BMJ Global Health

Higher risk of death from COVID-19 in low-income and non-White populations of São Paulo, Brazil

Sabrina L Li ,[1] Rafael H M Pereira ,[2] Carlos A Prete Jr ,[3] Alexander E Zarebski ,[4] Lucas Emanuel ,[2] Pedro J H Alves ,[2] Pedro S Peixoto ,[5] Carlos K V Braga ,[2] Andreza Aruska de Souza Santos ,[6] William M de Souza ,[4,7] Rogerio J Barbosa ,[8] Lewis F Buss ,[9] Alfredo Mendrone ,[10] Cesar de Almeida-Neto ,[10,11] Suzete C Ferreira ,[10,12] Nanci A Salles ,[10,12] Izabel Marcilio ,[13] Chieh-Hsi Wu ,[14] Nelson Gouveia ,[15] Vitor H Nascimento ,[3] Ester C Sabino ,[9] Nuno R Faria ,[4,9,16] Jane P Messina [1,17]

To cite: Li SL, Pereira RHM, Prete Jr CA, *et al*. Higher risk of death from COVID-19 in low-income and non-White populations of São Paulo, Brazil. *BMJ Global Health* 2021;**6**:e004959. doi:10.1136/bmjgh-2021-004959

Handling editor Seye Abimbola

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjgh-2021-004959).

SLL, RHMP and CAPJ contributed equally.

Received 6 January 2021
Revised 8 March 2021
Accepted 6 April 2021

For numbered affiliations see end of article.

Correspondence to
Sabrina L Li; lisabrinaly@gmail.com and Dr Rafael H M Pereira; rafael.pereira@ipea.gov.br

ABSTRACT

Introduction Little evidence exists on the differential health effects of COVID-19 on disadvantaged population groups. Here we characterise the differential risk of hospitalisation and death in São Paulo state, Brazil, and show how vulnerability to COVID-19 is shaped by socioeconomic inequalities.

Methods We conducted a cross-sectional study using hospitalised severe acute respiratory infections notified from March to August 2020 in the *Sistema de Monitoramento Inteligente de São Paulo* database. We examined the risk of hospitalisation and death by race and socioeconomic status using multiple data sets for individual-level and spatiotemporal analyses. We explained these inequalities according to differences in daily mobility from mobile phone data, teleworking behaviour and comorbidities.

Results Throughout the study period, patients living in the 40% poorest areas were more likely to die when compared with patients living in the 5% wealthiest areas (OR: 1.60, 95% CI 1.48 to 1.74) and were more likely to be hospitalised between April and July 2020 (OR: 1.08, 95% CI 1.04 to 1.12). Black and *Pardo* individuals were more likely to be hospitalised when compared with White individuals (OR: 1.41, 95% CI 1.37 to 1.46; OR: 1.26, 95% CI 1.23 to 1.28, respectively), and were more likely to die (OR: 1.13, 95% CI 1.07 to 1.19; 1.07, 95% CI 1.04 to 1.10, respectively) between April and July 2020. Once hospitalised, patients treated in public hospitals were more likely to die than patients in private hospitals (OR: 1.40%, 95% CI 1.34% to 1.46%). Black individuals and those with low education attainment were more likely to have one or more comorbidities, respectively (OR: 1.29, 95% CI 1.19 to 1.39; 1.36, 95% CI 1.27 to 1.45).

Conclusions Low-income and Black and *Pardo* communities are more likely to die with COVID-19. This is associated with differential access to quality healthcare, ability to self-isolate and the higher prevalence of comorbidities.

Key questions

What is already known?

- Black and *Pardo* (mixed ethnicity) Brazilians face higher risk of COVID-19 hospitalised death.
- Access to COVID-19 testing has been limited for low-income populations in São Paulo city.

What are the new findings?

- Individual and population-level risk of COVID-19 hospitalisation, death and adherence to non-pharmaceutical interventions vary by race and socioeconomic status.
- Low socioeconomic and/or Black and *Pardo* (Brazilians of mixed ethnic ancestries) communities have lower levels of social isolation and face higher risks of hospitalisation and death.

What do the new findings imply?

- The stark difference in COVID-19 mortality between public and private healthcare settings underscores the need for further investigation on the drivers of mortality in different hospital settings.
- While non-pharmaceutical interventions have been implemented in São Paulo and other states to slow down transmission, the effectiveness of these interventions among population groups varies with socioeconomic status.
- Healthcare workers, disadvantaged groups working in face-to-face occupations in crowded and segregated areas should be prioritised for vaccination.

INTRODUCTION

The COVID-19 pandemic has amplified the effects of social inequalities on exposure and death in low socioeconomic groups,[1] particularly in Brazil, where it has caused significant mortality.[2] The prevalence of COVID-19 mortality is partially driven by pre-existing non-communicable diseases, which are socially

clustered due to entrenched inequalities.[3] These inequalities are shaped by the social determinants of health,[4] which define a population's health based on the environments where they 'grow, live, work, and age', from birth.[5] Even when underlying health conditions are not present, the interactions of these social determinants disproportionately expose disadvantaged groups to COVID-19[4] and other conditions that could induce adverse chronic health conditions.[6] Furthermore, a review has found that disadvantaged groups are the most vulnerable to the psychosocial impacts of COVID-19,[7] which can aggravate the severity of COVID-19.[4]

Several studies that have been conducted in the context of high-income countries have mostly focused on the USA,[8–10] UK[11 12] and European countries,[13 14] which have consistently found that populations identified as non-White, of low socioeconomic status and those living in high poverty were associated with higher SARS-CoV-2 transmission and COVID-19 death. Few studies have addressed the uneven impact of COVID-19 by socioeconomic status and race in low and middle-income countries,[15] in part because national surveillance systems seldom collect or report this information.[16] In Brazil, higher risk of COVID-19 death has been found for Black and *Pardo* (mixed ethnicity) Brazilians, especially those who are identified as male with low socioeconomic status.[17 18] Nonetheless, there is still little information on how the differential health outcomes of COVID-19 are shaped by broader social inequalities that determine the capacity to self-isolate and non-pharmaceutical interventions (NPIs).

It is paramount to understand the potential social drivers of COVID-19 morbidity and mortality, particularly in countries with high inequality such as Brazil.[19] The first COVID-19 cases in Brazil were detected in São Paulo,[20] the most populous state and home to diverse racial groups. In the Brazilian context of politically polarised public health responses,[21] São Paulo has been severely affected by COVID-19[22] and access to testing has been limited for low-income populations.[23]

We conducted a multiscale analysis to investigate the risk of hospitalisation and death from severe acute respiratory infections (SARI), predominantly caused by COVID-19,[23] notified from March to August 2020, in the *Sistema de Monitoramento Inteligente de São Paulo* (SIMI-SP) database, for São Paulo state. We considered all SARI cases instead of only including patients who tested positive for COVID-19 to avoid the bias in access to SARS-CoV-2 testing towards higher socioeconomic classes in Brazil,[23] which allows us to better capture the disproportionate impact of social inequities on racial and socioeconomic groups. We examined differential risk by race and socioeconomic status, by combining multiple high-resolution data from mobile phones, government census and population surveys conducted during the epidemic. We assessed potential drivers of these inequalities by evaluating local levels of self-isolation, access to teleworking and prevalence of comorbidities.

METHODS

Data sources

SARI and patient information

Patient-level information on demographic characteristics, home address, hospitalisation and health outcomes was collected from the São Paulo State Health Secretariat SARI hospitalisations database (SIMI-SP).[24] SARI can be caused by SARS-CoV-2 and is defined by the Brazilian Ministry of Health as influenza-like syndrome plus one of the following: dyspnoea, persistent chest pain or hypoxia. We excluded all SARI cases that were confirmed to be caused by other respiratory viruses. All SARI cases and deaths are notified in the SIMI-SP database, regardless of hospitalisation.

We included all SARI related hospitalisations and deaths notified in São Paulo state between March 15 and August 29, 2020. Given that recent data is incomplete due to reporting delays[25] and to avoid biases, we limited our analysis to patients with symptoms onset between these dates (epidemiological weeks 10 – 35). We also included SARI cases with unknown etiology, as those are likely related to COVID-19 but not lab-confirmed due to low rates of COVID-19 testing in Brazil[26] and socioeconomic bias in testing.[23]

Zip code information was only available for cases reported in São Paulo state. Data were geocoded using the patient's self-reported home address or postal code with Galileo (www.img.com.br) and Google API. For our analysis, we aggregated these data to the census tract level (n=68 296), the smallest administrative unit reported by the Brazilian census for the purpose of spatial statistical analysis. In the state of São Paulo, 95% of the census tracts have a population size between 136 and 1347 individuals (median=724). Information on the health facility where each case was notified was linked to the National Registry of Health Facilities (Cadastro Nacional de Estabelecimentos de Saúde), which includes information on the mode of healthcare provision (public and private).

The race of patients was partially self-declared and partially identified by a health professional. Race was categorised as either 'White', 'Black', 'Asian' (East or Southeast Asian), '*Pardo*' (mixed ethnic ancestries with diverse skin colours)[27] or Indigenous. Race information was missing for 53 480 (23.9%) of retrieved SARI cases, and was imputed using the racial distribution of the census tract of residence (see online supplemental materials for details). About 0.1% of the population in São Paulo self-identified as Indigenous[28] and since only 166 Indigenous patients (0.07%) were recorded, they were not considered in our analysis.

Socioeconomic data

We obtained data on municipality-level socioeconomic factors from the latest population census (2010) compiled by the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística; IBGE). We selected indicators based on their relevance to the social determinants of health as defined

by the WHO's Commission on Social Determinants of Health,[29] which includes income and income distribution, education, employment and job security, and access to healthcare. We included household income per capita, population density and income inequality (Gini Index). We also determined the proportion of residents with a primary education or lower, employment to population ratio and the proportion of the working population without a formal labour market contract or social security. The road network distance from the centroid of each census tract to the nearest healthcare facility was computed considering all 830 facilities that hospitalised patients with SARI via the public healthcare system (Sistema Único de Saúde; SUS). Information on employment status and comorbidities during the epidemic was retrieved from the National Household Sample Survey (Pesquisa Nacional por Amostra de Domicílios (PNAD) COVID-19), a national telephone survey conducted by IBGE with over 1 888 560 interviews between May and September 2020. Details are described in online supplemental materials.

Seroprevalence data

To assess the broader risk of SARS-CoV-2 infection beyond hospitalisation, we adopted seroprevalence data collected as part of the national Covid-IgG study from blood donors aged 16 – 69 living in São Paulo city.[30] Given that samples were taken across the city, a population-weighted cluster sample of approximately 1000 blood donations were tested each month between February and August 2020 using a chemiluminescence assay that detects IgG against the SARS-CoV-2 nucleocapsid (N) protein (Abbott, Chicago, USA). Self-reported race and education level were recorded at the time of blood donation. To correct for differences in the age-sex distribution of blood donors compared with the population of São Paulo, we applied an age-sex normalisation to the measured prevalence. Details about the data collection methods can be found in Buss *et al.* 's [30] study. [31]

Daily mobility and NPIs

To assess the ability of populations to self-isolate at the local level, we used daily mobile phone data provided by In-Loco (https://www.inloco.com.br/covid-19)[32] for the greater metropolitan area of São Paulo (Região Metropolitana de São Paulo; RMSP). These data were aggregated using a hexagonal grid based on the global H3 index at resolution 8. Each cell has an edge of approximately 460 m and an area of 0.74 km^2 (https://h3geo.org/docs/core-library/restable). For each H3 cell, the social isolation index was measured as the number of people who did not leave their cell of residence during the day, divided by the number of residents in that cell. Each mobile phone was assigned to an H3 cell based on the owner's location of residence during the evening and their travel history. The racial composition and income level of each cell were determined using dasymetric interpolation (online supplemental materials).

We also used municipality and state-level data on NPIs from a continuous survey conducted between 13 May and 31 July 2020.[33] The survey had 13 questions related to the implementation and easing of social distancing measures, and responses were obtained from 612 mayors in São Paulo (94.8% of the total).

Data analysis

Probability of hospitalisation and death

We conducted an individual-level analysis to estimate the probability of reporting a SARI hospitalisation given a patient's race and average income level in their census tract of residence. Census tracts were grouped by quantiles of income per capita into six categories as presented in the results. Similarly, we determined the probability of death from SARI given a patient's race, income and administrative type of the health facility where the patient was hospitalised (public or private). Both probabilities were standardised by age and sex to account for demographic differences between groups. They were calculated for every month between March and August 2020. Probabilities for each age-sex group were estimated empirically using relative frequencies. ORs using White patients and the highest income level as reference groups were computed. CIs were calculated using bootstrapping. Details are in online supplemental materials.

Seroprevalence by socioeconomic status

We calculated the proportion of individuals by education and race category with detectable anti-SARS-CoV-2 antibodies between February and October 2020. 95% CIs were calculated by the exact binomial method and corrected for the specificity and sensitivity of the test.[31]

Socioeconomic drivers and hospitalisation risk

An ecological spatiotemporal regression analysis was conducted at the municipality level for São Paulo state (n=645 municipalities) to assess the monthly risk of hospitalisation and its association with socioeconomic factors between 1 March and 29 August 2020. To further understand the association between socioeconomic conditions and COVID-19 risk, we conducted the same analysis at the census tract level (n=30 815) for the RMSP, where the majority of cases were concentrated. The relative risk of hospitalisation was estimated using a hierarchical Bayesian model composed of a generalised log-linear model with spatially structured and unstructured random effects to account for spatial autocorrelation and time-varying random effect. The spatial structure is characterised by population movement between municipalities from 1 March to 15 August 2020 defined by In-Loco mobile geolocation data summarised elsewhere.[32] A detailed description of the model and interpretation, covariates and diagnostics can be found in online supplemental materials.

Population response to NPIs

We used an event study design[34] to examine how different socioeconomic groups changed their daily mobility levels in response to the introduction and relaxation of NPIs in the RMSP. We compared changes in mobility patterns of the population living in H3 cells with predominantly White versus predominantly Black or *Pardo* residents, as well as of the population living in areas of the wealthiest and poorest income quintiles. The daily isolation index from hexagons was regressed on a set of relative time dummies that indicated the number of days before and after the first NPI introduction in São Paulo state. Hexagon fixed effects controlled for unobserved time-invariant determinants of self-isolation while day fixed effects controlled for temporal shocks common to all hexagons. We further included an additional time-varying control variable, representing the number of days relative to the first confirmed SARI case in each hexagon, and a dummy variable indicating the period of NPI relaxation in each municipality. Sensitivity analyses were performed and discussed in online supplemental materials.

We employed a multinomial logistic regression to estimate the probability that employed individuals would be working face to face, teleworking, or taking paid or unpaid leave. Differences in the work status of individuals by race, education and occupation were calculated while controlling for age and sex. Thirty-five groups of employment occupations listed in the PNAD COVID-19 survey were aggregated to 10 SCO-08 one-digit occupational groups defined by the International Labour Organization. We further disaggregated health professionals and health technicians (online supplemental materials).

Comorbidities

Information on patient comorbidities was missing for approximately 61.6% of the cases in the SIMI-SP database. We estimated the incidence of comorbidities for the population of São Paulo state using the PNAD COVID-19 data. A binomial logistic regression was used to estimate the OR of having at least one comorbidity, by race and education attainment (preprimary, primary, secondary and tertiary), while controlling for age and sex for São Paulo state. The comorbidities considered were chronic obstructive pulmonary disease, diabetes, hypertension or cardiovascular disease such as myocardial infarction, angina or heart failure. CIs for the ORs were calculated taking into account PNAD's complex sample design.

RESULTS

SARIs capture COVID-19-related hospitalisations

Between 1 March and 29 August 2020, São Paulo state had the highest number of SARI hospitalisations per 100 000 habitants compared with all states in Brazil (figure 1A). A time series illustrating the number of SARI hospitalisations is presented in figure 1B. During this time, 232 540 patients were notified in the SIMI-SP database (figure 1C), from which 127 434 (54.8%) had a confirmed COVID-19 diagnosis and 103 360 (44.4%) were diagnosed with SARI of unknown or missing aetiology. From these, 223 455 were hospitalised (98.4%) or died without hospitalisation (1.6%). From the non-hospitalised cases, we only selected deaths; 54 108 patients died, of which 52.5% were White, 20.2% were *Pardo*, 6.13% were Black, 1.96% were Asian, 0.052% were Indigenous and 19.6% did not

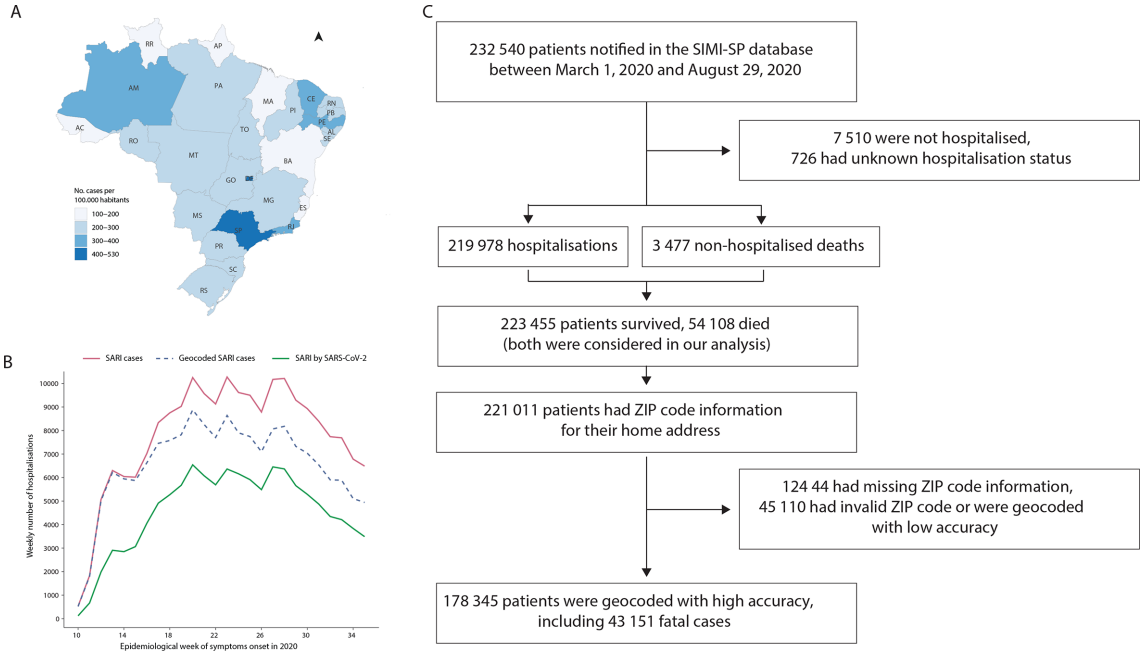

Figure 1 Severe acute respiratory infection (SARI) hospitalisations in São Paulo state. (A) Number of hospitalisations per 100 000 habitants by state in Brazil between 1 March and 29 August 2020. (B) Number of SARI hospitalisations for the state of São Paulo by week of symptom onset. (C) Flow chart of Sistema de Monitoramento Inteligente de São Paulo (SIMI-SP) data processing (Source: https://covid.saude.gov.br).

have race information. We geocoded 178 345 (79.8%) of all SARI cases considered in our analyses with high accuracy at either street address, route or neighbourhood level without compromising personal privacy.

Individual risk to hospitalisation and death varies by race, socioeconomic status and hospital type

During the first month of the COVID-19 epidemic (March) in Brazil, hospitalised patients were more likely to be White or Asian and come from census tracts with higher income per capita (figure 2A,B). During this period, people living in low-income areas were less likely (OR: 0.44, 95% CI 0.42 to 0.46) to be hospitalised with SARI compared with high-income areas. We found that point estimates of inequality levels do not change substantially when observations with missing race information are dropped, though CIs become smaller (online supplemental figure S1). This coincides with the early introduction of COVID-19 in Brazil, when the first infections occurred among higher income travellers returning from overseas.[20 23]

As the epidemic progressed from April onwards, patients were on average more likely to be from low-income census tracts (April to July, OR: 1.08, 95% CI 1.04 to 1.12), except for August, when patients were less likely

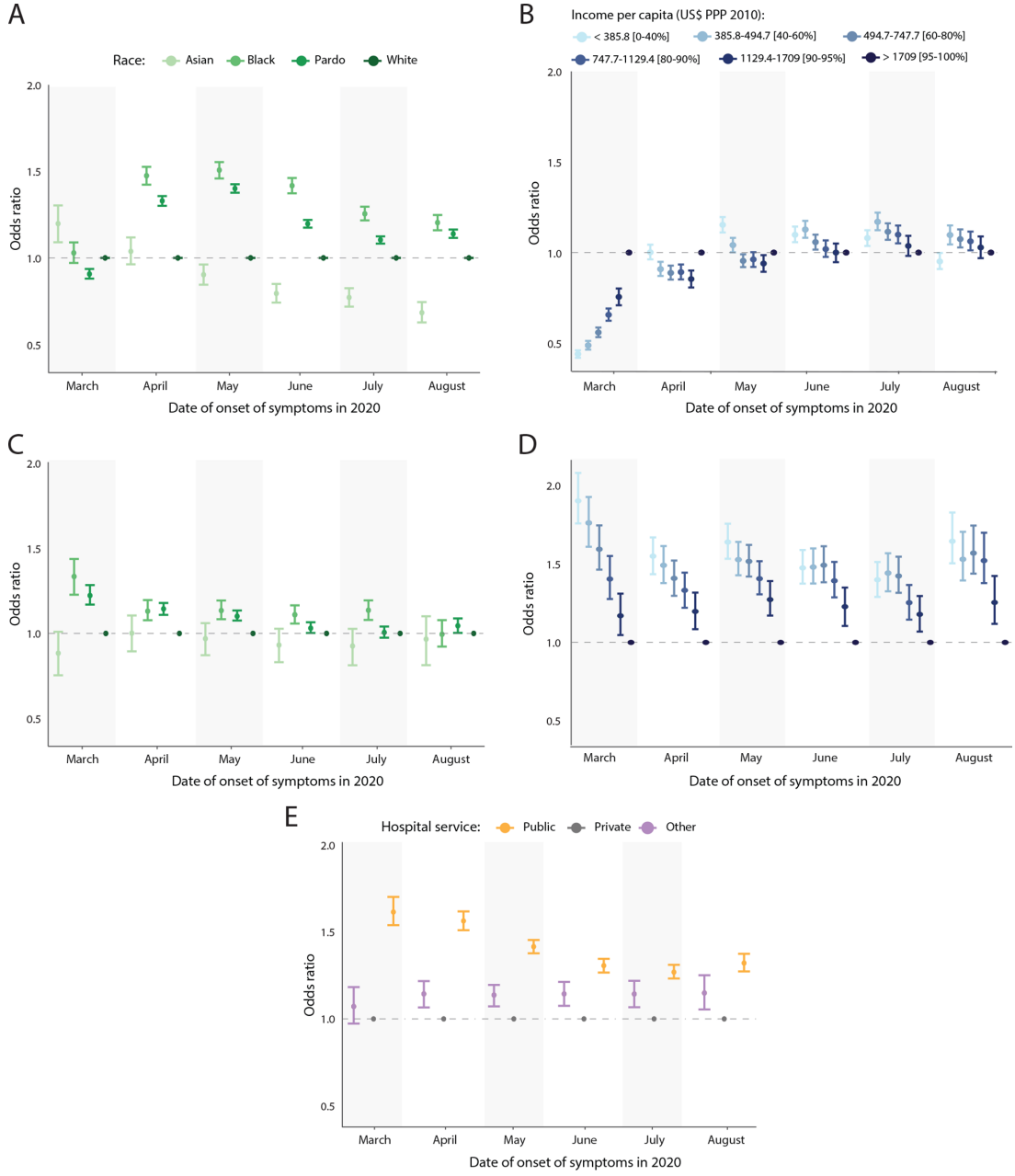

Figure 2 Individual-level hospitalisation and death risk by age-standardised OR. (A) OR for severe acute respiratory infection (SARI) hospitalisation by race. (B) OR for SARI hospitalisation by income. (C) OR for death among patients with SARI by race. (D) OR for death among patients with SARI by income. (E) OR for death among patients with SARI by hospital type. PPP, purchasing power parity.

to be from low-income census tracts (OR: 0.95, 95% CI 0.91 to 1.00). Similarly during this time period, Black Brazilians and *Pardos* became more likely to be hospitalised with SARI than Whites (OR: 1.41, 95% CI 1.37 to 1.46; OR: 1.26, 95% CI 1.23 to 1.28, respectively), while Asians became the least likely to be hospitalised (OR: 0.88, 95% CI 0.82 to 0.94). These results were further confirmed by our seroprevalence findings, where both crude and adjusted prevalence (for age, sex, sensitivity and specificity of the sex) showed that anti-SARS-CoV-2 antibodies were highest in Black blood donors and those with low educational attainment across all age groups (online supplemental figure S2).

Once hospitalised, Black and *Pardo* patients were more likely to die from SARI than White patients between March and August (OR: 1.14, 95% CI 1.07 to 1.21; 1.09, 95% CI 1.05 to 1.13, respectively) (figure 2C). This difference was more pronounced in March and decreased over time. Because our analysis does not control for comorbidities, these results indirectly reflect the differences in the incidence of comorbidities across racial groups. We found that patients living in the poorest census tracts were more likely to die from SARI compared with patients from the wealthiest tracts (OR: 1.60, 95% CI 1.48 to 1.74) (figure 2D). Likewise, patients treated in public hospitals were more likely to die than patients treated in private hospitals throughout the epidemic (OR: 1.40%, 95% CI 1.34% to 1.46%) (figure 2E). Racial differences in the probability of death decreased when considering only patients hospitalised at public health facilities but persisted among patients in private facilities (online supplemental figure S3).

Geographic variation in hospitalisation risk is driven by mobility and socioeconomic status

To understand the geographical variation in SARI hospitalisation, we estimated and mapped the relative risk of SARI hospitalisation at the municipality level (n=645 municipalities) for São Paulo state by month using a model with a spatial structure defined by human movement fluxes derived from anonymised mobile phone data (figure 3A) and covariates related to socioeconomic status (figure 3B). Overall, municipalities with higher levels of movement exchange with the RMSP had higher monthly risk of SARI hospitalisation (figure 3C). We found a lower risk of SARI by SARS-CoV-2 hospitalisation in municipalities with high income per capita (fixed effect=−0.87, 95% CI −1.12 to −0.62) and high proportion of adult residents with a primary education or lower (−0.90, 95% CI −1.12 to −0.68). Municipalities with fewer nearby public health facilities were also found to have lower risk of hospitalisation (−0.31, 95% CI −0.55 to −0.08). We also found a higher risk of SARI hospitalisation in municipalities with higher population density (0.31, 95% credible interval: 0.07–0.54).

We found that over time, the risk of SARI hospitalisation increased particularly in municipalities near and within the RMSP (greater metropolitan area of São Paulo), where 70% of the SARI cases reported for the state are concentrated (figure 4A). By mapping risk at the census tract level (n=30 815) for the RMSP, we detected increasing risk starting from São Paulo city (central region). By June, almost all of the census tracts in and near the city centre were found to have high relative risk, but this risk decreased by August.

Lower ability to self-isolate by disadvantaged groups

Differential risk to SARI in the RMSP was also associated with daily mobility levels. Before the implementation of NPIs on 13 March, mobility levels were similar across all socioeconomic groups (figure 4B,C). However, 14 days after the introduction of NPIs, isolation levels were 8.2% (95% CI 7.2% to 9.2%) higher in predominantly White areas compared with predominantly Black areas. Similarly, 27 days after the introduction of state-level NPIs, isolation levels were 11.2% (95% CI 10.6% to 11.9%) higher in the wealthiest than in the poorest areas. Overall, we detected a decreasing trend in isolation levels over time, and the magnitude of the differences in social isolation levels between areas with predominantly White

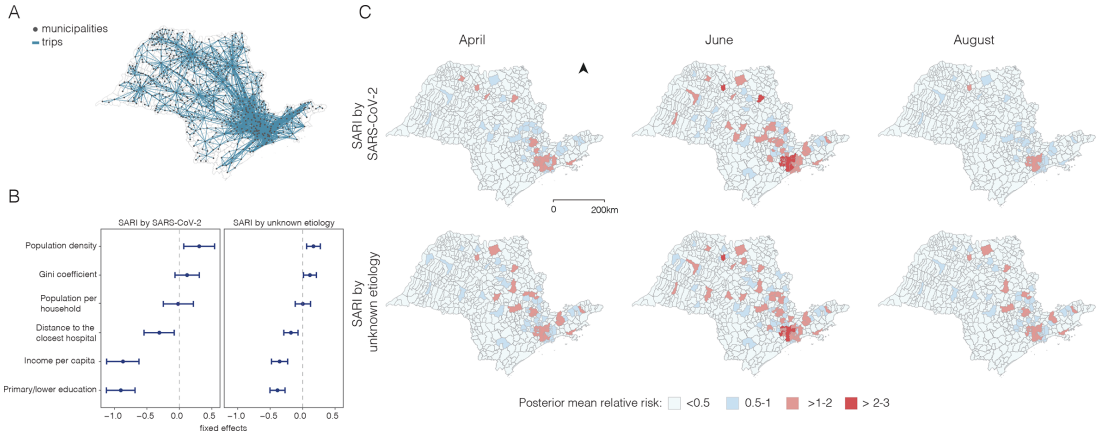

Figure 3 Hospitalisation risk by municipality in São Paulo state. (A) Human movement between municipalities based on InLoco mobile phone data retrieved from March to August 2020. (B) Fixed effects and 95% credible intervals for socioeconomic covariates. (C) Relative risk of severe acute respiratory infection (SARI) hospitalisation at the municipality level.

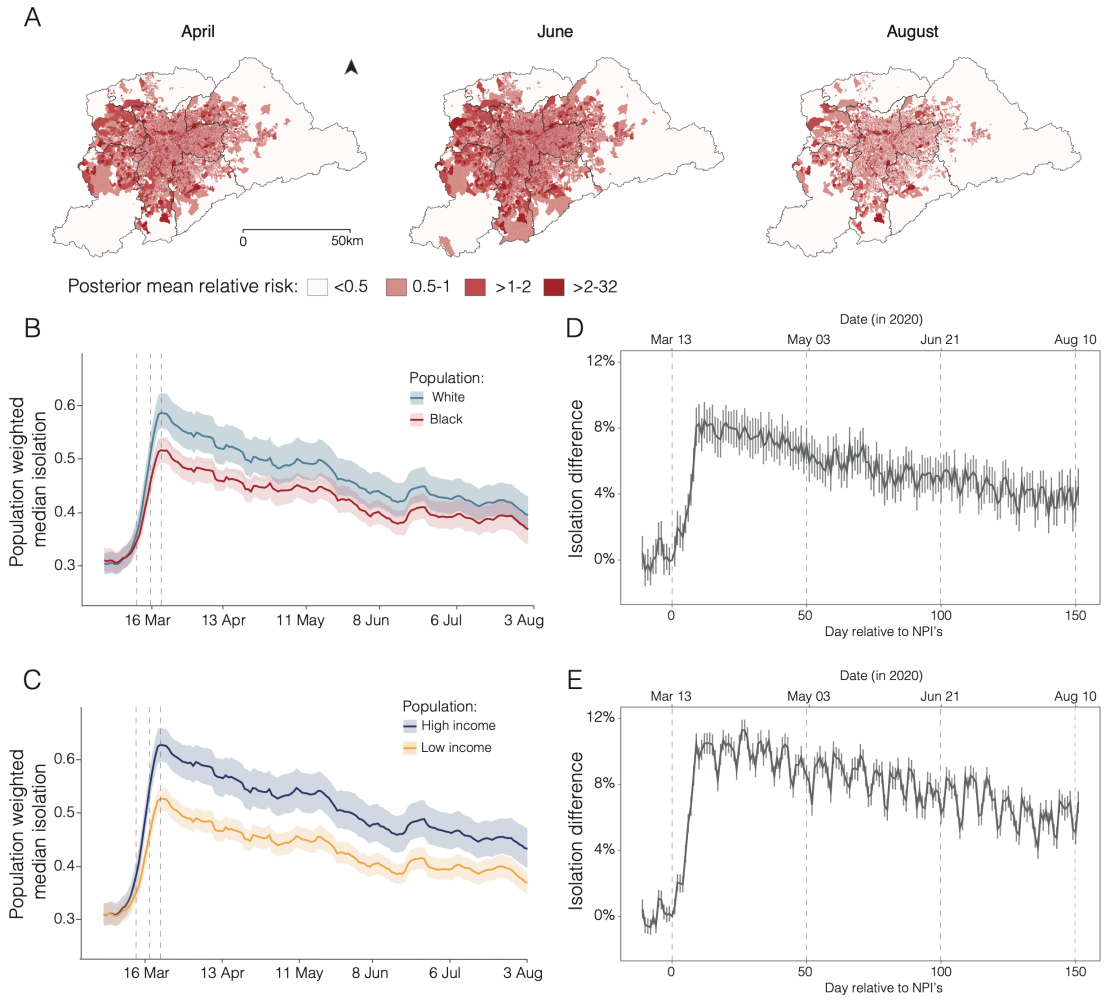

Figure 4 Differential risk based on varying ability to self-isolate in the Região Metropolitana de São Paulo (RMSP). (A) Relative risk of severe acute respiratory infection (SARI) hospitalisation for the RMSP. (B) Seven-day moving average of daily isolation levels by race. (C) Seven-day moving average of daily isolation levels by income. (D) Difference in daily social isolation by race after the introduction of non-pharmaceutical intervention (NPI). (E) Difference in daily social isolation by income after the introduction of NPIs. In panels (B) and (C), solid lines show population-weighted median isolation levels and shaded areas show population-weighted IQR (25%–75%). Dashed vertical lines indicate the dates of NPIs that enabled school closure (13 March was the state NPI) and non-essential activities (18 and 22 March, municipal and state NPIs, respectively).

and Black populations decreased to only 4.4% (95% CI 3.3% to 5.5%) 151 days after the introduction of the NPIs (figure 4D).

Finally, we investigated the differential risk to SARI based on workers' position in the labour market using data from the PNAD COVID-19 survey. After the introduction of NPIs, workers employed in low-skilled jobs or essential services were more likely to keep working face to face than workers in professional or managerial positions (online supplemental figure S4). Workers with pre-primary education were more likely to work in occupations that require in-person contact than workers with tertiary education (probability (PR): 0.89, 95% CI 0.87 to 0.90 compared with PR: 0.58, 95% CI 0.57 to 0.60) and less likely to work in occupations that allow teleworking (PR: 0.005, 95% CI 0.004 to 0.007 vs PR: 0.36, 95% CI 0.35 to 0.37, respectively) (figure 5A). When controlling for education and formal or informal employment, we found no substantial difference between racial groups in the probability of working face to face or teleworking (figure 5B). Nonetheless, because Black and Pardo populations are disproportionally employed in informal and low-skilled jobs, these racial groups were, in general, more likely to be working face to face during our study period.

Disadvantaged groups have more comorbidities

We found that population groups at risk of death from SARI were also more likely to have comorbidities known to aggravate COVID-19 severity. Compared with the population with tertiary education in São Paulo state, individuals with primary education or lower are more likely to have one or more comorbidities (OR: 1.36, 95% CI 1.27 to 1.45) (figure 5C). Similarly, Black individuals were also more likely to have one or more comorbidities than White individuals (OR: 1.29, 95% CI 1.19 to 1.39) (figure 5D). OR estimates for each health condition are summarised in online supplemental figure S5.

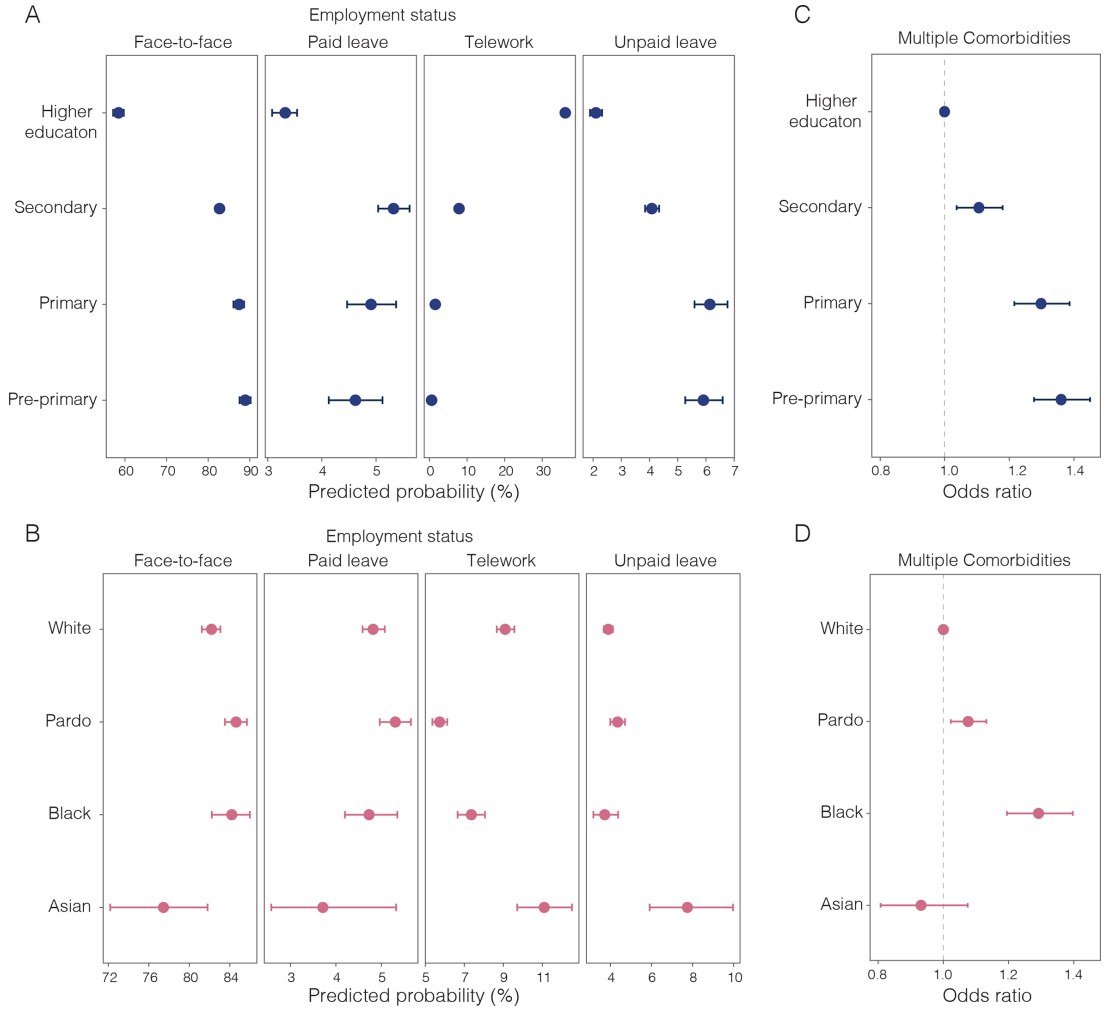

Figure 5 Inequalities in working conditions and comorbidities. (A) Probability of different working conditions by education attainment. (B) Probability of different working conditions by race. (C) OR (OR=1) of having one or more comorbidities by education attainment. (D) OR (OR=1) of having one or more comorbidities by race. Comorbidities considered include chronic obstructive pulmonary disease (COPD), diabetes, hypertension or cardiovascular disease such as infarction, angina and heart failure. Horizontal lines show 95% CIs (Source: Pesquisa Nacional por Amostra de Domicílios (PNAD) COVID-19/Instituto Brasileiro de Geografia e Estatística (IBGE),[17] July to September 2020).

DISCUSSION

Our study shows that socially disadvantaged groups are disproportionately more likely to be hospitalised and die from SARI. We find that the differential health outcomes can be explained by structural inequities linked to the incidence of comorbidities and to socioeconomic conditions, which limit the ability of low-income and non-White populations to socially isolate and reduce their access to quality health services.

Social and racial inequalities shape the risk of SARI hospitalisation and death. After the initial phase of international imports in Brazil,[23] Black or *Pardo* Brazilians and individuals residing in low-income areas were more likely to be hospitalised and die with SARI compared with White individuals and those from wealthier areas, which aligns with recent findings.[17 18] While these results report of severe cases of infection, we find similar results when looking broadly at the population level of COVID-19 infection. Our assessment of anti-SARS-CoV-2 antibodies in blood donors categorised by demographic background further confirms that Black Brazilians and those with lower socioeconomic status are disproportionately exposed to COVID-19.

Patients hospitalised in public health facilities were more likely to die than those in private health facilities. The uneven access to health services explains some but not all the inequality in the risk of death from COVID-19, since racial inequalities in death probability persist among patients within private hospitals. Potential factors influencing this inequality include higher comorbidities among poor Black patients and the lower access to private care among low-income individuals who are disproportionately Black. Other important factors include the disadvantage of having multiple comorbidities, which are more prevalent among Blacks and *Pardos* and those with lower education.

We found that hospitalisation risk is higher for populations living in municipalities that travel to and from

the RMSP, with low income per capita and high population density compared with the rest of São Paulo state. These populations mainly reside in the RMSP, which contains nearly half of the population in São Paulo state and where bias in testing is evident in regions of lower socioeconomic status[23] (online supplemental figure S6). The risk of SARI hospitalisation is particularly elevated in São Paulo city, where seroprevalence estimates from blood donors show that anti-SARS-CoV-2 antibodies were highest in older Black Brazilians and those with lower educational attainment.

We show that inequalities in the risk of SARI hospitalisation were partially explained by differential mobility responses to social distancing guidelines, similar to the UA.[35] In wealthier and predominantly White neighbourhoods, people were able to isolate more, faster, and sustain isolation for long periods of time. We also found that occupational factors played a key role in influencing individuals' ability to physically isolate. Among the working population, low-income and Black workers were less likely to receive a furlough from work or telework. Due to systemic inequalities in education and the labour market, these groups are disproportionately employed in precarious job positions with no social security and dependent on day-to-day income,[36] limiting their ability to socially distance through telework and reductions in daily mobility. The lack of capacity to self-isolate and curtail mobility in these occupations may increase exposure and facilitate SARS-CoV-2 transmission.

Our study has limitations that may have underestimated the level of inequality. First, geocoding cases may have discarded patients from poor census tracts where accuracy is limited.[37] Second, using data aggregated for various administrative levels has inherent limitations due to ecological fallacy and the modifiable areal unit problem.[38] Finally, the 2010 Brazilian population census and PNAD COVID-19 survey may have limited the capture of socioeconomic changes in the last decade and inclusion of extremely wealthy individuals.[39] Additionally, disadvantaged groups can be under-represented in health administrative records because of their lower access to healthcare. Given that São Paulo is the wealthiest state and has the most robust healthcare system in Brazil,[40] it is possible that the impact of inequalities is more severe in other states.

Our findings on the difference of SARI death risk reveal stark inequalities in access to healthcare. Health is a constitutional right and a state responsibility in Brazil.[41] The country's public healthcare system (SUS) is designed to provide universal health coverage without out-of-pocket costs, and regionally planned to improve spatial coverage of services. Yet, previous studies have found that access to COVID-19 health services tends to be lower in less developed regions in the country,[41 42] particularly among low-income and black communities.[43] Only 25% of Brazilians have access to private healthcare via health insurance, reflecting how inequality in access to quality healthcare is largely driven by income.[44] This leaves 75% of the population solely reliant on a chronically underfunded public healthcare system, which highlights a double disadvantage for low-income and non-White populations, who are more likely to be infected and deprived of care. Strengthening healthcare access and its capacity will be critical for reducing health inequities during this and forthcoming public health emergencies.[42]

Our findings on socioeconomic risk factors could help guide vaccine allocation in diverse settings to achieve equitable access. Ensuring that disadvantaged groups, especially those that have in-person occupations and live in crowded and deprived areas, receive vaccination will help prevent and slow down community transmission. While race is not a risk factor in itself, it is critical to consider systemic inequalities that lead Black and *Pardo* communities to be over-represented among low socioeconomic groups, to have higher rates of severe COVID-19 infection, and comorbidities that exacerbate their risk of death. Therefore, including disadvantaged populations among priority groups for vaccination could help reduce health inequities instead of exacerbating them.[45]

As shown in our study, the combination of social, racial and health inequalities in fostering high mortality risk among systematically marginalised communities exemplifies the syndemic nature of the COVID-19 pandemic.[3 46] The negative impact of the COVID-19 pandemic on population health is driven by the accumulation and interaction of two or more adverse conditions, often influenced by the social determinants of health.[47] Therefore, in order to reduce inequalities in COVID-19-related health risk, action is needed to address all factors that contribute to health inequalities. Prior to the start of the pandemic, these determinants have already shaped pre-existing living conditions of disadvantaged groups—such as poor education, precarious work, residential segregation and inadequate housing—which disproportionately impact their access to quality healthcare and expose them to the onset of comorbidities.[47 48] In order to reduce inequalities in COVID-19-related health risk, it is essential to understand what made them vulnerable in the first place by addressing the adverse social determinants that shape an individual's life course. While a key role is played by healthcare systems, action is also required from other stakeholders. COVID-19 control measures affect people differently based on varying levels of financial resources and support available to them, thus governments and industries must work together to address these inequalities by mobilising resources and tools[49] and by developing targeted interventions at both national[50] and local levels.[51]

Our study highlights the need for additional research to comprehend the effects of social and health inequalities during pandemics. First, an assessment of the inequality in access to quality care within public and private health facilities and its risk factors is needed to better understand mortality in different hospital settings. Second, our study shows that population response to NPIs can vary significantly based on social circumstances,

suggesting that future studies should also consider socioeconomic aspects when evaluating the effectiveness of NPIs. Third, more data are needed on whether social safety net programmes that are guaranteeing income for disadvantaged groups during the pandemic (eg, Brazil's emergency cash payment) may have enabled people to reduce their mobility. Nevertheless, our study has shown the impact of social inequities on COVID-19 hospitalisation and death, thus informing future research and policies related to the health impacts of COVID-19 in Latin America.

Author affiliations
[1]School of Geography and the Environment, University of Oxford, Oxford, UK
[2]Institute of Applied Economic Research, Brasília, Brazil
[3]Department of Electronic Systems Engineering, University of São Paulo, São Paulo, Brazil
[4]Department of Zoology, University of Oxford, Oxford, UK
[5]Department of Applied Mathematics, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil
[6]Oxford School of Global and Area Studies, Latin American Centre, University of Oxford, Oxford, UK
[7]Virology Research Center, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil
[8]Institute of Social and Political Studies (IESP), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil
[9]Departamento de Molestias Infecciosas e Parasitarias andInstituto de Medicina Tropical, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil
[10]Fundação Pró-Sangue Hemocentro de São Paulo, São Paulo, Brazil
[11]Disciplina de Ciências Médicas, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil
[12]Laboratory of Medical Investigation in Pathogenesis and Directed Therapy in Onco – Immuno – Hematology (LIM-31) HCFMUSP, University of São Paulo Medical School, São Paulo, Brazil
[13]Hospital das Clinicas da Faculdade de Medicina da Universidade de São Paulo, University of São Paulo, São Paulo, Brazil
[14]Mathematical Sciences, University of Southampton, Southampton, UK
[15]Department of Preventive Medicine, University of São Paulo Medical School, São Paulo, Brazil
[16]MRC Centre for Global Infectious Disease Analysis; and the Abdul Latif Jameel Institute for Disease and Emergency Analytics (J-IDEA), School of Public Health, Imperial College London, London, UK
[17]Oxford School of Global and Area Studies, University of Oxford, Oxford, UK

Twitter Sabrina L Li @sabrinalyli, Rafael H M Pereira @UrbanDemog and Carlos A Prete Jr @CarlosPrete1

Contributors SLL and RHMP conceived the research questions and designed the study. RHMP, CAPJ, LE, PJHA, CKVB and WMS collected the epidemiological and socioeconomic data. PSP managed mobility data while AASS curated the data on NPIs. LFB, AM, CAN, SCF, NAS and ECS collected the seroprevalence survey samples. SLL, RHMP, CAPJ, AEZ, LE, PJHA, CKVB and RJB conducted exploratory and statistical analyses and interpreted the results. AASS, WMS, RJB, IM, NG and VHN provided guidance on interpreting findings in the political, social and healthcare context of Brazil, while CHW and VHN provided guidance on statistical analysis. SLL, RHMP and CAPJ wrote the manuscript. ECS, NRF and JPM supervised the study. All authors read and revised the final manuscript.

Funding SLL is supported by the Oxford Martin Programme on Pandemic Genomics and the Canadian Social Sciences and Humanities Research Council (SSHRC) Doctoral Fellowship. CAPJ is supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Financial Code 001, Fundação Faculdade de Medicina (FFM), and the São Paulo Research Foundation (FAPESP 2019/21858-0). AEZ is supported by the Oxford Martin Programme on Pandemic Genomics. WMS is supported by the São Paulo Research Foundation (FAPESP 2017/13981-0 and 2019/24251-9). VHN is supported by the Brazilian National Council for Scientific and Technological Development (CNPq: 304714/2018-6). NRF is supported by a Wellcome Trust and Royal Society Sir Henry Dale Fellowship (204311/Z/16/Z). This project was supported by a Medical Research Council-São Paulo Research Foundation (FAPESP) CADDE partnership award (MR/S0195/1 and FAPESP 18/14389-0) (http://caddecentre.org/).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This project was approved by the Brazilian National Research Ethics Committee (CONEP CAAE-30178220.3.1001.0068).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon request. The data sets used and/or analysed during the current study are available from the corresponding authors on reasonable request.

ORCID iDs
Sabrina L Li http://orcid.org/0000-0002-1183-126X
Rafael H M Pereira http://orcid.org/0000-0003-2125-7465
Carlos A Prete Jr http://orcid.org/0000-0002-3907-423X
Alexander E Zarebski http://orcid.org/0000-0003-1824-7653
Lucas Emanuel http://orcid.org/0000-0002-3754-5755
Pedro J H Alves http://orcid.org/0000-0001-9340-030X
Pedro S Peixoto http://orcid.org/0000-0003-2358-3221
Carlos K V Braga http://orcid.org/0000-0002-6104-7297
Andreza Aruska de Souza Santos http://orcid.org/0000-0003-3585-8683
William M de Souza http://orcid.org/0000-0002-0025-8293
Rogerio J Barbosa http://orcid.org/0000-0002-6796-4547
Lewis F Buss http://orcid.org/0000-0002-9009-9301
Alfredo Mendrone http://orcid.org/0000-0002-3090-4575
Cesar de Almeida-Neto http://orcid.org/0000-0002-8490-4634
Suzete C Ferreira http://orcid.org/0000-0002-4190-0238
Nanci A Salles http://orcid.org/0000-0001-7580-4597
Izabel Marcilio http://orcid.org/0000-0002-2914-6535
Chieh-Hsi Wu http://orcid.org/0000-0001-9386-725X
Nelson Gouveia http://orcid.org/0000-0003-0625-0265
Vitor H Nascimento http://orcid.org/0000-0002-0543-4735
Ester C Sabino http://orcid.org/0000-0003-2623-5126
Nuno R Faria http://orcid.org/0000-0002-9747-8822
Jane P Messina http://orcid.org/0000-0001-7829-1272

REFERENCES

1 Ahmed F, Ahmed Na'eem, Pissarides C, *et al*. Why inequality could spread COVID-19. *Lancet Public Health* 2020;5:e240.
2 Walker PGT, Whittaker C, Watson OJ, *et al*. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* 2020;369:413–22.
3 Horton R. Offline: COVID-19 is not a pandemic. *Lancet* 2020;396:874.
4 Bambra C, Riordan R, Ford J, *et al*. The COVID-19 pandemic and health inequalities. *J Epidemiol Community Health* 2020;74:964–8.
5 Marmot M, Wilkinson R. *Social determinants of health*. OUP Oxford, 2005.
6 Bartley M. *Health inequality: an introduction to concepts, theories and methods*. John Wiley & Sons, 2016.
7 Dubey S, Biswas P, Ghosh R, *et al*. Psychosocial impact of COVID-19. *Diabetes Metab Syndr* 2020;14:779–88.
8 Emeruwa UN, Ona S, Shaman JL, *et al*. Associations between built environment, neighborhood socioeconomic status, and SARS-

CoV-2 infection among pregnant women in New York City. *JAMA* 2020;324:390–2.
9 Chen JT, Krieger N. Revealing the unequal burden of COVID-19 by income, Race/Ethnicity, and household crowding: US County versus ZIP code analyses. *J Public Health Manag Pract* 2021;27 Suppl 1, COVID-19 and Public Health: Looking Back, Moving Forward:S43–56.
10 Abedi V, Olulana O, Avula V. Racial, economic, and health inequality and COVID-19 infection in the United States. *J Racial Ethn Health Disparities* 2020:1–11.
11 UK Office for National Statistics. Coronavirus (COVID-19) related deaths by ethnic group, England and Wales - Office for National Statistics, 2020. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/coronavirusrelateddeathsbyethnicgroupenglandandwales/2march2020to10april2020 [Accessed 20 Nov 2020].
12 Niedzwiedz CL, O'Donnell CA, Jani BD, *et al*. Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank. *BMC Med* 2020;18:160.
13 Drefahl S, Wallace M, Mussino E, *et al*. A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. *Nat Commun* 2020;11:5097.
14 Dragano N, Rupprecht CJ, Dortmann O. Higher risk of COVID-19 hospitalization for unemployed: an analysis of 1,298,416 health insured individuals in Germany. *medRxiv* 2020:2020.06.17.20133918.
15 Argoty-Pantoja AD, Robles-Rivera K, Rivera-Paredez B, *et al*. COVID-19 fatality in Mexico's Indigenous populations. *Public Health* 2021;193:69–75.
16 Pan D, Sze S, Minhas JS, *et al*. The impact of ethnicity on clinical outcomes in COVID-19: a systematic review. *EClinicalMedicine* 2020;23:100404.
17 Baqui P, Bica I, Marra V, *et al*. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *Lancet Glob Health* 2020;8:e1018–26.
18 Ribeiro KB, Ribeiro AF, de Sousa Mascena Veras MA, *et al*. Social inequalities and COVID-19 mortality in the city of São Paulo, Brazil. *Int J Epidemiol* 2021:dyab022.
19 Facundo A, Chancel L, Thomas P. *World inequality report 2018*, 2017.
20 Jesus JGde, Sacchi C, Candido DdaS, *et al*. Importation and early local transmission of COVID-19 in Brazil, 2020. *Rev Inst Med Trop Sao Paulo* 2020;62:e30.
21 Henriques CMP, Vasconcelos W, *et al*. Crises dentro dA crise: respostas, incertezas E desencontros no combate à pandemia dA Covid-19 no Brasil. *Estud Av* 2020;34:25–44.
22 Rezende LFM, Thome B, Schveitzer MC, *et al*. Adults at high-risk of severe coronavirus disease-2019 (Covid-19) in Brazil. *Rev Saude Publica* 2020;54:50.
23 de Souza WM, Buss LF, Candido DdaS, *et al*. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nat Hum Behav* 2020;4:856–65.
24 Secretaria da saúde do estado de São Paulo. SIMI-SP: Pacientes internados POR Síndrome Respiratória Aguda Grave (SRAG). Available: https://www.saopaulo.sp.gov.br/planosp/simi/ [Accessed 25 Oct 2020].
25 Niquini RP, Lana RM, Pacheco AG, *et al*. Srag POR COVID-19 no Brasil: descrição E comparação de características demográficas E comorbidades CoM SRAG POR influenza E CoM a população geral. *Cad. Saúde Pública* 2020;36:e00149420.
26 Hasell J, Mathieu E, Beltekian D, *et al*. A cross-country database of COVID-19 testing. *Sci Data* 2020;7:345.
27 Telles E, Paschel T, Black WI. Who is black, white, or mixed race? how skin color, status, and nation shape racial classification in Latin America. *Am J Sociol* 2014;120:864–907.
28 IBGE - Instituto Brasileiro de Geografia e Estatística. *Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD) COVID-19. Microdados*, 2020.
29 Marmot M, Friel S, Bell R, *et al*. Closing the gap in a generation: health equity through action on the social determinants of health. *The Lancet* 2008;372:1661–9.
30 Buss LF, Prete CA, Abrahim CMM, *et al*. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* 2021;371:288–92.
31 Diggle PJ. Estimating prevalence using an imperfect test. *Epidemiol Res Int* 2011;2011:1–5.
32 Peixoto PS, Marcondes D, Peixoto C, *et al*. Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. *PLoS One* 2020;15:e0235732.
33 de Souza Santos AA, Candido DdaS, de Souza WM, *et al*. Dataset on SARS-CoV-2 non-pharmaceutical interventions in Brazilian municipalities. *Sci Data* 2021;8:73.
34 Weill JA, Stigler M, Deschenes O, *et al*. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proc Natl Acad Sci U S A* 2020;117:19658–60.
35 Chang S, Pierson E, Koh PW, *et al*. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 2021;589:82–7.
36 Lustig N, Pabon VM, Sanz F. The impact of COVID-19 Lockdowns and expanded social assistance on inequality, poverty and mobility in Argentina, Brazil, Colombia and Mexico. ECINEQ, Society for the study of economic inequality, 2020. Available: https://ideas.repec.org/p/inq/inqwps/ecineq2020-558.html [Accessed 20 Nov 2020].
37 Giest S, Samuels A. 'For good measure': data gaps in a big data world. *Policy Sci* 2020;53:559–69.
38 Duranton G, Overman HG. Testing for localization using Micro-Geographic data. *Rev Econ Stud* 2005;72:1077–106.
39 de SP. A distribuição de renda nas pesquisas domiciliares brasileiras: harmonização E comparação entre Censos, PNADs E POFs. *Rev Bras Estud Popul* 2015;32:165–88.
40 Paim J, Travassos C, Almeida C, *et al*. The Brazilian health system: history, advances, and challenges. *The Lancet* 2011;377:1778–97.
41 Castro MC, Massuda A, Almeida G, *et al*. Brazil's unified health system: the first 30 years and prospects for the future. *The Lancet* 2019;394:345–56.
42 Castro MC, de CLR, Chin T. Demand for hospitalization services for COVID-19 patients in Brazil. *medRxiv* 2020:2020.03.30.20047662.
43 Pereira RHM, Braga CKV, Servo LM, *et al*. Geographic access to COVID-19 healthcare in Brazil using a balanced float catchment area approach. *Soc Sci Med* 2021;273:113773.
44 Matijasevich A, Russo G. Covid-19 in Brazil has exposed socio-economic inequalities and underfunding of its public health system. *BMJ* 2020 https://blogs.bmj.com/bmj/2020/06/19/covid-19-in-brazil-has-exposed-deeply-rooted-socio-economic-inequalities-and-chronic-underfunding-of-its-public-health-system/
45 Emanuel EJ, Persad G, Kern A, *et al*. An ethical framework for global vaccine allocation. *Science* 2020;369:1309–12.
46 Singer M, Bulled N, Ostrach B, *et al*. Syndemics and the biosocial conception of health. *Lancet* 2017;389:941–50.
47 Abrams EM, Szefler SJ. COVID-19 and the impact of social determinants of health. *Lancet Respir Med* 2020;8:659–61.
48 Barber S, Diez Roux AV, Cardoso L, *et al*. At the intersection of place, race, and health in Brazil: residential segregation and cardio-metabolic risk factors in the Brazilian longitudinal study of adult health (ELSA-Brasil). *Soc Sci Med* 2018;199:67–76.
49 Ismail SJ, Tunis MC, Zhao L, *et al*. Navigating inequities: a roadmap out of the pandemic. *BMJ Glob Health* 2021;6:e004087.
50 Shadmi E, Chen Y, Dourado I, *et al*. Health equity and COVID-19: global perspectives. *Int J Equity Health* 2020;19:104.
51 Berkowitz RL, Gao X, Michaels EK, *et al*. Structurally vulnerable neighbourhood environments and racial/ethnic COVID-19 inequities. *Cities Health* 2020;31:1–4.

Higher risk of death from COVID-19 in low-income and non-White populations of São Paulo, Brazil

Supplementary Materials

Sabrina L Li, MSc [1*#], Rafael H M Pereira, DPhil [2*#], Carlos A Prete Jr, MSc [3*], Alexander E Zarebski, PhD [4], Lucas Emanuel, PhD [2], Pedro JH Alves, MSc [2], Pedro S Peixoto [5], PhD, Carlos K Braga, MSc [2], Andreza Aruska de Souza Santos, PhD [6], William M de Souza, PhD [4,7], Rogerio J Barbosa, PhD [8], Lewis F Buss, MD [9], Alfredo Mendrone-Junior, MD [10], Cesar de Almeida-Neto, MD [10,11], Suzete C Ferreira, PhD [10,12], Nanci A Salles, BSc [10,12], Izabel Marcilio, MD [13], Chieh-Hsi Wu, PhD [14], Nelson Gouveia, MD [15], Vitor H Nascimento, PhD [3], Ester C Sabino, MD [9] , Nuno R Faria, PhD [4,9,16], Jane P Messina, PhD [1,17]

1. School of Geography and the Environment, University of Oxford, Oxford, United Kingdom.
2. Institute for Applied Economic Research, Brasília, Brazil.
3. Department of Electronic Systems Engineering, University of São Paulo, São Paulo, Brazil.
4. Department of Zoology, University of Oxford, Oxford, United Kingdom.
5. Department of Applied Mathematics, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil.
6. Oxford School of Global and Area Studies, Latin American Centre, University of Oxford, Oxford, United Kingdom.
7. Virology Research Center, University of São Paulo, Ribeirão Preto, Brazil.
8. Institute of Social and Political Studies (IESP), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil.
9. Departamento de Molestias Infecciosas e Parasitarias & Instituto de Medicina Tropical da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.
10. Fundação Pró-Sangue Hemocentro de São Paulo.
11. Disciplina de Ciências Médicas, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.
12. Laboratory of Medical Investigation in Pathogenesis and Directed Therapy in Onco-Immuno – Hematology (LIM-31) HCFMUSP, University of São Paulo Medical School, São Paulo, Brazil.
13. Epidemiologic Surveillance Center, Hospital das Clinicas – University of São Paulo Medical School, São Paulo, Brazil.
14. Mathematical Sciences, University of Southampton, Southampton, United Kingdom.
15. Department of Preventive Medicine, University of São Paulo Medical School.

16. MRC Centre for Global Infectious Disease Analysis; and the Abdul Latif Jameel Institute for Disease and Emergency Analytics (J-IDEA), School of Public Health, Imperial College London, London, United Kingdom.
17. Oxford School of Global and Area Studies, University of Oxford, Oxford, United Kingdom.

*These authors contributed equally.

#Corresponding authors:

1. Sabrina L. Li, School of Geography and the Environment, University of Oxford, South Parks Rd, Oxford, OX1 3QY, United Kingdom. E-mail: lisabrinaly@gmail.com

2. Rafael H.M. Pereira, Institute for Applied Economic Research (Ipea), SBS - Quadra 1 - Bloco J - Ed. BNDES - CEP: 70076-900 - Brasília - DF – Brasil. E-mail: rafael.pereira@ipea.gov.br

Table of Contents

Table of Figures

Appendix A: Methodology details

1 Probability of hospitalisation and death

Probability of hospitalisation and death due to COVID-19 were calculated by income quantile, race or administrative type of the health facility (public or private). The same method was used to calculate the estimates for the different grouping criteria (i.e., income quantile, race, administrative type of the health facility). We use the term 'class' to refer to a specific subgroup for each of these criteria, i.e., the higher income class, in the case of analysis by income, or the public health facility class in case of analysis by type of health facility. It is worth noting that only hospitalisations or deaths contained in the SIVEP-Gripe dataset are considered SARI cases. We also discarded cases that were confirmed to be the results of other etiological agents.

The health facility where each patient was hospitalized is defined for all SARI patients in the SIVEP-Gripe dataset. Administrative types of health facilities were extracted from the National Registry of Health Facilities (CNES) database. We classified as "Public" all health facilities whose administrative type is "SUS" (Brazilian public health system), as "Private" all private health facilities, regardless of whether they accept patients with health insurance or not, and as "Other" all non-private health facilities whose administrative type is not marked as "SUS". In most cases, these are public health facilities that accept patients with health insurance. Health facilities with missing administrative type were discarded. 26·7% of all SARI patients were notified by health facilities classified as public, compared to 43·6% for private, 5.4% for "Other" and 21·7% for health facilities with missing administrative type.

Missing races are imputed based on the racial distribution of the corresponding census tracts. It is important to impute missing races instead of excluding them to avoid racial bias because different races may have different probabilities of being missing. Figures S1 shows SARI hospitalisation and death probabilities computed without race imputation. Even though the results with race imputation are more robust, similar probabilities of death and hospitalization are obtained if race imputation is not employed.

1.1 Probability of death

Let w_{ik} be the probability of the patient i belonging to a given class k in the grouping criterion of interest (income, race or administrative type of the health facility). When comparing private and public hospitals, w_{ik} can only be 1 or 0 because the health facility type is known for all patients. Similarly, w_{ik} can only be 1 or 0 for individuals whose races are not missing.

In the case of incomplete information, that is, individuals with unknown race and for the income comparisons, we assign a distribution over the possible values using estimates obtained from census data for each census tract. The probabilities w_{ik} are assigned as the proportion of the individuals in the corresponding census tract that belong to class k (that is, w_{ik} is given by the racial or income distribution of the census tract where the individual lives). In order to compute the probabilities of death and SARI for a given class, the weights w_{ik} are used to divide the contribution of each individual i between all classes k for $k = 1, \dots, K$, where K is the number of classes. Individuals whose census tract is not known were excluded from the income analysis, but only individuals with both unknown race and census tract were excluded from the racial analysis.

Let us denote the probability of an event A as $P(A)$. The probability of death given class, age, and sex $P(\text{death}|\text{class}, \text{age}, \text{sex})$ was estimated by dividing the number of SARI deaths for a given class and age-sex group by the number of SARI cases for the same class, age, and sex group. The probability of death given class (among SARI patients) can be obtained through

$$P(\text{death}|\text{class}) = \sum_{\text{age,sex}} P(\text{death}|\text{class}, \text{age}, \text{sex}) P(\text{age}, \text{sex}|\text{class}),$$

where $P(\text{age}, \text{sex}|\text{class})$ is the conditional age-sex distribution of SARI patients given their class. This distribution may differ significantly for different classes due to differences in the age composition of classes, especially when race is used as class: The proportion of residents of the state of São Paulo that are older than 70 years old is 6·1% for White, 2·9% for *Pardo*, 4·4% for Black and 11·4% for Asians according to the 2010 census. Since the probability of death depends heavily on age, in order to allow a fairer comparison of the burden of the disease according to income and race, we use an age-sex normalised death probability given class, which is the death probability given class that would be obtained if all classes had the same age-sex distribution. If this age-sex normalisation was not applied, the probability of death would be overestimated for white patients and underestimated for black patients, as the white population is older than the black population. This normalised distribution is obtained by substituting $P(\text{age}, \text{sex}|\text{class})$ by $P(\text{age}, \text{sex})$, the proportion of patients from an age-sex group regardless of the class:

$$P_{\text{normalised}}(\text{death}|\text{class}) = \sum_{\text{age,sex}} P(\text{death}|\text{class}, \text{age}, \text{sex}) P(\text{age}, \text{sex})$$

This age-sex standardisation was not employed for the health facility type because patients from public and private hospitals share similar age-sex distributions. The age of the patients was discretized into the following groups: 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+. It is worth noting that weighting estimates that depend on age to produce an age-standardised estimate was also done in other similar studies [1].

1.2 Probability of hospitalisation by SARI

The probability of hospitalisation by severe acute respiratory infection (SARI) given each class for each criterion (only income or race in this analysis) is computed for each age-sex group as the ratio between the number of recorded SARI cases for a given class, age and sex and the number of individuals of that class and age-sex group. The age-sex distributions of for population of São Paulo was extracted from the 2010 census. The same age-sex standardisation employed to compute the probabilities of death was used to estimate the probabilities of SARI, but as in the 2010 census, patients above 70 years old were aggregated into the same age group. Only the race and income quantile classes are used in this analysis.

1.3 Confidence intervals

Confidence intervals were estimated through bootstrapping using 1 000 realizations. For each realization, the same number of individuals in the dataset were randomly selected with replacement from the set of individuals in the dataset. Then, the desired estimates and odds ratios are obtained from each resampled dataset. For the racial analysis each selected individual i with unknown race had its race

imputed for each bootstrap iteration such that the probability of the imputed race being k is w_{ik}. The confidence intervals were obtained from the quantiles of the set of 1 000 bootstrapped parameters.

2 Geospatial analysis

2.1 Model description

We used a Bayesian hierarchical model to compute the relative risk of hospitalisation at the municipality level for São Paulo state (n= 645) and at the census tract level for the greater metropolitan area of São Paulo (GMSP: n=30 815). The number of observed cases Y_i in an area i is modelled using a Poisson distribution $Y_i \sim Poisson(\lambda_i)$ with mean $\lambda_i = E_i\,\mu_i$ where E_i is the expected number of cases in area i under a null model in which cases are uniformly distributed among the population, i.e., the number of cases in a given area is proportional to the population of that area. For each area i, this is given by $E_i = \frac{\sum_i Y_i}{\sum_i pop_i} \times pop_i$, where pop_i is the population in area i. The factor of μ_i describes the area-specific risk and models the additional variation in the observation process [2].

To quantify the uncertainty in the point estimates of the mean relative risk estimates, we mapped the posterior probability of elevated relative risk in each area (Appendix B - Figure S7). This is the posterior probability that a tract has an elevated risk of observing cases, formally P(μ_i > 1| data). For instance, a posterior probability of 0.6 in an area indicates a 60% chance that this area is at greater risk of observing cases.

2.2 Modelling relative risk

We fit a log-linear model to estimate the relative risk μ_i, which is modelled as the sum of an intercept and random effects. Random effects are broken into the spatial (A_i) and temporal components (B_i), as shown in Eq. (1·1):

$$\log(\mu_i) = \alpha + A_i + B_i \quad (1\cdot1)$$

$$\log(\mu_i) = \alpha + U_i + V_i + \gamma_t + \phi_t \quad (1\cdot2)$$

To account for existing spatial autocorrelation, we used a Besag-York-Mollié model (BYM) [3] to separate the spatial component into spatially structured U_i, and non-spatial, unstructured random effects, V_i, so ($A_i = U_i + V_i$), as shown in Eq. 1·2. In the BYM model, a conditional autoregressive (CAR) process is used to introduce correlation among the U_i for each tract. Given the U_i of neighbouring areas, the U_i has a normal distribution with mean equal to the average of the neighbours' U_i, and variance $s_i^2 = \frac{1}{\#N(i)\tau_U}$ where $\#N(i)$ is the number of areas that share boundaries with area i and τ_U is a precision parameter. The random effect, V_i follows a zero mean normal distribution with precision parameter, τ_V=$\frac{1}{\sigma_{v^2}}$ (where σ_{V^2} is the variance). Both random effects in the model capture extra-Poisson variability, and were expressed as the following:

$$U_i|\, U_{j\neq i} \sim \mathcal{N}\left(m_i, s_i^2\right), \quad V_i \sim \mathcal{N}(0, \sigma_v^2)$$

$$m_i = \frac{\sum_{j \in N(i)} U_j}{\#N(i)} \quad , \quad s_i^2 = \frac{\sigma_U^2}{\#N(i)} = \frac{1}{\#N(i)\tau_U}$$

To account for temporal structure in the data, we included the random effect ($B_t = \gamma_t + \phi_t$), which assumes that the number of cases observed in a given area depends on the number of cases observed in the given area in the previous month and a residual [2,4]. The temporal component includes γ_t, a temporally structured effect modelled dynamically using a random walk of order 1, and an unstructured temporal effect ϕ_t to account for independent time effects, which follows a zero mean normal distribution. Both are expressed as the following:

$$\gamma_t | \gamma_{t-1} \sim \mathcal{N}\left(\gamma_{t-1}, \frac{1}{\tau_\gamma}\right)$$
$$\phi_t \sim \mathcal{N}\left(0, \frac{1}{\tau_\phi}\right)$$

We adopted minimally informative prior distributions in R-INLA [3]. The log of the precision parameters adopted for the spatial effects, τ_U and τ_V , follows a gamma distribution with shape 1 and rate 0·0005. The precision parameter for both the structured and unstructured temporal effects τ_γ and τ_ϕ also follows a gamma distribution, with shape 1 and rate 0·001. The prior default distributions in R-INLA, which are the recommended settings [4], were also used for the precision parameters of both U_i, V_i, γ_t, and ϕ_t.

2.3 Ecological regression

To evaluate the effects of socioeconomic covariates on the risk of hospitalisation at the municipality level, we reformulated our model expressed by Eq. 1·3 by adding a fixed effect, which we refer as the ecological regression model. The log of the relative risk is given by the following:

$$\log(\mu_i) = \alpha + U_i + V_i + \gamma_t + \phi_t + X'_{ik}\,\beta \quad (1{\cdot}3)$$

where X'_{ik} is the i th row and kth column of covariates matrix X based on the socioeconomic covariates for each municipality and month, and β is the regression parameter modelled as fixed effects with normal priors ($\beta \sim \mathcal{N}(0, 100)$) .

Here, we define U_i using a graphical structure in R-INLA which describes the connections between municipalities by looking at estimates of the level of human mobility between them in the state of São Paulo. We considered two municipalities to be connected if there were at least 550 journeys between them. This threshold was selected to ensure that the sparsity of the connectivity matrix was similar to the nearest neighbour matrix described in Section 3·2. The number of origin-destination journeys between municipalities are retrieved from processed mobile geo-location data obtained from the In Loco company described elsewhere [5]. The prior default distributions in R-INLA were used for the precision parameters of both U_i, V_i, γ_t, and ϕ_t.

2.4 Computational method

We carried out model fitting with R-INLA which uses an Iterated Nested Laplace Approximation (INLA) based on a combination of analytical approximations and numerical integration to estimate

posterior distributions [6]. INLA was designed as an efficient alternative to Markov Chain Monte Carlo (MCMC), which is both computationally and time-intensive when applied to a large amount of data. It can be suitably applied to latent Gaussian models including generalised linear models to spatial and spatio-temporal models.

2.5 Model evaluation

We evaluated our model by plotting the empirical relative risk in each area against the fitted risk determined by our model (Figure A1). The empirical relative risk was calculated by weighting the total number of observed cases in a given municipality, with municipality level population as a proportion of the entire state of São Paulo (defined as offset). A density plot illustrating the distribution of empirical versus predicted risk was also created to assess model fit (Figure A2).

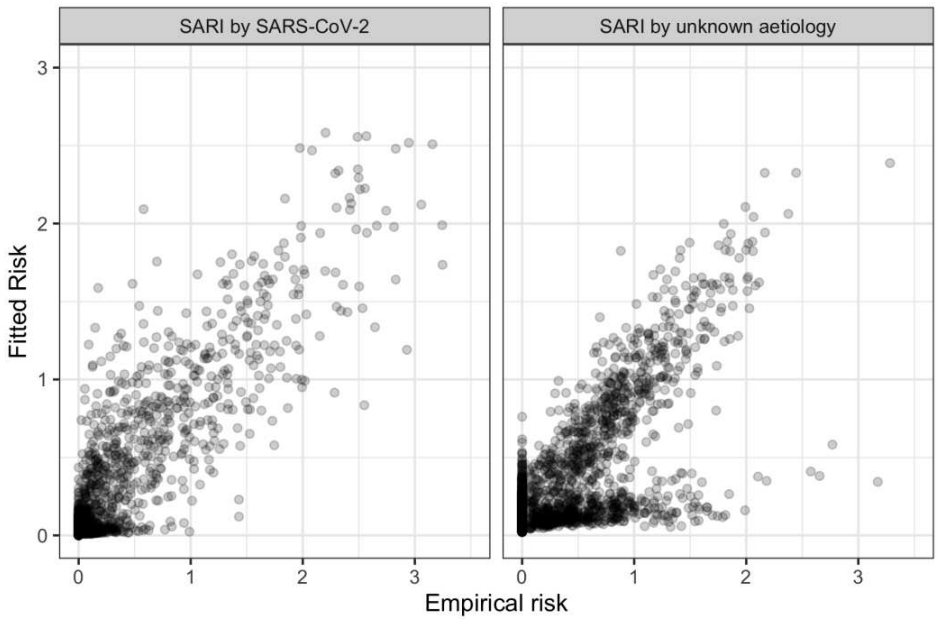

Figure A1. Empirical vs. fitted risk

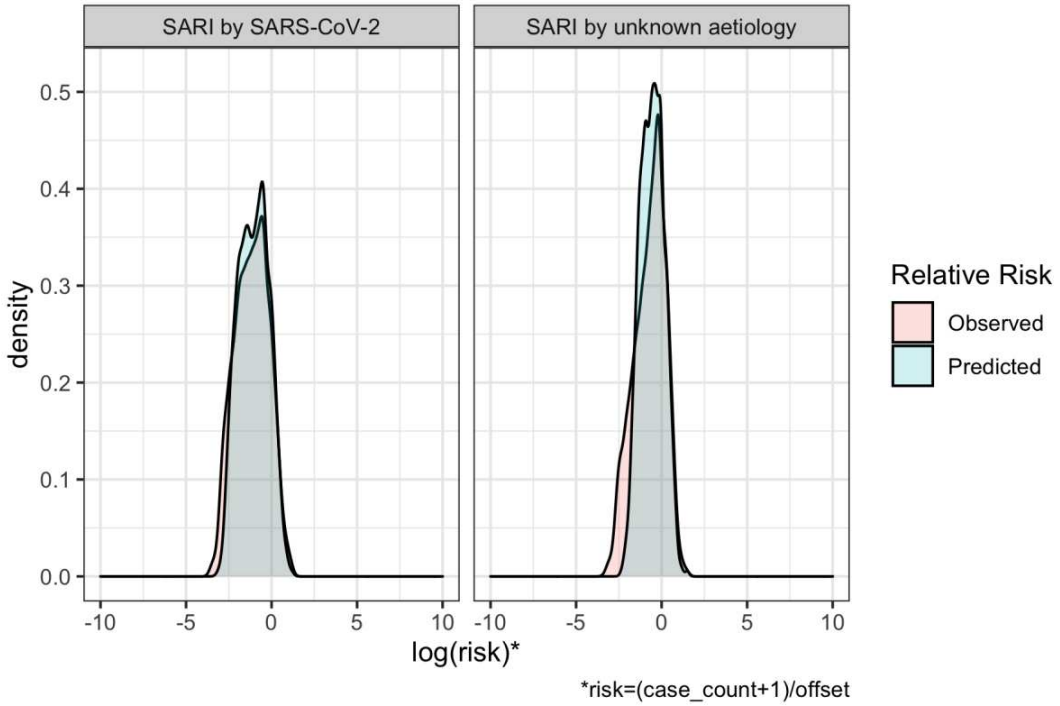

Figure A2. The estimated value of the log-risk for each area closely matches the observed risk in each area.

2.6 Covariates

2.6.1 2010 census

Data on socioeconomic covariates used in the ecological regression at the municipality level were obtained from the 2010 population census compiled by the Brazilian Institute of Geography and Statistics (IBGE) [7].These data include the average household income per capita, housing density (number of private households/km^2), sanitation conditions (proportion of population with access to piped water, access to sewage network, and/or access to septic tank), percentage of urban and rural population, and income inequality measured with the Gini index.

To our knowledge, this is the most complete and recent socioeconomic dataset available for Brazil at our required spatial resolution since no census was carried out in 2020. To assess the variability of some of these covariates over a ten year period, we obtained the same socioeconomic data from the 2000 IBGE census. We compared the same covariates by plotting them at the municipality level for a small sample (in this case, we decided to extract the 39 municipalities in RMSP) (Figure A3). We observed a general trend of increase in income per capita, household density, and number of residents per household, at the municipality level, between 2000 and 2010. Given the inconsistent variation in population access to piped water, septic tank, and sewage network by municipality over the 10-year time frame, we decided to omit these variables from the analysis.

To improve accuracy, we decided to compute population density (people/km^2) at the census tract level, then aggregating it for each municipality, instead of using household density. This was calculated by dividing the total resident population by the area of each census tract using data retrieved from the 2010

census. For large census tracts (with area greater than 0·12 km^2), we considered the occupied area rather than the total area of the census tract. This was done to improve the accuracy of population density estimates in large census tracts, particularly for those located in rural areas, by focusing only on human habitats. Human occupied areas were identified based on population counts from a fine regular grid of 200 meters, which was generated by IBGE for the 2010 census using fieldwork and satellite imagery data.

We checked for multicollinearity by assessing the correlation between variables and computing the variation inflation factor (VIF) of each covariate. Based on these tests, we removed the proportion of informal workers and unemployment from our final model to avoid multicollinearity with the variables on household density and income per capita.

2.6.2 Distance to the nearest health facility

We have also computed mean distance to the nearest health facility of each municipality. To do this, we calculated the road network distance from the centroid of each census tract to the nearest healthcare facility in R with the dodgr package [8]. We considered all the 830 healthcare facilities registered in the SIVE-Gripe database in São Paulo state that hospitalised SARI patients via the Unified Health System (SUS).

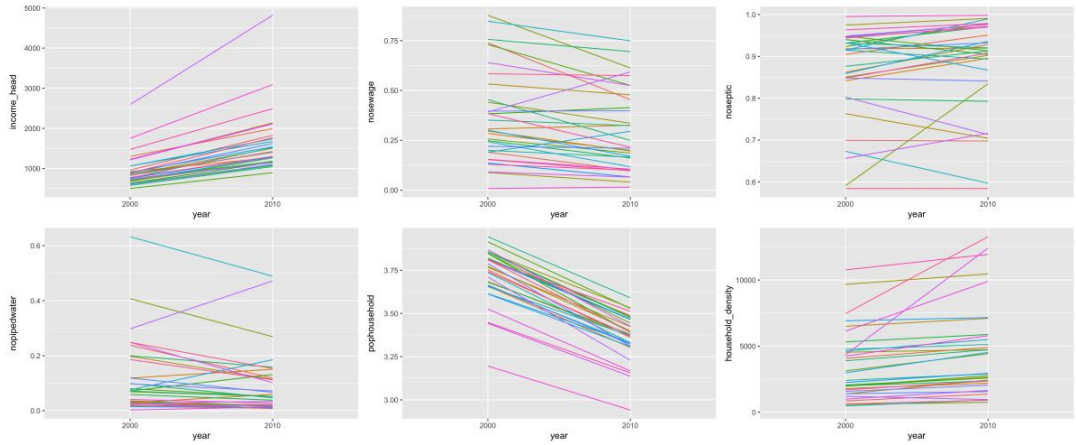

Figure A3. Trends in covariate values between 2000 and 2010. Each coloured line represents a municipality within the GMSP (n= 39).

2.7 Spatial joins for the GMSP

To improve the accuracy of modelling small areas for the RMSP, we performed spatial joins of census tracts to ensure that each census tract had at least one SARI case. Spatial joins were performed by selecting the census tract with the smallest population and joining it to its nearest neighbour (determined by the shortest distance between centroids). This was repeated until all tracts had at least one case of SARI.

3 Probability of working conditions

We used PNAD COVID-19 data from May to September 2020. During this period, 1 888 560 individuals were interviewed, of whom 171 480 were living in São Paulo state. For each individual, the PNAD COVID-19 survey also collects self-reported data on access to COVID-19 testing and comorbidities. From July, August, and September 2020, interviewees were also asked about comorbidities.

In the PNAD COVID-19 survey, work status is a categorical variable that can assume four values: face-to-face, telework, paid leave, and unpaid leave. We model the conditional probabilities using a multinomial logistic regression.

$$\mathrm{P}(y_i = m|\boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta}_m)}{\sum_{j=1}^{4}\exp(\boldsymbol{x}_i\boldsymbol{\beta}_j)}$$

$\mathrm{P}(y_i = m|\boldsymbol{x}_i)$ represents the probability that, for the individual i, the variable Work Status (y) will assume a particular value m, given a $k \times 1$ vector $\boldsymbol{x}_i$ of explanatory variables and an intercept. The letter j varies from 1 to 4 and indexes the four categories of Work Status. $\boldsymbol{\beta}_j$ is a vector of coefficients for the category j. The reference category ($j = 1$) is Face-to-Face work and, by construction, $\boldsymbol{\beta}_1 = \mathbf{0}$. So, the estimated vector of parameters $\widehat{\boldsymbol{\beta}}$ is $3k \times 1$ dimensional:

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{\boldsymbol{T}} \\ \widehat{\boldsymbol{\beta}}_{\boldsymbol{P}} \\ \widehat{\boldsymbol{\beta}}_{\boldsymbol{U}} \end{bmatrix}$$

Where $\underbrace{\widehat{\boldsymbol{\beta}}_{\mathrm{T}}}_{k\times 1} = \widehat{\boldsymbol{\beta}}_{\text{Telework|Face-to-face}}$, $\underbrace{\widehat{\boldsymbol{\beta}}_{\mathrm{P}}}_{k\times 1} = \widehat{\boldsymbol{\beta}}_{\text{Paid leave|Face-to-face}}$, and $\underbrace{\widehat{\boldsymbol{\beta}}_{\mathrm{U}}}_{k\times 1} = \hat{\beta}_{\text{Unpaid leave|Face-to-face}}$.

We estimated four different nested models, by adding explanatory variables stepwise. Model 0 has race, sex (a dummy variable) and age (a 3rd degree orthogonal polynomial). Model 1 adds education, Model 2 adds occupation (ISCO-08 1-digit groups – see Table **A2**) and a dummy variable for (in)formality, Model 3 adds dummy variables for the months the observations were collected in PNAD COVID-19.

In order to calculate predicted probabilities, we let the variable of interest (e.g., race) to vary and set all other explanatory variables to their grand mean. This way, all the probability differences between categories (e.g. White, Black, *Pardo*, and Asian) are due to changes in the variable of interest.

Models were estimated taking into account the Complex Sample Design of PNAD COVID-19. So, the Variance-Covariance Matrix of the Coefficients already take into account heteroskedasticity and autocorrelation between observations. Confidence intervals for the predicted probabilities were calculated by parametric bootstrapping. By the Central Limit Theorem, the estimated coefficients follow a Multivariate Normal Distribution:

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\beta})$$

Using $\widehat{\boldsymbol{\beta}}$ as the mean and $\underbrace{\widehat{\boldsymbol{\Sigma}}_{\beta}}_{3k\times 3k}$ as the variance-covariance matrix, we simulated 2 000 samples of coefficients. We organized the simulated coefficients for each category of dependent variable in separate matrices: $\underbrace{\widehat{\mathbf{B}}_{\mathrm{T}}^{sim}}_{2000\times k}$, $\underbrace{\widehat{\mathbf{B}}_{\mathrm{P}}^{sim}}_{2000\times k}$, and $\underbrace{\widehat{\mathbf{B}}_{\mathrm{U}}^{sim}}_{2000\times k}$. For a given vector of observations $\boldsymbol{x}$, a distribution of fitted values can be calculated as:

$$\underbrace{\boldsymbol{x}}_{1\times k}\underbrace{\left(\widehat{\boldsymbol{\beta}}_{\mathrm{T}}^{sim}\right)^{T}}_{k\times 2000} = \underbrace{\widehat{\boldsymbol{\mu}}_{\mathrm{T}}^{sim}}_{1\times 2000}$$

$$\underbrace{\boldsymbol{x}}_{1\times k}\underbrace{\left(\widehat{\boldsymbol{\beta}}_{\mathrm{P}}^{sim}\right)^{T}}_{k\times 2000} = \underbrace{\widehat{\boldsymbol{\mu}}_{\mathrm{P}}^{sim}}_{1\times 2000}$$

$$\underbrace{\boldsymbol{x}}_{1\times k}\underbrace{\left(\widehat{\boldsymbol{\beta}}_{\mathrm{U}}^{sim}\right)^{T}}_{k\times 2000} = \underbrace{\widehat{\boldsymbol{\mu}}_{\mathrm{U}}^{sim}}_{1\times 2000}$$

And, by definition $\underbrace{\widehat{\boldsymbol{\mu}}_{\mathrm{F}}^{sim}}_{1\times 2000} = \underbrace{\mathbf{0}}_{1\times 2000}$. A distribution of predicted probabilities given $\boldsymbol{x}$ is given by:

$$\underbrace{\mathbf{P}(y_i = m|\boldsymbol{x}_i)}_{1\times 2000} = \frac{\exp\left(\widehat{\boldsymbol{\mu}}_m^{sim}\right)}{\sum_{j=1}^{4}\exp\left(\widehat{\boldsymbol{\mu}}_j^{sim}\right)}$$

A 95% confidence interval is obtained if we take the quantiles 0,025 and 0,975 of this vector of predicted probabilities.

4 Event-study analysis

We used an event study model to investigate how people from different socioeconomic groups changed their daily isolation levels after the implementation of state non-pharmaceutical intervention (NPI). The model was conducted using an ecological analysis where both socioeconomic characteristics of the population and daily isolation levels are spatially aggregated on the H3 hexagonal grid at resolution 8. Each cell has an edge of approximately 460 meters and an area of 0·74 km^2. Hexagonal H3 cells were then ranked by income based on quintiles of average income per capita. Cells were also categorised as predominantly White when at least 60% of the population self-declared White and predominantly Black when at least 60% of the population self-declared Black or *Pardo*.

The racial composition and income level of each cell were determined using dasymetric interpolation of the 2010 census tract data in two steps. First, data on income and race were passed to a finer regular grid of 200 meters and linked with population count by finding the aerial intersection and population size of each cell. This was reaggregated from the regular grid to the hexagonal grid. Hexagonal H3 cells were then ranked by income based on quintiles of average income per capita. Cells were categorised as predominantly Black when at least 60% of the population self-declared Black or *Pardo*, and likewise for White.

In the event study model, our treated group is composed of hexagons predominantly of White population (race analysis) and the 20% wealthiest hexagons (income analysis). Conversely, the comparison groups were composed of hexagons with predominantly Black population and the 20% poorest hexagons. Our specification includes indicators for pre-and post-treatment effects, as follows:

$$Y_{id} = \left[\sum_{\tau=-12}^{-2} \beta_\tau I(t_{id} - t^* = \tau) + \sum_{\tau=0}^{151} \beta_\tau I(t_{id} - t^* = \tau)\right] + X'_{id}\Theta + \omega_d + \mu_i + \varepsilon_{id}$$

where Y_{id} is the outcome (daily isolation level) observed for hexagon i at day d; the indicator $I(t_{id} - t^* = \tau)$ measures the time (in days) relative to the day of the state NPI implementation on date t^*. We set the coefficient β_{-1} (March 12) equal to zero to use the day immediately prior to the state NPI implementation as the reference. X'_{id} represents the set of hexagons covariates: a dummy variable indicating the beginning of NPI flexibilization period in each municipality, and a time-varying variable with the number of days relative to the first confirmed case of SARI in each hexagon (equates to 0 for days prior to the first case). μ_i is a hexagon fixed effect that controls non-parametrically for time-invariant hexagon factors, such as hexagons' fixed geographical aspects (e.g., urban infrastructure, proximity to healthcare facilities, urban density etc), ω_d is a day fixed effect that controls non-parametrically for aggregate shocks and other policies common to all hexagons at a specific moment in time, and ε_{id} is an idiosyncratic error term. All observations are weighted by the size of resident population in each hexagon [7]. Finally, we clustered standard errors at the hexagon level to make estimations robust to serial correlation and heteroskedasticity [9].

This method allows us to more formally test for pre-trends in outcome variables in the pre-period. The identifying assumption is that the time trend in the mobility level in treated areas would have a similar trend as the one observed in similar nontreated areas in the absence of the policy intervention. Coefficient estimates of β_τ; with $\tau < 0$ (representing the change, in percentage points, in the outcome each day pre-intervention) serves as a direct test of the plausibility of the identifying assumption. If hexagons have similar trends before the date of state declaration and diverge only after policy, it provides strong evidence that such changes were caused by the state NPI adoption rather than an unobservable factor.

4.1 Summary statistics

Table 1 presents summary statistics for the main variables used in our analysis. Columns (1)-(3) shows the number of observations, the mean and the standard deviation for the treated group, while columns (4)-(6) exhibit the same statistics for the comparison group. Panel A shows the summary statistics for the race estimation sample, while Panel B presents the summary statistics for the income estimation sample.

Table A1 - Summary Statistics of hexagon characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A – Race sample estimation					
	Black			White		
	Obs.	Mean	Std. Dev	Obs.	Mean	Std. Dev
Daily isolation level	16 187	0·418	0·084	225 231	0·457	0·100
Days after first SARI case	16 187	45·47	45·23	225 231	59·16	47·14
Days with NPI flexibilization	16 187	0·397	0·489	225 231	0·387	0·487
	Panel B – Income sample estimation					
	Low Income			High Income		
	Obs.	Mean	Std. Dev	Obs.	Mean	Std. Dev

Daily isolation level	97 139	0·419	0·088	90 383	0·486	0·103
Days after first SARI case	97 139	47·55	45·81	90 383	63·40	47·44
Days with NPI flexibilization	97 139	0·396	0·489	90 383	0·381	0·486

Notes. This table polls all days of data per group in each sample estimation (from March 1 to August 11). Data are at the hexagon-by-day level.

4.2 Model sensitivity analysis

We tested how the results of the event study by running the regression with and without covariates and testing for different number of days after the introduction of NPIs. Table A2 shows the results including only one treatment variable so it represents the effect of the average treatment every day. With 99% of statistical significance in all scenarios, the results indicate that the effect tends to increase up until the 120th day. The effects of NPI on isolation levels is consistently positive and significant in all estimates. The inclusion of covariates does not significantly change the coefficients nor the confidence intervals.

Table A2 – Sensitivity analysis of NPI effects on isolation levels

Number of days after NPI	**Analysis by income**					
	Without covariates			With covariates		
	Mean	Min 95	Max 95	Mean	Min 95	Max 95
10	0·055***	0·053	0·058	0·054***	0·052	0·057
30	0·084***	0·082	0·088	0·081***	0·079	0·085
90	0·087***	0·084	0·091	0·084***	0·081	0·088
120	0·083***	0·080	0·087	0·082***	0·079	0·086
150	0·079***	0·076	0·083	0·078***	0·075	0·082
Number of days after NPI	**Analysis by race**					
	Without covariates			With covariates		
	Mean	Min 95	Max 95	Mean	Min 95	Max 95
10	0·039***	0·036	0·044	0·038***	0·035	0·043
30	0·064***	0·060	0·069	0·061***	0·057	0·066
90	0·062***	0·058	0·068	0·059***	0·055	0·064
120	0·059***	0·055	0·064	0·056***	0·053	0·061
150	0·055***	0·051	0·060	0·053***	0·049	0·058

Note. *** p <0·01

5 Crosswalk: PNAD COVID-19 Occupations to ISCO-08 (1-digit)

The crosswalk from PNAD COVID-19 occupational codes to ISCO-08 1-digit codes was done in two steps. First, we applied the conversion rule presented in Table A1 to the occupational codes of occupation categories in PNAD COVID-19 (variable C007C). Then we classified the values 2 and 3 of variable C007 as "Military" (once these occupations are not registered in C007). We further disaggregated Health Professionals and Technicians into separate occupational categories.

Table A2 – Crosswalk: Variable C007C to ISCO-08 1-digit groups

PNAD-COVID-19 Occ. Code	PNAD COVID-19 Occupations (English Label)	ISCO-08 (1-digit)	ISCO-08 Label
1	Domestic worker, daily cleaner, cook (in private households),	9	Elementary Occupations
2	Janitor, cleaning assistant, etc. (in public or private company),	9	Elementary Occupations
3	Office clerk	4	Clerical Support Workers
4	Secretary, receptionist	4	Clerical Support Workers
5	Telemarketing operator	4	Clerical Support Workers
6	Merchant (owner of bars or shops etc.)	5	Services and Sales Workers
7	Store salesperson	5	Services and Sales Workers
8	Home seller, sales representative, catalog seller	5	Services and Sales Workers
9	Street vendors	9	Elementary Occupations
10	Cook and waiter (for restaurants, companies)	5	Services and Sales Workers
11	Baker, butcher and confectioner	5	Services and Sales Workers
12	Farmer, animal breeder, fisherman, forester and gardener	6	Skilled Agricultural, Forestry and Fishery Workers
13	Agricultural labourers	9	Elementary Occupations
14	Drivers (ride hailing apps, taxi, van, mototaxi, bus)	8	Plant and Machine Operators and Assemblers
15	Truck driver	8	Plant and Machine Operators and Assemblers
16	Courier services by motorcycle	9	Elementary Occupations
17	Delivery of goods (restaurant, pharmacy, store, Uber Eats, IFood, Rappy etc.)	9	Elementary Occupations
18	Bricklayer, stonemasons, painter, electrician, carpenter	7	Craft and Related Trades Workers
19	Mechanic of vehicles, industrial machineries etc.	7	Craft and Related Trades Workers
20	Craftsman, dressmaker and shoemaker	7	Craft and Related Trades Workers
21	Hairdresser, manicure and other beauty occupations	5	Services and Sales Workers
22	Machine operator, assembler in the industry	8	Plant and Machine Operators and Assemblers
23	Production assistant, loading and unloading	8	Plant and Machine Operators and Assemblers
24	Teachers and professors (kindergarten, elementary, high school or higher education)	2	Professionals
25	Pedagogue, teacher of languages, music, art and tutoring	2	Professionals
26	Health professionals	2	Professionals
27	Health Technician	3	Technicians and Associate Professionals
28	Babysitters and personal caretakers	5	Services and Sales Workers
29	Security, vigilant, other guard security services	5	Services and Sales Workers
30	Civil police	5	Services and Sales Workers
31	Doorman or Porter	9	Elementary Occupations
32	Artist, religious (priest, pastor, etc.)	2	Professionals
33	Director, manager, political or commissioned position	1	Managers
34	Other higher-level profession (lawyer, engineer, accountant, journalist, etc.)	2	Professionals
35	Other mid-level technician or professional	3	Technicians and Associate Professionals
36	Others	10	OTHER

6 Classification of non-classified SARI cases as COVID-19 cases

We considered all SARI cases with missing or unknown etiology as COVID-19 cases. This approach leads to an estimated number of COVID-19 cases that is closer to the actual number than taking into account only confirmed COVID-19 cases[10]. Figure A4 shows the number of daily SARI cases and deaths that were confirmed as COVID-19, confirmed by cause of a different etiology, or with unknown or missing etiology (non-classified SARI cases) for the state of São Paulo. In March 2020, both confirmed and non-classified SARI cases and deaths increased to a much higher level than what was previously recorded. This suggests that most non-classified SARI cases and deaths are actually caused by COVID-19, and that even severe COVID-19 cases suffer from undernotification in Brazil.

Figure A4. SARI cases and deaths in São Paulo that were confirmed by cause: COVID-19, other etiologies, and unknown etiology.

7 References

1. Dowd JB, Andriano L, Brazel DM, Rotondi V, Block P, Ding X, et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. Proc Natl Acad Sci. 2020;117:9696–8.

2. Lawson AB. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology, Second Edition. CRC Press; 2013. 398 p.

3. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math. 1991;43:1–20.

4. Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. Stat Med. 2000;19:2555–67.

5. Peixoto PS, Marcondes D, Peixoto C, Oliva SM. Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. PLOS ONE. 2020;15:e0235732.

6. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B Stat Methodol. 2009;71:319–92.

7. Censo Demográfico | IBGE [Internet]. [cited 2020 Nov 20]. Available from: https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html?=&t=o-que-e

8. Padgham M. dodgr: An R Package for Network Flow Aggregation. Findings. 2019;6945.

9. Bertrand M, Duflo E, Mullainathan S. How Much Should We Trust Differences-In-Differences Estimates? Q J Econ. 2004;119:249–75.

10. de Souza WM, Buss LF, Candido D da S, Carrera J-P, Li S, Zarebski AE, et al. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. Nat Hum Behav. 2020;4:856–65.

11. Buss LF, Prete CA, Abrahim CMM, Mendrone A, Salomon T, Almeida-Neto C de, et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. Science [Internet]. 2020 [cited 2020 Dec 8]; Available from: https://science.sciencemag.org/content/early/2020/12/07/science.abe9728

Appendix B: Figures

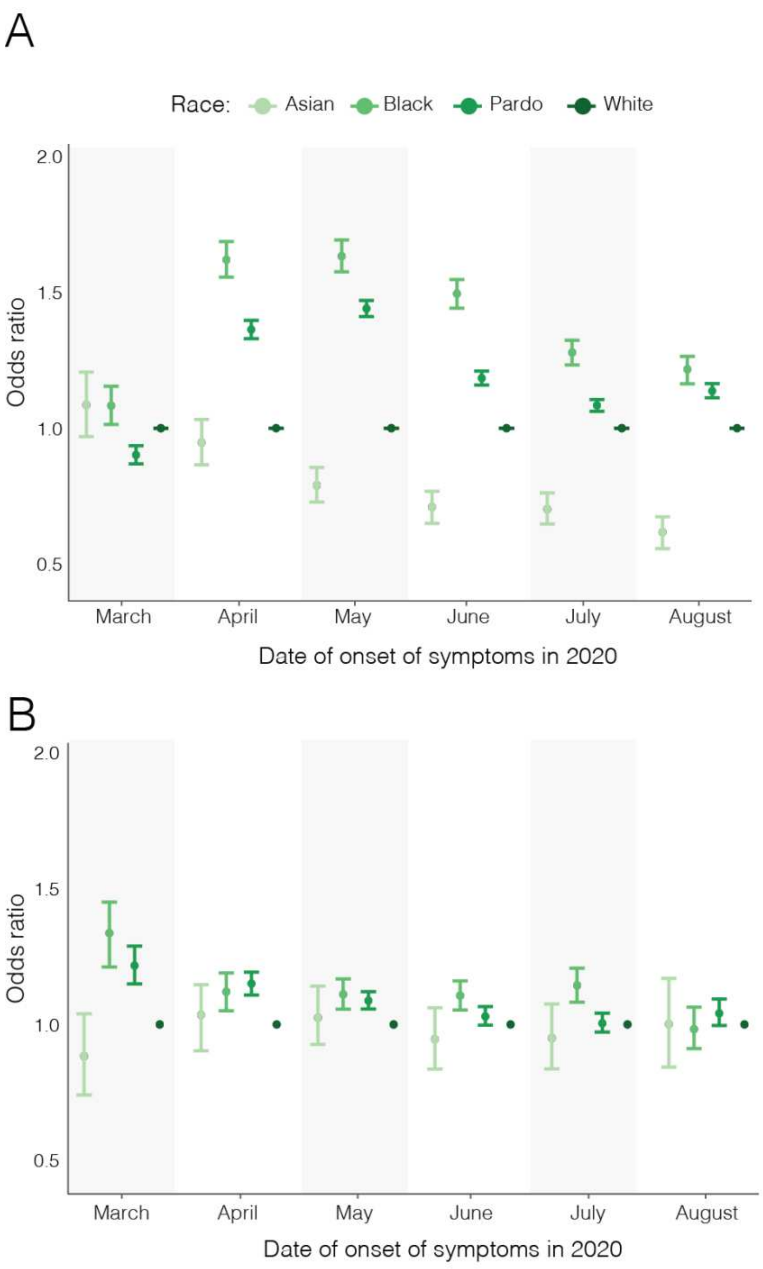

Figure S 1. Individual level hospitalisation and death risk by age-standardised odds ratio (OR) after discarding data with missing race information. **A**, OR for SARI hospitalisation by race and **B**, OR for death among SARI patients by race.

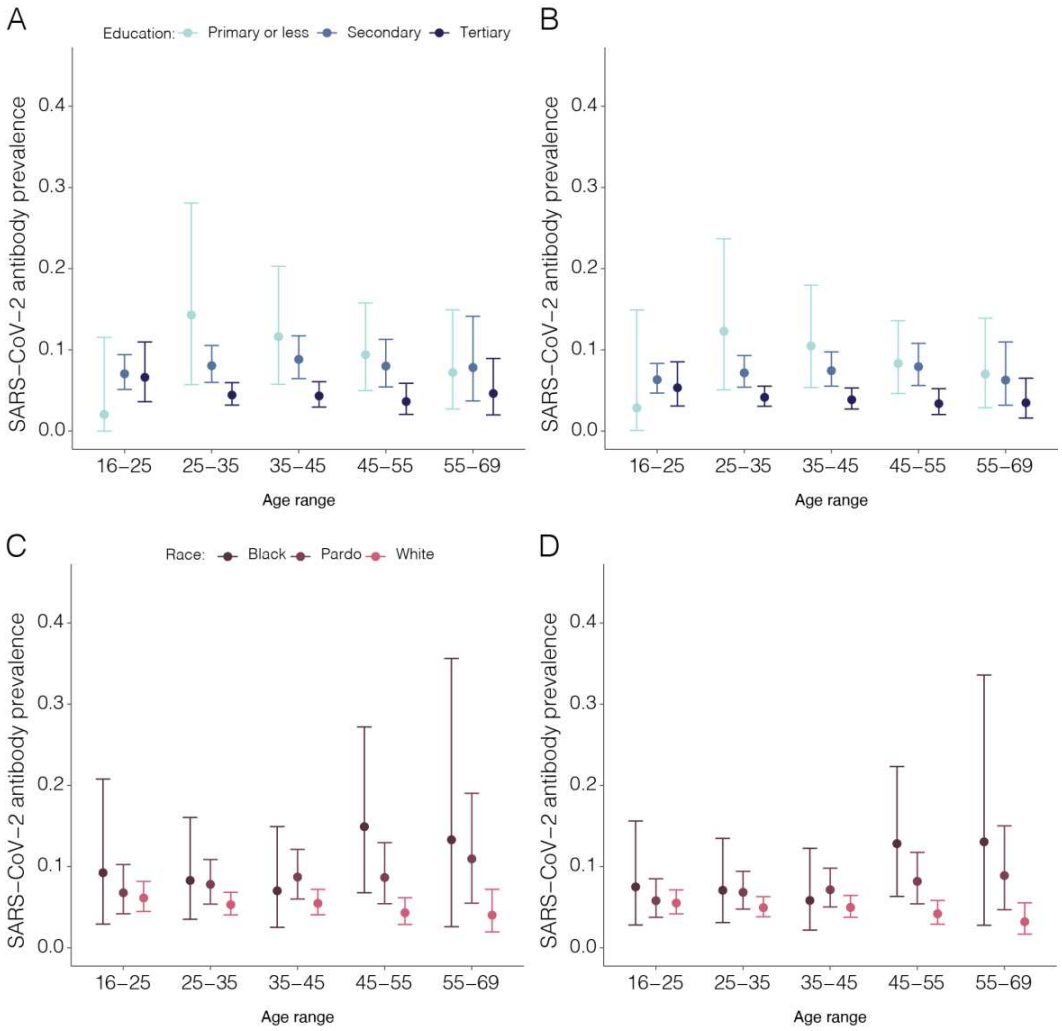

Figure S 2. Prevalence of anti-SARS-CoV-2 IgG antibodies among blood donors in São Paulo, Brazil, according to self-reported race and education attainment. **A and C**, crude prevalence and **B and D**, prevalence corrected by age, sex, specificity and sensitivity are shown. Source: Covid-IgG study [11].

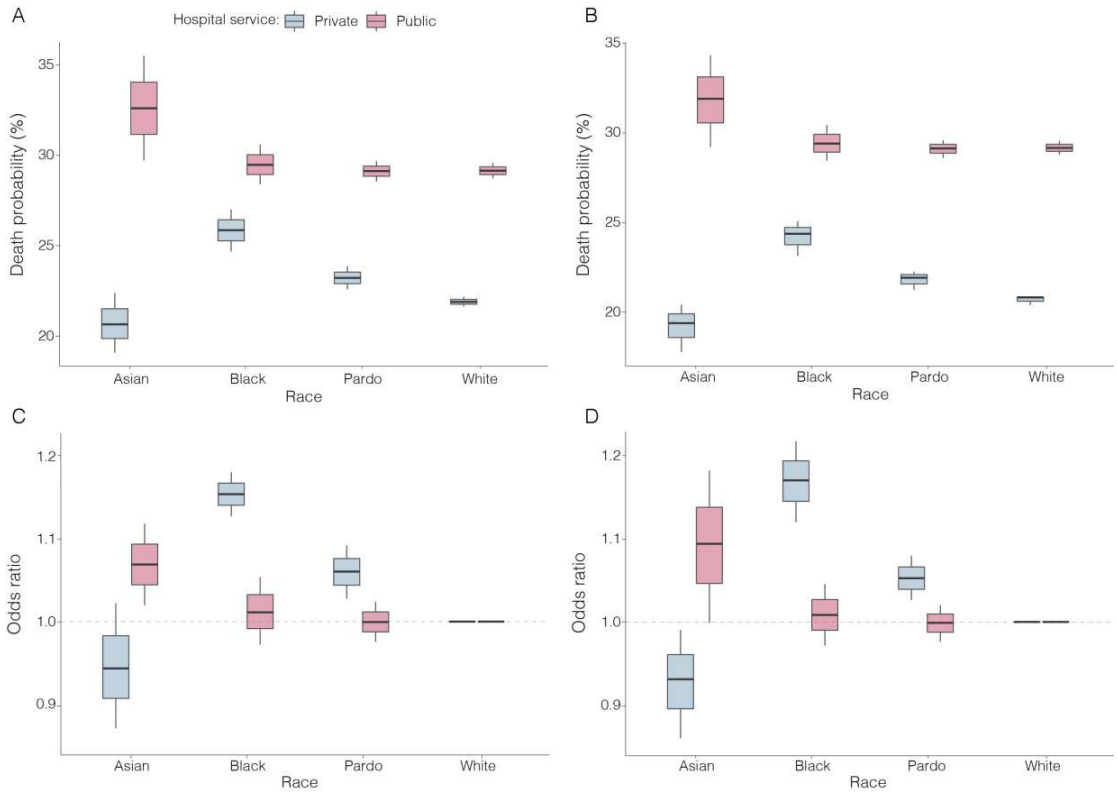

Figure S 3. **A and C,** Death probability and odds ratio of death by race in each hospital type using data without missing race information. **B and D,** with race imputed. Source: SIMI-SP.

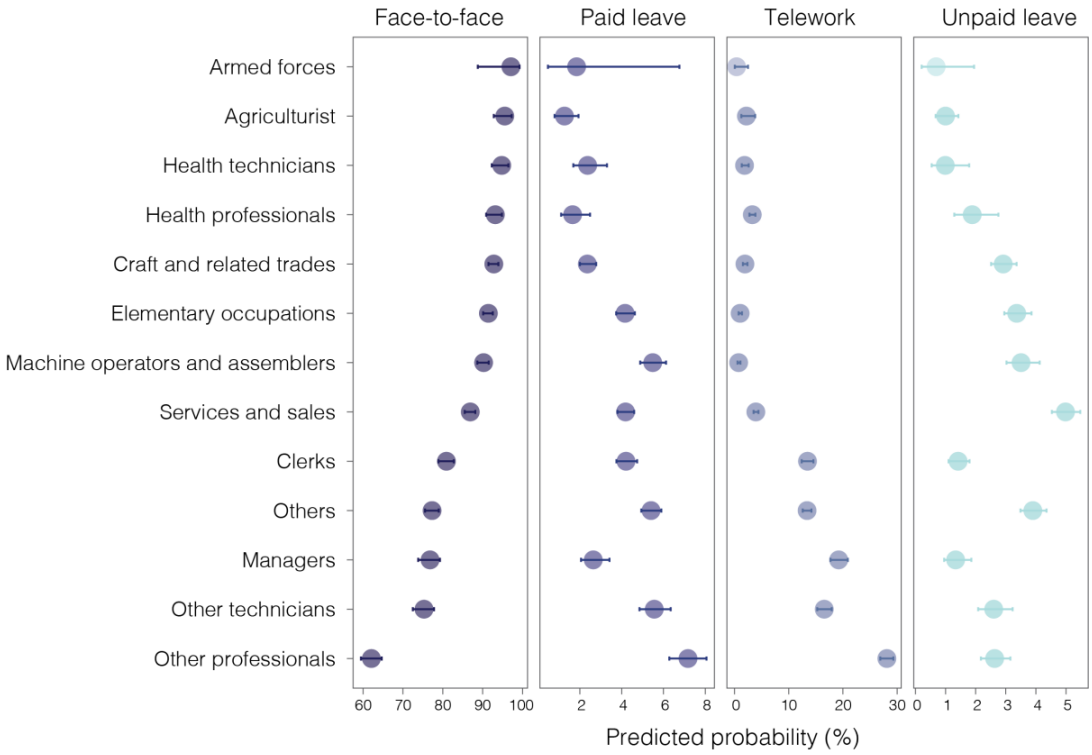

Figure S 4. Probability of working condition by occupation type between May and September 2020 in São Paulo state. Source: PNAD COVID-19 (IBGE).

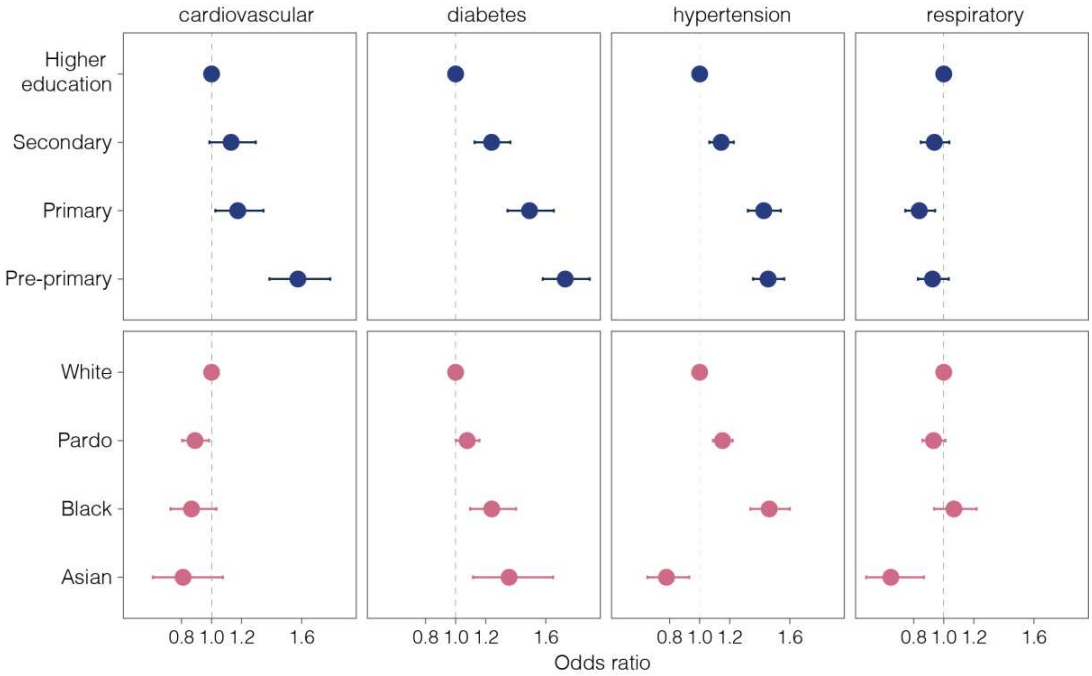

Figure S 5. Odds ratio of having been diagnosed with a comorbidity, by race and education attainment in São Paulo State, 2020. Source: PNAD COVID-19(IBGE).

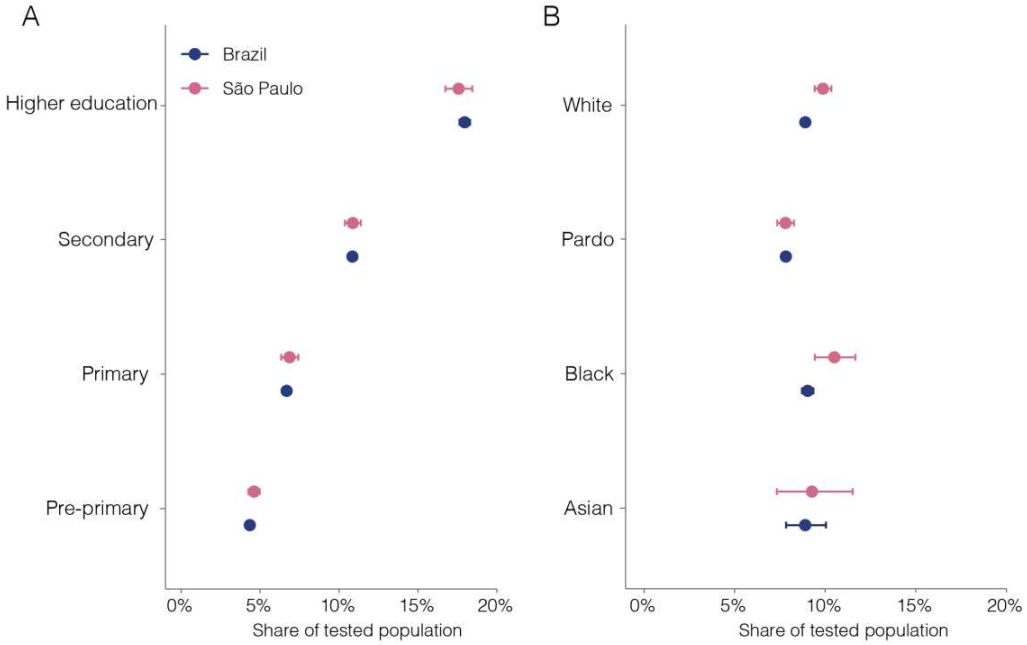

Figure S 6. Proportion of population that reported of being tested for COVID-19 between July and September 2020 by **A,** education attainment and **B,** race in Brazil and São Paulo state. Source: PNAD COVID-19 (IBGE).

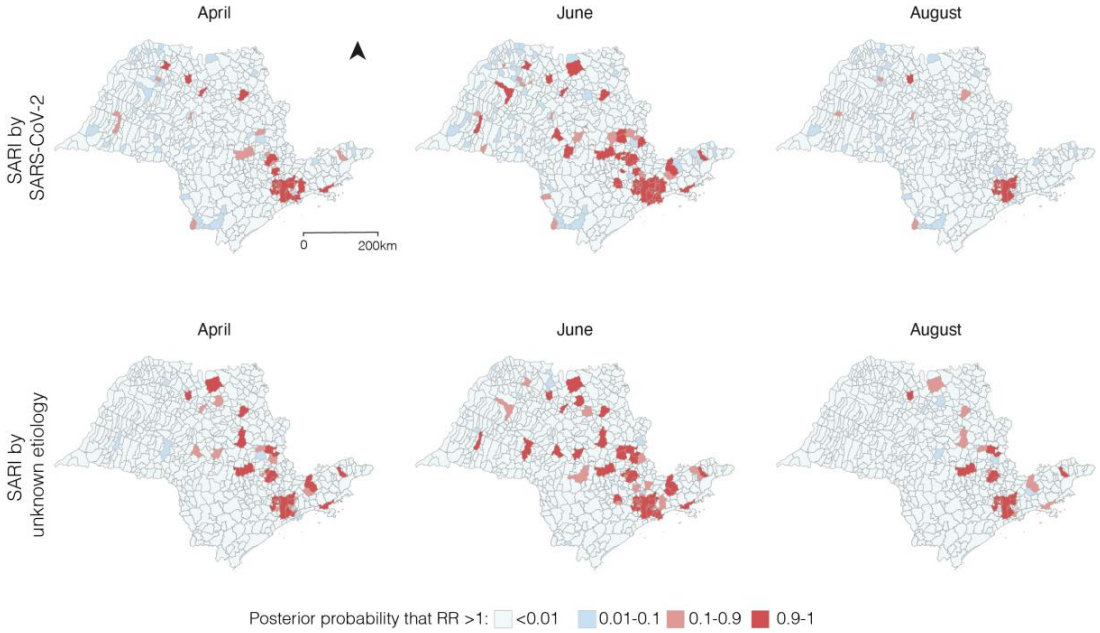

Figure S 7. Posterior probability of elevated relative risk at the municipality level for São Paulo state.