



Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil

William Marciel de Souza^{1,29}, Lewis Fletcher Buss^{2,29}, Darlan da Silva Candido^{2,3,29}, Jean-Paul Carrera^{1,3,4,29}, Sabrina Li^{5,29}, Alexander E. Zarebski^{1,3}, Rafael Henrique Moraes Pereira^{1,6}, Carlos A. Prete Jr⁷, Andreza Aruska de Souza-Santos^{1,8}, Kris V. Parag^{1,9}, Maria Carolina T. D. Belotti^{1,7}, Maria F. Vincenti-Gonzalez¹⁰, Janey Messina^{1,5,11}, Flavia Cristina da Silva Sales^{1,2}, Pamela dos Santos Andrade^{1,2}, Vitor Heloiz Nascimento^{1,7}, Fabio Ghilardi², Leandro Abade^{1,3}, Bernardo Gutierrez^{1,3,12}, Moritz U. G. Kraemer^{1,3,13,14}, Carlos K. V. Braga⁶, Renato Santana Aguiar¹⁵, Neal Alexander¹⁶, Philippe Mayaud¹⁷, Oliver J. Brady^{15,18}, Izabel Marcilio^{1,19}, Nelson Gouveia^{1,20}, Guangdi Li²¹, Adriana Tami¹⁰, Silvano Barbosa de Oliveira²², Victor Bertollo Gomes Porto^{1,22}, Fabiana Ganem²², Walquiria Aparecida Ferreira de Almeida^{1,22}, Francieli Fontana Sutile Tardetti Fantinato²², Eduardo Marques Macário²³, Wanderson Kleber de Oliveira²³, Mauricio L. Nogueira^{1,24}, Oliver G. Pybus^{1,3}, Chieh-Hsi Wu^{1,25,29}, Julio Croda^{1,23,26,27,28,29}✉, Ester C. Sabino^{1,29} and Nuno Rodrigues Faria^{1,2,3,9,29}✉

The first case of COVID-19 was detected in Brazil on 25 February 2020. We report and contextualize epidemiological, demographic and clinical findings for COVID-19 cases during the first 3 months of the epidemic. By 31 May 2020, 514,200 COVID-19 cases, including 29,314 deaths, had been reported in 75.3% (4,196 of 5,570) of municipalities across all five administrative regions of Brazil. The R_0 value for Brazil was estimated at 3.1 (95% Bayesian credible interval = 2.4–5.5), with a higher median but overlapping credible intervals compared with some other seriously affected countries. A positive association between higher per-capita income and COVID-19 diagnosis was identified. Furthermore, the severe acute respiratory infection cases with unknown aetiology were associated with lower per-capita income. Co-circulation of six respiratory viruses was detected but at very low levels. These findings provide a comprehensive description of the ongoing COVID-19 epidemic in Brazil and may help to guide subsequent measures to control virus transmission.

COVID-19 is a severe acute respiratory infection (SARI) that emerged in early December 2019 in Wuhan, China¹. The outbreak was declared a public health emergency of international concern by the World Health Organization on 30 January 2020. COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), an enveloped, single-stranded positive-sense RNA virus that belongs to the *Betacoronavirus* genus and Coronaviridae family². SARS-CoV-2 is closely related genetically to bat-derived SARS-like coronaviruses³. Human-to-human transmission occurs primarily via respiratory droplets and direct contact, similar to human influenza viruses, SARS-CoV and Middle East respiratory syndrome coronavirus⁴. The most commonly reported clinical symptoms are fever, dry cough, fatigue, dyspnoea, anosmia, ageusia, or some combination of these^{1,4,5}. As of 16 June 2020, more than 7.9 million cases have been confirmed worldwide, resulting in 434,796 deaths⁶.

Brazil declared COVID-19 a national public health emergency on 3 February 2020⁷. After the development of a national emergency plan and the early establishment of molecular diagnostic facilities across Brazil's network of public health laboratories, the country

reported its first confirmed COVID-19 case on 25 February 2020, in a traveller returning to São Paulo from northern Italy⁸. São Paulo is the largest city in South America and no other Brazilian city receives a greater proportion of international flights⁹. Currently, Brazil has one of the fastest-growing COVID-19 epidemics in the world, now accounting for 1,864,681 cases and 72,100 deaths, comprising over 55% of the total number of reported cases in Latin America and the Caribbean (as of 14 July 2020)⁶. About 21% of Latin American and Caribbean populations are estimated to be at risk of severe COVID-19 illness¹⁰. The region has been experiencing large outbreaks, with growing epidemics in Brazil, Peru, Mexico, Chile, Colombia, Panama and possibly Venezuela and Nicaragua, amid growing concerns about testing capacity for COVID-19 (refs. ^{11–14}). Preparedness for laboratory surveillance of SARS-CoV-2 in Latin America is centred around a network of national reference influenza surveillance laboratories that is facing several challenges, including a shortage of reagents and equipment¹⁵.

Conscious of the challenges associated with surveillance since the beginning of the epidemic in Brazil, here we focus on two main objectives. First, we contextualize the Brazilian SARS-CoV-2

A full list of affiliations appears at the end of the paper.

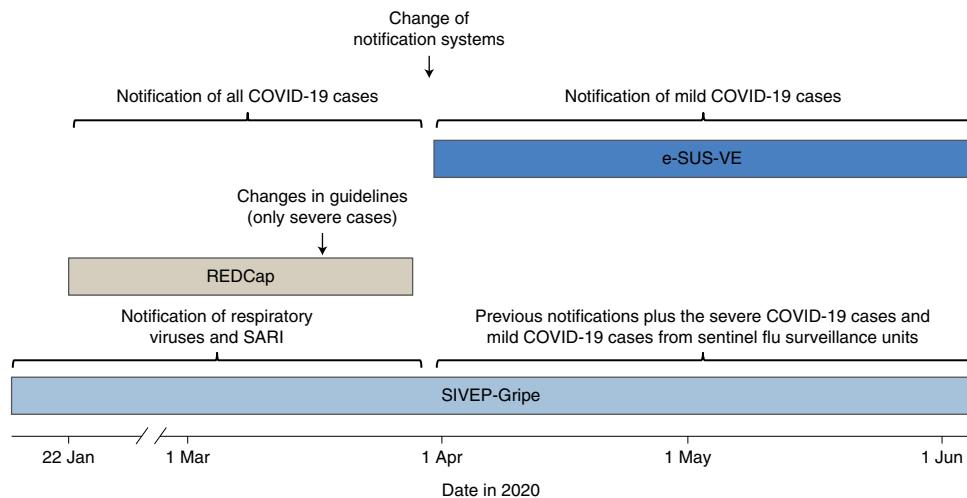


Fig. 1 | Timeline of national COVID-19 reporting systems in Brazil. The REDCap system operated between late January and 25 March 2020. Aggregated numbers from e-SUS-VE and SIVEP-Gripe data for mild and hospitalized COVID-19 cases, respectively, are updated on a daily basis on the Portal do COVID-19 website (<https://covid.saude.gov.br/>).

epidemic by comparing local transmission dynamics with those observed in other selected countries. Second, we use geospatial data related to confirmed COVID-19 cases and SARI cases with unknown aetiology to evaluate the relationship between socioeconomic factors and COVID-19 distribution.

Results

Contextualizing COVID-19 data reporting systems in Brazil. On 22 January 2020—more than 1 month before the first case in Brazil—the Brazilian Ministry of Health implemented the REDCap platform to report prospective suspected, probable and confirmed COVID-19 cases (see Methods for case definitions), as part of an early response to the pandemic¹⁶. By 27 March 2020, the REDCap system was discontinued (Fig. 1). Since then, mild COVID-19 cases started to be reported on e-SUS Vigilância Epidemiológica (e-SUS-VE), a new national COVID-19 reporting system, and hospitalized COVID-19 cases started to be recorded on a pre-existing Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP-Gripe) system. The SIVEP-Gripe system has been in use since 2009 (having been implemented in response to the 2009 influenza H1N1 pandemic) and has since centralized the reporting of respiratory viruses and SARI for the Brazilian Ministry of Health (Fig. 1). Both e-SUS-VE and SIVEP-Gripe include suspected and confirmed COVID-19 cases as reported by public health and private services (primary and emergency care). These two reporting systems (e-SUS-VE and SIVEP-Gripe) are inter-related on the Portal do COVID-19 website (<https://covid.saude.gov.br/>), which summarizes daily the aggregated counts from both platforms.

SARS-CoV-2 reporting in Brazil. We analysed a total of 514,200 SARS-CoV-2 cases from the Portal do COVID-19 website (SIVEP-Gripe and e-SUS-VE databases combined) that were confirmed by molecular diagnostic and clinical epidemiological criteria by 31 May 2020 (see Methods). Cases were reported in 75.3% (4,196 of 5,570) of municipalities across all five administrative regions of Brazil and included 206,555 (40.2%) recovered patients and 29,314 fatal (17.5%) COVID-19 cases (Fig. 2a). We further analysed a total of 1,468 confirmed cases from the REDCap system, including 342 imported cases with associated travel history information. After excluding individuals who travelled to multiple countries before entering Brazil ($n=56$) and who had an unknown country of origin ($n=16$), the self-reported countries of infection for cases

acquired abroad until 19 March 2020 were the United States (28.6%; $n=76$), Italy (24.4%; $n=65$), the United Kingdom (10.5%; $n=28$) and Spain (8.3%; $n=22$) (Extended Data Fig. 1). The first reported case (SPBR1) was reported on 25 February 2020 in the municipality of São Paulo, the fourth most populous urban area worldwide. Following the first reports of COVID-19 in Brazil's largest population centres, SARS-CoV-2 subsequently spread to municipalities with smaller population sizes (Fig. 2b). Until 31 May 2020, most confirmed cases and deaths were reported in the states of São Paulo (109,698 cases and 7,615 deaths), Rio de Janeiro (53,388 cases and 5,344 deaths), Ceará (48,489 cases and 3,010 deaths) and Amazonas (41,378 cases and 2,052 deaths), which together account for 49.2% of all cases and 61.5% of deaths in Brazil (Fig. 2c).

Basic reproduction number of SARS-CoV-2 in Brazil and comparison countries. To estimate the basic reproduction number (R_0) of SARS-CoV-2 in Brazil, daily confirmed cases in São Paulo, Rio de Janeiro, Ceará and Amazonas states were compiled from Ministry of Health data (for specification of the time windows used in the analyses, see Extended Data Fig. 2). For comparison, we compiled time series of confirmed cases in several European countries from the Johns Hopkins Coronavirus Resource Center (<https://coronavirus.jhu.edu/>; see also Extended Data Fig. 3). We found that São Paulo, Rio de Janeiro and Amazonas were characterized by similar R_0 values of 2.9 (95% Bayesian credible interval (BCI)=2.2–5.1), 2.9 (95% BCI=2.2–4.9) and 2.6 (95% BCI=2.0–4.5), respectively. However, for Ceará, the estimated R_0 was considerably lower at 1.9 (95% BCI=1.5–3.0) (Fig. 3 and Extended Data Fig. 1). This finding could be a result of the small window between the first reported cases and the early implementation of non-pharmaceutical interventions (NPIs) in this state (Supplementary Table 1 and Extended Data Fig. 2). On a national scale, the estimated R_0 for Brazil was slightly higher than that of the Brazilian states considered in this study, with a median of 3.1 (95% BCI=2.4–5.5), and also slightly higher than R_0 values estimated for other severely affected countries: Spain (2.6; 95% BCI=2.0–4.6); France (2.5; 95% BCI=1.9–4.4); the United Kingdom (2.6; 95% BCI=2.0–5.1); and Italy (2.5; 95% BCI=2.0–4.4) (Fig. 3). While the incidence curves for European countries have consistently flattened and declined since the implementation of NPIs (suggesting that the R_0 value has fallen below 1), Brazil's daily incidence curve has continued to increase (Fig. 2a and Extended Data Fig. 4).

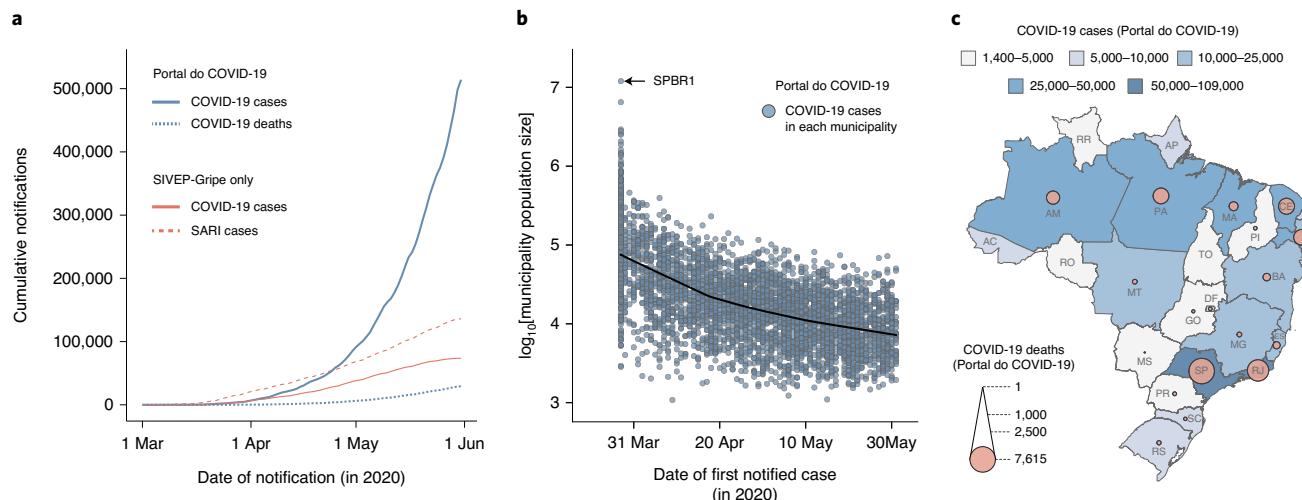


Fig. 2 | COVID-19 epidemiology in Brazil. **a**, Numbers of COVID-19 cases (blue solid line) and deaths (blue dashed line) reported to the Ministry of Health (Portal do COVID-19 website), along with numbers of COVID-19 confirmed cases (salmon solid line) and cases of SARI with unknown aetiology (salmon dashed line) reported to the SIVEP-Gripe database. **b**, First COVID-19 cases by date and Brazilian municipal population size based on the Ministry of Health data, from 28 March 2020. Each circle represents the first confirmed COVID-19 case in the municipality ($n = 4,196$ Brazilian municipalities). SPBR1 is the first detected SARS-CoV-2 infection in Brazil¹⁸. **c**, Map coloured according to the number of confirmed COVID-19 cases per state reported to the Ministry of Health (Portal do COVID-19 website). Circle sizes are proportional to the number of reported COVID-19 deaths in each federal unit. 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.

SARIs mostly reflect COVID-19 cases. In the early phase of the COVID-19 epidemic in Brazil, we analysed the results for other respiratory pathogens tested in Brazil as part of a differential diagnosis by the Central Public Health Laboratories and National Influenza Centres (Brazilian Ministry of Health), obtained from a REDCap platform¹⁷ designed for COVID-19. The respiratory viruses most frequently identified between 7 January 2020 and 27 March 2020, in patients with a suspected but negative diagnosis of COVID-19, were influenza A virus (347 (14.3%) of 2,429 tested cases), influenza B virus (251 (10.3%) of 2,429) and human rhinovirus (136 (5.6%) of 2,429). We found co-detection of SARS-CoV-2 with six other respiratory viruses, the most frequent of which were influenza A (11 (0.5%) of 2,429) and human rhinovirus (6 (0.2%) of 2,429) (Extended Data Fig. 7).

The SIVEP-Gripe system started reporting hospitalized COVID-19 cases in early March 2020 (epidemiological week 10) (Fig. 4). In this system, the number of tested cases is unavailable. We found that the peak of influenza confirmed cases ($n = 447$) occurred at epidemiological week 12 (15–21 March 2020). During the same week 12, we detected an 8.5-fold increase in total cases attributed to SARS-CoV-2 ($n = 3,789$) and a 9.9-fold increase in total cases reported as SARI with unknown aetiology ($n = 4,424$) (Fig. 4). From 2 January to 31 May 2020, a total of 2,136 influenza cases and 272 cases caused by other respiratory pathogens, including human respiratory syncytial virus, human rhinovirus, adenovirus and metapneumovirus, were reported in the SIVEP-Gripe database. The low observed incidence of influenza and other respiratory viruses may have been influenced by limited testing for these viruses during this period. Although NPIs may have an impact in reducing influenza virus transmission, this does not necessarily reflect a lower co-circulation of other respiratory viruses¹⁸.

Socioeconomic differences are associated with COVID-19 diagnosis. Until 31 May 2020, a total of 73,648 COVID-19 confirmed cases and 168,001 SARI cases with unknown aetiology were reported in the SIVEP-Gripe system. We hypothesized that the 2.3-fold

increase of SARI cases with unknown aetiology was associated with differential access to healthcare due to socioeconomic factors.

We focused on the Metropolitan Region of São Paulo (MRSP), which has a population of 23 million inhabitants across six sub-regions (Central, West, North, East, Southeast and Southwest) and 39 municipalities (Fig. 5a). To test this hypothesis, we obtained per-capita income at the census tract level (typically 150–300 households) in the MRSP, based on the residential address of each case. We then linked this information to each patient's final diagnosis outcome: confirmed case of COVID-19 or SARI with unknown aetiology. While the income distribution of SARI cases with unknown aetiology was similar to that of all residents of the MRSP over the whole period (Fig. 5b), we observed that the income distribution of individuals with COVID-19 confirmed by laboratory and clinical criteria was initially higher than that of all MRSP residents and decreased over time towards similar levels by epidemiological week 21 (Fig. 5b). Importantly, we found that the log odds of one or more confirmed COVID-19 cases per census tract increased with per-capita income in epidemiological weeks 12 and 22 (likelihood ratio test P value < 0.001 ; Fig. 5b and Supplementary Table 2). This provides statistical evidence of an association between confirmed COVID-19 diagnosis and per-capita income, suggesting a socioeconomic difference in access to COVID-19 diagnosis in the MRSP. For reference, we also provide a map of per-capita income (Fig. 5a) and population density in each census tract (Extended Data Fig. 8).

We conducted a geospatial analysis to understand the distribution of relative risk of observing a COVID-19 case or SARI case with unknown aetiology in the MRSP, using a Bayesian method and adjusted for spatial and non-spatial effects as defined by the Besag–York–Mollie model¹⁹ (Fig. 5). Our estimates show an increase in the relative risk of COVID-19 diagnosis in higher-income census tracts between epidemiological weeks 12 and 21, especially in the central region of the MRSP (Fig. 5a,c). We observed a similar trend in the relative risk of SARI cases with unknown aetiology among residents of the central region. However, there was also an increased probability

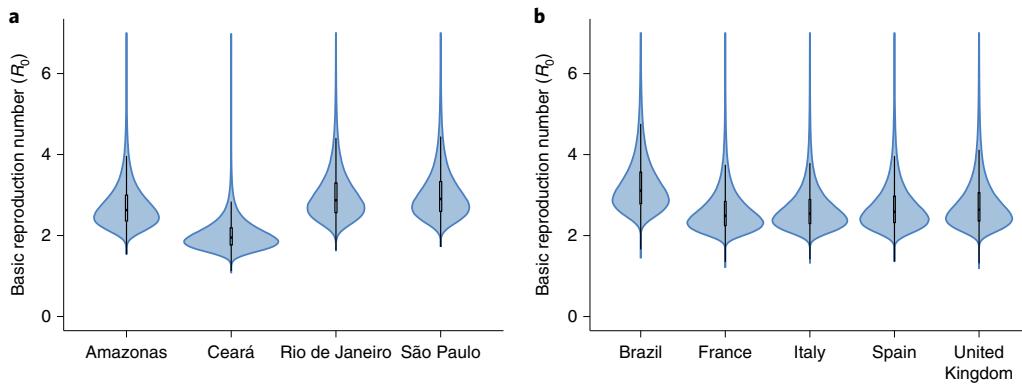


Fig. 3 | Estimated R_0 values for four Brazilian states and selected countries. Left: R_0 values for the Amazonas, Ceará, Rio de Janeiro and São Paulo states. Right: R_0 for Brazil, France, Italy, Spain and the United Kingdom. Violin plots of posterior samples for the basic reproduction number, the box plots show the median, first, and third quartiles. The whiskers extend to the most extreme value less than 1.5 times the interquartile range beyond the quartile. The daily numbers of infections used in each analysis can be found in Extended Data Figs. 3 and 4. Daily numbers of infections and prior distributions can be found in Extended Data Figs. 5 and 6.

of SARI cases with unknown aetiology in the southwest, west, north and south sub-regions, where income per capita is typically lower. Overall, the relative risk of SARI cases with unknown aetiology is more spatially widespread in the MRSP than that of confirmed COVID-19 cases (Fig. 5c).

The relative risk of SARI cases with unknown aetiology compared with confirmed COVID-19 cases in the central region of the MRSP decreased through time, probably as a response to several NPIs implemented throughout the state of São Paulo (see Supplementary Table 1). By week 16 (1 month after the start of the NPIs in São Paulo), we detected an increased risk particularly of SARI cases with unknown aetiology outside the central region of the MRSP, especially in the southwest region. SARI cases with unknown aetiology risk were also high in the east region. By week 21, the risk remained high throughout the central region and the risk of SARI cases with unknown aetiology decreased in the east region, possibly as a result of interventions targeting the reduction of SARS-CoV-2 transmission.

Demographics and characteristics of COVID-19 hospitalized and fatal cases in Brazil. Analysis of the age–sex structure of 67,180 confirmed COVID-19 cases reported on the SIVEP-Gripe system revealed a high proportion (44,027 (65.5%) of 67,180) of confirmed COVID-19 infections in middle- or older-aged individuals (≥ 50 years of age) and a lower proportion (1,454 (2.2%) of 67,180) in younger age groups (≤ 20 years of age) (Fig. 6a). The median age was 59 years (interquartile range = 44–72). The majority (38,654 (57.5%) of 67,180) were male. Similarly, 59% (14,498 of 24,519) of COVID-19 deaths were in men, and 85% (20,916 of 24,519) were in people aged ≥ 50 years. A total of 2.95% (1,983 of 67,180) cases were reported as nosocomial transmission, defined as a COVID-19 case acquired after hospitalization. Overall, 116 newborns (≤ 1 month old), 381 infants (≥ 1 –12 months old), 518 children (≥ 1 –12 years old) and 258 adolescents (≥ 12 –17 years of age) were diagnosed with COVID-19. In addition, 740 patients were pregnant (61 in the first trimester, 172 in the second trimester, 447 in the third trimester and 60 with missing gestational age).

By 31 May 2020, 91% (67,042 of 73,649) of patients with COVID-19 reported in the SIVEP-Gripe system had been hospitalized. Of these, 30.3% (22,332 of 73,649) were admitted to an intensive care unit (ICU). The median length of ICU stay for patients with COVID-19 was 5 d (interquartile range = 2–10 d; range = 0–65 d), based on the ICU admission and discharge dates of 8,240 confirmed cases. Most symptoms reported by patients with COVID-19

were a cough (56,681 (85.2%) of 66,514 without missing data), fever (51,312 (79.6%) of 65,310) and dyspnoea (51,312 (76.6%) of 65,310) (Fig. 6b). These three symptoms comprise part of the case definition of SARI in Brazil. In addition, 68% (40,806 of 60,400) of individuals with COVID-19 were hypoxic (O_2 saturation $< 95\%$), reflecting the overall severity of cases reported on SIVEP-Gripe (as shown in Fig. 1). The most prevalent comorbidities were cardiovascular disease (23,085 (66.5%) of 34,693 without missing data) and diabetes (17,271 (54.5%) of 31,672) (Fig. 6a). Among the patients with COVID-19, older age groups tended to have a higher proportion of comorbidities than younger age groups in different outcomes (Fig. 6c). The proportions of the general Brazilian population with cardiovascular disease and diabetes are 4.2 and 6.2%, respectively²⁰. A total of 83.7% (17,921 of 21,414 with complete comorbidity information) of individuals with confirmed COVID-19 had at least one comorbidity (see Supplementary Table 2 for information on data completeness).

Discussion

While the COVID-19 epidemic in Brazil continues to grow, details of its transmission potential and clinical and epidemiological characteristics remains poorly understood. We estimate a higher median transmission potential (R_0) of SARS-CoV-2 of 3.1 (2.4–5.5) in Brazil compared with Italy, the United Kingdom, France, and Spain, which have point estimates of R_0 varying from 2.5–2.6; however, the credible intervals overlap substantially. We have also observed rapid spread of COVID-19 through the country, with more populated and better-connected municipalities being affected earlier, and less populated municipalities being affected at a later stage of the epidemic. In the São Paulo metropolitan region, we found a higher risk of diagnosed COVID-19 cases in census tracts with higher per-capita income during the early phase of the COVID-19 epidemic but also as the weeks progressed. This contrasts with the wider spread of SARI cases among sub-regions with lower per-capita income. Our results provide new insights into the Brazilian COVID-19 epidemic and highlight the high transmission potential of SARS-CoV-2 in the country, the role of its large urban centres and the lack of lockdown and the challenges in reporting and non-equitable access to testing/diagnostics as factors potentially contributing to the rapid and sustained spread of the epidemic in Brazil.

Recent estimates of R_0 at the beginning of the COVID-19 epidemic in Brazil have suggested that an infected individual would infect on average three or four others²¹. The credible intervals of our estimates broadly overlap with these observations and are lower compared with previously published estimates for Brazil²². As a

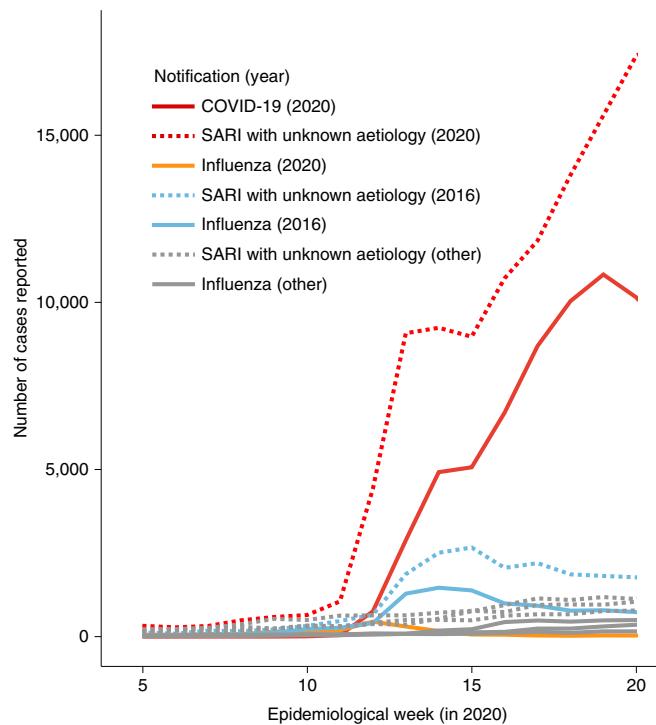


Fig. 4 | Reports of COVID-19 and SARI with unknown aetiology and influenza. The red and orange lines indicate cases reported in 2020 (solid red, COVID-19; solid orange, influenza; dashed red, SARI with unknown aetiology). The blue lines indicate cases reported in 2016 for influenza (solid blue line) and SARI with unknown aetiology (dashed blue line). Grey lines indicate influenza (solid line) and SARI cases with unknown aetiology (dashed line) for 2017, 2018 and 2019 combined.

comparison, the reproduction number in Peru has been estimated at around 2.3 (2.0–2.5)²³. Since the start of the epidemic in Brazil, several types of NPI have been adopted with varied success by the country's 27 federal units and 5,596 municipalities. Virus transmission seems to have dropped substantially in most affected states²¹ and also in the city of São Paulo²⁴. However, the estimated reproduction number remains above 1 (refs. ^{21,24}). Thus, only mitigation (and not suppression) of the epidemic has been achieved so far, which has been linked to substantial excess deaths due to poorer healthcare available^{25,26}. Closer surveillance of viral transmission at the local scales and an assessment of the impact of the different control measures on COVID-19 transmission will help to determine an optimal mitigation strategy to minimize infections and reduce healthcare demand in Brazil. Moreover, continued monitoring of the genetic diversity of the virus lineages circulating in Brazil²⁴ will be important, as recent data suggest that virus diversity may play a role in virus transmissibility^{27,28}.

We found that 65.5% of reports in the SIVEP-Gripe system, which includes most severe COVID-19 cases, are from patients aged ≥ 50 years of age. This observation is remarkably similar to current estimates for Latin America¹⁰, where 65% of the individuals ≥ 50 years of age have been estimated to be at high risk of severe COVID-19, defined as individuals with at least one condition who would require hospitalization if infected. Moreover, we found that 57 and 59% of the severe COVID-19 cases and deaths (respectively) reported in SIVEP-Gripe were male, and that the most frequent comorbidities were cardiovascular disease and diabetes. Overall, 84% of SIVEP-Gripe reports had at least one underlying condition. Of these, 21% ($n=9,471/45,480$) were included in the working age bracket (16–65 years of age). Moreover, only 2.6% ($n=1,892/73,673$) of the COVID-19 confirmed cases reported in the SIVEP-Gripe

system included occupation information. Information on socio-economic determinants, as well as occupation and race/ethnicity, are critical²⁹ as this allows the prioritization of control efforts; for example, towards healthcare workers and patients attending hospitals³⁰ or work settings³¹.

Our data uncover a socioeconomic bias in testing and diagnostics in current surveillance guidelines and suggest that the number of reported confirmed case counts may substantially underestimate the number of cases in the general population, particularly in regions of lower socioeconomic status. Socioeconomic differences are associated with access to healthcare³² and should be taken into account when designing targeted interventions. We found that the proportion of SARI cases with unknown aetiology versus confirmed COVID-19 cases has increased across the entire country (as of 15 June 2020, the number of reported SARI cases with unknown aetiology was nearly twofold greater than the number of confirmed COVID-19 cases). Based on clinical and epidemiological grounds, it is likely that many SARI cases with unknown aetiology are caused by SARS-CoV-2. In order to rigorously establish the contribution of non-SARS-CoV-2 infections to the SARI cases, we would need additional denominator data to understand the level of testing for these viruses (that is, the negative test results). Our findings with regards to socioeconomic bias are likely to apply to other states and regions of Brazil and highlight the importance of scaling up surveillance and laboratory capacity within Latin America. Indeed, the largest Brazilian serosurvey conducted to date suggests that undetected cases may be seven times higher than reported cases³³.

We further show that SARI cases with unknown aetiology are associated with lower socioeconomic status in the MRSP. The socioeconomic disparities observed here were particularly evident at the beginning of the outbreak (Fig. 5b). This can be explained in part by: (1) the high proportion of early cases in returning travellers with higher income and better access to private laboratories for diagnostics; and (2) the more limited access to freely available diagnostic screening. For example, between 25 February and 18 March 2020, two-thirds (586 (66.9%) of 876) of diagnostic tests were performed in private medical laboratories where costs varied typically between 300 and 690 Brazilian Reais (for context, the current minimum monthly salary is 1,045 Brazilian Reais). Thus, the true burden of the epidemic in lower-income neighbourhoods is probably underestimated. In New York City, for example, poorer neighbourhoods have been found to have a higher disease burden, which is driven in part by the movement of essential workers using public transport during the pandemic³⁴. Data-driven analyses are urgently needed to help tackle health inequities during the ongoing epidemic in Brazil. Strategies to evaluate and control transmission should consider differential access to COVID-19 diagnosis for lower-income populations, changes in reporting systems and delays in reporting, which are key to accurately determining rates of epidemic growth³⁵. Innovative infectious disease surveillance approaches such as those obtained from aggregated mobility data, when used properly, could help support public health actions across the COVID-19 epidemic^{36–39}.

Epidemics of COVID-19 and influenza seem to have occurred simultaneously in Brazil (Fig. 4 and Extended Data Fig. 7) and symptoms overlap between the two infections. We detected co-circulation of eight other respiratory viruses, the most common of which were influenza A and B and human rhinovirus. We also detected multiple co-detection of SARS-CoV-2 with other respiratory viruses, such as influenza A and B and human metapneumovirus, which have also been reported elsewhere^{40,41}. Although, co-infections with other respiratory viruses have been reported in other countries^{40,42,43}, no difference in clinical disease severity between cases with and without viral co-infection has been observed thus far⁴⁴. The co-circulation of other respiratory pathogens highlights the need to scale up laboratory and molecular screening of SARS-CoV-2 and other respiratory

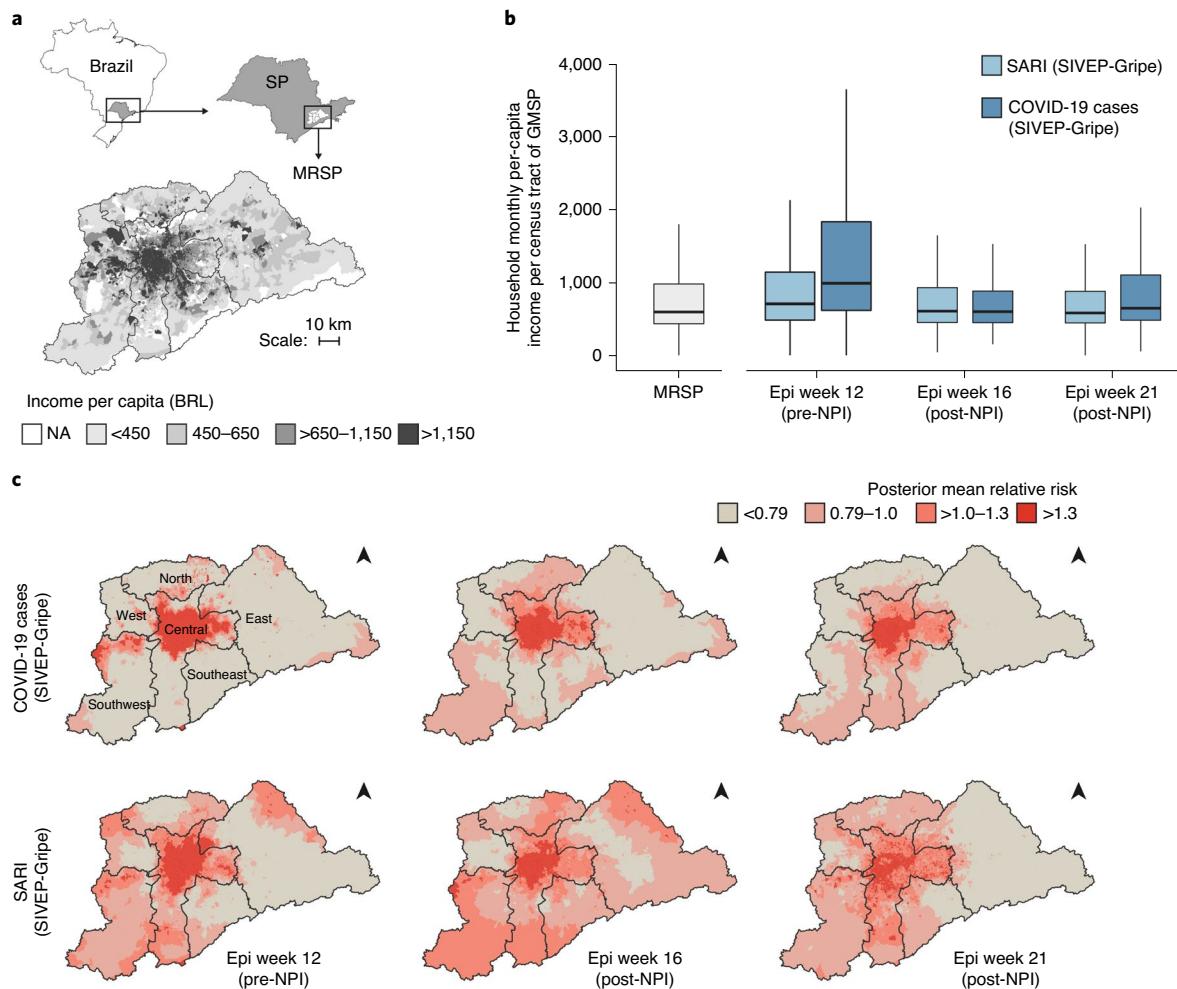


Fig. 5 | COVID-19 diagnosis and socioeconomic factors in the MRSP. **a**, Spatial distribution of income per capita of MRSP based on the census tract of residence. BRL, Brazilian reais. NA, not applicable. **b**, Distribution of household per-capita income based on the census tract of residence for COVID-19 cases and SARI cases with unknown aetiology. The distribution of average per-capita income for MRSP as a whole, weighted by population size, is shown on the left. Per-capita income distributions are presented in box plots, where the horizontal line inside the box represents the median per-capita income level, and the box edges show the per-capita income within the first and third interquartile range. The whiskers represent the per-capita income range. Epi, epidemiological. **c**, Posterior mean relative risk of COVID-19 confirmed diagnosis (top) and SARI cases with unknown aetiology (bottom) for epidemiological week 12 (before implementation of NPI in São Paulo state) and weeks 16 and 21 (after implementation of NPI in São Paulo state) (see Methods for details).

viruses in public laboratories across Brazil¹⁵. Continued molecular and genomic surveillance will be important to determine patterns of virus transmission and to guide public health measures in forthcoming phases of the epidemic^{24,45–47}.

There are several limitations to this study. First, detailed individual-level data were only available for the REDCap and SIVEP-Gripe systems, in which many cases had incomplete documentation, particularly regarding comorbidities. Second, our socio-economic analysis was based partially on ecological inference, using the per-capita income in the census tract of residence (rather than the actual income of the patients), and assuming the same denominator for each census tract (~300 households). We emphasize that our spatial analysis is prone to methodological constraints caused by ecological fallacy and the modifiable areal unit problem. These constraints are inherent to any spatial analysis of aggregated data. Despite the above-mentioned limitation, census tracts correspond to small areas of analysis, of no more than 300 households but often fewer than that. Social science literature on Brazil not only highlights the country's socioeconomic inequality but also how it is spatially pronounced.

For this reason, census tracts remain a useful tool with which to infer per-capita income in the absence of individual-level data. In addition, our databases were predominantly composed of hospitalized patients with COVID-19, and we were unable to evaluate the rate of hospitalization among the different socioeconomic statuses. In the future, robust modelling of the relationships between socioeconomic factors and disease severity will require a data collection system with detailed information on symptoms/signs and comorbidities both in severe and non-severe cases. Finally, our retrospective study focused predominantly on symptomatic patients who presented or were referred to health services for testing. Therefore, we are unable (and do not attempt) to describe the full spectrum of disease, nor can we describe the full epidemiological picture of this epidemic.

In conclusion, we have provided a comprehensive assessment of COVID-19 reporting and transmission in Brazil. Our findings provide important context for diagnostic screening and healthcare planning, and for future precision studies focusing on the impacts of non-pharmaceutical and pharmaceutical interventions, and the effects of social health determinants on COVID-19 transmission.

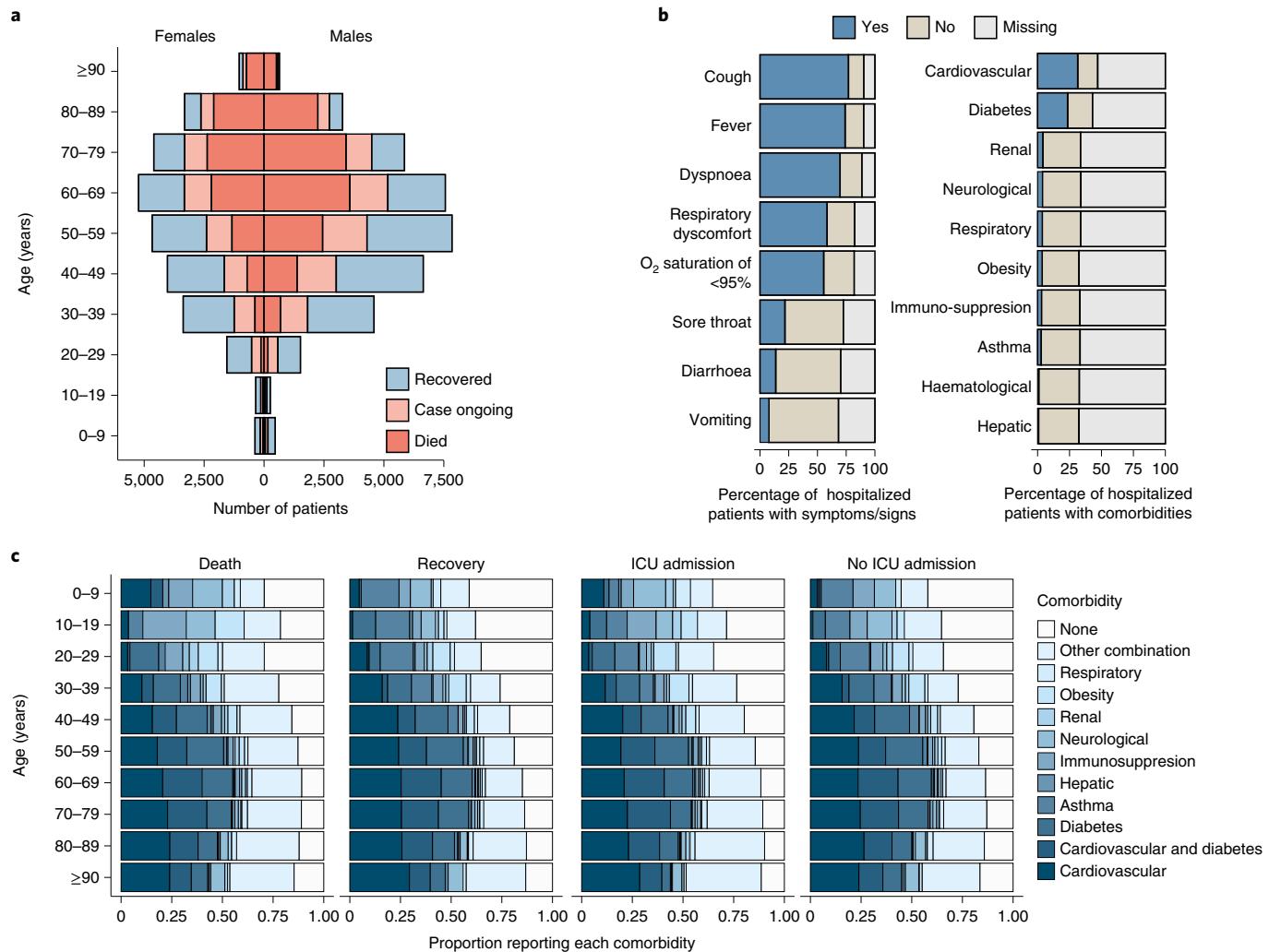


Fig. 6 | Age-sex structure and clinical features of confirmed COVID-19 cases reported in the SIVEP-Gripe system. **a**, Numbers of patients with ongoing COVID-19, or who have recovered or died from the disease, by age and sex. Ongoing cases were those still active on the SIVEP-Gripe database and without a recorded clinical outcome (death or recovered). **b**, Symptoms, signs and comorbidities of hospitalized individuals with confirmed COVID-19. **c**, Comorbidities among confirmed COVID-19 cases according to age and outcome ($n=15,720$ confirmed COVID-19 cases with complete comorbidity and outcome (death or recovery) information; $n=19,409$ confirmed COVID-19 cases with complete information on comorbidities and ICU admission). Horizontal axes show the proportion of patients in each age/outcome stratified for each of the comorbidities recorded.

Methods

Ethical approval and case definitions. This retrospective national study was supported by the Brazilian Ministry of Health and ethical approval was provided by the national ethical review board (Comissão Nacional de Ética em Pesquisa; protocol number CAAE 30127020.0.0000.0068).

A patient presenting with an acute respiratory syndrome (fever and at least one sign/symptom of respiratory illness) and: (1) a history of travel to a location with community transmission of COVID-19; or (2) contact with a confirmed or probable COVID-19 case in the 14 d preceding symptom onset; or (3) absence of an alternative diagnosis that completely explained the clinical presentation⁶ was considered to have suspected COVID-19.

Initially, a traveller was suspected to have COVID-19 only when arriving from China, although the definition of suspected cases associated with travel later included Japan, Singapore, South Korea, North Korea, Thailand, Vietnam and Cambodia (21 February 2020), then also Italy, Germany, Australia, the United Arab Emirates, the Philippines, France, Iran and Malaysia (25 February 2020), then also the United States, Canada, Switzerland, the United Kingdom and four additional countries (3 March 2020). From 9 March 2020 onwards, the Ministry of Health decided to start testing all hospitalized patients with severe respiratory symptoms, regardless of their travel history.

Contact with a confirmed or probable COVID-19 case was defined as face-to-face or direct contact with someone known to have COVID-19, or direct contact in a healthcare setting. Moreover, patients reporting travel to an affected country in the preceding 14 d were considered imported

cases. Cases not meeting this criterion were considered to be due to local transmission.

Suspected COVID-19 cases were confirmed by laboratory testing (that is, molecular diagnostics with real-time quantitative PCR), or by clinical epidemiological criteria. In the latter case, the classification was used when laboratory testing was inconclusive or unavailable, as recommended by the Brazilian Ministry of Health guidelines dated 6 April 2020⁴⁸, and by the World Health Organization interim guidance dated 25 March 2020⁴⁹.

Individual-level reporting of COVID-19 and SARI cases with unknown aetiology from Brazil. To investigate individual-level diagnostic and demographic data, self-reported travel history, place of residence and likely place of infection, differential diagnoses for other respiratory pathogens, as well as clinical details, including comorbidities, we collected three epidemiological data sources: (1) $n=67,344$ suspected and $n=1,468$ confirmed cases reported to the REDCap database from 25 February to 25 March 2020; (2) $n=73,637$ confirmed SIVEP-Gripe cases from 1 March to 31 May 2020 (available at <http://shiny.hmg.saude.gov.br/dataset>); and (3) $n=514,200$ confirmed cases from aggregated data released daily at the Portal do COVID-19 (Brazilian Health Ministry) from 25 February to 31 May 2020 (available at <https://covid.saude.gov.br>). The SIVEP-Gripe system reports cases of SARI, which can be defined as an acute respiratory infection with onset, within the past 10 d, of fever ($\geq 38^{\circ}\text{C}$) and cough, and typically requires hospitalization (see also Fig. 1a).

Basic reproduction number estimation. We estimated the basic reproduction number (R_0) for SARS-CoV-2 using time series of confirmed COVID-19 cases at the national and state (São Paulo, Rio de Janeiro, Ceará and Amazonas) level (Extended Data Fig. 1). To avoid the impact of NPIs on R_0 estimates, only data points up to 14 d after the implementation of the strictest interventions were used. As lockdown was not imposed in Brazil, the strictest measure was considered to be the closure of non-essential commerce. For European countries, the date of lockdown was used as the NPI date. NPI dates for Brazilian states were collected from state decrees. For Brazil as a whole, the NPI date for São Paulo state was used, as by that point most states in Brazil had already closed non-essential commerce. For the European countries, lockdown dates were collected from <https://www.covid19healthsystem.org/mainpage.aspx>.

To test the estimation routine and provide international context, this analysis was replicated on equivalent time series from Italy, Spain, France and the United Kingdom. Aggregated epidemiological data from the United States and China were not included due to possible heterogeneity within each country. Daily counts of confirmed cases were modelled with a negative binomial distribution with a mean equal to a fixed portion, ρ , of the total daily number of cases in an exponential model of incidence. The functional form of the incidence model is $\rho R_0 i_0 e^{(R_0 - 1)t}$, where ρ is the probability of an infection being counted in the time series, R_0 is the basic reproduction number, i_0 is the rate at which individuals cease to be infectious. We assume that the observed number of cases on day n was drawn from a negative binomial observation where the mean is $\mu(n)$ and the variance $\sigma^2 = \mu + \mu^2/k$, with fixed size parameter k (dispersion parameter). The product of ρ and i_0 is denoted ξ . Since the probability of being observed and the initial condition only appear as the product ξ in the likelihood, there is an identifiability problem preventing the estimation of ρ and i_0 individually, and consequently we only consider their product, ξ . Although in this model it is theoretically possible to estimate both R_0 and γ , in practice this is difficult, so we use an informative prior to constrain γ to a priori plausible values. The factor of $\rho R_0 \gamma$ accounts for the partial observation of the incidence. In this analysis, the delay between infection and reporting was not accounted for.

Since ρ and i_0 only appear together, they were unidentifiable, and we combine them into a single parameter, ξ . This identifiability issue prevents us from estimating the prevalence without additional information to inform either i_0 or ρ . The analysis was carried out in a Bayesian framework with an uninformative prior distribution on R_0 and an informative prior on the removal rate. All other parameters had weakly informative prior distributions (see Supplementary Information). The informative prior ensures that an individual is infectious for an average of 5–14 d (ref. ⁵⁰) (Supplementary Information and Figs. 5 and 6). Standard diagnostics were used to check whether the Markov chain Monte Carlo samples were satisfactory. Full details of the model used, the estimation process and convergence of Markov chain Monte Carlo chains can be found in the Supplementary Information.

Geospatial analysis of COVID-19 cases and socioeconomic status. The average household per-capita income for the MRSP was retrieved at the census tract level from the 2010 census (<https://censo2010.ibge.gov.br/>). We geocoded 24,063 COVID-19 cases and 32,914 SARI cases with unknown aetiology from MRSP, which were reported until 28 May 2020. The geocoding was based on self-reported residential addresses or postal codes using the Galileo algorithm⁵¹ and coordinates were confirmed using Google API.

To elucidate the distribution of COVID-19 cases and SARI cases with unknown aetiology, we mapped the mean relative risk of COVID-19 and SARI with unknown aetiology at the census tract level for MRSP for three epidemiological weeks (12, 16 and 21) (Extended Data Fig. 9). As the observation process was a confounding process and without additional assumptions (for example, covariates), we cannot disentangle an increase in prevalence from an increase in case ascertainment. The cumulative number of cases in each tract was modelled as a Poisson random variable with a mean specified by the expected number of cases under a null model adjusted by tract specific risk due to spatial and non-spatial effects: the Besag–York–Mollie model¹⁹. Estimates of the risk of COVID-19 diagnosis or SARI cases with unknown aetiology were obtained using approximate Bayesian methods (integrated nested Laplace approximation). A complete specification of the model and the computational methodology can be found in the Supplementary Information.

The association between final diagnostic category (COVID-19 or SARI with unknown aetiology) and socioeconomic status in the subset of cases in the MRSP with geocoded residential information was evaluated using logistic regression models. We focused on the cases in epidemiological weeks 12, 16 and 22. Within each of those weeks, if a census tract reported any COVID-19 or SARI with unknown aetiology, we calculated the proportion of the number of COVID-19 cases. Since most census tracts reported only one case each week, the proportion of COVID-19 cases for each census tract was mostly either 0 or 1 in a given week. For this reason, we defined two categories: (1) the census tract only reported SARI of unknown aetiology (that is, no COVID-19 cases); or (2) the census tract reported at least one COVID-19 case during the week. We used these two categories as the binary response, and applied logistic regression models to investigate whether income per capita was associated with this response. The analyses were adjusted

by the logarithm of the population sizes and the longitude and latitude coordinates of the census tracts. The analyses were performed individually for each of epidemiological weeks 12, 16 and 22. Further analysis details can be found in the Supplementary Information.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Datasets of clinical and laboratory data presented in the current study from the SIVEP-Gripe/Portal do COVID-19 database are available at <https://datadryad.org/stash/share/xj7kX8675lwvLzrnnPn9ebEfjNoOB38aXBTTQqfGBhE>. The REDCap database and geolocation information are available from the corresponding authors upon request and ethical approval.

Code availability

The custom code used in this study is available at <https://datadryad.org/stash/share/xj7kX8675lwvLzrnnPn9ebEfjNoOB38aXBTTQqfGBhE>.

Received: 8 July 2020; Accepted: 15 July 2020;
Published online: 31 July 2020

References

- Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species *Severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-020-0695-z> (2020).
- Lu, R. et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).
- Guan, W.-J. et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa2002032> (2020).
- Livingston, E. & Bucher, K. Coronavirus disease 2019 (COVID-19) in Italy. *J. Am. Med. Assoc.* <https://doi.org/10.1001/jama.2020.4344> (2020).
- Coronavirus Disease (COVID-2019) Situation Reports* (World Health Organization, 2020); <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
- Croda, J. et al. COVID-19 in Brazil: advantages of a socialized unified health system and preparation to contain cases. *Rev. Soc. Bras. Med. Trop.* **53**, e20200167 (2020).
- Jesus, J. G. et al. Importation and early local transmission of COVID-19 in Brazil, 2020. *Rev. Inst. Med. Trop. SP* **62**, e30 (2020).
- Candido, D. S. et al. Routes for COVID-19 importation in Brazil. *J. Travel Med.* **27**, taaa042 (2020).
- Clark, A. et al. Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. *Lancet Glob. Health* [https://doi.org/10.1016/S2214-109X\(20\)30264-3](https://doi.org/10.1016/S2214-109X(20)30264-3) (2020).
- Burki, T. COVID-19 in Latin America. *Lancet Infect. Dis.* **20**, 547–548 (2020).
- Cimerman, S., Chebabo, A., Cunha, C. A. D. & Rodriguez-Morales, A. J. Deep impact of COVID-19 in the healthcare of Latin America: the case of Brazil. *Braz. J. Infect. Dis.* **24**, 93–95 (2020).
- Ezequiel, G. E. et al. The COVID-19 pandemic: a call to action for health systems in Latin America to strengthen quality of care. *Int. J. Qual. Health Care* <https://doi.org/10.1093/intqhc/mzaa062> (2020).
- Miller, M. J., Loaiza, J. R., Takyar, A. & Gilman, R. H. COVID-19 in Latin America: novel transmission dynamics for a global pandemic? *PLoS Negl. Trop. Dis.* **14**, e0008265 (2020).
- Andrus, J. K. et al. Perspectives on battling COVID-19 in countries of Latin America and the Caribbean. *Am. J. Trop. Med. Hyg.* <https://doi.org/10.4269/ajtmh.20-0571> (2020).
- Croda, J. H. R. & Garcia, L. P. Immediate health surveillance response to COVID-19 epidemic. *Epidemiol. Serv. Saude* **29**, e2020002 (2020).
- Harris, P. A. et al. The REDCap consortium: building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).
- Influenza Update* (WHO, 2020); https://www.who.int/influenza/surveillance_monitoring/updates/latest_update_GIP_surveillance/en/
- Besag, J., York, J. & Mollié, A. Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Stat. Math.* **43**, 1–20 (1991).
- Pesquisa Nacional de Saúde 2013: Percepção do Estado de Saúde, Estilos de Vida e Doenças Crônicas*. Brasil, Grandes Regiões e Unidades da Federação (IBGE, 2015).
- Mellan, T. A. et al. Report 21: estimating COVID-19 cases and reproduction number in Brazil. Preprint at medRxiv <https://doi.org/10.1101/2020.05.09.20096701> (2020).

22. Caicedo-Ochoa, Y., Rebollar-Sánchez, D. E., Penalosa-Rallón, M., Cortes-Motta, H. F. & Mendez-Fandino, Y. R. Effective reproductive number estimation for initial stage of COVID-19 pandemic in Latin American countries. *Int. J. Infect. Dis.* **95**, 316–318 (2020).
23. Munayco, C. V. et al. Early transmission dynamics of COVID-19 in a Southern Hemisphere setting: Lima-Peru February 29th–March 30th, 2020. *Infect. Dis. Model.* <https://doi.org/10.1016/j.idm.2020.05.001> (2020).
24. Da Silva Candido, D. et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* <https://doi.org/10.1126/science.abb2161> (2020).
25. Ferguson, N. et al. *Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand* (Imperial College COVID-19 Response Team, 2020).
26. Walker, P. G. T. et al. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* <https://doi.org/10.1126/science.abc0035> (2020).
27. Korber, B. et al. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.04.29.069054> (2020).
28. Zhang, L. et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.06.12.148726> (2020).
29. Khalatbari-Soltani, S., Cumming, R. G., Delpierre, C. & Kelly-Irving, M. Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. *J. Epidemiol. Commun. Health* <https://doi.org/10.1136/jech-2020-214297> (2020).
30. Rivett, L. et al. Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. *eLife* **9**, <https://doi.org/10.7554/eLife.58728> (2020).
31. Park, S. Y. et al. Coronavirus disease outbreak in call center, South Korea. *Emerg. Infect. Dis.* **26**, <https://doi.org/10.3201/eid2608.201274> (2020).
32. Pereira, R. H. et al. *Mobilidade Urbana e o Acesso ao Sistema Único de Saúde para Casos Suspeitos e Graves de COVID-19 nas Vinte Maiores Cidades do Brasil* Nota Técnica No. 14 (Diretoria de Estudos e Políticas Regionais, Urbanas e Ambientais, IPEA, 2020).
33. Silveira, M. et al. Repeated population-based surveys of antibodies against SARS-CoV-2 in Southern Brazil. Preprint at *medRxiv* <https://doi.org/10.1101/2020.05.01.20087205> (2020).
34. Sy, K. T. L., Martinez, M. E., Rader, B. & White, L. F. Socioeconomic disparities in subway use and COVID-19 outcomes in New York City. Preprint at *medRxiv* <https://doi.org/10.1101/2020.05.28.20115949> (2020).
35. Dehning, J. et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* <https://doi.org/10.1126/science.abb9789> (2020).
36. Buckee, C. O. et al. Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146 (2020).
37. De Oliveira, S. B. et al. Monitoring social distancing and SARS-CoV-2 transmission in Brazil using cell phone mobility data. Preprint at *medRxiv* <https://doi.org/10.1101/2020.04.30.20082172> (2020).
38. Kraemer, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
39. Nouvellet, P. et al. *Report 26: Reduction in Mobility and COVID-19 Transmission* (Imperial College COVID-19 Response Team, 2020).
40. Wu, X. et al. Co-infection with SARS-CoV-2 and influenza A virus in patient with pneumonia, China. *Emerg. Infect. Dis.* <https://doi.org/10.3201/eid2606.200299> (2020).
41. Kim, D., Quinn, J., Pinsky, B., Shah, N. H. & Brown, I. Rates of co-infection between SARS-CoV-2 and other respiratory pathogens. *J. Am. Med. Assoc.* <https://doi.org/10.1001/jama.2020.6266> (2020).
42. Cuadrado-Payan, E. et al. SARS-CoV-2 and influenza virus co-infection. *Lancet* **395**, e84 (2020).
43. Zheng, X. et al. Co-infection of SARS-CoV-2 and influenza virus in early stage of the COVID-19 epidemic in Wuhan, China. *J. Infect.* <https://doi.org/10.1016/j.jinf.2020.05.041> (2020).
44. Asner, S. A. et al. Clinical disease severity of respiratory viral co-infection versus single viral infection: a systematic review and meta-analysis. *PLoS ONE* **9**, e99392 (2014).
45. Black, A., MacCannell, D. R., Sibley, T. R. & Bedford, T. Ten recommendations for supporting open pathogen genomic analysis in public health. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0935-z> (2020).
46. Deng, X. et al. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science* <https://doi.org/10.1126/science.abb9263> (2020).
47. Lu, J. et al. Genomic epidemiology of SARS-CoV-2 in Guangdong province, China. *Cell* **181**, 997–1003.e9 (2020).
48. *Coronavirus COVID-19 Diretrizes para Diagnóstico e Tratamento da COVID-19* (Ministério da Saúde do Brasil, 2020).
49. *COVID-19 Coding in ICD-10* (WHO, 2020); <https://www.who.int/classifications/icd/COVID-19-coding-icd10.pdf?ua=1>
50. Wölfel, R. et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* <https://doi.org/10.1038/s41586-020-2196-x> (2020).
51. Medel, C. H., Catalan, C. C., Vidou, M. A. F. & Perez, E. S. The Galileo ground segment integrity algorithms: design and performance. *Int. J. Navigation Observation* <https://doi.org/10.1155/2008/178927> (2008).

Acknowledgements

We thank M. Gome, L. Bastos and L. M. Carvalho (MAVE) for useful discussions on SIVEP-Gripe, and we thank L. Matkin (Oxford) for technical support. This work was supported by a FAPESP (2018/14389-0) and Medical Research Council and CADDE partnership award (MR/S01951/1) (<http://caddecentre.org/>). W.M.S. is supported by the São Paulo Research Foundation, Brazil (2017/13981-0 and 2019/24251-9). N.R.F. is supported by a Wellcome Trust and Royal Society Sir Henry Dale Fellowship (204311/Z/16/Z). O.J.B. was funded by a Sir Henry Wellcome Fellowship funded by the Wellcome Trust (206471/Z/17/Z). V.H.N. and C.A.P. were supported by FAPESP (2018/12579-7). A.E.Z. and B.G. were supported by Oxford Martin School. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

W.M.S., L.F.B., D.d.S.C., R.H.M.P., C.A.P., J.C., J.-P.C., V.H.N., A.E.Z., J.M., F.C.S.S., P.d.S.A., F. Ghilardi, A.A.S.-S., B.G., C.-H.W., S.L., N.G., S.B.O., K.V.P., M.C.T.D.B., V.B.G.P., C.K.V.B., F. Ganem, W.A.F.A., F.F.S.T.F., E.M.M. and W.K.O. collected the epidemiological, spatial and clinical data and processed the statistical data. N.R.F., W.M.S., L.F.B., C.-H.W., J.-P.C., D.d.S.C., R.H.M.P., J.M., E.C.S., P.M., S.L., L.A., A.A.S.-S., G.L., A.T., M.F.V.-G., M.U.G.K., R.S.A., N.A., P.M., O.J.B., I.O.M.S., N.G., G.L., O.G.P., A.E.Z., M.L.N. and J.C. interpreted the results and wrote the manuscript. All authors read and revised the final manuscript. W.M.S., L.F.B., S. L., J.C., A. E. Z. and N.R.F. summarized the epidemiological and clinical data.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41562-020-0928-4>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41562-020-0928-4>.

Correspondence and requests for materials should be addressed to J.C. or N.R.F.

Peer review information Primary Handling Editor: Stavroula Koustas.

Reprints and permissions information is available at www.nature.com/reprints.

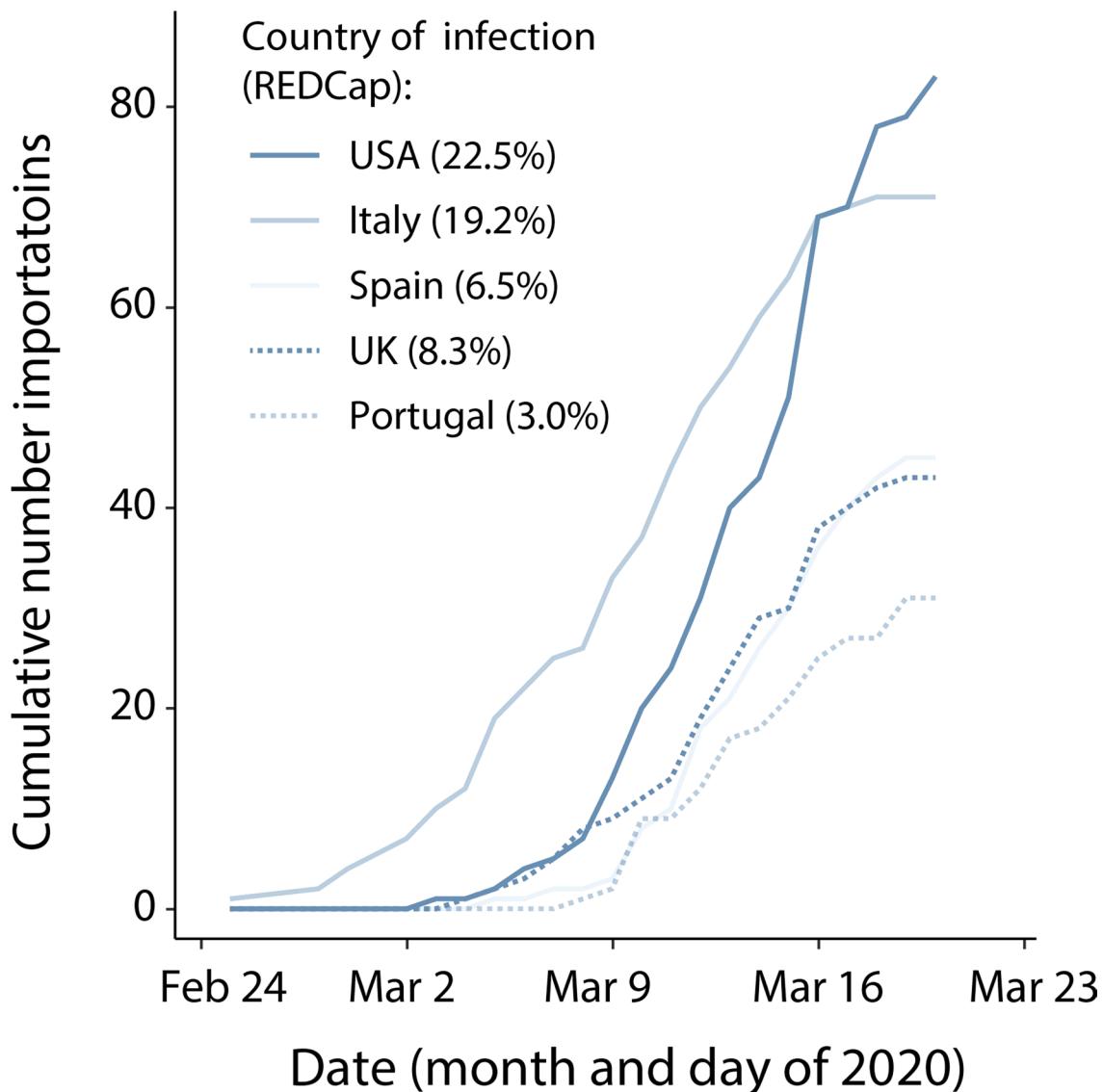
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

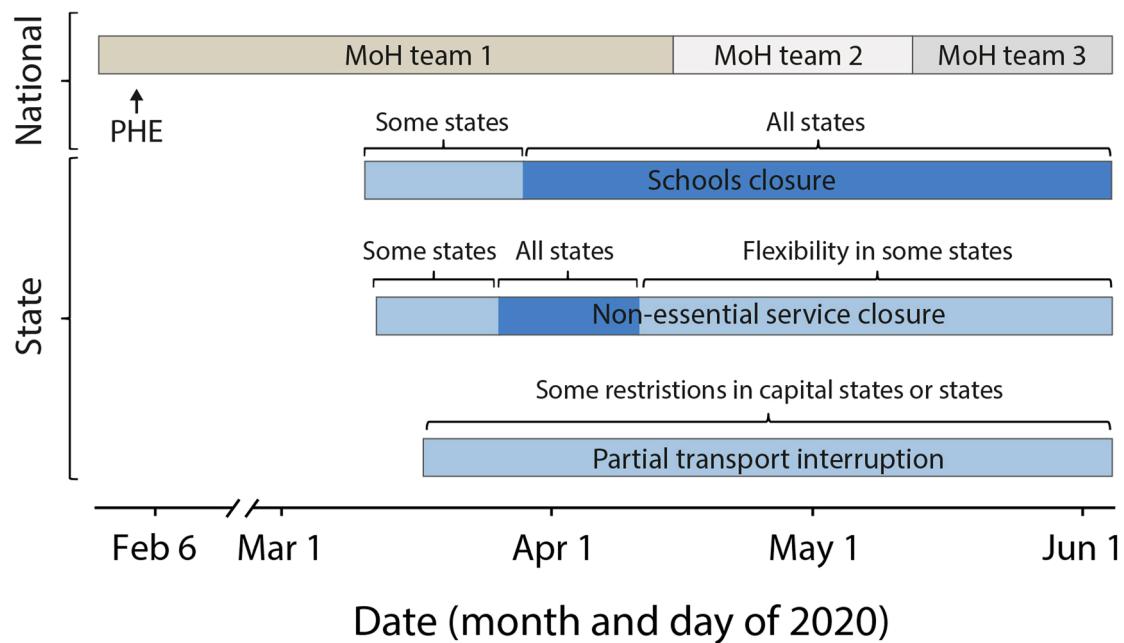
¹Centro de Pesquisa em Virologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil. ²Instituto de Medicina Tropical, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil. ³Department of Zoology, University of Oxford, Oxford, UK. ⁴Department of Research in Virology and Biotechnology, Gorgas Memorial Institute of Health Studies, Panama City, Panama. ⁵School of Geography and the Environment, University of Oxford, Oxford, UK. ⁶Institute for Applied Economic Research (IPEA), Brasília, Brazil. ⁷Escola Politécnica, Universidade de São Paulo, São Paulo, Brazil. ⁸Brazilian Studies Programme, Latin American Centre, University of Oxford, Oxford, UK. ⁹MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London, UK. ¹⁰Department of Medical Microbiology and Infection Prevention, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. ¹¹Oxford School of Global and Area Studies, University of Oxford, Oxford, UK. ¹²School of Biological and Environmental Sciences, Universidad San Francisco de Quito (USFQ), Quito, Ecuador. ¹³Harvard Medical School, Harvard University, Boston, MA, USA. ¹⁴Boston Children's Hospital, Boston, MA, USA. ¹⁵Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. ¹⁶MRC Tropical Epidemiology Group, Department of Infectious Disease Epidemiology, Faculty of Epidemiology

and Population Health, London School of Hygiene and Tropical Medicine, London, UK.¹⁷Department of Clinical Research, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK.¹⁸Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.¹⁹Núcleo de Vigilância Epidemiológica do Hospital das Clínicas da Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil.²⁰Departamento de Medicina Preventiva, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil.²¹Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China.²²Secretariat of Health Surveillance, Department of Immunization and Communicable Diseases, Brazilian Ministry of Health, Brasília, Brazil.²³Secretariat of Health Surveillance, Brazilian Ministry of Health, Brasília, Brazil.²⁴Faculdade de Medicina de São José do Rio Preto, São José do Rio Preto, Brazil.²⁵Mathematical Sciences, University of Southampton, Southampton, UK.²⁶Laboratório de Pesquisa em Ciências da Saúde, Universidade Federal da Grande Dourados, Dourados, Brazil.²⁷Fundação Oswaldo Cruz, Campo Grande, Brazil.²⁸Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, CT, USA.²⁹These authors contributed equally: William Marciel de Souza, Lewis Fletcher Buss, Darlan da Silva Cândido, Jean-Paul Carrera, Sabrina Li, Chieh-Hsi Wu, Julio Croda, Ester C. Sabino, Nuno Rodrigues Faria.

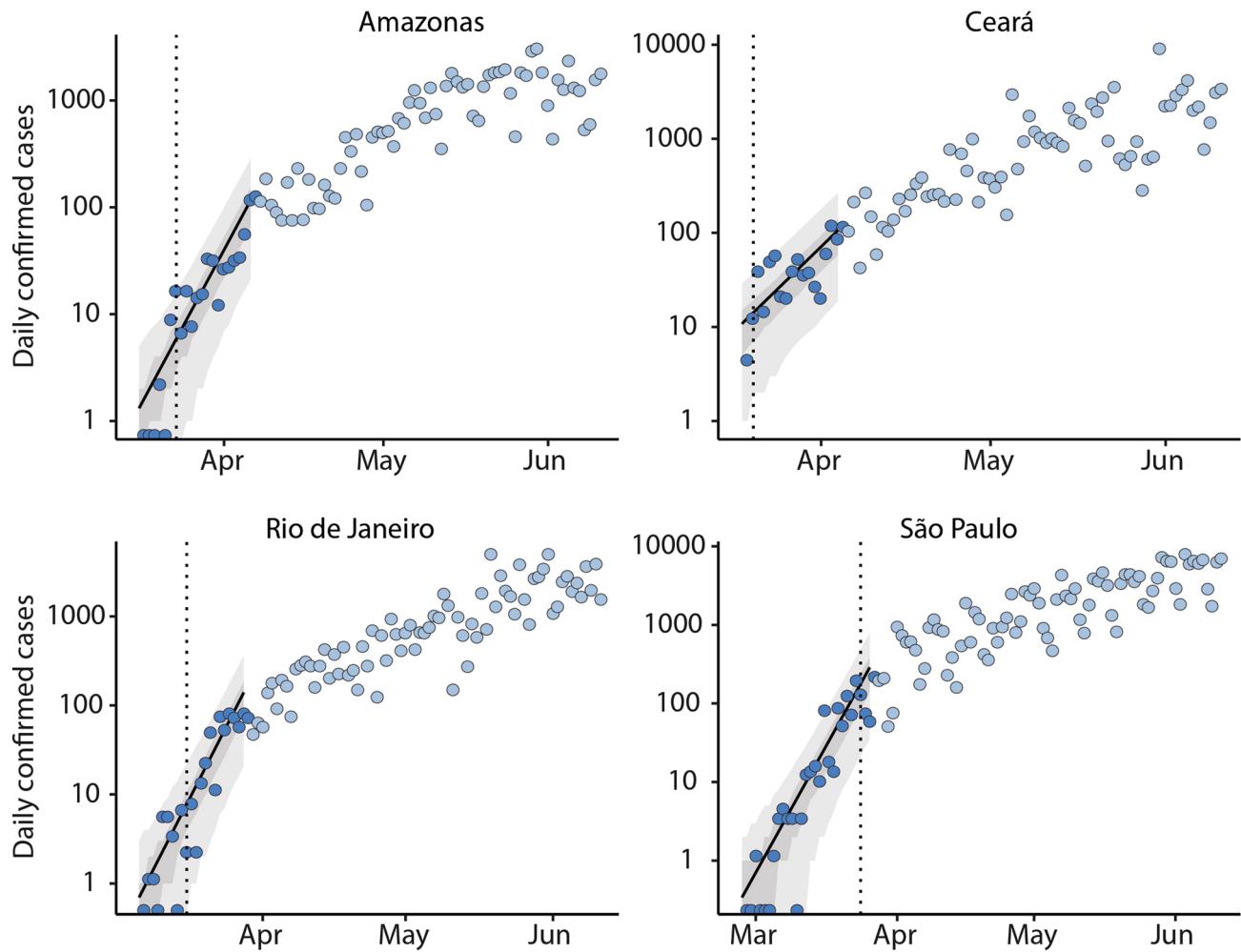
✉e-mail: juliocroda@gmail.com; nfaria@ic.ac.uk



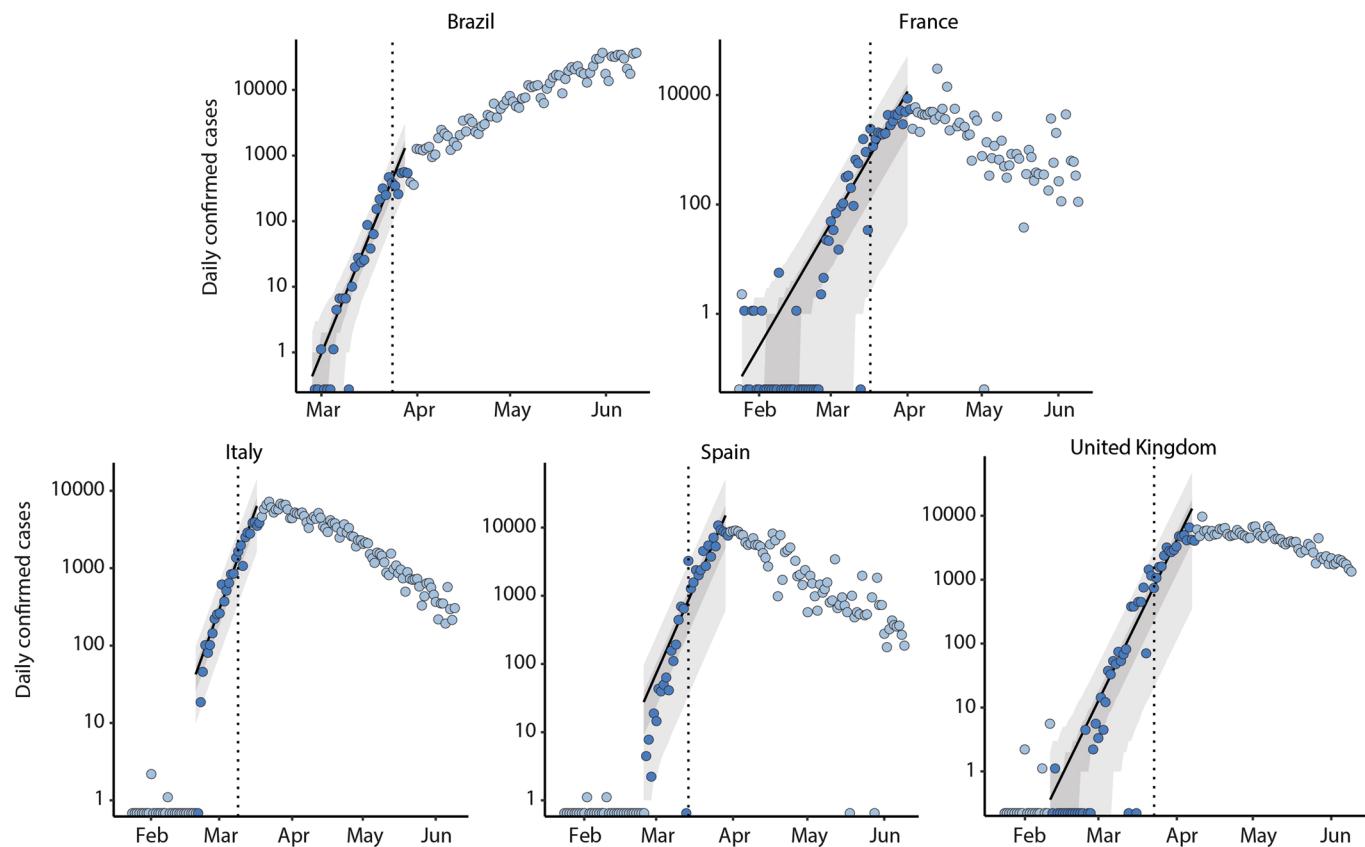
Extended Data Fig. 1 | Imported cases by self-reported country of infection from REDCap database. Percentage indicates proportion of cases acquired outside of Brazil between 25 February and 19 March (n=342) by unambiguously identified country of infection as recorded in REDCap database (see also Fig. 1).



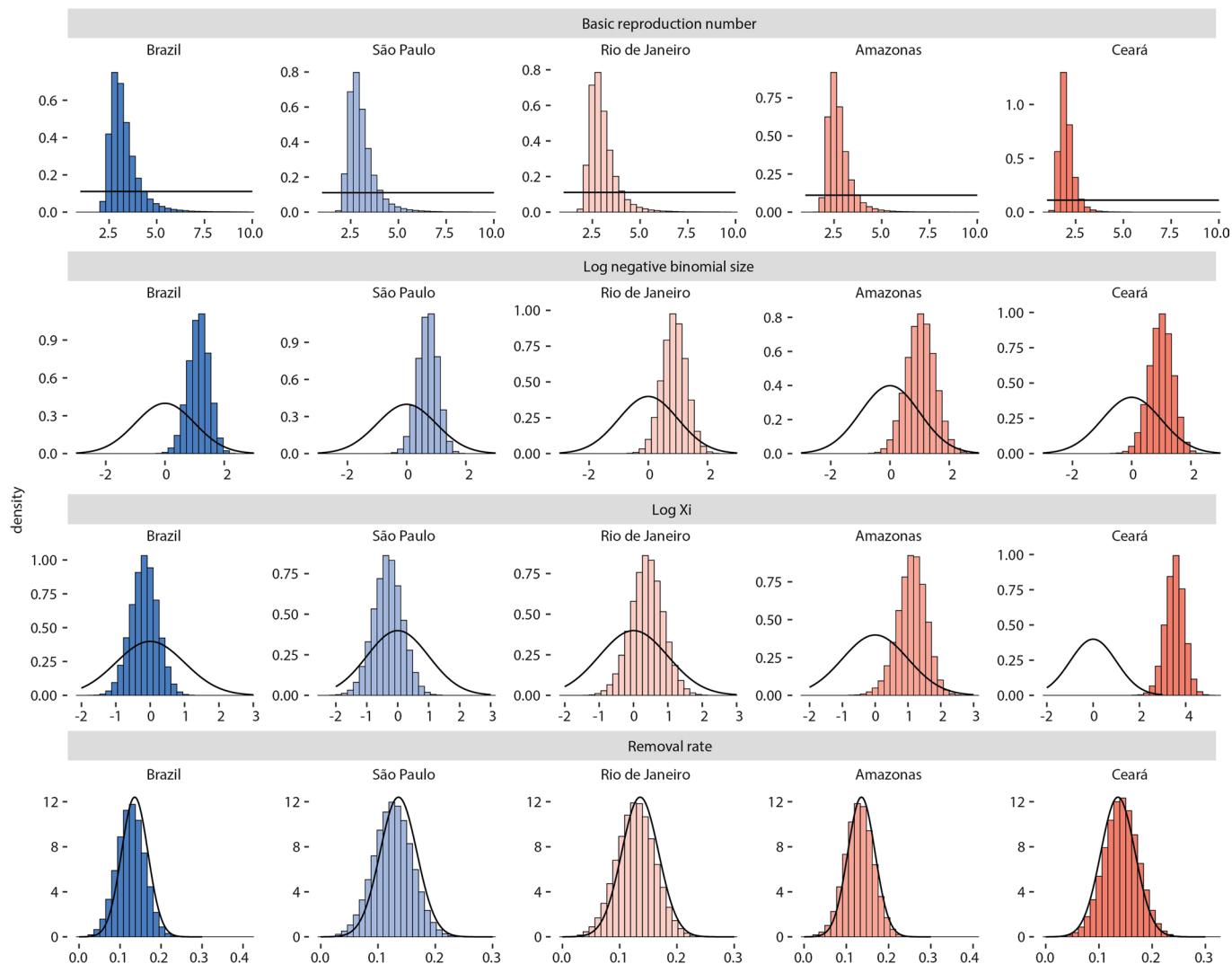
Extended Data Fig. 2 | Non-pharmaceutical interventions taken during the first three months of the epidemic in Brazil. Time of implementation of measures for COVID-19 control in Brazil. PHE = declaration of Public Health Emergency of International Concern. MoH = Ministry of Health. Data on non-pharmaceutical interventions compiled from state official decrees can be found in Supplementary Table 1.



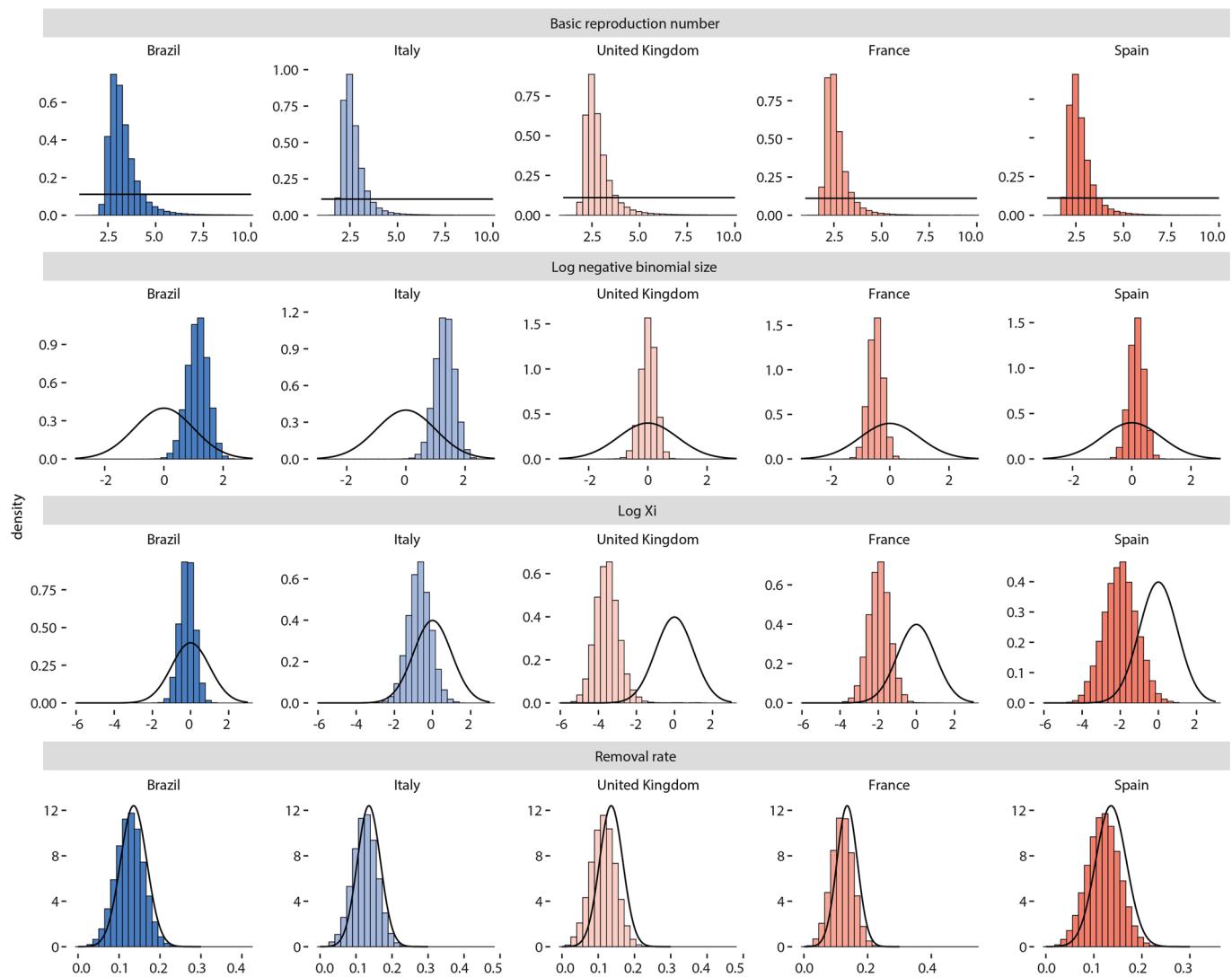
Extended Data Fig. 3 | Daily number of infections used for the R_0 estimations of confirmed cases of Brazil and European countries (France, Italy, Spain, and United Kingdom). The dashed vertical line indicates when the non-pharmaceutical intervention (NPI) was implemented. The dark blue dots were used to estimate R_0 . The shaded region is the model fit for those data points. The light blue dots included how the time series continued. They were included to show the effects of NPI.



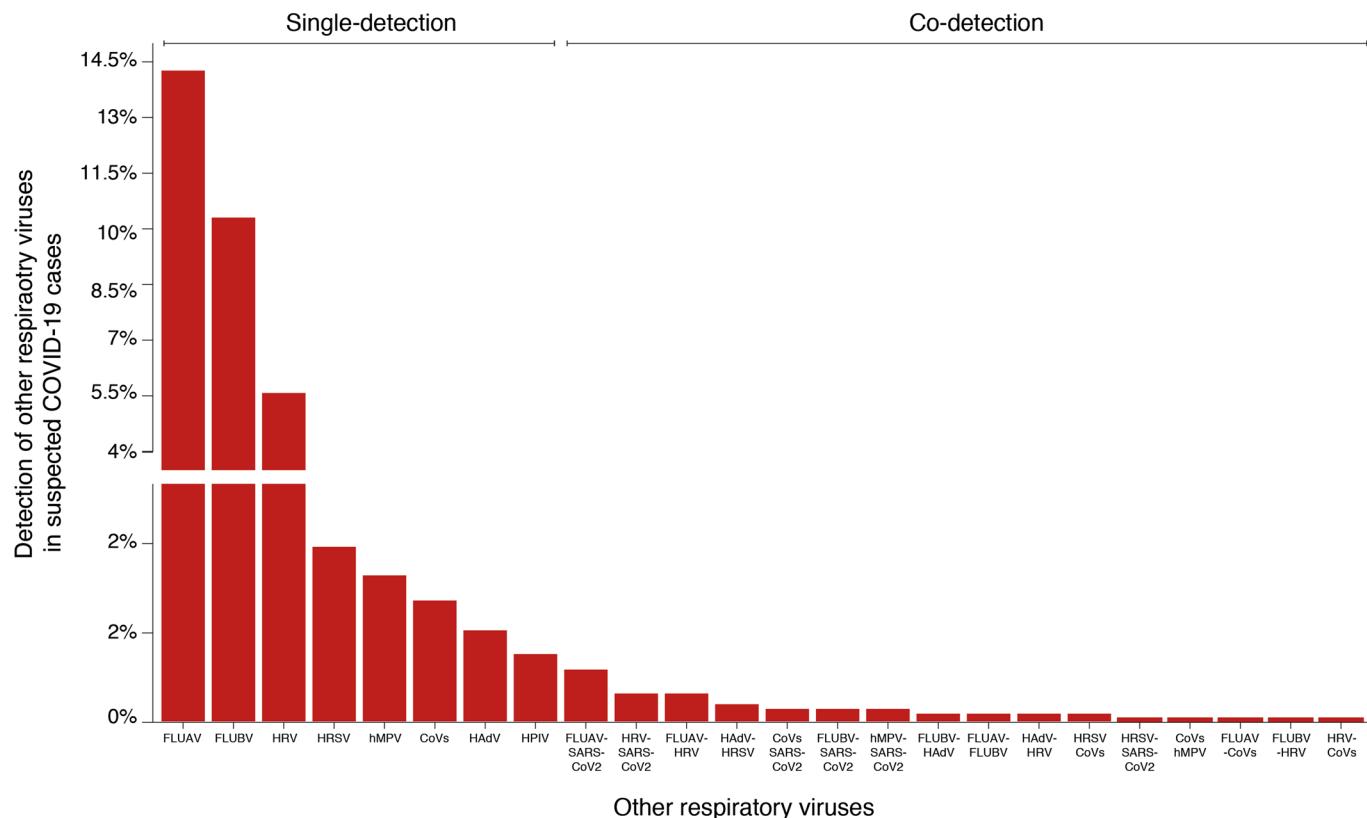
Extended Data Fig. 4 | Daily number of infections used for the R₀ estimations of confirmed cases in states of Amazonas, Ceará, Rio de Janeiro, and São Paulo. The dashed vertical line indicates when the NPI was implemented. The dark blue dots were used to estimate R₀. The shaded region is the model fit for those data points. The light blue dots included how the time series continued. They were included to show the effects of NPI.



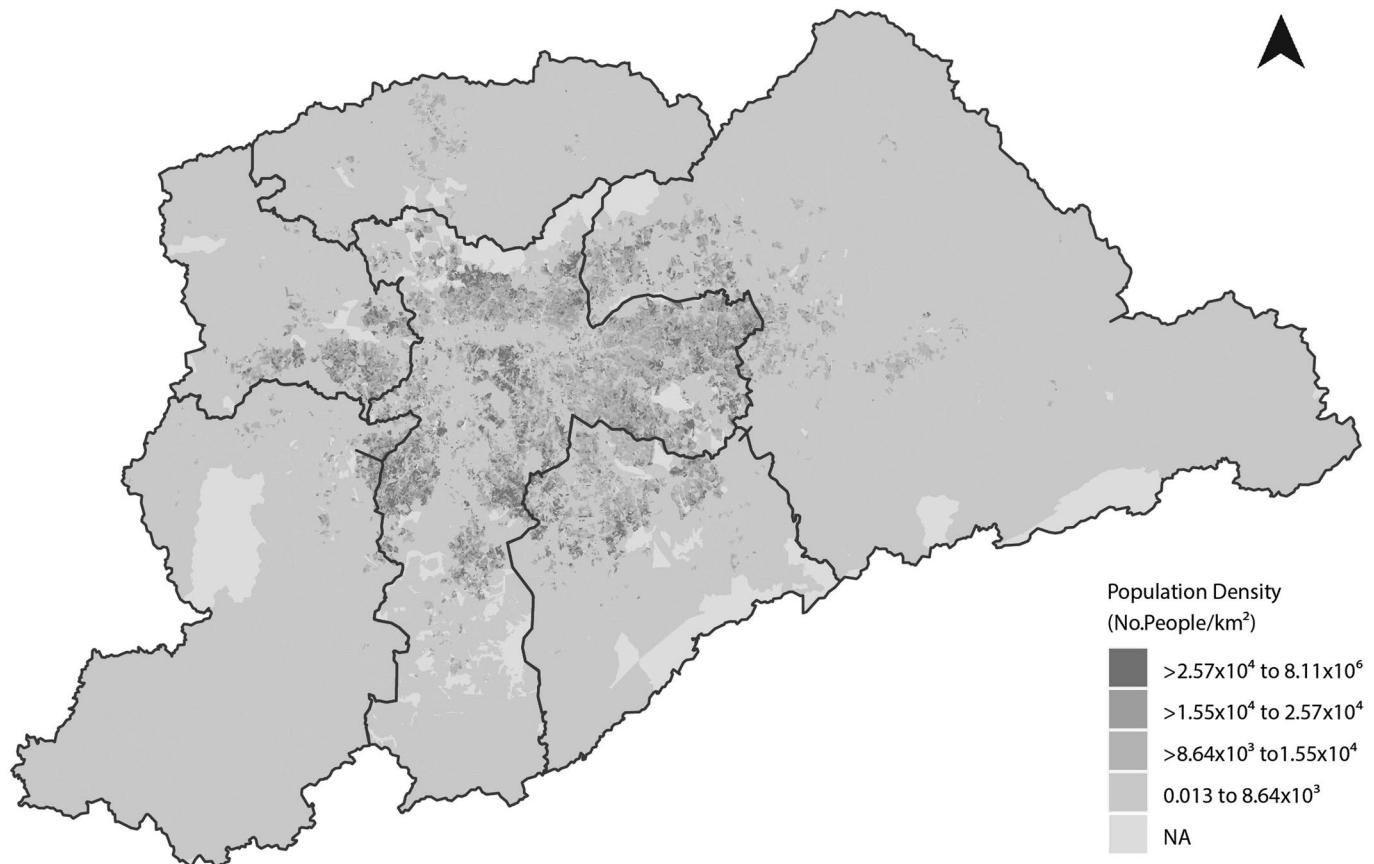
Extended Data Fig. 5 | The prior/posterior plots for the different parameters in the analysis of the time series from all of Brazil, and states of São Paulo, Rio de Janeiro, Amazonas, and Ceará. The histogram is of the posterior samples and the solid line shows the prior density about those values. From top to bottom, they are basic reproduction number, the log of the size of the negative binomial distribution, ξ , and removal rate.



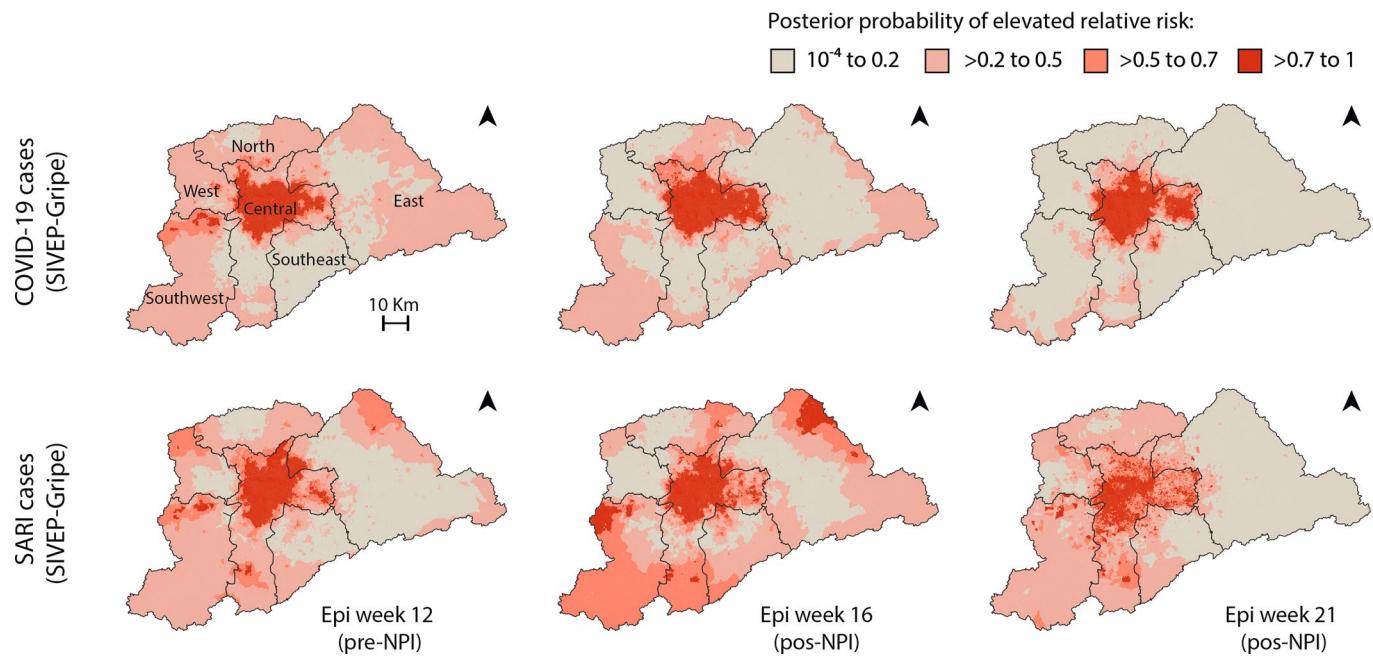
Extended Data Fig. 6 | The prior/posterior plots for the different parameters in the analysis of the time series of Brazil, Italy, the United Kingdom, France, and Spain. The histogram is of the posterior samples and the solid line shows the prior density about those values. From top to bottom, they are basic reproduction number, the log of the size of the negative binomial distribution, ξ , and removal rate.



Extended Data Fig. 7 | Diagnosis of other respiratory viruses in 2,429 suspected COVID-19 cases reported to Brazilian Ministry of Health between February 25 to March 25, 2020. influenza A virus (FLUAV), influenza B virus (FLUBV), human rhinovirus (HRV), human respiratory syncytial virus (HRSV), human metapneumovirus (hMPV), human adenovirus (HAdV), human parainfluenza viruses 1-4 (HPIV), and CoVs (that is, human coronavirus 229E, OC43, NL63 and HKU1).



Extended Data Fig. 8 | Map of the population density in each census tract in the Metropolitan Region of São Paulo. NA=not applicable.



Extended Data Fig. 9 | COVID-19 diagnosis and socio-economic factors in the Metropolitan Region of São Paulo. Posterior probability of elevated relative risk of COVID-19 for confirmed diagnosis (upper panels) and SARI cases with unknown aetiology (lower panels) for epidemiological weeks 12 (pre-implementation of non-pharmaceutical interventions in São Paulo state), and weeks 16 and 21 (post-implementation of non-pharmaceutical interventions in São Paulo state).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Datasets of clinical and laboratory data presented in the current study will be made available from the corresponding authors upon request and ethical approval.
Data analysis	Custom code used in this study can be found in GitHub repository (Link available upon acceptance of the publication).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Custom code used in this study can be found in GitHub repository (Link available upon acceptance of the publication). Datasets of clinical and laboratory data presented in the current study will be made available from the corresponding authors upon request and ethical approval.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size Not applicable as this is an observational study.

Data exclusions We have not excluded data.

Replication Not applicable as this is an observational study.

Randomization Not applicable as this is an observational study.

Blinding Not applicable as this is an observational study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|-------------------------------|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | Antibodies |
| <input checked="" type="checkbox"/> | Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | Animals and other organisms |
| <input type="checkbox"/> | Human research participants |
| <input checked="" type="checkbox"/> | Clinical data |
| <input checked="" type="checkbox"/> | Dual use research of concern |

Methods

- | | |
|-------------------------------------|------------------------|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | ChIP-seq |
| <input checked="" type="checkbox"/> | Flow cytometry |
| <input checked="" type="checkbox"/> | MRI-based neuroimaging |

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

We have used data from COVID-19 cases notified to Brazilian Ministry of Health as suspected COVID-19 infections until the May 31, 2020. The characteristics of the dataset used in our study are described in detail in Materials and Methods (see also Data Availability statement above).

Recruitment

Not applicable.

Ethics oversight

National ethical review board (Comissão Nacional de Ética em Pesquisa), Brazil (CAAE 30127020.0.0000.0068).

Note that full information on the approval of the study protocol must also be provided in the manuscript.