Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study

Pedro Baqui, Ioana Bica*, Valerio Marra, Ari Ercole, Mihaela van der Schaar*

Summary

Background Brazil ranks second worldwide in total number of COVID-19 cases and deaths. Understanding the possible socioeconomic and ethnic health inequities is particularly important given the diverse population and fragile political and economic situation. We aimed to characterise the COVID-19 pandemic in Brazil and assess variations in mortality according to region, ethnicity, comorbidities, and symptoms.

Methods We conducted a cross-sectional observational study of COVID-19 hospital mortality using data from the SIVEP-Gripe (*Sistema de Informação de Vigilância Epidemiológica da Gripe*) dataset to characterise the COVID-19 pandemic in Brazil. In the study, we included hospitalised patients who had a positive RT-PCR test for severe acute respiratory syndrome coronavirus 2 and who had ethnicity information in the dataset. Ethnicity of participants was classified according to the five categories used by the Brazilian Institute of Geography and Statistics: Branco (White), Preto (Black), Amarelo (East Asian), Indígeno (Indigenous), or Pardo (mixed ethnicity). We assessed regional variations in patients with COVID-19 admitted to hospital by state and by two socioeconomically grouped regions (north and central-south). We used mixed-effects Cox regression survival analysis to estimate the effects of ethnicity and comorbidity at an individual level in the context of regional variation.

Findings Of 99 557 patients in the SIVEP-Gripe dataset, we included 11 321 patients in our study. 9278 (82·0%) of these patients were from the central-south region, and 2043 (18·0%) were from the north region. Compared with White Brazilians, *Pardo* and Black Brazilians with COVID-19 who were admitted to hospital had significantly higher risk of mortality (hazard ratio [HR] 1·45, 95% CI 1·33–1·58 for *Pardo* Brazilians; 1·32, 1·15–1·52 for Black Brazilians). *Pardo* ethnicity was the second most important risk factor (after age) for death. Comorbidities were more common in Brazilians admitted to hospital in the north region than in the central-south, with similar proportions between the various ethnic groups. States in the north had higher HRs compared with those of the central-south, except for Rio de Janeiro, which had a much higher HR than that of the other central-south states.

Interpretation We found evidence of two distinct but associated effects: increased mortality in the north region (regional effect) and in the *Pardo* and Black populations (ethnicity effect). We speculate that the regional effect is driven by increasing comorbidity burden in regions with lower levels of socioeconomic development. The ethnicity effect might be related to differences in susceptibility to COVID-19 and access to health care (including intensive care) across ethnicities. Our analysis supports an urgent effort on the part of Brazilian authorities to consider how the national response to COVID-19 can better protect *Pardo* and Black Brazilians, as well as the population of poorer states, from their higher risk of dying of COVID-19.

Funding None.

Introduction

The COVID-19 pandemic has created an unprecedented worldwide strain on health care. Although early reports from east Asia and Europe meant that Brazil was well positioned to implement non-pharmaceutical interventions, Brazilians, like those in many low-income and middle-income countries, have limited access to testing and social security.[1,2] Difficulties in testing make it harder to assess the growth of the pandemic, while limited access to social security can lead a sizable fraction of society to not engage in physical distancing. This situation has been further complicated by an unstable federal government[3] that has failed to support measures such as physical distancing and attempted to downplay the gravity of the pandemic, as has been well publicised in the media.[4] Worryingly, as of June 23, 2020, Brazil ranks second worldwide in total number of COVID-19 cases and deaths, with a high estimated rate of transmission (effective reproduction number [R_t] 1·44).[5]

Worldwide, substantial interest is being given to the emerging societal inequities of the impact of COVID-19, and evidence is emerging of variability in the impact of the disease across ethnicities in various settings, including in the UK,[6–8] the USA,[9,10] and Norway.[11] Brazil's population is diverse, comprising many races and ethnic groups. The Brazilian Institute of Geography and Statistics (IBGE) racially classifies the Brazilian population into five categories. This IBGE classification

Lancet Glob Health 2020

Published Online
July 2, 2020
https://doi.org/10.1016/S2214-109X(20)30285-0

See Online/Comment
https://doi.org/10.1016/2214-109X(20)30314-4

For a Portuguese translation of the abstract see Online for appendix 1

*Contributed equally

Núcleo de Astrofísica e Cosmologia (P Baqui PhD, V Marra PhD) **and Departamento de Física** (V Marra), **Universidade Federal do Espírito Santo, Vitória, ES, Brazil; Department of Engineering Science, University of Oxford, Oxford, UK** (I Bica MPhil); **The Alan Turing Institute, London, UK** (I Bica, Prof M van der Schaar PhD); **Department of Medicine, University of Cambridge, Cambridge, UK** (A Ercole PhD); **Cambridge Centre for Artificial Intelligence in Medicine, Cambridge, UK** (Prof M van der Schaar, A Ercole); **Department of Applied Mathematics and Theoretical Physics and Department of Population Health, University of Cambridge, Cambridge, UK** (Prof M van der Schaar); **and Department of Electrical and Computer Engineering, University of California Los Angeles, Los Angeles, CA, USA** (Prof M van der Schaar)

Correspondence to:
Dr Valerio Marra, Departamento de Física, Universidade Federal do Espírito Santo, 29075–910, Vitória, ES, Brazil
marra@cosmo-ufes.org

Research in context

Evidence before this study

Brazil is a highly ethnically and socioeconomically diverse country. The severe impact of COVID-19, coupled with an unstable federal government, might make the country particularly susceptible to outcome inequities. Although the issue of the disproportionate effect of COVID-19 on ethnic groups has been debated in the Brazilian media, quantitative or systematic studies assessing the ethnic and regional variation in mortality are needed. We searched PubMed, Google Scholar, medRxiv, and bioRxiv on May 18, 2020, for studies published in English describing the disproportionate effect of COVID-19 on ethnic groups. We used the search terms "COVID-19", "Brazil", "ethnicity", and related synonyms, and found no studies matching our search criteria.

Added value of this study

We found that *Pardo* (mixed ethnic ancestry) and Black Brazilians admitted to hospital with COVID-19 had significantly higher mortality than that of White Brazilians, the comparator group. Particularly, *Pardo* ethnicity was the second most important risk factor for death after age. We also found that COVID-19 mortality increased in socioeconomically similar northern regions compared with central and southern regions, and that mortality risk was very high in Rio de Janeiro compared with that of its neighbouring states.

Implications of all the available evidence

Our results have serious social implications: *Pardo* and Black Brazilians have, on average, less economic security, are less likely to be able to stay at home and work remotely, and comprise a substantial proportion of health and care workers. We hope that this analysis assists the authorities in better directing and aligning their response to COVID-19 to protect *Pardo* and Black Brazilians from their higher risk of death from COVID-19. Our results also indicate that the states in the north and northeast macroregions are more vulnerable to the COVID-19 pandemic, an issue that merits additional urgent attention by the federal government of Brazil.

is based on colour and, as in international practice, individuals are asked to self-identify as either *Branco* (White), *Preto* (Black), *Amarelo* (East Asian), *Indígeno* (Indigenous), or *Pardo*. The term *Pardo* is a particularly complex one and is used in Brazil to refer to people of mixed ethnic ancestries: *Pardo* Brazilians represent a diverse range of ethnic backgrounds. In the 2010 census, the Brazilian population was 47·5% *Branca*, 43·4% *Parda*, 7·5% *Preta*, 1·1% *Amarela*, and 0·4% *Indígena*. For this study, we will use the Portuguese term *Pardo* and use the English terms White to mean *Branco*, Black to mean *Preto*, East Asian to mean *Amarelo*, and Indigenous to mean *Indígeno*.

Brazil is an important and interesting country in which to study the impact of COVID-19, partly because of the combination of the severity of the outbreak, governmental failure to implement non-pharmaceutical interventions, and complex social and ethnic societal composition. In this study, we analysed COVID-19 hospital mortality from the prospectively collected SIVEP-Gripe (*Sistema de Informação de Vigilância Epidemiológica da Gripe*) respiratory infection registry data,[12] which are maintained by the Ministry of Health for the purposes of recording cases of severe acute respiratory syndrome (SARS) across both public and private hospitals. Using this rich dataset, we characterised the COVID-19 pandemic in Brazil, particularly regarding risk factors related to comorbidities, symptoms, and ethnicity, similarly to previous analyses done in countries such as the UK.[8,13,14]

Methods

Study design and population

Our analysis is based on the SIVEP-Gripe public dataset.[12] As of May 18, 2020, SIVEP-Gripe contained epidemiological data for 99 557 patients from different states. Each entry has 138 features, including symptoms, age, sex, ethnicity, and comorbidities. We included in the study patients who had a positive RT-PCR test for SARS coronavirus 2 (SARS-CoV-2); who had their ethnicity recorded in the dataset, because we were interested in the relation between ethnicity and health risk; and who were admitted to hospital. The date of COVID-19 diagnosis spans the time interval from Feb 27 to May 4, 2020.

Brazil is divided geopolitically into five macroregions: north, comprising the states of Acre, Amapá, Amazonas, Pará, Rondônia, Roraima, and Tocantins; northeast, comprising Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte, and Sergipe; central-west, comprising Distrito Federal, Goiás, Mato Grosso, and Mato Grosso do Sul; southeast, comprising Espírito Santo, Minas Gerais, Rio de Janeiro, and São Paulo; and south, comprising Paraná, Rio Grande do Sul, and Santa Catarina.

For descriptive purposes, we chose to dichotomise the data into two maximally contrasting regions on the basis of similar education (literacy, higher education, and school drop-out rates), income (per-capita gross domestic product, salary, and poverty level), and health (life expectancy, child mortality, and food security). Living conditions, such as population density, overcrowding, and public transport use, are not included but are expected to correlate with the above socioeconomic factors. Ethnicity was not considered at this point. The two contrasting regions that we consider are the central-south region, comprising the central-west, southeast, and south macroregions; and the north region, comprising the north and northeast

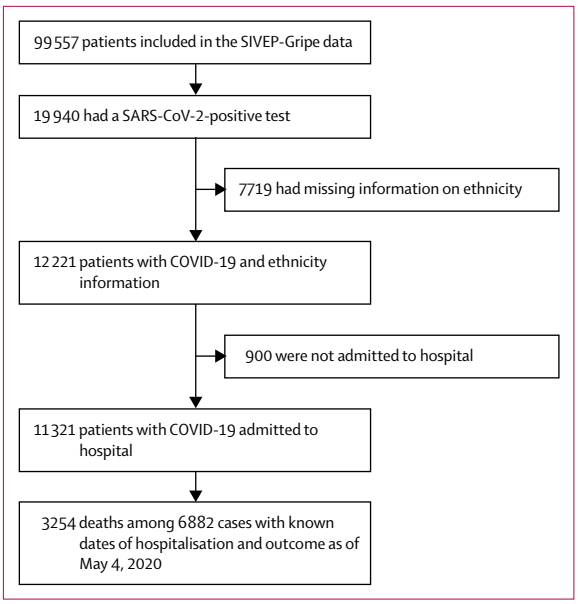

Figure 1: **Flowchart of SIVEP-Gripe data used in this study**
SARS-CoV-2=severe acute respiratory syndrome coronavirus 2. SIVEP-Gripe=*Sistema de Informação de Vigilância Epidemiológica da Gripe*.

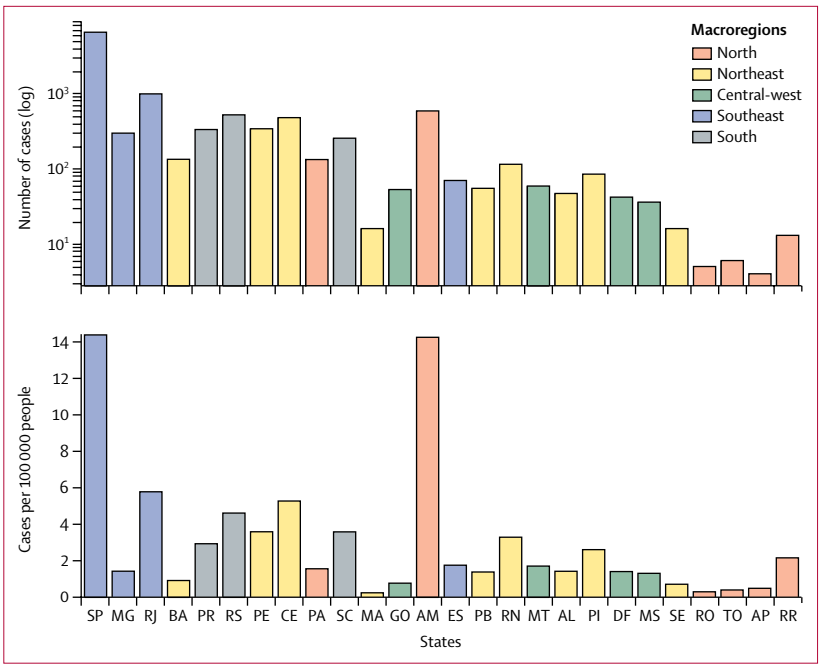

Figure 2: **Distribution of patients among Brazilian states according to absolute number of cases and number of cases per 100 000 people**
n=11 321. States are ordered according to their population, larger on the left. No patients from Acre were included in the dataset of 11 321 patients admitted to hospital. AL=Alagoas. AM=Amazonas. AP=Amapá. BA=Bahia. CE=Ceará. DF=Distrito Federal. ES=Espírito Santo. GO=Goiás. MA=Maranhão. MG=Minas Gerais. MS=Mato Grosso do Sul. MT=Mato Grosso. PA=Pará. PB=Paraíba. PE=Pernambuco. PI=Piauí. PR=Paraná. RJ=Rio de Janeiro. RN=Rio Grande do Norte. RO=Rondônia. RR=Roraima. RS=Rio Grande do Sul. SC=Santa Catarina. SE=Sergipe. SP=São Paulo. TO=Tocantins.

macroregions. This is the customary division when considering northern and southern Brazil and is socioeconomically and quantitatively well justified (appendix 2, p 16). Although White and *Pardo* Brazilians together comprise most of Brazil's population, with approximately equal proportions, their distribution varies considerably by macroregion. The population in the south macroregion is 78% White and 17% *Pardo*, whereas the north macroregion's population is 23% White and 67% *Pardo*.[15]

The SIVEP-Gripe data include information on co-morbidities and symptoms. Missingness is described in appendix 2 (p 2). We interpreted missing values as the absence of comorbidities or symptoms. Missing values are also present for intensive care unit (ICU) admissions. In this case, we considered missing values as non-admissions to ICU.

Statistical analysis

Our analysis used descriptive statistics to quantify the COVID-19 pandemic in Brazil, and mixed-effects Cox regression to investigate the importance of record-level risk factors and estimate hazard ratios (HRs). We used patient-level clinical features, namely age group, sex, ethnic group, and comorbidities, as fixed effects, with state as a random effect (similar to a UK analysis on COVID-19-related ICU mortality).[16] For the categorical variables of age group and ethnic group, we used younger than 40 years as a reference category for age group and White for ethnic group. We also considered as additional fixed effects the number of ICU beds, ventilators, and nurses per 100 million inhabitants in each state (appendix 2, p 3). We did not find evidence[17] for statistically significant violation of the proportional hazards assumption (p=0·11).

To check a possible lead-time and outcome ascertainment bias, we considered a split on the basis of fewer than 7 days or 7 days or more between symptoms and outcome and another split where we considered fewer than 14 days and 14 days or more between symptoms and outcomes.

See **Online** for appendix 2

Role of the funding source

There was no funding source for this study. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

The SIVEP-Gripe dataset comprised 99 557 patients. 19 940 had a positive test for SARS-CoV-2 and, of these, 12 221 had their ethnicity recorded. 900 of these patients were not admitted to hospital, resulting in a base dataset of 11 321 patients (figure 1). 9278 (82·0%) of 11 321 patients were from the central-south region, and 2043 (18·0%) were from the north region. São Paulo, Rio de Janeiro, and Amazonas had the highest number of cases, both absolute and per 100 000 people (figure 2). Of 6882 patients with known outcome as of May 4, 2020, 3254 died.

	Survivors (n=4043)	Non-survivors (n=3328)
North (n=1350)	479 (35·5%)	871 (64·5%)
Age (years)	46·9 (19·3)	65·3 (16·0)
Sex		
Men (n=795)	261 (32·8%)	534 (67·2%)
Women (n=555)	218 (39·3%)	337 (60·7%)
Ethnic group		
White (n=225)	89 (39·6%)	136 (60·4%)
Pardo (n=1049)	366 (34·9%)	683 (65·1%)
Black (n=51)	16 (31·4%)	35 (68·6%)
East Asian (n=17)	5 (29·4%)	12 (70·6%)
Indigenous (n=8)	3 (37·5%)	5 (62·5%)
Comorbidities		
Cardiovascular disease (n=420)	95 (22·6%)	325 (77·4%)
Asthma (n=35)	22 (62·9%)	13 (37·1%)
Diabetes (n=371)	74 (19·9%)	297 (80·1%)
Pulmonary disease (n=51)	15 (29·4%)	36 (70·6%)
Obesity (n=58)	13 (22·4%)	45 (77·6%)
Immunosuppression (n=49)	28 (57·1%)	21 (42·9%)
Renal disease (n=69)	13 (18·8%)	56 (81·2%)
Liver disease (n=17)	4 (23·5%)	13 (76·5%)
Neurological disease (n=33)	7 (21·2%)	26 (78·8%)
Central-south (n=6021)	3564 (59·2%)	2457 (40·8%)
Age (years)	52·2 (16·6)	67·0 (15·8)
Sex		
Men (n=3495)	2039 (58·3%)	1456 (41·7%)
Women (n=2526)	1525 (60·4%)	1001 (39·6%)
Ethnic group		
White (n=4108)	2548 (62·0%)	1560 (38·0%)
Pardo (n=1355)	728 (53·7%)	627 (46·3%)
Black (n=425)	215 (50·6%)	210 (49·4%)
East Asian (n=126)	69 (54·8%)	57 (45·2%)
Indigenous (n=7)	4 (57·1%)	3 (42·9%)
Comorbidities		
Cardiovascular disease (n=2083)	936 (44·9%)	1147 (55·1%)
Asthma (n=244)	158 (64·8%)	86 (35·2%)
Diabetes (n=1521)	641 (42·1%)	880 (57·9%)
Pulmonary disease (n=336)	115 (34·2%)	221 (65·8%)
Obesity (n=266)	130 (48·9%)	136 (51·1%)
Immunosuppression (n=260)	104 (40·0%)	156 (60·0%)
Renal disease (n=320)	87 (27·2%)	233 (72·8%)
Liver disease (n=66)	25 (37·9%)	41 (62·1%)
Neurological disease (n=283)	84 (29·7%)	199 (70·3%)

Data are n (%) or mean (SD). For this table, we considered patients for which the outcome was known but not the corresponding dates. Therefore, the total number of survivors and non-survivors (n=7371) is larger than that reported in figure 1 (6882).

Table 1: **Demographic characteristics and coexisting conditions among survivors and non-survivors of COVID-19**

Tables 1 and 2 show the demographic and comorbidity data among survivors and non-survivors of COVID-19 and their ethnic composition at each stage of the COVID-19 trajectory. Survivors in both north and central-south regions were younger and more likely to be White and women, whereas non-survivors were older and more likely to be Black and *Pardo* (results regarding other ethnicities were more difficult to interpret because of low numbers). Almost all comorbidities were more common in non-survivors in the north than in the central-south, suggesting structural health disparities. This is further evidenced by a substantially larger percentage of non-survivors in the north than in the central-south (table 1). The proportion of patients admitted to hospital who died (table 2) revealed a similar pattern, with a higher proportion in the north than in the central-south (suggesting a regional effect) and a higher proportion in Black and *Pardo* Brazilians than in other ethnic groups (suggesting an ethnicity effect).

We plotted comorbidity distributions by number of comorbidities and ethnicity for survivors and non-survivors, excluding Indigenous patients because of small numbers (figure 3A, B). We observed a substantial asymmetry between north and central-south, with more non-survivors in the north than in central-south. Furthermore, White Brazilians without comorbidities were the group less likely to die. Additionally, we plotted the number of symptoms per patient (fever, cough, sore throat, shortness of breath, respiratory discomfort, arterial oxygen saturation <95%, diarrhoea, and vomiting) by ethnicity for survivors and non-survivors (figure 3C, D). Most patients presented to hospital with three to six symptoms, suggesting that, in this dataset, more severe presentations (patients presenting with several severe symptoms) are the ones being tested for COVID-19.

Regarding distributions of survivors and non-survivors according to age and ethnicity, we observed a pattern of increasing deaths with age (figure 3E, F). In the north region, the pattern of younger patients having a higher likelihood of survival was even more pronounced than in the central-south region. Except in the oldest age group, *Pardo* and Black Brazilians appeared to be less likely to survive COVID-19 compared with White Brazilians, with the difference being more pronounced in the central-south region.

Assessing HRs for all clinical features (fixed effects) considered in the Cox model (figure 4A), we observed that, compared with White Brazilians, *Pardo* and Black Brazilians admitted to hospital had significantly higher risk of mortality (HR 1·45, 95% CI 1·33–1·58 for *Pardo* Brazilians; 1·32, 1·15–1·52 for Black Brazilians; appendix 2, p 18). Notably, *Pardo* ethnicity was the second most important risk factor for death after age. We found substantial variations in HR between the different states in Brazil considered (figure 4B). The states in the north region tended to have higher HRs than those in the central-south region, further justifying our approach of

	Brazilian population*	Hospital admission	ICU admission	Death	Death/ hospitalisation	Death (not ICU)	Death (ICU)
North (n=2043)							
White	27·8%	342 (16·7%)	127 (19·4%)	136 (15·6%)	39·8%	69 (14·9%)	67 (16·5%)
Pardo	61·5%	1567 (76·7%)	481 (73·8%)	683 (78·4%)	43·6%	368 (79·3%)	315 (77·4%)
Black	8·8%	85 (4·2%)	26 (4·0%)	35 (4·0%)	41·2%	20 (4·3 %)	15 (3·7%)
East Asian	1·2%	36 (1·8%)	13 (2·0%)	12 (1·4%)	33·3%	5 (1·1 %)	7 (1·7%)
Indigenous	0·7%	13 (0·6%)	5 (0·8%)	5 (0·6%)	38·5%	2 (0·4%)	3 (0·7%)
Central-south (n=9278)							
White	58·7%	6291 (67·8%)	2344 (69·4%)	1560 (63·5%)	24·8%	616 (60·3%)	944 (65·8%)
Pardo	33·2%	2112 (22·8%)	731 (21·6%)	627 (25·5%)	29·7%	278 (27·2%)	349 (24·3%)
Black	6·8%	667 (7·2%)	220 (6·5%)	210 (8·6%)	31·5%	108 (10·6%)	102 (7·1%)
East Asian	1·1%	195 (2·1%)	82 (2·4%)	57 (2·3%)	29·2%	19 (1·9%)	38 (2·7%)
Indigenous	0·3%	13 (0·1%)	2 (0·1%)	3 (0·1%)	23·1%	1 (0·1%)	2 (0·1%)

Data are % or n (%). ICU=intensive care unit. *Census values of the Brazilian population.

Table 2: **Ethnic composition of patients at each stage of the COVID-19 trajectory**

splitting Brazil into two sets. This finding also corresponded with the regional effect previously discussed (appendix 2, p 18).

Cox regression results were qualitatively similar for north and central-south regions, metropolitan and rural subgroups, and public-predominant and private-predominant health-care subgroups, suggesting robustness to differences in outbreak start and outliers (appendix 2, pp 3–11).

Discussion

We present, to our knowledge, the most extensive study of COVID-19 hospital survival in Brazil. We found that survivors were more likely to be younger,[18] be women,[19] and have fewer comorbidities,[20] in keeping with worldwide findings. However, we also report several other important sociodemographic trends specific to Brazil.

We found significant regional variation in both case characteristics and outcomes. The high number of cases in São Paulo, Rio de Janeiro, and Amazonas (figure 2) are noteworthy. These regions are important ports of entrance to Brazil. Amazonas hosts the Free Economic Zone of Manaus and most of the international flights route through São Paulo and Rio de Janeiro: in 2019, 7·7 million international passengers landed in São Paulo and 2·2 million in Rio de Janeiro (additional details in appendix 2, p 17). Additionally, both São Paulo and Rio de Janeiro are characterised by a particularly high population density, and the outbreak coincided with the rainy season (associated with respiratory infections) in Amazonas.

The finding of a higher comorbidity burden in patients admitted to hospital in the north is concordant with a lower life expectancy in this region,[21] mirroring differences in the average age of survivors and non-survivors between north and central-south and the substantially larger percentage of non-survivors in the north. Survivors in both regions were more likely to be White, and White Brazilians were more likely to be admitted to ICU than *Pardo* Brazilians. Therefore, the increased death rate of *Pardo* Brazilians might be partly due to non-ICU admission, raising concerns regarding the organisation of public and private medical resources.

Also noteworthy is the finding that the distribution of patients with COVID-19 admitted to hospital between the central-south (82·0% of patients) and north (18·0%) regions in our data is discordant with the population sizes of these regions (64% in the central-south and 36% in the north).[15] This discrepancy highlights national heterogeneity and might, at least partly, be due to intrinsically lower hospitalisation rates in the north region, the disproportionate impact of COVID-19 in populous areas such as São Paulo and Rio de Janeiro (both in the central-south region), or both.

The disproportionately large percentage of survivors with no comorbidities in the central-south region (figure 3) is remarkable. We can speculate that this might be due to differences in comorbidity ascertainment either because of structural differences in the way data are collected (perhaps comorbidity data were less available from patients who were sicker at the time of presentation) or because patients with less severe disease, perhaps with concerns regarding the risks posed by their comorbidities, presented to hospital more readily in the central-south region. White and *Pardo* Brazilians had a similar number of comorbidities in the north and central-south populations. Therefore, comorbidities seem unlikely to be associated with ethnicity in our study, but rather they might be associated with regional socioeconomic development (education, income, and health).

However, an interplay between ethnic and regional socioeconomic factors is apparent in the lower likelihood of survival of younger *Pardo* and Black Brazilians compared with that of White Brazilians (figure 3), with the difference being more pronounced in the central-south. For context, the typical life expectancy at birth in Brazil is 76·0 years (as of 2017),[22] compared with

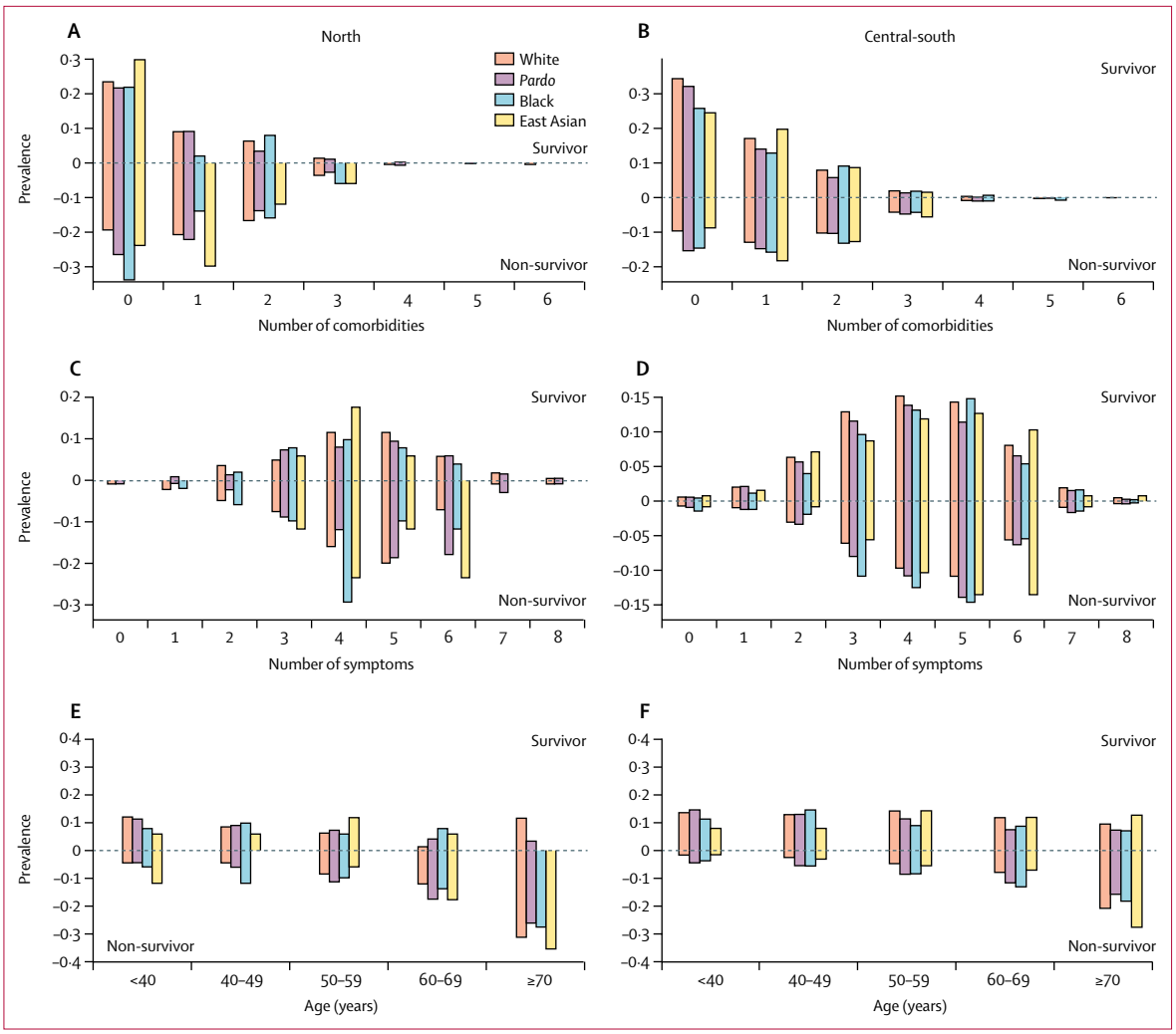

***Figure 3*: Distributions of ethnicity according to number of comorbidities (A, B), symptoms (C, D), and age (E, F)**
The normalisation is such that all the fractions of a given ethnicity add to unity (to adjust for differences in ethnic prevalence). We exclude Indigenous patients for clarity because of their small numbers in the study population.

80·9 years in Europe.[23] Average life expectancy varies by region, being higher in the central-south (79·4 years in Santa Catarina) than in the north (70·9 years in Maranhão), providing a baseline for our findings on mortality between regions.

ICU access might be a factor for regional and ethnic variations in mortality, with White Brazilians more likely to be admitted to ICU once hospitalised. Although White Brazilians were more likely to survive overall, we observed similar proportions between White and *Pardo* ethnicities when comparing total hospitalisations with deaths after ICU admission. The distribution of comorbidities, symptoms, and age did not show strong ethnic variations, especially between *Pardo* and White Brazilians (figure 3). The greater proportion of deaths without admission to ICU for *Pardo* Brazilians is noteworthy and likely to reflect higher levels of access to private health care for White Brazilians compared with that for *Pardo* Brazilians, because ICU admission policies are known to differ between public and private hospital settings.[24] Private health care serves only 25% of the Brazilian population and total spending is similar to that of public health care, implying that, on average, a patient in a private hospital costs three times more than one in a public hospital.[25] The proportions of the different ethnicities admitted to ICU with COVID-19 were similar to those in the full 2019 SIVEP-Gripe dataset,[12] suggesting that this is not a specific feature of COVID-19 treatment (appendix 2, p 18).

Survival analysis showed that, after age, the most important factor for hospital mortality was being of *Pardo* or, to a lesser extent, Black ethnicity compared with White ethnicity (figure 4). The other risk factors largely replicated worldwide findings, although we found that being male was perhaps slightly less of a risk factor in our study than in other series.[19] This ethnic inequity has

important social roots and implications: compared with White Brazilians, *Pardo* and Black Brazilians have, on average, less economic security, live in contagion-prone conditions, are less likely to be able to stay at home and work remotely, and comprise a substantial proportion of health and care workers, making them disproportionately the most vulnerable to COVID-19.[2] Additionally, *Pardo* and Black Brazilians tend to be more exposed to COVID-19 risk factors such as indoor pollution and the availability of water, which has been identified as a potential risk factor elsewhere.[26] As a proxy for the availability of clean water and general sanitation, sewerage cover is also congruent with our findings (appendix 2, p 16).

We observed substantial variation in HR by region. The states in the north region tended to have higher HRs than those in the central-south region, concordant with the larger percentage of non-survivors in the north (table 1). Incorporating the number of ICU beds and ventilators and nurses per 100 million inhabitants for each region as proxies for physical availability of health-care resources did not qualitatively change our result (appendix 2, p 3), suggesting a more fundamental difference in health-care access and trajectory of disease.

Rio de Janeiro, despite high standards of education, income, and health, had one of the highest HRs, similar to those in the Pernambuco and Amazonas states. States in the north had HRs higher than 1 and a high proportion of *Pardo* Brazilians (appendix 2, p 18). Again, Rio de Janeiro was an exception, having a much higher HR than that of the neighbouring state of São Paulo and an ethnicity profile similar to that of the north region states. Additionally, the HR for the metropolitan area of Rio de Janeiro was twice that of the similar local rural area (appendix 2, pp 8–9); we did not observe similar differences between other metropolitan areas and their rural neighbouring regions. Furthermore, the disparity in HR between public and private health-care mortality in Rio de Janeiro was the highest in Brazil, suggesting that access to high quality health care in the metropolitan area together with a large *Pardo* community were important drivers of outcome in the city.

Many Black Brazilians might identify themselves as *Pardo* Brazilians.[27] For this reason, it is reasonable to consider Black and *Pardo* populations together. Indeed, as seen from our analysis, both ethnic groups shared higher percentages of non-survivors and higher HRs than those of other ethnic groups. The results of our analysis can then be interpreted according to the interplay of regional and ethnicity effects. We can speculate that the regional effect is due to expected variations in number of comorbidities (or poorly controlled comorbidities) and general health-care access, which we might expect to have a notable impact in regions where socioeconomic levels are lower, such as the north. We similarly postulate that the ethnicity effect was driven by the greater susceptibility of contracting COVID-19 and reliance on publicly funded health care and reduced ICU access of *Pardo* and Black communities.

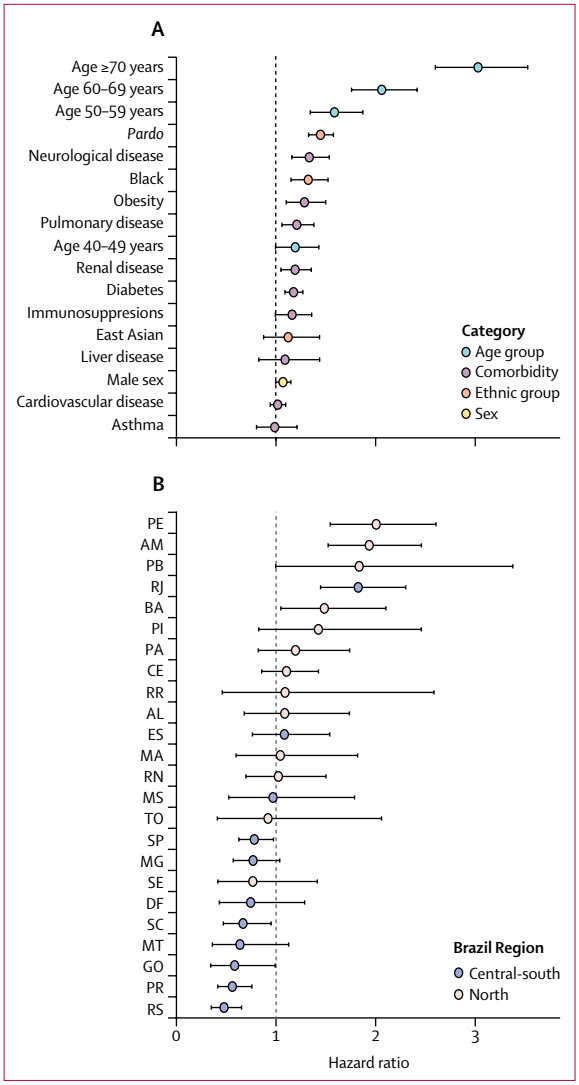

***Figure 4*: Risk of mortality for all clinical features (fixed effects; A) and all states in Brazil (random effects; B) considered in the fitted multivariate mixed-effects Cox model**

Error bars represent 95% CIs. No patients from Acre, Amapá, and Rondônia were included in the dataset of 6882 patients with known outcome. AL=Alagoas. AM=Amazonas. BA=Bahia. CE=Ceará. DF=Distrito Federal. ES=Espírito Santo. GO=Goiás. MA=Maranhão. MG=Minas Gerais. MS=Mato Grosso do Sul. MT=Mato Grosso. PA=Pará. PB=Paraíba. PE=Pernambuco. PI=Piauí. PR=Paraná. RJ=Rio de Janeiro. RN=Rio Grande do Norte. RR=Roraima. RS=Rio Grande do Sul. SC=Santa Catarina. SE=Sergipe. SP=São Paulo. TO=Tocantins.

For most states, the regional and ethnicity effects were correlated, resulting in a larger cumulative mortality. Indeed, lower socioeconomic development was correlated with a larger *Pardo* and Black population in the north region. In this, Rio de Janeiro was an outlier, with an ethnic composition (ethnicity effect) similar to that of states in the north region, but with high levels of development (regional effect) more akin to those of central-south states.

Our study had some limitations that need discussion. Limitations and possible biases from case ascertainment cannot be ruled out, in common with all observational and database research. Ethnicity information was missing in 7719 (38·7%) of 19940 patients. This is comparable to a large UK dataset that has been used in COVID-19 studies (26%).[8] We observed that the percentage of White Brazilians admitted to hospital was lower than the corresponding population percentage in the north and higher than that in the central-south, with the opposite occurring for *Pardo* Brazilians (table 2). This discrepancy could indicate that COVID-19 spreads differently through the ethnic groups within the two regions, rather than it being an effect of missing ethnicity data. Indeed, in the central-south, COVID-19 spread initially among White Brazilians (especially in populous São Paulo and Rio de Janeiro) who tend to travel internationally more frequently, whereas in the North, more precarious living and working conditions might have been more important for the spread of the disease.

We have limited our analysis to patients who were admitted to hospital because testing in the community is more likely to be biased according to local factors. However, we cannot be sure that the availability of testing was homogeneous even in this population. Indeed, the fact that a large proportion of patients that have tested positive in the SIVEP-Gripe dataset are admitted to hospital could show that testing, at least as far as this dataset is concerned, was done only when symptoms were severe, indicating in turn that the number of COVID-19 cases in Brazil is likely to be much higher than that suggested by available data.[28,29]

For **SIVEP-Gripe data** see http://plataforma.saude.gov.br/coronavirus/dados-abertos/

For the **analysis code** see http://doi.org/10.5281/zenodo.3903812

Health-seeking behaviour can vary with ethnicity and region; late presentation to hospital could be an important determinant of hospital outcome. We were not able to consider this in our analysis because data for physiological severity at hospital presentation were not available. However, a COVID-19 study in the UK did not show an important effect of physiological severity,[16] at least for ICU mortality, suggesting a high degree of homogeneity at admission.

The analysis of early data is important if findings are to be actionable but introduces the possibility of lead-time and outcome ascertainment bias. A sensitivity analysis excluding patients outside 7 or 14 days between symptoms and outcomes yielded qualitatively similar results (appendix 2, pp 12–15).

Although we have focused on hospital mortality, it is important to note that we did not have data on out-of-hospital mortality (which may be substantial) and neither could we robustly address the question of access to hospital services by region, ethnicity, or socioeconomic status. As such, an assessment of hospital mortality is likely to substantially underestimate the true impact of COVID-19, and we could plausibly assume that inequities in health-care availability would be further amplified in patients who are not hospitalised. In other words, we might sensibly assume that *Pardo* Brazilians are at an even higher risk than the findings of this study might suggest. Urgent work is needed to understand deaths occurring in the community.

In conclusion, we present evidence suggesting a higher risk of death among *Pardo* and Black Brazilians and in the north region of Brazil. As of June 22, 2020, the Brazilian federal administration has not supported non-pharmacological interventions such as physical distancing. Our results suggest that major metropolitan areas might be particularly affected, and that it is highly plausible that viral spread might be particularly rapid in these settings. Urgent work is needed to understand the impact of the basic reproduction number in these areas and testing should be increased. However, even without this detail, our observations motivate application of non-pharmacological interventions, at least in such areas. Across the rest of the country, urgent political attention should be directed towards understanding and alleviating societal, educational, and financial barriers to health-care access, because these might lead to delayed presentation to hospital in socioeconomically disadvantaged groups.

Contributors

MvdS conceived the research question. All authors designed the study and analysis plan. VM obtained the epidemiological and socioeconomic data. PB did the analysis with descriptive statistics. IB did the analysis with Cox regression. VM drafted the initial version of the manuscript. AE oversaw the clinical review of the methods and manuscript. All authors critically reviewed early and final versions of the manuscript.

Declaration of interests

We declare no competing interests.

Data sharing

SIVEP-Gripe data are publicly available. Our analysis code is available online.

References

1 Hasell J, Mathieu E, Beltekian D, et al. Coronavirus (COVID-19) testing. 2020. https://ourworldindata.org/coronavirus-testing (accessed June 13, 2020).

2 Tavares F, Betti G. Vulnerability, poverty and COVID-19: risk factors and deprivations in Brazil. 2020. https://www.researchgate.net/publication/340660228_Vulnerability_Poverty_and_COVID-19_Risk_Factors_and_Deprivations_in_Brazil (accessed May 10, 2020).

3 The Lancet. COVID-19 in Brazil: "So what?". *Lancet* 2020; **395:** 1461.

4 Phillips T. Brazil: Bolsonaro's defiance of distancing criticized by health minister. 2020. The Guardian. https://www.theguardian.com/world/2020/apr/13/brazil-bolsonaro-coronavirus-covid-19-social-distancing (accessed June 13, 2020).

5 Imperial College COVID-19 response team. Short-term forecasts of COVID-19 deaths in multiple countries. 2020. https://mrc-ide.github.io/covid19-short-term-forecasts/ (accessed June 23, 2020).

6 Kirby T. Evidence mounts on the disproportionate effect of COVID-19 on ethnic minorities. *Lancet Respir Med* 2020; **8:** 547–48.

7 Pareek M, Bangash MN, Pareek N, et al. Ethnicity and COVID-19: an urgent public health research priority. *Lancet* 2020; **395:** 1421–22.

8 Williamson E, Walker AJ, Bhaskaran KJ, et al. Open-SAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. *medRxiv* 2020; published online May 7. DOI:10.1101/2020.05.06.20092999 (preprint).

9 US Centers for Disease Control and Prevention. COVID-19 in racial and ethnic minority groups. 2020. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html (accessed May 10, 2020).

10 APM research lab staff. The color of coronavirus: COVID-19 deaths by race and ethnicity in the U.S. 2020. https://www.apmresearchlab.org/covid/deaths-by-race (accessed May 10, 2020).

11 Cookson C, Milne R. Nations look into why coronavirus hits ethnic minorities so hard. 2020. https://www.ft.com/content/5fd6ab18-be4a-48de-b887-8478a391dd72 (accessed May 10, 2020).

12 Ministry of Health. SIVEP-Gripe public dataset. 2020. http://plataforma.saude.gov.br/coronavirus/dados-abertos/ (accessed May 10, 2020; in Portuguese).

13 Docherty AB, Harrison EM, Green CA, et al. Features of 16,749 hospitalised UK patients with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol. *medRxiv* 2020; published online April 28. DOI:10.1101/2020.04.23.20076042 (preprint).

14 Brat GA, Weber GM, Gehlenborg N, et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE Consortium. *medRxiv* 2020; published online April 30. DOI:10.1101/2020.04.13.20059691 (preprint).

15 Brazilian Institute of Geography and Statistics. Tabela 2094—População residente por cor ou raça e religião. 2010. https://sidra.ibge.gov.br/tabela/2094#/n1/all/n2/all/n3/all/v/1000093/p/last201/c86/allxt/c133/0/d/v1000093201/l/v,p+c86,t+c133/resultado (accessed May 10, 2020).

16 Qian Z, Alaa AM, van der Schaar M, Ercole A. Between-centre differences for COVID-19 ICU mortality from early data in England. *Intens Care Med* (in press).

17 Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81:** 515–26.

18 The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly* 2020; **2:** 113–22.

19 Wenham C, Smith J, Morgan R. COVID-19: the gendered impacts of the outbreak. *Lancet* 2020; **395:** 846–48.

20 Halpin DMG, Faner R, Sibila O, Badia JR, Agusti A. Do chronic respiratory diseases or their treatment affect the risk of SARS-CoV-2 infection? *Lancet Respir Med* 2020; **8:** 436–38.

21 Wikipedia. Lista de unidades federativas do Brasil por expectativa de vida. 2017. https://pt.wikipedia.org/wiki/Lista_de_unidades_federativas_do_Brasil_por_expectativa_de_vida (accessed May 10, 2020).

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

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

24 Costa NdR. A disponibilidade de leitos em unidade de tratamento intensivo no SUS e nos planos de saúde diante da epidemia da COVID-19 no Brasil. 2020. http://www.ensp.fiocruz.br/portal-ensp/informe/site/arquivos/ckeditor/files/DISPONIBILIDADE%20DE%20UTI%20NO%20BRASIL_27_03_2020(1).pdf (accessed May 10, 2020).

25 de Carvalho T. Saúde pública: um panorama do Brasil. 2018. https://www.politize.com.br/panorama-da-saude/ (accessed June 11, 2020).

26 Díaz de León-Martínez L, de la Sierra-de la Vega L, Palacios-Ramírez A, Rodriguez-Aguilar M, Flores-Ramírez R. Critical review of social, environmental and health risk factors in the Mexican indigenous population and their capacity to respond to the COVID-19. *Sci Total Environ* 2020; **733:** 139357.

27 Bailey SR, Telles EE. Multiracial versus collective black categories: examining census classification debates in Brazil. *Ethnicities* 2006; **6:** 74–101.

28 Cintra PHP, Fontinele Nunes F. Estimative of real number of infections by covid-19 on Brazil and possible scenarios. *medRxiv* 2020; published online May 12. DOI:10.1101/2020.05.03.20052779 (preprint).

29 Alves D, Gaete R, Miyoshi N, Carciofi B, Olveira L, Sanchez T. Estimativa de casos de COVID-19. 2020. https://ciis.fmrp.usp.br/covid19-subnotificacao/ (accessed May 10, 2020).