

Using a delay-adjusted case fatality ratio to estimate under-reporting

Status: in-progress | First online: 22-03-2020 | Last update: 23-03-2020

Authors: [Timothy W Russell](#)^{*}, Joel Hellewell¹, Sam Abbott¹, Christopher I Jarvis, Kevin van Zandvoort, Ruwan Ratnayake, [CMMID nCov working group](#), Stefan Flasche, Rosalind Eggo, W John Edmunds & Adam J Kucharski.^{*} corresponding author¹ contributed equally

This study has not yet been peer reviewed.

Aim

To estimate the percentage of symptomatic COVID-19 cases reported in different countries using case fatality ratio estimates based on data from the ECDC, correcting for delays between confirmation-and-death.

Methods Summary

- In real-time, dividing deaths-to-date by cases-to-date leads to a biased estimate of the case fatality ratio (CFR), because this calculation does not account for delays from confirmation of a case to death, and under-reporting of cases.
- Using the distribution of the delay from hospitalisation-to-death for cases that are fatal, we can estimate how many cases so far are expected to have known outcomes (i.e. death or recovery), and hence adjust the naive estimates of CFR to account for these delays.
- The adjusted CFR does not account for under-reporting. However, the best available estimates of CFR (adjusting or controlling for under-reporting) are in the 1% - 1.5% range. We assume a baseline CFR, taken from a large study in China, of 1.38% (95% crl: 1.23–1.53%)[1]. If a country has an adjusted CFR that is higher (e.g. 20%),

it suggests that only a fraction of cases have been reported (in this case, $\frac{1.38}{20} = 6.9\%$ cases reported approximately).

Current estimates for percentage of symptomatic cases reported for countries with greater than ten deaths

Figure
Table

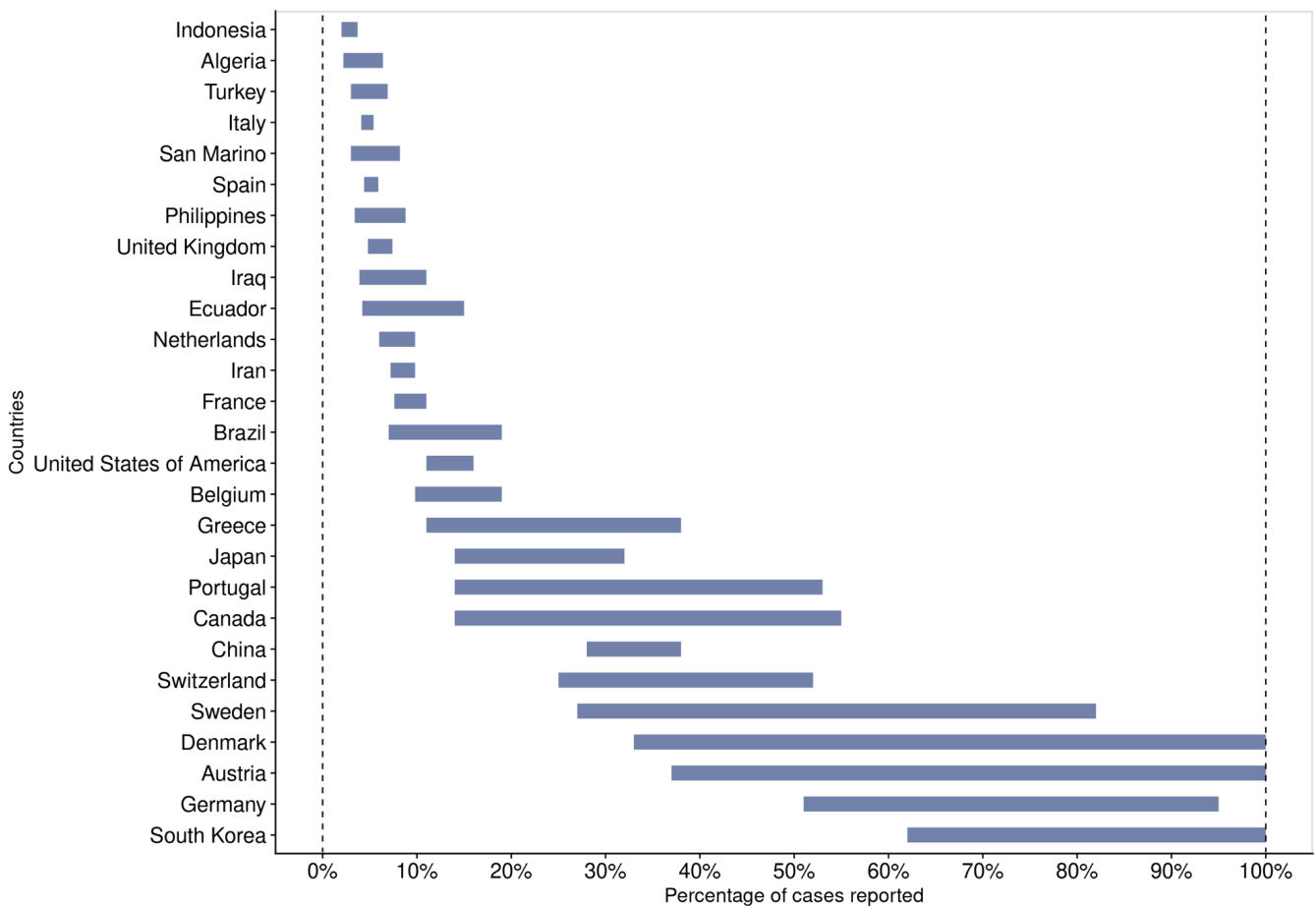


Figure 1: Plotting the estimates for the proportion of symptomatic cases reported in different countries using cCFR estimates. Dark blue shading is the 25% - 75% confidence range and light blue is 2.5% - 97.5% confidence range

Adjusting for outcome delay in CFR estimates

During an outbreak, the naive CFR (nCFR), i.e. the ratio of reported deaths date to reported cases to date, will underestimate the true CFR because the outcome (recovery or death) is not known for all cases [2]. We can therefore estimate the true denominator for the CFR (i.e. the number of cases with known outcomes) by accounting for the delay from confirmation-to-death [4].

We assumed the delay from confirmation-to-death followed the same distribution as estimated hospitalisation-to-death, based on data from the COVID-19 outbreak in Wuhan, China, between the 17th December 2019 and the 22th January 2020, accounting right-censoring in the data as a result of as-yet-unknown disease outcomes (Figure 1, panels A and B in [5]). The distribution used is a Lognormal fit, has a mean delay of 13 days and a standard deviation of 12.7 days [5].

To correct the CFR, we use the case and death incidence data to estimate the number of cases with known outcomes [3,4]:

$$u_t = \frac{\sum_{i=0}^t \sum_{j=0}^{\infty} c_{i-j} f_j}{\sum_{i=0}^t c_i},$$

where

u_t represents the underestimation of the known outcomes [2–4] and is used to scale the value of the cumulative number of cases in the denominator in the calculation of the cCFR,

c_t is the daily case incidence at time, t

and

f_t is the proportion of cases with delay of t between confirmation and death.

Approximating the proportion of symptomatic cases reported

At this stage, raw estimates of the CFR of COVID-19 correcting for delay to outcome, but not under-reporting, have been calculated. These estimates range between 1% and 1.5% [1,4,6]. We assume a CFR of 1.38% (95% crl 1.23% - 1.53%), taken from a recent large study [1], as a baseline CFR. We use it to approximate the potential level of under-reporting in each country. Specifically, we perform the calculation $\frac{1.38\%}{cCFR}$ of each country to estimate an approximate fraction of cases reported.

Limitations

Implicit in assuming that the under-reporting is

$$\frac{1.38\%}{cCFR}$$

for a given country is that the deviation away from the assumed 1.38% CFR is entirely down to under-reporting. In reality, burden on healthcare system is a likely contributing factor to higher than 1.38% CFR estimates, along with many other country specific factors.

The following is a list of the other prominent assumptions made in our analysis:

- We assume that people get tested upon hospitalisation. A few examples where this is not the case are Germany and South Korea, where people can get tested earlier.
- We assume that hospitalisation to death from early Wuhan is representative of the all the other countries (by using the distribution parameterised using early Wuhan data) and that all countries have the same risk and age profile as Wuhan.
- Severity of COVID-19 is known to increase with age. Therefore, countries with older populations will naturally see higher death rates. We are extending this analysis to adjust for the age distribution for countries with more than five reported deaths and where age distribution data is available.

- All results are linked and biased by the baseline CFR, assumed at 1.38% [1].
- The under-reporting estimate is very sensitive to the baseline CFR, meaning that small errors in it lead to large errors in the estimate for under-reporting.

Code and data availability

The code is publically available at

https://github.com/thimotei/CFR_calculation. The data required for this analysis is a time-series for both cases and deaths, along with the corresponding delay distribution. We scrape this data from [ECDC](#), using the NCoVUtils package [7].

References

- 1 Verity R, Okell LC, Dorigatti I *et al*. Estimates of the severity of covid-19 disease. *medRxiv* 2020.
- 2 Kucharski AJ, Edmunds WJ. Case fatality rate for ebola virus disease in west africa. *The Lancet* 2014;**384**:1260.
- 3 Nishiura H, Klinkenberg D, Roberts M *et al*. Early epidemiological assessment of the virulence of emerging infectious diseases: A case study of an influenza pandemic. *PLoS One* 2009;**4**.
- 4 Russell TW, Hellewell J, Jarvis CI *et al*. Estimating the infection and case fatality ratio for covid-19 using age-adjusted data from the outbreak on the diamond princess cruise ship. *medRxiv* 2020.
- 5 Linton NM, Kobayashi T, Yang Y *et al*. 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* 2020;**9**:538.
- 6 Guan W-j, Ni Z-y, Hu Y *et al*. Clinical characteristics of coronavirus

disease 2019 in china. *New England Journal of Medicine* 2020.

7 Abbott S MJ Hellewell J. NCoVUtils: Utility functions for the 2019-ncov outbreak. *doi:105281/zenodo3635417* 2020.