Perspective

Role of Ascertainment Bias in Determining Case Fatality Rate of COVID-19

Maia P. Smith[*], [iD]

Department of Public Health and Preventive Medicine, St. George's University School of Medicine, True Blue, Grenada, West Indies

ARTICLE INFO

ABSTRACT

Case fatality rate (CFR) is used to calculate mortality burden of COVID-19 under different scenarios, thus informing risk-benefit balance of interventions both pharmaceutical and nonpharmaceutical. However, observed CFR is driven by testing: as more low-risk cases are identified, observed CFR will decline. This report quantifies test bias by modeling observed CFR as log-log-linear function of test density (tests per population) in 163 countries. CFR declined almost 20% (e.g. from 5% to 4%) for each doubling of test density ($p < 0.0001$); this association did not vary by continent (interaction $p > 0.10$) although at any given test density CFR was higher in Europe or North America than in Asia or Africa. This effect of test density on observed CFR is adequate to hide all but the largest true differences in case survivorship. Published estimates of CFR should specify test density, and comparisons should correct for it such as by applying the provided model.

1. INTRODUCTION

Case Fatality Rate (CFR) is used to calculate mortality burden of COVID-19 under different scenarios and, thus, to inform the risk-benefit balance of interventions. These risks range from the financial (e.g. job loss) to the medical (e.g. missed prenatal care [1]) or even the emotional (e.g. impaired mood homeostasis during lockdown) [2]. Accurate estimates of CFR are thus essential to allow policymakers to balance risks against benefits.

However, observed CFR is driven by availability of testing. If tests are reserved for individuals at high risk of complications, many low-risk cases will be missed. Since asymptomatic or less-symptomatic cases often outnumber those with severe symptoms [3], observed CFR is likely to be inflated. Indeed, seroprevalence data [4] suggest that CFR in the US is usually below 1% and often below 0.50%, far below the 9.7% currently reported [5]. However, the seroprevalence study did not distinguish active cases of COVID-19 from recovered ones and thus missed any deaths that occurred after testing. Thus, true CFR of resolved cases of COVID-19 is unknown.

While ascertainment bias cannot be eliminated, its association with observed CFR can be quantified. This report does so by establishing the shape of the association between observed CFR and geography-level test density, and using this model to suggest a plausible range for the true CFR of COVID-19.

2. MATERIALS AND METHODS

2.1. Study Design and Data Collection

Using the software SAS (Cary, NC), I estimate associations between observed CFR in resolved cases of COVID-19, and test density, in every geography for which data were available and at least one death had occurred, from a publicly-available real-time counter [5] for July 3, 2020.

2.2. Data Availability

All data are publicly available.

2.3. Data Analysis

I calculate linear associations between CFR (ratio between total deaths and total resolved cases) and test density per million

*Email: Maia.phillips.smith@gmail.com

residents, weighted by total cases. CFR, test density, and total cases were each approximately log-normally distributed and thus were log-transformed for normality. Locally-weighted regression was used to estimate the shape of the association, identify any inflection points, and choose a statistical model.

Based on the observed shape of the association, a regression was run predicting CFR from test density. Exploratory analyses examined both confounding and effect modification by continent (Africa, Asia, Europe, North America, South America, and Oceania). Backward selection at $p = 0.10$ was used to create the final model.

3. RESULTS

In countries with complete data ($n = 163$) median CFR was 4.08% (25th and 75th percentiles, 2.13–6.93) Median CFR by continent ranged from 1.43 (Oceania) to 8.18 (North America). Median test density was 24,115 tests per million residents (25th and 75th percentiles, 6194–64,227).

Median CFR in the bottom quartile of test density was 5.36%; this dropped to 4.04% in the middle half, and 2.92 in the top quartile.

For each doubling of test density, CFR dropped by almost 20% ($p < 0.0001$, Table 1). The decline was near-linear on the log–log scale. For a given test density, CFR varied between continents by a factor of up to 3.9 (Asia vs. Europe; Table 1 and Figure 1) but its association with test density did not (p for interaction >0.10).

In the model containing an interaction between test density and continent, the estimated effect of test density was nonsignificantly stronger than that in the simple model presented here. (0.739 vs. 0.805). However, the interaction term was not statistically significant; and in this model the main effect of continent also lost significance, and most slope estimates were extreme and unstable. Thus the interaction was dropped in the final model.

The final model yielded the following regression equation:

$$CFR = 41.4 * 0.805^{\text{base 2 log of test density}} * \text{continent}$$

Table 1 | Case fatality rate of COVID-19 declines with test density $R^2 = 0.40$

	Parameter estimate	95% confidence interval		p
		Bottom	**Top**	
Intercept[a]	41.4	16.9	101	<0.0001
Test density, doubling	0.805	0.758	0.854	<0.0001
Continent[b]				<0.0001
Africa	1.033	0.696	1.533	0.87
Asia	1 (reference)	1	1	–
Europe	3.936	2.753	5.628	<0.0001
North America	3.427	2.236	5.252	<0.0001
Oceania	1.244	0.375	4.125	0.72
South America	2.366	1.458	3.838	0.0005
Interaction: test density * continent	(removed)			0.26

[a]Intercept corresponds to a test density of one test per million inhabitants. [b]p-values in the top row are for the global null (all continents equal). Those for each continent are pairwise comparison with the reference (Asia). Model fit using log–log regression and shown in Figure 1. Parameter estimates have been reverse-transformed for interpretability and are multiplicative, not additive.
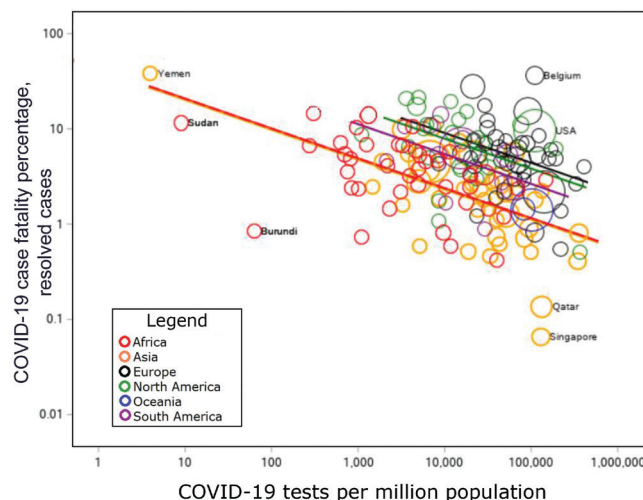

Figure 1 | Case fatality rate of COVID-19 declines with test density. Least-squares regression lines shown. Bubble size for each location corresponds to total cases. Selected points are labeled.

where 'continent' is a class-level nominal variable with 'Asia' as reference.

For a test density of 4096 (2^{12}) tests per million people, the model predicts the following CFR for Africa:

$$CFR = 41.4 * 0.805^{12} * 1.033$$
$$= 41.4 * 0.074 * 1.033$$
$$= 3.17\%.$$

For the same test density, the model predicts for Europe:

$$CFR = 41.4 * 0.805^{12} * 3.936$$
$$= 41.4 * 0.074 * 3.936$$
$$= 12.1\%.$$

If test density were to increase to 1,048,576 tests per million people (2^{20}; just over one per person) the predicted CFR for Europe would be 2.13%, and that for Africa would be 0.56%.

4. DISCUSSION

Observed CFR of COVID-19 declines with increasing test density, suggesting ascertainment bias. Published estimates of CFR, and thus of total mortality burden in different scenarios, may be too high. At a density of one test per person, this model predicts that CFRs will average about 0.5% in Asia and Africa and 2% in North America and Europe: higher than that estimated by Ioanniddis et al. [4] in the USA from seroprevalence but below most values in the current dataset.

As test density increases and re-testing becomes common, the decline in CFR with each additional test will likely plateau. However, no such plateau occurred within the range of these data. Based on this model, we suggest that saturation has not been reached and CFR may continue to decline. As testing becomes more available, many mild cases of COVID-19 will be found.

However, increases in testing may also reveal serious cases who did not have symptoms previously associated with COVID-19. Although initial screening focused on those with respiratory

symptoms, COVID-19 is now being shown to have potentially severe cardiovascular effects such as thromboembolism [6]. Those who died of these symptoms, such as by myocardial infarction or stroke, may not be recognized as COVID-19 cases, and thus CFR will be underestimated.

5. PUBLIC HEALTH IMPLICATIONS

Bias in estimated CFR will prevent accurate balancing of the risks and benefits of proposed interventions. This study suggests a true CFR between 0.5% and 2% for test densities not too far outside the range of our data. Bias can be minimized by combining the presented model with universal postmortem testing.

CONFLICTS OF INTEREST

The author declares no conflicts of interest.

REFERENCES

[1] Khalil A, von Dadelszen P, Draycott T, Ugwumadu A, O'Brien P, Magee L. Change in the incidence of stillbirth and preterm delivery during the COVID-19 pandemic. JAMA 2020;324;705–6.

[2] Taquet M, Quoidbach J, Fried EI, Goodwin GM. Mood homeostasis before and during the coronavirus disease 2019 (COVID-19) lockdown among students in the Netherlands. JAMA Psychiatry 2021;78;110–12.

[3] Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 Infection: a narrative review. Ann Intern Med 2020;173;362–7.

[4] Ioannidis JPA. Infection fatality rate of COVID-19 inferred from seroprevalence data. Bull World Health Organ 2021;99;19F–33F.

[5] Worldometers.info. COVID-19 Coronavirus Pandemic. 2020. Available from: https://www.worldometers.info/coronavirus/.

[6] Puntmann VO, Carerj ML, Wieters I, Fahim M, Arendt C, Hoffmann J, et al. Outcomes of cardiovascular magnetic resonance imaging in patients recently recovered from coronavirus disease 2019 (COVID-19). JAMA Cardiol 2020;5;1265–73.