Predicting the number of COVID-19 cases from the reported number of deaths

3 authors:

Åke Brännström
Umeå University
70 PUBLICATIONS 2,058 CITATIONS
SEE PROFILE

Henrik Sjödin
Umeå University
10 PUBLICATIONS 110 CITATIONS
SEE PROFILE

Joacim Rocklöv
Umeå University
200 PUBLICATIONS 5,169 CITATIONS
SEE PROFILE

Predicting the number of COVID-19 cases from the reported number of deaths

Åke Brännström[1,2], Henrik Sjödin[3], and Joacim Rocklöv[3]

1. Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden
2. Evolution and Ecology Program, International Institute for Applied Systems Analysis, Laxenburg, Austria
3. Department of Public Health and Clinical Medicine, Section of Sustainable Health, Umeå University, Umeå, Sweden

The new corona virus disease 2019 (COVID-2019) is rapidly spreading through the world. With insufficient testing, available case data may underestimate the total number of infections. We statistically estimated the cumulative number of cases with confidence intervals from the reported number of deaths, assuming that the number of infections grow exponentially with a constant doubling time. We took the doubling time and the COVID-19 case fatality rate to be 4 and 0.8% respectively. The results are shown in Figure 1. The predicted number of cases, though not the confidence intervals, can be adjusted to any other case fatality rate, *p,* by multiplying with 0.008/*p*. Our findings indicate that the number of unreported infections is likely at least one order of magnitude higher than the reported cases, in particular in the United States and the United Kingdom. This finding is supported by a recent mass screening for SARS-CoV-2 in Iceland which found that the actual case numbers may be about 20 times higher than the officially reported number [1].

In computing the estimates, we assumed that the latency time and time from onset to death are gamma distributed with an average time of 5.6 days and 15.0 days, respectively, and a standard deviation of 3.9 days and 6.9 days, respectively [2]. Assuming that these times are statistically independent, we determined the average time from infection to death as 20.6 days with a standard deviation of 7.9 days. Under these assumptions, we found that the expected number of deaths are given by $pk(\mathrm{T})N$ where p is the COVID-19 case fatality rate, N is the cumulative number of cases, and $k(T)$ is a function dependent on the doubling time. From this relationship, we estimated the cumulative number of cases as $\tilde{N}=pk(T)/d$ where d is the observed number of deaths.

To determine indicative confidence intervals, we assumed that the number of new cases t days ago was $\tilde{N}$ exp(-log(2)t/4)- $\tilde{N}$ exp(-log(2)(t-1)/4), rounded off to the nearest integer. From the COVID-19 case-fatality rate and the distribution of time from infection to death, we determined for one of these individuals the probability that he/she had died before the current date. Assuming statistical independence, the number of dead individuals at the current day is approximately normally distributed. We determined the 2.5% and 97.5% quantiles, $d_{0.025}$ and $d_{0.975}$ and estimate a 95% confidence interval for the number of current cases as $pk(T)/d_{0.025}$ to $pk(T)/d_{0.975}$

A limitation of our estimates is that it rests on an assumption of exponential growth with a constant doubling time that may not be satisfied even in an early epidemic, for example due to imported cases or recent interventions that have reduced the growth rate. The results are sensitive to changes in the doubling time and the average time from infection to death. Nevertheless, we believe our estimates offer guidance. Combined with other, independent, estimates of the number of cases, our methodology can also be used to estimate of the COVID-19 case-fatality rate as $p=dN/k(T)$.

References

1. First deCode COVID-19 findings are in. https://icelandmonitor.mbl.is/news/news/2020/03/16/first_decode_covid_19_findings_are_in/
2. Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S. M., ... & Nishiura, H. (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.

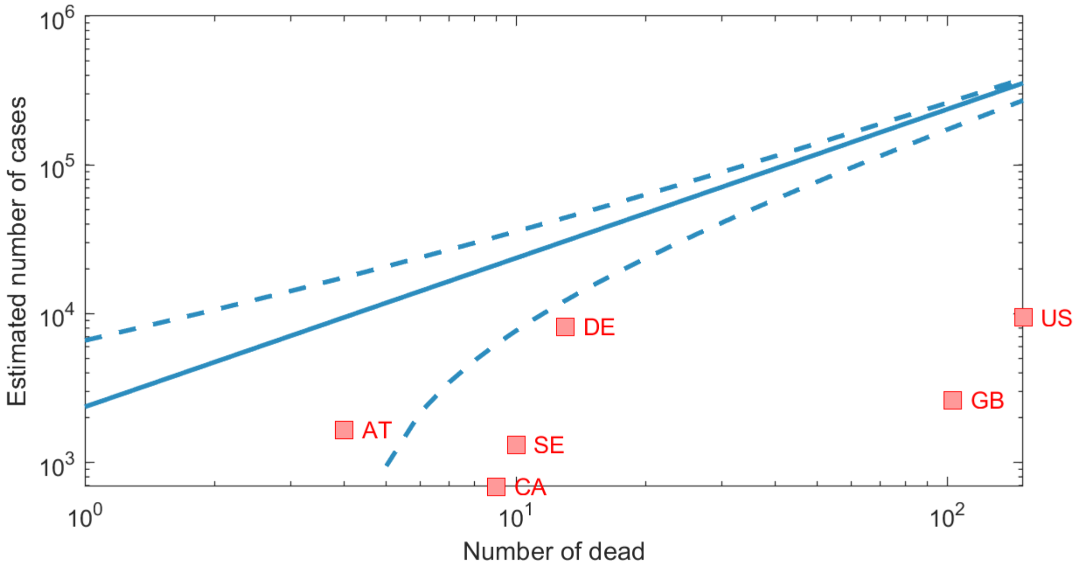

Figure 1 | Predicted number of COVID-19 cases based on the reported number of deaths. The predicted cumulative number of COVID-19 cases (solid line) depending on the number of reported deaths, with 95% indicative confidence intervals (dashed line), here based on a doubling time of 4 days and a case fatality rate of 0.8%. Square markers indicate the reported deaths and cases in selected countries based on World Health Organization data from March 19.