regions or an anomaly with the database. In order to remedy this issue, the average number of new cases from the day prior to and the day after a particular date would be used to smoothen the number of new cases on each date. If there were two consecutive days with no reported cases when there should have been, the number of new cases on the first date would be the average of the new number of cases the day before and two days later. The number of cases on the second date would then be the average of the number of newly reported cases of the previous day and the day after. A similar process was done for daily new deaths reported. However, since there is a higher chance of no deaths actually being reported on a single day, this would only be done on days where either the previous or next day's deaths were greater than 10. It should also be noted that this situation with two consecutive days having no new cases when there reasonably should have been new cases occurred only seven times in all of the data analyzed.

In order to find the lag between the newly reported daily cases

and the newly reported daily deaths of region i , L_i was defined as the estimated lag for region i to maximize the estimated correlation r_i^l between series X_i^t and Y_i^t with a lag of l . The correlation r_i^l with lag l for region i was found using equation (1).

$$r_i^l = \frac{\sum_{t=1}^{n_i-l} (X_i^t - \bar{X}_i(l))(Y_i^{t+l} - \bar{Y}_i(l))}{\sqrt{\sum_{t=1}^{n_i-l} (X_i^t - \bar{X}_i(l))^2 \times \sum_{t=1}^{n_i-l} (Y_i^{t+l} - \bar{Y}_i(l))^2}} \quad \forall i \in I; \forall l \in \{1, 2, \dots\} \quad (\text{eq. 1})$$

where $\bar{X}_i(l) = \frac{\sum_{t=1}^{n_i-l} X_i^t}{n_i-l}$ and $\bar{Y}_i(l) = \frac{\sum_{t=1+l}^{n_i} Y_i^t}{n_i-l}$.

Based on r_i^l , the estimated lag L_i was found using equation (2).

$$L_i = \underset{l \in \{1, 2, \dots\}}{\operatorname{argmin}} r_i^l, \quad \forall i \in I, \quad (\text{eq. 2})$$

The following Results subsection will show that r_i^l is unimodal regarding l for each region i . The mean, standard deviation, median, and interquartile range of L_i across all regions were then found to give an overall picture about the disease regarding the lag.

It is widely believed that the severity and fatality rate of Covid-19 are related to ages. In order to investigate the relationship between L_i and the A_i , the percentage of cases in people 60 years or older within region i , a simple linear regression analysis was carried out for the two variables with A_i as the independent variable and L_i as the dependent variable. Residual and normality plots were plotted to validate the linear regression. If the regression model has a power lower than the significant level of 0.10, then the relationship between the two variables is considered statistically significant.

Results

Lag between newly reported daily Covid-19 cases and deaths

Following the criteria set up above, data from 25 regions were analyzed. However, the U.K., Ukraine, United States of America, Ecuador, Denmark, and Indonesia either had their highest correlation between new cases and deaths below 0.60, or the lag found for them was 0 days. A low correlation would seem to indicate incorrect or poorly gathered data since more cases should ultimately lead to more deaths. A lag of 0 days would give a similar impression since patients generally do not die the same day their case is reported.¹⁴ It is also possible that limits on data reporting caused deaths and cases to be reported on the same day in those regions. Therefore, those regions were excluded from the analysis. Specifically, for the United States of America, it is also possible that the data could not show any strong correlation due to the nation's large population size spread across a broad geographical region. However, multiple other regions within the country that were known hotspots, such as New Jersey or New York City, were included in the analysis. Table 1 shows the data used for all the remaining 19 regions. Looking at the correlation r_i^l between the series X_i^t and Y_i^t for multiple regions i in Figure 1, it is apparent that r_i^l is unimodal over l . The average L_i for all regions i was 8.053 days with a standard deviation of approximately 4.116 days. The median L_i was 7 days, while the interquartile range was 6 days. Norway was an outlier in this data, with a lag of twenty days.

Relationship between lag and age of Covid-19 cases

There were even fewer countries or states that could be analyzed for a relationship between the percentage of Covid-19 cases

above 60 years old and the number of days lag. Kuwait, South Africa, and Hungary were not included due to a lack of data. This means that sixteen regions were analyzed in total. It should be noted that for New York City and Ireland, data were only available for cases above 65 years old. As such, these two countries are highlighted in Figure 2. The data were still utilized as is because they would be reasonably close to the percentage of cases above 60 years of age, and it was important to have more data points in this linear regression analysis. In France, only the data of those who had at any point been hospitalized since March 1st or were currently hospitalized as of July 15th was available. This means that there is a chance that this data is flawed. However, these data were still used. To ensure a linear regression can be properly carried out, the residual plot and normal probability plot should indicate that the residual terms follow a Normal (or Gaussian) distribution. Indeed, the residual plot has no apparent pattern to it, and the Normal probability plot is approximately linear. Carrying out a linear regression analysis yields the least-squares regression line of $\text{lag} = 14.015 - 0.153 (\% \text{ cases above } 60)$, which is shown in Figure 2. This equation indicates that with every percent increase in A_i for a region i , L_i will decrease by 0.153 days. Conversely, this means that for every percent decrease in A_i for a region i , L_i will increase by 0.153 days. Additionally, the p-value of this analysis was approximately 0.066, lower than the alpha level of 0.10. This fact indicates that the correlation between the percent of cases above 60 years old on the lag is statistically significant.

Discussion

This research aimed to find two things: what the lag time between newly reported daily cases and daily deaths was for the Covid-19 disease and what the relationship, if any, was between the percent of reported Covid-19 cases in people older than 60 years old and the lag. For the first question, it was found that the average lag between new daily cases and deaths was about 8.053 days. This means that when a surge of cases happens in a region, it will usually take about 8.053 days before a surge in deaths begins to be reported. In regard to the second question, it was found that with a p-value of 0.066, lower than the alpha level of 0.10, a relationship does indeed exist between the percent of confirmed Covid-19 cases above 60 years old and the length of the lag. This relationship happens to be a negative one, where the higher the percentage of older people becoming sick with Covid-19, the shorter the delay between new daily cases and deaths. Such a relationship

also means that as a larger proportion of younger people becomes infected, which corresponds to a smaller proportion of older people becoming infected, the lag will increase.

Previous research has estimated the time from the onset of symptoms to death, but this does not accurately reflect itself in the data provided to the public.⁸⁻¹⁰ This is because it takes time between getting tested and receiving the results of that test back.¹⁵ Additionally, deaths are not always reported the day when they happen, further warping the data.¹⁴ Another thing that has been done to offer more insight into the relationship between the daily confirmed cases and deaths is to visualize the lag between the two variables. An animation showcased in a working paper from the

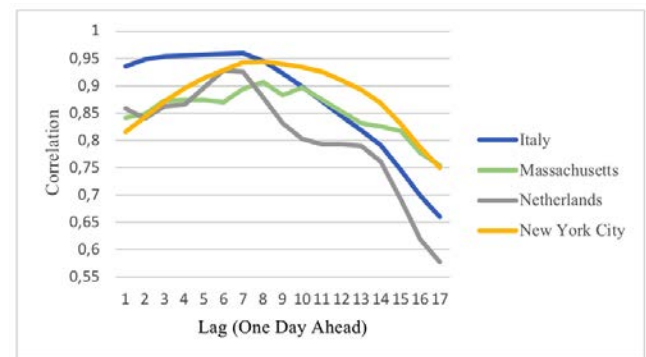


Figure 1. Correlation between new daily cases and new daily deaths with lag L .

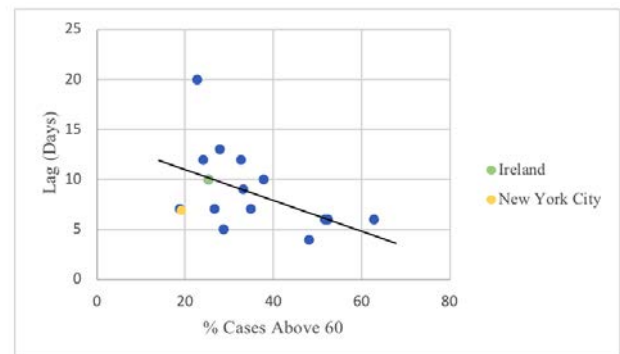


Figure 2. Lag vs % cases above 60.

Table 1. Lag and % of Cases Above 60 in each Region

Region	Lag (days)	% cases above 60	Region	Lag (days)	% cases above 60
Italy	6	52.4	Switzerland	12	32.6
Spain	6	51.5	Hungary	7	-
France	6	62.9**	N.Y.C.	7	18.9*
Norway	20	22.6	South Africa	5	-
Massachusetts	7	34.9	Connecticut	10	37.8
Netherlands	5	28.6	Kuwait	2	-
Ireland	10	25.2*	Belgium	4	47.9
Canada	9	33.3	Portugal	5	28.5
Romania	7	27.0	Germany	13	28.0
New Jersey	12	24.1			

*Data was only available for people 65 years or older; **data does not include all Covid-19 cases, just those that were hospitalized at any point in time.

Harvard Center for Population and Development Studies shows how it takes approximately four to six weeks for new daily deaths to increase after new daily cases do so. However, this visualization relied on data solely from states in the United States, giving a less comprehensive view of the entire situation.¹¹ Moreover, little to no research seems to have been done in the past relating the lag between cases and deaths to the ages of the people who have become infected with the virus.

Having a clear understanding of the data about Covid-19 is crucial for the public's perception of the disease. Specifically, in the United States, mixed messaging from government and health officials from the onset of this pandemic in the country has caused a seemingly uncontrollable spread of the virus. Most recently, the U.S. has seen a second surge in cases, one that is much larger than the first one, but some officials still attempt to persuade the public that things are getting better.^{6,7} They point to the fact that these new cases are mostly from younger people, who are not as seriously impacted by this disease, and that the deaths have not substantially increased.

However, this research clearly indicates how it may take several days from the start of a surge in cases before a surge of deaths begins. This in itself, however, is not enough to evaluate the current situation in the United States specifically. This is because of the changing demographics of the people becoming infected with Covid-19. Many reports, as well as comments made by health officials, indicate that a higher proportion of younger people are testing positive for Covid-19 compared to the initial surge in cases back in April.^{16,17} Taking a look at Florida as an example, back in early April, the age group with the largest percentage of Covid-19 cases was the 45 to 54 age group.¹⁸ As of July 27th, 25 to 45-year old people comprise the plurality of total cases.¹⁹ As such, not only will there be a lag between a spike in new Covid-19 cases and deaths, the lag could be longer than what occurred in April. This fact means that officials should not begin celebrating too early at the seemingly stagnate death count and give the public false reassurances about this disease. This may only make the public more complacent and further exacerbate and prolong the pandemic in the United States.

For various countries around the world, the same things said above still hold true. While many nations have contained Covid-19 so far, it is possible that they will see a resurgence of cases in the fall.^{20,21} This second wave of cases may very well affect a different demographic of people, but even if it does not, health and government officials must remain vigilant about the severity of the surge. Understanding the lag between new daily cases and daily deaths means that if testing indicates a significant increase in Covid-19 cases, measures should be put in place again to stem that increase. Officials cannot wait until they begin seeing an increase in deaths to do so, or else it may be too late.

Of course, this study has its limitations. Foremost of them is the fact that the data of reported daily cases and deaths is likely not accurate. This is because it is nearly impossible for the entire population of a region cannot be repeatedly tested. As a result, there is no way to know precisely how many people have been infected with the virus. Without accurate knowledge of who has been infected with Covid-19, health professionals may not know for sure if a person died as a result of the disease. Another issue with this data is that regions may be manipulating it in some way to make the situation seem better than it is. While there was an attempt to limit this possibility through a selection process, it is difficult to know for certain if a region had nefarious intentions with data manipulation. Additionally, the quickly evolving nature of an ongoing pandemic means that data used in this research is

outdated. This research only used data that was available as of July 15, 2020. Besides possible errors in the data used for this study, there is also the issue of not using a very representative sample of countries. Since there was a need for accurate data and a fair number of cases and deaths for accurate analysis, the regions chosen were mostly European countries or states in the United States of America. This situation means that due to genetic differences, possible differences in the virus itself, differences in societal structures, and differences in economic and medical capabilities, it may not be accurate to extrapolate the results of this study to all regions in the world.

Conclusion

This study shows that there is an average lag of 8.053 days between new reported daily cases and daily deaths. Additionally, with a p-value of approximately 0.066, there is good reason to believe that a higher percent of confirmed Covid-19 cases in people over 60 years of age leads to a shorter lag between daily cases and deaths. This then means that a lower percent of confirmed Covid-19 cases in people over 60 years of age, which would correspond with a higher percent of confirmed Covid-19 in younger people, will likely result in a longer lag. As a result, the recent surge in cases without an increase in deaths in the United States cannot be dismissed since it is likely that it will take some time before deaths begin to spike as well. Additionally, with a higher proportion of younger people becoming infected with this disease now, this lag may be even longer than the average found above. As such, the people and the government of the United States and other countries in the world cannot and should not become complacent to this disease or else risk the situation becoming even worse.

Correspondence: Raymond Jin, Farragut High School, 11336 Hawkstowe Ln, Knoxville, TN, 37934, USA. Tel. +1.865.230-6647. E-mail: raymond.jin102@gmail.com

Key words: Covid-19; lag between reported new cases and deaths; age.

Acknowledgements: This study could not have been done without the support of my family.

Contributions: RJ, collected, organized, and analyzed all of the data utilized in this article and drew the conclusions from the results of the data analysis.

Conflict of interest: The author declares no conflicts of interest.

Funding: The author declares that there was no outside funding for the writing of this paper.

Data Availability: All data is available on public databases with no access restrictions.

Received for publication: 1 December 2020.

Accepted for publication: 8 March 2021.

©Copyright: the Author(s), 2021

Licensee PAGEPress, Italy

Journal of Public Health Research 2021;10:2049

doi:10.4081/jphr.2021.2049

This work is licensed under a Creative Commons Attribution NonCommercial 4.0 License (CC BY-NC 4.0).

References

1. World Health Organization. Archived: WHO Timeline - COVID-19. 2020. Accessed: 2020 Aug 3]. Available from: <https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19>
2. World Health Organization. Novel Coronavirus – China. 2020. Accessed: 2020 Aug 3. Available from: <https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/>
3. World Health Organization. Naming the coronavirus disease (COVID-19) and the virus that causes it. Accessed: 2020 Nov 30. Available from: [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it)
4. Johns Hopkins Coronavirus resource Center [Internet]. COVID-19 map. Accessed: 2020 Jul 30. Available from: <https://coronavirus.jhu.edu/map.html>
5. Our world in data [Internet]. Coronavirus pandemic (COVID-19). 2020. Accessed: 2020 Nov 30]. Available from: <https://ourworldindata.org/coronavirus>
6. NPR [Internet]. Some people “have the sniffles”: Trump downplays the coronavirus’s severity. 2020 Jul 19. Accessed: 2020 Aug 3. Available from: <https://www.npr.org/2020/07/19/892787298/some-young-people-have-the-sniffles-trump-downplays-the-coronavirus-severity>
7. Miami Herald [Internet]. DeSantis downplays COVID surge, suggests it was more prevalent than state admitted. 2020. Accessed: 2020 Aug 3. Available from: <https://www.miamiherald.com/news/coronavirus/article244025297.html>
8. World Health Organization. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). 2020. Available from: <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>
9. Yang X, Yu Y, Xu J, , et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* 2020;8:475-81.
10. Wilson N, Kvalsvig A, Barnard LT, Baker MG. Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality. *Emerg Infect Dis* 2020;26:1339–441.
11. Harvard Center for Population and Development Studies Working Paper Series [Internet]. Visualizing the lagged connection between COVID-19 cases and deaths in the United States: An animation using per capita state-level data (January 22, 2020 – July 8, 2020). Available from: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1266/2020/07/HCPDS-WP_19_4_testa-et-al_Visualizing-Lagged-Connection-Between-COVID-19-Cases-and-Deaths-in-US_final_07_10_with-cover.pdf
12. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA* 2020;323:1775–6.
13. Centers for Disease Control and Prevention. Severe outcomes among patients with Coronavirus disease 2019 (COVID-19) - United States, February 12-March 16, 2020.. Available from: https://www.cdc.gov/mmwr/volumes/69/wr/mm6912e2.htm?s_cid=mm6912e2_w
14. Centers for Disease Control and Prevention. Provisional death counts for Coronavirus disease 2019 (COVID-19). 2020. Accessed: 2020 Aug 3. Available from: <https://www.cdc.gov/nchs/nvss/vsrr/covid19/>
15. Centers for Disease Control and Prevention.. Test for Current Infection. 2020. Accessed: 2020 Aug 3. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/testing/diagnostic-testing.html>
16. Vox [Internet]. Why more young people are getting sick in the latest Covid-19 outbreaks. 2020. Accessed: 2020 Aug 3. Available from: <https://www.vox.com/2020/7/18/21328358/covid-19-cases-by-age-florida-arizona-texas-miami>
17. Center for Infectious Disease Research and Policy. COVID-19 cases among US young adults spike. 2020. Accessed: 2020 Aug 3. Available from: <https://www.cidrap.umn.edu/news-perspective/2020/06/covid-19-cases-among-us-young-adults-spike>
18. Florida Division of Emergency Management [Internet]. Coronavirus: Summary of persons being monitored, persons under investigation, and cases 2020 Apr. Available from: <https://www.floridadisaster.org/globalassets/covid19/dailies/covid-19-data---daily-report-2020-04-08-0954.pdf>
19. Florida Division of Emergency Management [Internet]. COVID-19: Summary of Persons Being Monitored, Persons Tested, and Cases. 2020 Jul. Available from: https://www.floridadisaster.org/globalassets/covid19/dailies/state_reports_2020_0728.pdf
20. John Hopkins Medicine [Internet]. Coronavirus second wave? Why cases increase 2020. Accessed: 2020 Aug 3. Available from: <https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus>
21. Wise J. Covid-19: Risk of second wave is very real, say researchers. *BMJ* 2020;369:m2294.