A Bayesian estimate of the COVID-19 infection fatality rate in Brazil based on a random seroprevalence survey

Valerio Marra[1,*] and Miguel Quartin[2,†]

[1] *Núcleo de Astrofísica e Cosmologia & Departamento de Física, Universidade Federal do Espírito Santo, Vitória, ES, Brazil*

[2] *Instituto de Física & Observatório do Valongo, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil*

(Dated:)

We infer the infection fatality rate (IFR) of SARS-CoV-2 in Brazil by combining three datasets. We compute the prevalence via the population-based seroprevalence survey EPICOVID19-BR, which tested 89000 people in 3 stages over a period of 5 weeks. This randomized survey selected people of 133 cities (accounting for 35.5% of the Brazilian population) and tested them for IgM/IgG antibodies making use of a rapid test. We estimate the time delay between the development of antibodies and subsequent fatality using the public SIVEP-Gripe dataset. The number of fatalities is obtained using the public Painel Coronavírus dataset. We obtain the IFR via Bayesian inference for each survey stage and 27 federal states. In particular, we include the effect of fading IgG levels by marginalizing over the time T after contagion at which the test gives a negative result. We adopt a flat broad prior on the interval $[40, 80]$ days. We infer a country-wide average IFR of 0.85% (95% CI: 0.76–0.99%).

The infection fatality rate (IFR) is one of the most important quantities of any new disease. An accurate estimate of both the case fatality rate (CFR) and IFR is thus usually a challenge before the end of a pandemic.[1] Nevertheless, the IFR has direct implications on the amount of resources and effort that should be allocated to prevent the spread of the disease and on steer policy-making in general. For instance using the United States as reference, Perlroth et al.[2] concluded that a CFR below 1% makes school-closures and social distancing not cost-effective.

In order to estimate the IFR one needs not only an estimate of the number of deaths, but also of the total infected population, and then to compare both within the same time period. It is, therefore, a difficult task as many cases are asymptomatic or develop only mild symptoms and are often unaccounted for. It is also hampered due to the lack of testing in many countries.[3]

The total number of deaths during an epidemic can be biased by the mislabeling of undiagnosed fatalities. To circumvent this possibility, one can rely on statistical estimates from the study of the excess deaths in a given period of time. In the case of COVID-19 this method is being pursued by many groups,[4–6] including the mainstream media,[7–9] as a method which is complementary to the officially reported numbers. However, this approach invariably suffers from important modeling uncertainties.[5] This may be especially true during the current pandemic which has seen an unprecedented amount of disruption of economic activity and social behavior, which includes a large fraction of the population undertaking social distancing measures.[10]

One of the first detailed analysis of the IFR of COVID-19 was based on around 70 thousand clinically diagnosed cases in China. After adjusting for demography and under-ascertainment Verity et al. arrived at the estimate of 1.38% (95% CI: 1.23–1.53%).[11] A similar study based only on the Wuhan province found a CFR of 1.4% (95% CI: 0.9–2.1%).[12] In France, a study recently modeled both death and hospital data and estimated the IFR to be 0.5% (95% CI: 0.3–0.9%).[13] Another model-based investigation arrived at an IFR of 0.8% (95% CI: 0.45–1.25%).[14] In Brazil, the focus of this work, the IFR was recently forecast with models. Results varied substantially between two different groups. A Brazilian team found that it should be much lower than the first estimates, around 0.3%.[15] On the other hand, a report by the group at Imperial College London estimated much higher values for the 16 Brazilian states they considered,[16] which, combined, suggest an overall IFR of 0.9%.

The incompatible estimates above highlight the inherent uncertainty in modeling a new disease that has caused such an unprecedented change in lifestyle worldwide. This is the main reason why one should rely on seroprevalence estimates in order to estimate the IFR of COVID-19. In a recent study, relying on antibody screening of blood donors, the IFR was estimated to be much lower, less than 0.21% at 95% CL.[17] Such an approach is, however, limited by the fact that blood donors may not be representative of the population. In particular all donors are younger than 70 and healthy. The ideal approach to circumvent the limitations above is to conduct random serology studies in the population. One such study – conducted in Geneva, Switzerland, with 2766 participants – found that for every reported COVID-19 case there were another 10.6 unreported ones,[18] a large discrepancy which again stresses the difficulties that models have to deal with. The same group reported an IFR of 0.64% (95% CI: 0.38–0.98%).[19] A much larger survey with 61075 participants was conducted in Spain, but IFR estimates were not reported.[20]

A meta-analysis of 36 seroprevalence studies performed

* Contributed equally

† Contributed equally; Corresponding author: mquartin@if.ufrj.br

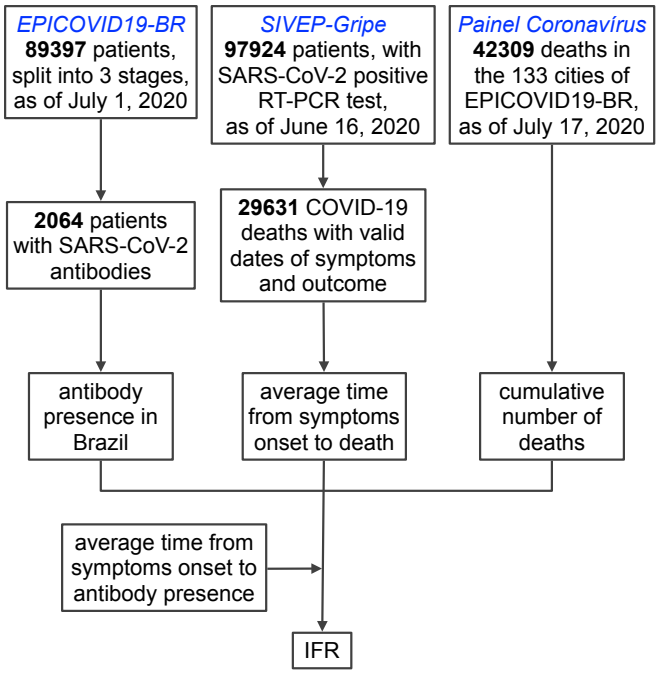

Figure 1. Flowchart of data used in this study.

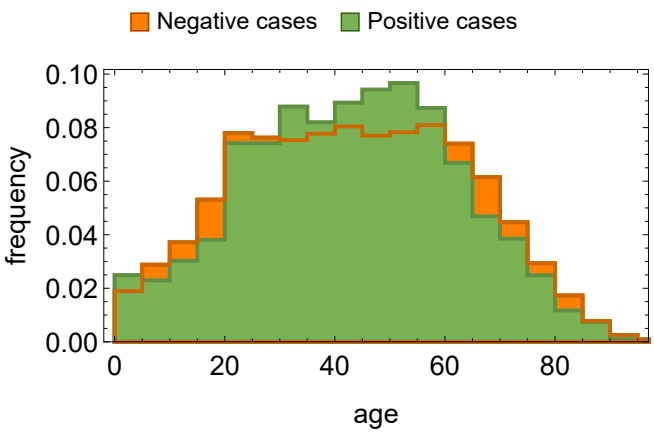

Figure 2. Age distributions relative to the EPICOVID19-BR survey (split according to positive and negative test results).

Table I. Results for Brazil, in percentages (maximum of probability distribution and 95%CI).

Brazil	date	antibody prevalence	IFR
Round 1	17 May 2020	2.59(2.03–3.21)	0.80(0.64–1.04)
Round 2	5 June 2020	3.78(3.22–4.40)	0.87(0.70–1.04)
Round 3	22 June 2020	3.80(3.27–4.37)	1.10(0.77–1.30)
All	–	–	0.85(0.76–0.99)

by Ioannidis[21] found that the IFR values ranged from 0.00% to 1.31%, and among 32 different locations the median IFR was 0.24%. Another meta-analysis of 25 IFR studies found an IFR of 0.68% (95% CI: 0.53–0.82%).[22] These results hint at a possible large variation in IFR values around the globe, although data from different countries were reported to be highly heterogeneous.

In Brazil, a large random seroprevalence study was performed by the EPICOVID19-BR team[23–25] which aimed to test 250 individuals in each of the 133 selected large sentinel cities. It has so far been carried out in 4 stages using the Wondfo lateral flow test for immunoglobulin M and G antibodies against SARS-CoV-2. Here, we consider data relative to the first three stages. The first stage was conducted between May 14 and 21, 2020, but did not reach its target number of samples, and in only 90 of the 133 cities at least 200 tests were performed. The total number of tests in all cities was 25025. Round 2 was conducted from June 4 to 7 and reached over 200 tests in 120 cities. Considering all cities a total of 31165 individuals were tested. Round 3 was performed between June 21 and 24 and made over 200 tests in all 133 cities for a total of 33207 tests. The total number of tests in all rounds was 89397, see Figure 1. Figure 2 depicts the age distributions for the first three rounds of the EPICOVID19-BR survey, split according to positive and negative test results. As can be seen, the survey tested people of all age groups and prevalence seems not to correlate with age: the mean ages of COVID-19 positive and negative Brazilians are 42.7 and 42.9 years, respectively.

The COVID-19 pandemic has strongly affected Brazil.[26] The federal government response has been heavily criticized,[27] and in September 2020 the number of confirmed cases and deaths crossed 4 million and 140 thousand, respectively, second only to the USA in the raw number of deaths. Furthermore, strong ethnic end regional variations in hospital mortality were found, casting doubts on the availability of public health care for the sections of society that cannot afford private care.[28] This daring situation motivates even further the need for an estimation of the IFR which is as accurate as possible in order to trigger an adequate political response to the crisis.

As summarized by Figure 1, in order to estimate the IFR we make use of three complementary datasets. We compute the percentage $p_a(t)$ of Brazilians that have been infected by SARS-CoV-2 at the city, state and Brazilian levels via the EPICOVID19-BR data. We robustly correct for false positive and negative rates and combine prevalences from different cities without neglecting the non-Gaussian nature of the distributions (details in the Supplementary Materials). The result is shown in Figure 3 and in Table I (the federal state acronyms explanation and full numerical tables can be found in the Supplementary Materials). We note a sharp increase in prevalence between rounds 1 and 2, and a subsequent stabilization between rounds 2 and 3. The state of Pará (PA) exhibits a sharp decrease in prevalence in round 3.

We obtain the number of fatalities via the public Painel

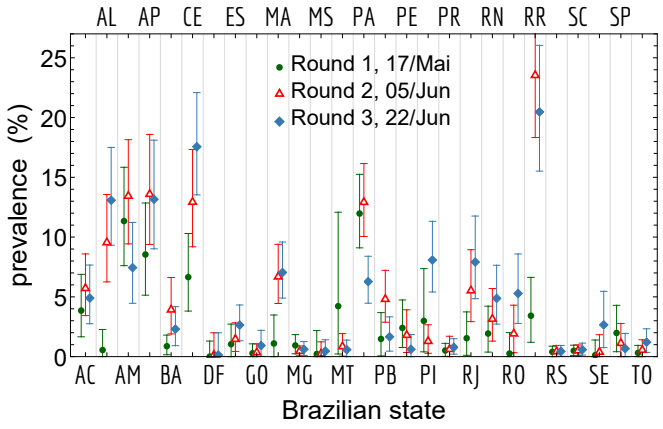

Figure 3. Prevalence (maximum posterior and 95% CI) of COVID-19 antibodies in each of the 27 Brazilian states in the 3 rounds of the EPICOVID19-BR survey.

Coronavírus dataset.[29] Painel Coronavírus is the Brazilian reference to keep track of the pandemic at the federal level and provides the deaths by COVID-19 with their geographic location. Long *et al.*[30] reported that IgG levels fade in recovered patients on a timescale of a few months, which was also suggested by the results relative to the first 2 rounds of EPICOVID19-BR.[25] Moreover, preliminary results from the recent fourth round of EPICOVID19-BR exhibit a large decrease in seroprevalence in the country,[31] which is consistent with a short window of detectability. For this reason we consider here a detectability window T and thus the number of fatalities relative only to such a window, which is equivalent to assume a sharp drop of IgG levels after T days. In order to take into account the uncertainty on T, we marginalize our results with respect to this parameter adopting a broad flat prior on the interval $[40, 80]$ days (details in the Supplementary Materials).

We cannot compute the IFR directly via the ratio of p_d and p_a because, at a given time $\bar{t}$, there are patients that developed antibodies but did not die yet from the disease.[32] In order to estimate the time delay τ_{ad} between the development of antibodies and subsequent fatality we use the public SIVEP-Gripe dataset ("Sistema de Informação da Vigilância Epidemiológica da Gripe"), a prospectively collected respiratory infection registry data that is maintained by the Ministry of Health for the purposes of recording cases of Severe Acute Respiratory Syndrome (SARS) across both public and private hospitals. The SIVEP-Gripe dataset contains the dates of symptoms onset and death for patients with SARS-CoV-2 positive RT-PCR test, together with their geographic location, which allow us to estimate the time delay τ_{sd} between the development of symptoms and subsequent fatality. We also make use of an empirical distribution between the first symptoms and the development of antibodies[33] to estimate the mean time-delay τ_{sa} between both events. Together, these estimates allow us to obtain the time-delay $\tau_{ad} \simeq \tau_{sd} - \tau_{sa}$. For the whole

Table II. Time scales used in the analysis.

Time scales for SARS-CoV-2	Mean (days)
τ_{cs} (contagion → symptoms)	5.5
τ_{sa} (symptoms → antibody)	5.8
τ_{ca} (contagion → antibody)	11.5
τ_{Δ} (symptoms → severe sympt.)	2
Time scales specific to Brazil	**Mean (days)**
τ_{sd}^{sivep} (severe sympt. → death)	14.1
τ_{sd} (symptoms → death)	16.1
τ_{cd} (contagion → death)	21.8
τ_{ad} (antibodies → death)	10.3

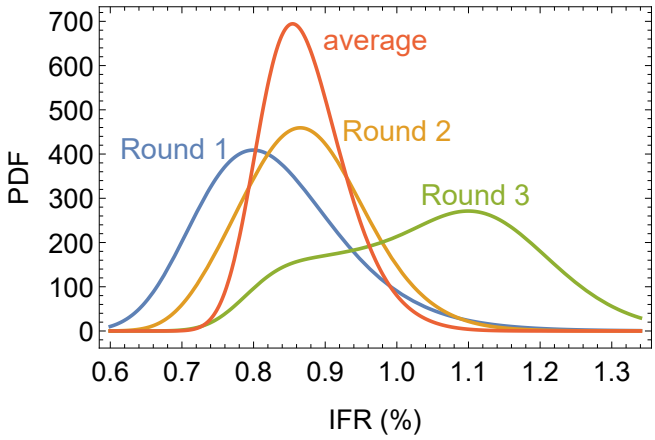

Figure 4. IFR posterior PDF for Brazil for each of the 3 rounds and all rounds combined.

Brazil we find $\tau_{ad} \simeq 10.3$ days. Table II summarize all the estimated time-delays which are used in our calculations (details in the Supplementary Materials).

Using this combined information we can then compute the IFR at the state and country levels:

$$\text{IFR} = \frac{p_d(\bar{t} + \tau_{ad})}{p_a(\bar{t})}, \tag{1}$$

where $\bar{t}$ is the time of a given EPICOVID19-BR phase. The results for Brazil are given in Table I and Figure 4, the ones for the states (combining all rounds) in Figure 5. We note significant statistical tension in the data of Roraima (RR). We, therefore, consider its IFR estimate unreliable, but due to its small population it has an insignificant impact on the IFR estimates at the country-level. The numerical results for all the states and for the three rounds separately can be found in the Supplementary Materials. The confidence intervals are computed by combining the statistical sources of error and including the non-Gaussian nature of the distributions. Since we marginalize the posterior on the IFR over the IgG fading time T, our estimation is robust against the uncertainty on T and the confidence interval includes the effect of the correlation between IFR and T.

Our overall estimate of the IFR of 0.85% (95% CI: 0.76–0.99%) is in agreement with some, but not all, of

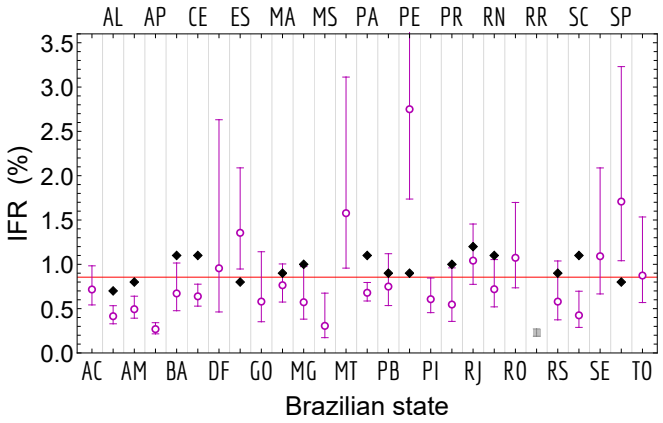

Figure 5. Combined IFR using all 3 rounds (maximum posterior and 95%CI). The black dots represent model-based results by the Imperial College COVID-19 Response Team.[16] The horizontal red line is the IFR estimate for Brazil given in Table I.

the previous world estimates discussed earlier. In particular, at the country level, our combined estimate agrees with the one by the Imperial College COVID-19 Response Team,[16] even though at the state level we find several disagreements between their values and our 95% CIs, see Figure 5.

Our estimate features a small 7% standard deviation including the uncertainty on the IgG fading time T, but it may suffer from the following systematic biases. First, not all COVID-19 related deaths may be registered in Painel Coronavírus. One expects this to happen for out-of-hospital fatalities and be stronger in the poorest areas with a less present health care infrastructure. As we are analyzing the 133 large sentinel cities that entered the EPICOVID19-BR survey this bias is not expected to be sizable. Its effect is, nonetheless, the underestimation of the IFR.

A second potential bias comes from the fact that the time in Painel Coronavírus is not the actual time of death but rather the time of notification. In order to alleviate this issue and also average out oscillations due to weekends, we smooth the $\mathrm{d}n_d/\mathrm{d}t$ data according to a forward 7-day moving average (details in the Supplementary Materials).

Third, the SIVEP-Gripe dataset is biased towards cases with severe symptoms. Indeed, there is a significant number of cases that are hospitalized when symptoms are notified (see Supplementary Materials). We took this into account via a delay parameter $\tau_\Delta = \tau_{sd} - \tau_{sd}^{\mathrm{sivep}} = 2 \pm 1$ days (see Table II) which models the time that a patient takes to go from symptoms onset to severe symptoms (details in the Supplementary Materials). Had we set $\tau_\Delta = 0$, we would have obtained for the IFR in Brazil a value of 0.83% (95% CI: 0.75–0.96%), an only 2% lower estimate.

Finally, the participants of the study may not be fully representative of the whole population of Brazil. Indeed the overall IFR we computed is relative to the 133 large cities that were tested by the EPICOVID19-BR survey. These cities amount to 35.5% of the Brazilian population, but one may speculate that the IFR could be different in smaller cities and rural or poorer areas.

It is well known that the IFR of COVID-19 depends on the patient's health and age,[34] and one expects significant country-by-country variations of the population IFR. Our IFR estimate should, therefore, be contextualized to the Brazilian population. To this end, a reasonable proxy for the overall health of a country is life expectancy, and the lower socioeconomic development of Brazil is reflected into a lower life expectancy as compared to, for example, Europe—76.0 years as compared with 80.9 years, as of 2017.[35,36]

As new medications and treatment protocols for the disease are discovered and become available it is hoped that the IFR will decrease. Since our data comes from the first months of the pandemic, our results therefore also set a baseline for future comparisons of the fight against COVID-19 in Brazil.

Concluding, we hope that our careful evaluation of the IFR in Brazil will help reinforce, at the federal, state and municipal levels, the seriousness of the COVID-19 pandemic and the urgency of taking the proper actions in order to reduce its societal and economic impact.

[1] B. Y. Lee, L. A. Haidari, and M. S. Lee, "Modelling during an emergency: the 2009 h1n1 influenza pandemic," *Clinical Microbiology and Infection* **19** no. 11, (2013) 1014–1022.

[2] D. J. Perlroth, R. J. Glass, V. J. Davey, *et al.*, "Health Outcomes and Costs of Community Mitigation Strategies for an Influenza Pandemic in the United States," *Clinical Infectious Diseases* **50** no. 2, (2010) 165–174.

[3] H. Ritchie, E. Ortiz-Ospina, D. Beltekian, *et al.*, "Coronavirus (COVID-19) Testing," 2020. https://ourworldindata.org/coronavirus-testing. (accessed August 7, 2020).

[4] D. M. Weinberger, J. Chen, T. Cohen, *et al.*, "Estimation of Excess Deaths Associated With the COVID-19 Pandemic in the United States, March to May 2020," *JAMA Internal Medicine* (07, 2020) .

[5] S. H. Woolf, D. A. Chapman, R. T. Sabo, D. M. Weinberger, and L. Hill, "Excess Deaths From COVID-19 and Other Causes, March-April 2020," *JAMA* **324** no. 5, (08, 2020) 510–513.

[6] S. Vandoros, "Excess mortality during the covid-19 pandemic: Early evidence from england and wales," *Social Science & Medicine* **258** (2020) 113101.

[7] J. Aron and J. Muellbauer, "A pandemic primer on excess mortality statistics and their comparability across countries," 2020. https:

//ourworldindata.org/covid-excess-mortality. (accessed July 21, 2020).

[8] The Economist, "Tracking covid-19 excess deaths across countries," 2020. https://www.economist.com/graphic-detail/2020/07/15/tracking-covid-19-excess-deaths-across-countries. (accessed July 21, 2020).

[9] J. Katz, D. Lu, and M. Sanger-Katz, "Tracking covid-19 excess deaths across countries," 2020. https://www.nytimes.com/interactive/2020/05/05/us/coronavirus-death-toll-us.html. (accessed July 22, 2020).

[10] Brazilian Ministry of Economy, "Impactos Econômicos da COVID-19 ," 2020. https://www.gov.br/economia/pt-br/centrais-de-conteudo/publicacoes/notas-informativas/2020/nota-impactos-economicos-da-covid-19.pdf/view. (accessed August 7, 2020).

[11] R. Verity, L. C. Okell, I. Dorigatti, *et al.*, "Estimates of the severity of coronavirus disease 2019: a model-based analysis," *The Lancet Infectious Diseases* **20** no. 6, (2020) 669–677.

[12] J. T. Wu, K. Leung, M. Bushman, *et al.*, "Estimating clinical severity of covid-19 from the transmission dynamics in wuhan, china," *Nature Medicine* **26** no. 4, (2020) 506–510.

[13] H. Salje, C. Tran Kiem, N. Lefrancq, *et al.*, "Estimating the burden of sars-cov-2 in france," *Science* **369** no. 6500, (2020) 208–211.

[14] L. Roques, E. K. Klein, J. Papaix, A. Sar, and S. Soubeyrand, "Using early data to estimate the actual infection fatality ratio from covid-19 in france," *Biology* **9** no. 5, (2020) 97.

[15] S. Dana, A. B. Simas, B. A. Filardi, *et al.*, "Brazilian modeling of covid-19 (bram-cod): a bayesian monte carlo approach for covid-19 spread in a limited data set context," *medRxiv* **2020.04.29.20081174** (2020) .

[16] T. A. Mellan, H. H. Hoeltgebaum, S. Mishra, *et al.*, "Report 21: Estimating covid-19 cases and reproduction number in brazil," *Imperial College London* (2020) 1–24.

[17] C. Erikstrup, C. E. Hother, O. B. V. Pedersen, *et al.*, "Estimation of SARS-CoV-2 infection fatality rate by real-time antibody screening of blood donors," *Clinical Infectious Diseases* (2020) .

[18] S. Stringhini, A. Wisniak, G. Piumatti, *et al.*, "Seroprevalence of anti-sars-cov-2 igg antibodies in geneva, switzerland (serocov-pop): a population-based study," *The Lancet* **396** no. 10247, (2020) 313–319.

[19] J. Perez-Saez, S. A. Lauer, L. Kaiser, *et al.*, "Serology-informed estimates of sars-cov-2 infection fatality risk in geneva, switzerland," *The Lancet Infectious Diseases* (2020) .

[20] M. Pollán, B. Pérez-Gómez, R. Pastor-Barriuso, *et al.*, "Prevalence of sars-cov-2 in spain (ene-covid): a nationwide, population-based seroepidemiological study," *The Lancet* (2020) .

[21] J. Ioannidis, "The infection fatality rate of covid-19 inferred from seroprevalence data," *medRxiv* (2020) .

[22] G. Meyerowitz-Katz and L. Merone, "A systematic review and meta-analysis of published research data on covid-19 infection-fatality rates," *medRxiv* **2020.05.03.20089854** (2020) .

[23] Epicovid19-BR, "Epicovid19," 2020. http://www.epicovid19brasil.org/. (accessed Sep 17, 2020).

[24] P. Hallal, F. Hartwig, B. Horta, *et al.*, "Remarkable variability in sars-cov-2 antibodies across brazilian regions: nationwide serological household survey in 27 states," *medRxiv* **2020.05.30.20117531** (2020) .

[25] Hallal, Pedro C and others, "SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys," *The Lancet Global Health* **0** (2020) 0.

[26] D. S. Candido, I. M. Claro, J. G. de Jesus, W. M. Souza, *et al.*, "Evolution and epidemic spread of sars-cov-2 in brazil," *Science* (2020) .

[27] T. Lancet, "Covid-19 in brazil:"so what?"," *Lancet (London, England)* **395** no. 10235, (2020) 1461.

[28] Baqui, Pedro and Bica, Ioana and Marra, Valerio and Ercole, Ari and van der Schaar, Mihaela, "Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study," *The Lancet Global Health* **8** (2020) 1018–26.

[29] Brazilian Ministry of Health, "Painel Coronavírus," 2020. https://covid.saude.gov.br. (accessed July 18, 2020).

[30] Q.-X. Long, X.-J. Tang, Q.-L. Shi, *et al.*, "Clinical and immunological assessment of asymptomatic sars-cov-2 infections," *Nature Medicine* (2020) 1–5.

[31] K. Gaythorpe, N. Imai, G. Cuomo-Dannenburg, *et al.*, "4ª fase do EPICOVID19 mostra desaceleração do coronavírus no Brasil," 2020. http://www.epidemio-ufpel.org.br/site/content/sala_imprensa/4-fase-do-epicovid19-mostra-desaceleracao-do-coronavirus-no-php?noticia=3149. (accessed September 19, 2020).

[32] D. Baud, X. Qi, K. Nielsen-Saines, D. Musso, L. Pomar, and G. Favre, "Real estimates of mortality following COVID-19 infection," *The Lancet infectious diseases* **20** (2020) 773.

[33] Q.-X. Long, B.-Z. Liu, H.-J. Deng, *et al.*, "Antibody responses to SARS-CoV-2 in patients with COVID-19," *Nature Medicine* **26** no. 6, (2020) 845–848.

[34] The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, "The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) – China, 2020," *China CDC Weekly* **2** (2020) 113.

[35] Brazilian Institute of Geography and Statistics, "Tábua completa de mortalidade para o Brasil – 2017," 2018. ftp://ftp.ibge.gov.br/Tabuas_Completas_de_Mortalidade/Tabuas_Completas_de_Mortalidade_2017/tabua_de_mortalidade_2017_analise.pdf. (accessed May 10, 2020).

[36] Eurostat, "Life expectancy at birth in the EU: men vs. women," 2017. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20190725-1. (accessed May 10, 2020).

ACKNOWLEDGEMENTS

MQ's research is partially supported by the Brazilian research agencies CNPq and FAPERJ. VM's research is partially supported by Brazilian research agency CNPq. **Funding:** This project has not received any funding. **Author contributions:** Both authors contributed equally for this work. **Competing inter-**

ests: The authors declare no competing interests. **Data and materials availability:** The Wolfram Mathematica codes and the data used for this work are available at: www.github.com/mquartin/covid19-ifr-br.