**Explaining differences in reported COVID-19 fatality rates among countries: the hidden positivity rate explanation.**

**DRAFT VERSION- WORKING PAPER -20/04/2020**

**ABSTRACT**

Reported fatality rate (number of deaths/number of confirmed cases) due to COVID-19 virus greatly differs among and within countries. Using data on the number of tests for COVID-19 positivity we show that these discrepancies can be traced back, for the most part, to differences about the ratio between number of reported cases and number of actual “hidden” cases.

**Gabriele Pinto**\*

\*Ministry of Economics and Finance, Ragioneria Generale Dello Stato, Sapienza University of Rome, PhD student in Economics.

*The views expressed here are those of the authors and do not necessarily reflect the official policy or position of their affiliations.*

As is common knowledge, the fatality rate ratio (the ratio between number of confirmed cases and number of deaths),

,

due to COVID-19 cannot be informatively compared between countries as both the numerator (number of deaths cases) and the denominator (number of confirmed cases) might suffer from measurement error due to differences in the way countries identify the two cases. The issue[[1]](#footnote-1) is related to the way each country identifies the confirmed cases (the denominator), which is innately linked to the number of tests each country makes as well as the protocols applied to identify individuals to be tested. Since in most countries, individuals who more vulnerable, more at risk (such as physicians or health workers) or with evident symptoms are more likely to be tested – this inflates the estimate of the Fatality Rat

In other words, since the sample is not casual and suffers from a selection problem, the reported fatality rate is in most of the cases higher than the “actual” fatality rate (the fatality rate where at the denominator we have the “real” number of confirmed cases and not only those officially reported[[2]](#footnote-2)). This issue is also one of the reasons why policymakers and experts are considering the option of running serological tests on a randomly selected sample[[3]](#footnote-3).

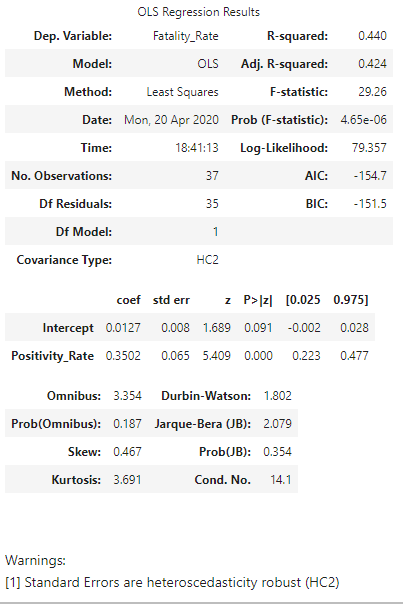
Before results from random samples for each country will be available, we propose a strategy to partially “clean” the fatality rate from the distortion due to the different ways countries identify positive cases. In fact, if the distortion due to a different proportion in the ratio between confirmed cases and “actual” cases is aligned in each country, the fatality rate could then be *more* informatively compared.

Our analysis is simple and intuitive. We take as a proxy of the ratio confirmed/actual cases, the ratio of the confirmed cases, over the total number of tests run in each country[[4]](#footnote-4). We collect the number of confirmed and deaths cases from the John Hopkins Resource Center (Du & Gardner, 2020), while the data on the number of tests are from Our World in Data (Joe Hasell, 2020).

To measure the distortion, we run a simple linear model as in:

,

where the estimated residuals will measure the remaining differences. Our estimation of the *“true”* fatality rate for country *c,* then will be given by the sum of the estimated intercept () and the estimated residual .

Our results report an estimate of the “baseline” fatality rate of 1.27 per cent, which is very close to the estimated fatality rate given by (Verity & Al., 2020) of 1.38 per cent. Almost half of the variation (R-squared= 0.44) in the reported fatality is captured by our proxy. To prove our strategy is robust, in the replication material attached we reported the same procedure applied to regional data from Italy.

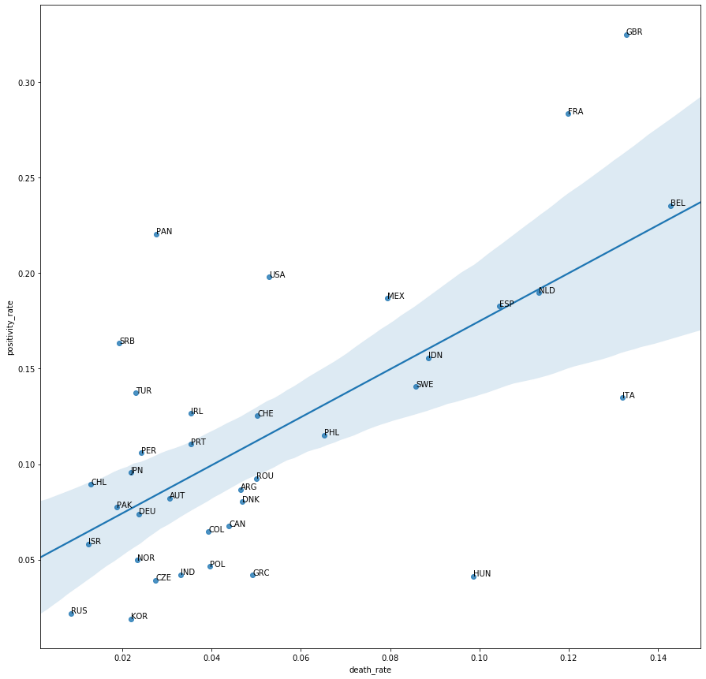


Figure 1: Scatterplot positivity rate vs Fatality Rate

Table 1: Regression Results

**Conclusion**

Using an extremely simple procedure, we showed how the differences in the death rates between countries due to distortions in the measurement can be partially cleaned using data on the number of people tested in each country. Our results suggest that this distortion can explain at least 44 per cent of the variation. The procedure used here can be useful for researchers using epidemic data on COVID-19 to capture part of this measurement error. If on the one side we advice researchers to be extremely careful when trying to explain the large differences we observe on the reported fatality rates between countries, we also propose an extremely simple strategy to tackle the issue. The procedure can also be improved in its capacity to absorb measurement error disturbances by adding more country-specific information.

Replication notebook and data available here: <https://github.com/gabrielepinto/politicalcovid19/tree/master/explaining_difference_fatality_rate>

# References

Du, D. E., & Gardner, H. (2020). An interactive web-based dashboard to track COVID-19 in real time. . *Lancet Infect Dis; published online Feb 19. https://doi.org/10.1016/S1473-3099(20)30120-1.*

Joe Hasell, E. O.-O. (2020, 04 20). *Our World in Data , COVID TESTING.* Retrieved from Our World in Data: https://ourworldindata.org/covid-testing

Verity, R., & Al. (2020). Estimates of the severity of coronavirus disease: a model-based analysis, published 30/03/2020 , https://doi.org/10.1016/S1473-3099(20)30243-7. *The Lancet*.

1. There are of course also problems with the numerator, as many fatalities due to COVID-19 might not be reported by national authorities for several reasons (e.g. limited capacity for autopsy). [↑](#footnote-ref-1)
2. E.g. the number of cases that would results from a “mass” population testing. [↑](#footnote-ref-2)
3. In some countries this has already been done. [↑](#footnote-ref-3)
4. [↑](#footnote-ref-4)