A Bayesian estimate of the COVID-19 infection fatality ratio 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*

We infer the infection fatality ratio (IFR) of SARS-CoV-2 in Brazil by combining three datasets. We compute the prevalence via the population-based seroprevalence survey EPICOVID19-BR. For the fatalities we obtain the absolute number using the public Painel Coronavírus dataset and the age-relative number using the public SIVEP-Gripe dataset. The time delay between the development of antibodies and subsequent fatality is estimated via the SIVEP-Gripe dataset. We obtain the IFR via Bayesian inference for each survey stage and 27 federal states. We include the effect of fading IgG antibody levels by marginalizing over the time after contagion at which the test gives a negative result with a flat prior on the interval $[40, 80]$ days. We infer a country-wide average IFR (maximum posterior and 95% CI) of 0.97% (0.82–1.14%) and age-specific IFR: 0.028% (0.024–0.036%) [$<$ 30 years], 0.21% (0.17–0.25%) [30–49 years], 1.06% (0.88–1.31%) [50–69 years], 2.9% (2.5–3.7%) [$\geq$ 70 years].

INTRODUCTION

The infection fatality ratio (IFR) – the ratio between the number of deaths from a disease and the number of infected individuals (irrespective of developing symptoms) – is one of the most important quantities of any new disease. An accurate estimate of the IFR is usually a challenge before the end of a pandemic, being subject to many possible sources of biases.[1,2] 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.[3] concluded that an IFR 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.[4]

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,[5–7] including the mainstream media,[8–10] as a method which is complementary to the officially reported numbers. However, this approach invariably suffers from important modeling uncertainties.[6] 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.[11]

One of the first detailed analyses 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%).[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] A meta-analysis of 25 IFR studies found an IFR of 0.68% (95% CI: 0.53–0.82%).[15] A compartmental model for IFR in China and Europe using surveillance data resulted in the following estimates: Hubei, China 2.9% (95% CI: 2.4%–3.5%); Switzerland 0.5% (95% CI: 0.4%–0.6%); Lombardy, Italy 1.4% (95% CI: 1.1%–1.6%).[16] Finally, relying on antibody screening of Danish blood donors – all younger than 70 and healthy – the IFR was estimated to be less than 0.21% at 95% CL.[17] 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, the focus of this work, the IFR was recently estimated 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%.[18] On the other hand, a report by the group at Imperial College London estimated much higher values for the 16 Brazilian states they considered,[19] which, combined, suggest an overall IFR of 0.9%.

The incompatible estimates above highlight the need for a careful consideration of the biases in the IFR estimates for COVID-19. A possible solution to mitigate biases is to base the estimation of the IFR on a large representative random serology study 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 unre-

* Contributed equally

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

Table I. Cumulative IFR for Brazil (maximum of probability distribution and 95%CI).

Brazil	date	antibody prevalence (%)	IFR (%)
Round 1	17 May 2020	2.62(2.04–3.23)	0.96(0.77–1.26)
Round 2	5 Jun 2020	3.79(3.22–4.40)	0.97(0.75–1.17)
Round 3	22 Jun 2020	3.81(3.27–4.39)	1.16(0.76–1.37)
All ages	all rounds	–	0.97(0.82–1.14)
$<$ 30 years	all rounds	–	0.028(0.024–0.036)
30–49 years	all rounds	–	0.21(0.17–0.25)
50–69 years	all rounds	–	1.06(0.88–1.31)
$\geq$ 70 years	all rounds	–	2.9(2.5–3.7)

ported ones,[20] 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%).[21] A much larger survey with 61075 participants was conducted in Spain, but IFR estimates were not reported.[22]

A large random seroprevalence study was performed in Brazil by the EPICOVID19-BR team[23–25] which aimed to test 250 individuals in each of the 133 selected large sentinel cities. It was an extension to the whole country of a smaller regional set of surveys in the Rio Grande do Sul state[26] and 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.

The COVID-19 pandemic has strongly affected Brazil.[27] The federal government response has been heavily criticized,[28] and in December 2020 the number of confirmed cases and deaths crossed 7 million and 180 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.[29] 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.

RESULTS AND DISCUSSION

Figure 1 summarizes how the three datasets used in this work where combined in order to estimate the IFR. Our estimations of the IFR for Brazil are given in Table I and Figure 2, while the ones for the individual states (combining all rounds) in Figure 3. We note significant statistical tension in the data of Roraima (RR) (details in the Supplementary Materials). 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 (see Materials and Methods).

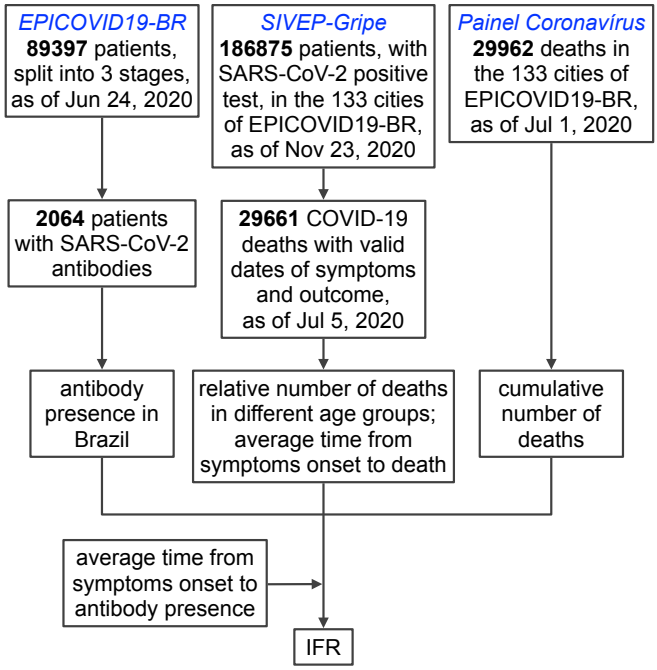

Figure 1. **Flowchart of data used in this study.**

Our overall estimate of the IFR of 0.97% (95% CI: 0.82–1.14%) is in agreement with some, but not all, of 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,[19] even though at the state level we find several disagreements between their values and our 95% CIs, see Figure 3.

It is well known that the IFR of COVID-19 strongly depends on age. As the Painel Coronavírus dataset does not provide age information, we adopt the Painel Coronavírus dataset for the absolute number of fatalities and the much richer but hospital-specific SIVEP-Gripe dataset for the relative number of fatalities for the var-

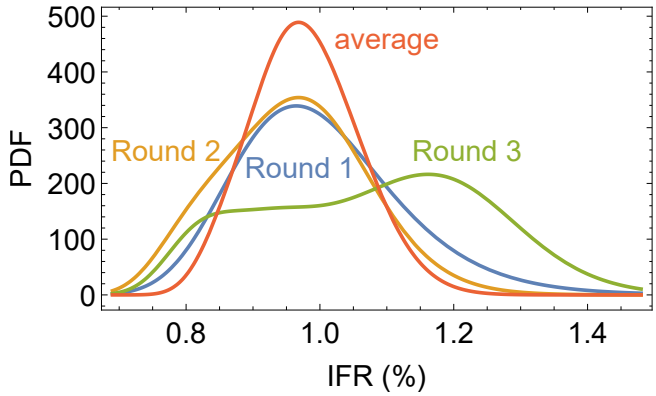

Figure 2. **IFR posterior PDF for Brazil for each of the 3 rounds and all rounds combined.**

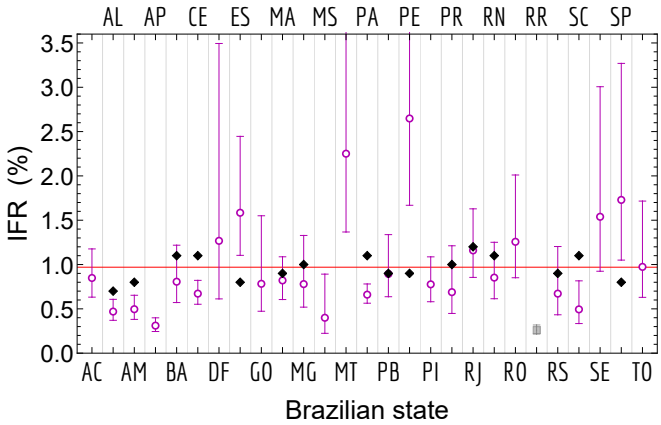

Figure 3. **Combined IFR using all 3 rounds**. Shown is the maximum posterior and 95%CI. The black dots represent model-based results by the Imperial College COVID-19 Response Team.[19] The horizontal red line is the IFR estimate for Brazil given in Table I.

ious age bins. We henceforth split the total population into 4 age groups (in years): <30, 30–49, 50–69 and ≥70. We find that not only the IFR but also the prevalence increases with age. For Round 3, for instance, we find $p_a = 3.7\%, 4.9\%, 6.4\%, 8.4\%$ for the age groups <30, 30–49, 50–69 and ≥70, respectively. For Round 1 and 2 the pattern is similar. The result for the IFR is summarized in Table I. The trend of a rapid increase of the IFR with age agrees well with previous findings[30,31] and urges authorities to promote measures to protect elderly people from SARS-CoV-2 exposure.

In Figure 4, instead of marginalizing over T, we show the Brazilian IFR as a function of T. For $T > 80$ days the results converge and the IFR remains unchanged. We also find that the effect of neglecting the time delay τ_{ad} between the development of antibodies and subsequent fatality results in an underestimate of the IFR of around 0.14%, roughly a 1.7 standard-deviation shift. For instance, the country-wide IFR for all ages changes from 0.97% (95% CI: 0.82–1.14%) to 0.83% (95% CI: 0.74–0.95%).

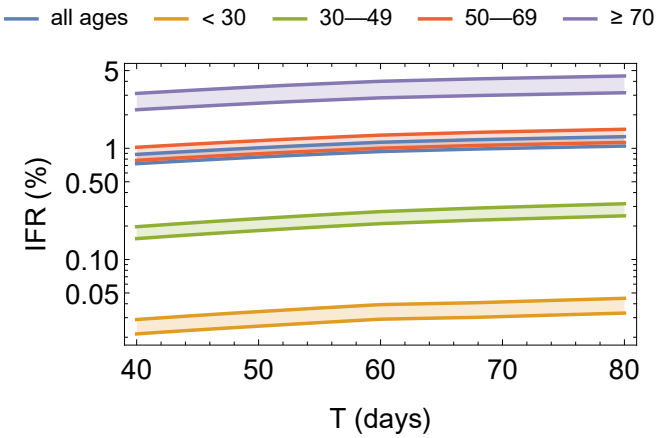

Figure 4. **IFR for Brazil for all rounds combined as a function of IgG fading time T and age.** Shown is the 95% CI using the logarithmic scale. For $T > 80$ the results converge and the IFR remains unchanged.

Our estimate features a small 8% 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. Some reports indeed claim a non-negligible number of undiagnosed respiratory deaths in 2020.[32] However, as we are analyzing the 133 large sentinel cities that entered the EPICOVID19-BR survey this bias is not expected to be significant. Its effect is, nonetheless, the underestimation of the IFR.

Regarding the computation of the IFR as a function of age, a potential bias could arise from the fact that the age distribution of the SIVEP-Gripe dataset may not be representative of the one of the overall population as some age groups may have a higher tendency to present themselves at the hospital and enter the SIVEP-Gripe dataset.

Another 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 20-day moving average which corresponds to shifting on average the time of deaths by 10 days earlier. This figure is justified by a direct comparison between Painel Coronavírus and SIVEP-Gripe and by estimates of the delay between time of death and notification[33,34] (details in the Supplementary Materials).

Next, 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 a relative 2% lower estimate.

Finally, the participants of the study may not be fully representative of the whole population of Brazil. The EPICOVID19-BR survey took several measures to mitigate enrollment bias, as discussed by Hallal et al.[25] For instance, for each city 25 census tracts were selected and in each one 10 households were selected at random. In each one a resident was also selected at random from a listing of members. If the selected individual refused participation another resident was selected at random and if they also refused (or if the residents were absent) the team moved into the neighboring household on the right. As discussed in the Supplementary Materials the age and ethnic distribution of the participants follow well the distributions in Brazil, with a small undersampling of under 20s and people of white ethnicity. Nevertheless since the survey considered only 133 large sentinel cities the overall IFR we computed is relative to these cities, which amount to 35.5% of the Brazilian population. One may nevertheless speculate that the IFR could be different in smaller cities and rural or poorer areas.

The IFR of COVID-19 depends not only on the patient's age, as discussed earlier, but also on their health.[35] 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.[36,37]

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.

MATERIALS AND METHODS

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[39] 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

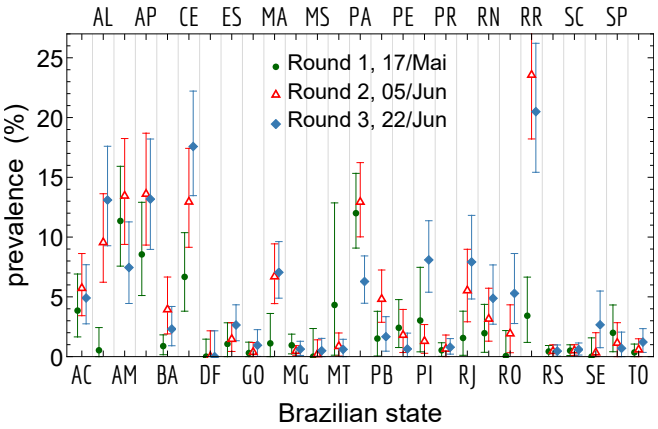

Figure 5. **Prevalence of COVID-19 antibodies in each of the 27 Brazilian states in the 3 rounds of the EPICOVID19-BR survey.** Shown is the maximum posterior and 95% CI. The state acronyms are AC: Acre, AL: Alagoas, AM: Amazonas, AP: Amapá, BA: Bahia, CE: Ceará, DF: Distrito Federal, ES: Espírito Santo, GO: Goiás, MA: Maranhão, MG: Minas Gerais, MS: Mato Grosso do Sul, MT: Mato Grosso, PA: Pará, PB: Paraíba, PE: Pernambuco, PI: Piauí, PR: Paraná, RJ: Rio de Janeiro, RN: Rio Grande do Norte, RO: Rondônia, RR: Roraima, RS: Rio Grande do Sul, SC: Santa Catarina, SE: Sergipe, SP: São Paulo, TO: Tocantins.

Figure 5 and in Table I (the 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 absolute number of fatalities via the public Painel Coronavírus dataset.[40] 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. Reports from Long *et al.*[41] and more recently by Grzelak *et al.*[42] show 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,[43] 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. The upper extremum of 80 days is justified by comparing the seroprevalence of the fourth and third rounds of EPICOVID19-BR (details in the Supplementary Materials). The lower limit was chosen in a conservative way as half of the upper limit.

Table II. Time scales used in the analysis (in days).

Time scales for SARS-CoV-2	mean	std. dev.
τ_{cs} (contagion $\to$ symptoms)	5.5	2.4
τ_{sa} (symptoms $\to$ antibody)	5.8	4.6
τ_{ca} (contagion $\to$ antibody)	11.5	5.2
τ_{Δ} (sympt. $\to$ severe sympt.)	2	1
Time scales specific to Brazil	**mean**	**std. dev.**
τ_{sd}^{sivep}(severe sympt. $\to$ death)	15.5	11.5
τ_{sd} (symptoms $\to$ death)	17.5	11.5
τ_{cd} (contagion $\to$ death)	23.2	11.7
τ_{ad} (antibodies $\to$ death)	9.7	10.5

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.[44] 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 and age, 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[45] 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 Brazil we find $\tau_{ad} \simeq 9.7$ 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 finally 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.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Code availability

The Wolfram Mathematica codes and the data used for this work are available at:
www.github.com/mquartin/covid19-ifr-br.

Acknowledgements

VM's research is partially supported by the Brazilian research agencies CNPq and FAPES. MQ's research is partially supported by the Brazilian research agencies CNPq and FAPERJ. This project has not received any funding.

Author contributions

VM and MQ both contributed equally to this work.

Competing interests

The authors declare no competing interests.

Supplementary Materials

Materials and Methods
Sections S1 to S7
Figs. S1 to S9
Table S1

[1] B. Y. Lee, L. A. Haidari, M. S. Lee, Modelling during an emergency: the 2009 H1N1 influenza pandemic. *Clinical Microbiology and Infection* **19**, 1014-1022 (2013).

[2] M. Lipsitch, C. .A. Donnelly, C. Fraser, I. M. Blake, A. Cori *et al.*, Potential Biases in Estimating Absolute and Relative Case-Fatality Risks during Outbreaks. *PLOS Neglected Tropical Diseases* **9**, 1-16 (2015).

[3] 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**, 165-174 (2010).

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

[5] 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* **180**, 1336-1344 (2020).

[6] S. H. Woolf, D. A. Chapman, R. T. Sabo, D. M. Weinberger, L. Hill, Excess Deaths From COVID-19 and Other Causes, March-April 2020. *JAMA* **324**, 510-513 (2020).

[7] S. Vandoros, Excess mortality during the Covid-19 pandemic: Early evidence from England and Wales. *Social Science & Medicine* **258**, 113101 (2020).

[8] J. Aron, J. Muellbauer, A pandemic primer on excess mortality statistics and their comparability across countries (2020); https://ourworldindata.org/covid-excess-mortality [accessed 21 July 2020].

[9] 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 21 July 2020].

[10] J. Katz, D. Lu, 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 22 July 2020].

[11] 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 7 August 2020].

[12] 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**, 669-677 (2020).

[13] H. Salje, C. Tran Kiem, N. Lefrancq *et al.*, Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208-211 (2020).

[14] L. Roques, E. K. Klein, J. Papaix, A. Sar, S. Soubeyrand, Using early data to estimate the actual infection fatality ratio from COVID-19 in France. *Biology* **9**, 97 (2020).

[15] G. Meyerowitz-Katz, L. Merone, A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *International Journal of Infectious Diseases*, **101**, 138-148 (2020).

[16] A. Hauser, M. J. Counotte, C. C. Margossian, G. Konstantinoudis, N. Low *et al.*, Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe. *PLOS Medicine* **17**, 1-17 (2020).

[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* 10.1093/cid/ciaa849 (2020).

[18] 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).

[19] T. A. Mellan, H. H. Hoeltgebaum, S. Mishra *et al.*, Report 21: Estimating COVID-19 cases and reproduction number in Brazil. *Imperial College London* 10.25561/78872 (2020).

[20] 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**, 313-319 (2020).

[21] 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* 10.1016/S1473-3099(20)30584-3 (2020).

[22] 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* 10.1016/S0140-6736(20)31483-5 (2020).

[23] Epicovid19-BR, EpiCovid-19 no Brasil (2020); http://www.epicovid19brasil.org/ [accessed 17 Sep 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] P. C. Hallal *et al.*, SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys. *The Lancet Global Health* **8**, 1390-1398 (2020).

[26] M. F. Silveira *et al.*, Population-based surveys of antibodies against SARS-CoV-2 in Southern Brazil. *Nature Medicine* **26**. 1196-1199 (2020).

[27] 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* **369**, 1255-1260 (2020).

[28] The Lancet, COVID-19 in Brazil:"So what?". *The Lancet* **395**, 1461 (2020).

[29] P. Baqui, I. Bica, V. Marra, A. Ercole, M. van der Schaar, Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *The Lancet Global Health* **8**, 1018-1026 (2020).

[30] M. O'Driscoll, G. Ribeiro Dos Santos, L. Wang *et al.*, Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* 10.1038/s41586-020-2918-0 (2020).

[31] A. T. Levin, W. P. Hanage, N. Owusu-Boaitey *et al.*, Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Eur J Epidemiol* **35**, 1123-1138 (2020).

[32] R. Lopes (Folha de São Paulo), Brasil já pode ter superado 220 mil mortes por Covid-19, mostra levantamento (2020); https://www1.folha.uol.com.br/equilibrioesaude/2020/12/brasil-ja-pode-ter-superado-220-mil-mortes-por-covid-19-mostra-levantamento.shtml [accessed 26 December 2020].

[33] G. Sá Pessoa (UOL), Cálculo de novas mortes por covid-19 inclui óbitos com 16 dias de atraso (2020); https://noticias.uol.com.br/saude/ultimas-

noticias/redacao/2020/04/09/calculo-de-novas-mortes-por-covid-19-inclui-obitos-com-12-dias-de-atraso.htm [accessed 27 November 2020].

[34] P. Duran (CNN), São Paulo usa dados de coronavírus com atraso de 10 dias; MS, com até 80 dias (2020); https://www.cnnbrasil.com.br/nacional/2020/08/15/governo-de-sp-usa-dados-de-coronavirus-com-atraso-de-10-dias-ms-com-ate-80-dias [accessed 27 November 2020].

[35] 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**, 113 (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 10 May 2020].

[37] 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 10 May 2020].

[38] Brazilian Institute of Geography and Statistics, Pesquisa Nacional de Saúde 2013 (2014); https://biblioteca.ibge.gov.br/visualizacao/livros/liv91110.pdf [accessed 27 November 2020].

[39] S. Greenland, Basic Methods for Sensitivity Analysis of Biases. *International Journal of Epidemiology* **25**, 1107-1116 (1996).

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

[41] Q.-X. Long, X.-J. Tang, Q.-L. Shi *et al.*, Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature Medicine* **26**, 1200-1204 (2020).

[42] L. Grzelak, A. Velay, Y. Madec *et al.*, Sex differences in the decline of neutralizing antibodies to SARS-CoV-2. *medRxiv* 2020.11.12.20230466 (2020).

[43] Epicovid19-BR, Quarta etapa da pesquisa nacional mostra desaceleração do coronavírus no Brasil (2020); http://www.epicovid19brasil.org/wp-content/uploads/2020/09/press_release_4.pdf [accessed 19 September 2020].

[44] D. Baud *et al.*, Real estimates of mortality following COVID-19 infection. *The Lancet infectious diseases* **20**, 773 (2020).

[45] 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**, 845-848 (2020).

[46] S. A. Lauer, K. H. Grantz, Q. Bi *et al.*, The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine* **172**, 577-582 (2020).

[47] H. P. C. Cintra and F. N. Fontinele, Estimative of real number of infections by COVID-19 in Brazil and possible scenarios. *Infectious Disease Modelling* **5**, 720-736 (2020).

[48] Imperial College COVID-19 response team, Short-term forecasts of COVID-19 deaths in multiple countries (2020); https://mrc-ide.github.io/covid19-short-term-forecasts [accessed 26 June 2020].

[49] K. Gaythorpe, N. Imai, G. Cuomo-Dannenburg *et al.*, Report 8: Symptom progression of COVID-19. *Imperial College London* 10.25561/77344 (2020).

[50] Brazilian Institute of Geography and Statistics, Population (2020); https://www.ibge.gov.br/en/statistics/social/population [accessed 10 Aug 2020].

[51] H. E. de Melker, N. J. D. Nagelkerde and M. A. E. Conyn-Van Spaendonck, Non-Participation in a Population-Based Seroprevalence Study of Vaccine-Preventable Diseases, *Epidemiology and Infection* Vol. 124, No. 2 (Apr., 2000), pp. 255-262