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

Pedro Baqui[1,*], Ioana Bica[2,3,*], Valerio Marra[1,4,+], Ari Ercole[5] and Mihaela van der Schaar[3,6,7]

1 Núcleo de Astrofísica e Cosmologia (Cosmo-ufes), Universidade Federal do Espírito Santo, 29075-910, Vitória, ES, Brazil

2 Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

3 The Alan Turing Institute, 96 Euston Rd, London NW1 2DB, UK

4 PPGCosmo & PPGFis & Departamento de Física, Universidade Federal do Espírito Santo, 29075-910, Vitória, ES, Brazil

5 University of Cambridge Department of Medicine, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK

6 Centre for Mathematical Sciences, University of Cambridge, Wilberforce Rd, Cambridge CB3 0WA, UK

7 Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA 90095 USA

*These authors contributed equally to this work.

+ Corresponding author

Background Brazil is rapidly ascending the ranks of countries with the highest number of COVID-19 cases and deaths. Understanding possible socioeconomic and ethnic health inequities is of particular importance given the diverse population and fragile political and economic situation.

Methods We performed a cross-sectional observational study of COVID-19 hospital mortality using observational data from the SIVEP-Gripe dataset to quantify the COVID-19 pandemic in Brazil. We assess regional variations in hospitalised COVID-19 patients by state socioeconomically grouping. Survival analysis was used to estimate the effects of ethnicity and comorbidity at an individual level in the context of regional variation.

Findings Compared to *branco* Brazilians, hospitalised *pardo* and *preto* Brazilians have significantly higher risk of mortality (hazard ratio [95% CI]=1.47[1.33-1.58] and 1.32[1.15-1.52] respectively). *Pardo* ethnicity was the second most important risk factor (after age). Comorbidity was more common in hospitalised Brazilians in the country's North region, with similar proportions between the various ethnic groups. States in the North had a higher hazard ratio compared to the Central-South although Rio de Janeiro was an outlier.

Interpretation We find evidence for two distinct, but correlated, effects: Increased mortality in the North (regional effect), and for the *pardo* and *preto* 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 may be related to differences in vulnerability to COVID-19 and access to healthcare access (including intensive care) across ethnicities. Our analysis motivates an urgent effort on the part of Brazilian authorities to consider how the national response to COVID-19 can better protect *pardo* and *preto* Brazilians as well as the population of poorer states from their higher death risk.
Funding None.

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 regime, may make it 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/systematic studies assessing the ethnic and regional variation in mortality are lacking.
Added value of this study
We found that *pardo* and *preto* Brazilians hospitalised with COVID-19 have statistically significant higher mortality compared to a *branco* comparator group. In particular, *pardo* ethnicity was the second most important risk factor after age. We also found that COVID-19 mortality increases in socioeconomically comparable Northern regions, and that Rio de Janeiro has an exceptionally high risk compared to its neighbouring states.
Implications of all the available evidence
Our results have serious social implications: *pardo* and *preto* Brazilians have, on average, less economic security, are less likely to be able to stay at home and work remotely, and also comprise a significant proportion of health and care workers. We hope that this analysis assists the authorities in better directing and aligning their response to COVID-19 in order to protect *pardo* and *preto* Brazilians from their higher death risk. 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 further urgent attention by the federal government.

INTRODUCTION

The COVID-19 pandemic has created an unprecedented worldwide strain on healthcare. Whilst early reports from East Asia and Europe meant that Brazil was well-positioned to implement non-pharmaceutical interventions, Brazilians, like those in many developing countries, have limited access to testing and social security.[1,2] The former makes it difficult to assess the growth of the pandemic, while the latter prevents a sizable fraction of society from engaging in physical distancing. This has been further complicated by an unstable federal government[3] that has failed to support measures such as social distancing and attempted to downplay the gravity of the pandemic as has been well publicised in the media.[4] Worryingly, as of May 19th, Brazil ranks fourth worldwide for total COVID-19 cases and sixth for deaths, with the highest estimated rate of transmission in the world (R_0 = 2.81),[5] second only to the USA for the daily increase of confirmed cases and deaths.

Worldwide there is substantial interest in the emerging societal inequity of the impact of COVID-19, and there is emerging evidence to suggest variability in the impact of the disease across ethnicities in a variety of settings including in the UK[6–8] USA[9,10] and Norway.[11] Brazil's population is particularly diverse, comprising many races and ethnic groups. The Brazilian Institute of Geography and Statistics (IBGE) racially classifies the Brazilian population in five categories (percentages as of 2010): *branca* (47.7%), *parda* (43.1%), *preta* (7.6%), *amarela* (1.1%) and *indígena* (0.4%). We will use the Portuguese terms throughout this work. This IBGE classification is based on color and, as in international practice, individuals are asked to self identify as either: *Branco* ("white"), *preto* ("black"), *amarelo* ("yellow"), *indígeno* ("Amerindians") 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. While *branco* and *pardo* Brazilians together comprise the majority of the population, with approximately equal proportions, their distribution varies considerably regionally. The population in the South macroregion is 78% *branca* and 17% *parda*, while the North macroregion's population is 23% *branca* and 67% *parda*.[12]

Brazil is a particularly important and interesting country in which to study the impact of COVID-19, due in part to the combination of the severity of the outbreak, governmental failure to implement non-pharmaceutical interventions, and complex social and ethnic societal composition. In this work, we analyse COVID-19 hospital mortality from the prospectively collected SIVEP-Gripe ("Sistema de Informação da Vigilância Epidemiológica da Gripe") respiratory infection registry data, which is maintained by the Ministry of Health for the purposes of recording cases of Severe Acute Respiratory Syndrome (SARS) across both public and private hospitals. Using this rich dataset, we characterise the COVID-19 pandemic in Brazil, particularly with regard to risk factors related to comorbidities, symptoms and ethnicity, similarly to previous analyses in countries such as the UK.[8,13–15]

METHODS

Our analysis is based on the SIVEP-Gripe public dataset.[16] As of the time of access, this contains epidemiological data for 99,557 patients from different states. Each entry has 138 features, including symptoms, age, sex, ethnicity and comorbidities. Applying the condition

of having tested positive (RT-PCR) for SARS-CoV-2 leaves data for 19,940 patients. As we are interested in the relation between ethnicity and health risk, we then consider only data with ethnicity recorded, leaving data for 12,221 patients. Furthermore, we consider only the subset that was hospitalised: this is our base dataset of 11,321 patients, see Fig.1. The date of diagnosis spans the time interval from 27th February 2020 to 4th May 2020. Our analysis employs descriptive statistics to quantify the COVID-19 pandemic in Brazil, and Cox regression to estimate hazard ratios (HR) for factors including education, income and health indexes such as ICU availability and public / private healthcare staffing.

Brazil is divided geopolitically into 5 macroregions:

- North: Acre (AC), Amapá (AP), Amazonas (AM), Pará (PA), Rondônia (RO), Roraima (RR), Tocantins (TO);
- Northeast: Alagoas (AL), Bahia (BA), Ceará (CE), Maranhão (MA), Paraíba (PB), Pernambuco (PE), Piauí (PI), Rio Grande do Norte (RN), Sergipe (SE);
- Central-West: Distrito Federal (DF), Goiás (GO), Mato Grosso (MT), Mato Grosso do Sul (MS);
- Southeast: Espírito Santo (ES), Minas Gerais (MG), Rio de Janeiro (RJ), São Paulo (SP);
- South: Paraná (PR), Rio Grande do Sul (RS), Santa Catarina (SC).

For descriptive purposes, we chose to dichotomise the data into two maximally contrasting regions based on 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 utilisation, are not included but are expected to correlate with the above socio-economic factors. Ethnicity was not considered at this point.

The two regions that we consider are:

- “Central-South” comprising Central-West, Southeast and South macroregions (9,278 patients; 81%),
- “North” comprising North and Northeast macroregions (2,043 patients; 19%).

This is the customary division when considering Northern and Southern Brazil and is socioeconomically quantitatively well justified (supplementary material, page 16).

The SIVEP-Gripe data include information on comorbidities and symptoms. Missingness is described in the supplementary materials (page 2). We naturally interpret missing values as the absence of comorbidities or symptoms. Missing values are also present for ICU admissions. In this case, we consider missing values as non-admissions to ICU.

We used mixed-effects Cox regression to investigate the effects of record-level risk factors. We used patient-level clinical features, namely age group, sex, ethnic group, and comorbidities, as fixed effects, with state as a random effect (similarly to a recent UK analysis[17]). For the categorical variables age group and ethnic group, we used Age < 40 and *branco* as reference categories, respectively. We did not find evidence[18] for statistically significant violation of the proportional hazards assumption (p=0.11).

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

Fig. 2 shows the distribution of COVID-19 cases across states. SP, RJ, and AM have the highest number of cases, both absolute and per capita. Table I and II shows demographic and comorbidity data among survivors and non-survivors of COVID-19 and ethnic composition at each stage of the COVID-19 trajectory. Survivors in both North and Central-South regions are younger and more likely to be white and female, while non-survivors are more likely to be *preto* and *pardo*, respectively (results regarding other ethnicities are more difficult to interpret due to the lower numbers). Almost all comorbidities are more common in non-survivors in the North compared to Central-South suggesting structural health disparity. This is further evidenced by the significantly larger percentage of non-survivors in the North.

The total number of hospitalised patients to the number of deaths (Table II) reveals a similar trend with higher mortality in the North (regional effect) and in *preto* and *pardo* Brazilians (ethnicity effect).

Fig. 3 (panels A and B) shows comorbidity distributions by ethnicity and comorbidity count for survivors and non-survivors (excluding *indígeno* patients due to small numbers). There is a notable North/Central-South asymmetry, with more non-survivors in the former. Fig. 3 (panels C and D) shows symptom count (fever, cough, sore throat, shortness of breath, respiratory discomfort, SpO_2<95%, diarrhoea, and vomiting) and ethnicity for survivors and non-survivors. Most patients present between 3 and 6 symptoms, suggesting again that in this dataset, it is the more severe presentations being tested for COVID-19.

Fig. 3 (panels E and F) presents distributions according to age and ethnicity for survivors and non-survivors. In the North, the trend of younger patients having a higher likelihood of survival is even more pronounced than for the Central-South region. Additionally, younger *pardo* and *preto* Brazilians appear to be less likely to survive the virus compared to *branco* Brazilians, with the difference being more pronounced in the Central-South region.

Fig. 4 (panel A) shows HRs for all clinical features (fixed effects) considered in the Cox model. Compared to *branco* Brazilians, hospitalised *pardo* and *preto* Brazilians have significantly higher risk of mortality. Notably, *pardo* ethnicity was the second most important risk factor (after age). Fig. 4 (panel B) shows HRs for all different states in Brazil considered. Substantial between-states variation is apparent. The states in the North region tend to have higher HRs than those in the Central-South region, further justifying our approach of splitting Brazil into two sets. This also corresponds with the regional effect discussed earlier.

DISCUSSION

We present, to our knowledge, the most extensive study of COVID-19 hospital survival in Brazil. We show that survivors are younger[19] with female preponderance,[20] and tend to have

fewer comorbidities,[21] in keeping with worldwide findings. However, we also report a number of other important socio-demographic trends specific to Brazil.

There is significant regional variation in both case-mix and outcome. The high number of cases in SP, RJ and AM (Fig. 2) are noteworthy. These regions are important ports of entrance to Brazil: AM hosts the Free Economic Zone of Manaus and most of the international flights route through SP and RJ: in 2019, 7.7 million international passengers landed in SP, and 2.2 million in RJ (further details in the supplementary materials, page 17). Furthermore, both SP and RJ are characterised by a particularly high population density; the outbreak coincided with the rainy season (associated with respiratory infections) in AM.

The finding of a higher comorbidity burden in hospitalised patients in the North region is concordant with a lower life expectancy[22] mirroring differences in the average age of survivors/non-survivors between these regions and the significantly larger percentage of non-survivors in the North. Survivors were more likely to be *branco*, and *branco* Brazilians were more likely to be admitted to ICU than *pardo* Brazilians. In other words, the increased death rate of *pardo* Brazilians seems to be due in part to non-ICU admission, raising concerns regarding the organisation of public and private medical resources.

It is also noteworthy that the distribution of hospitalised COVID-19 patients between the Central-South (81%) and North (19%) regions in our data is discordant with the population sizes of these regions (64% and 36%, respectively).[12] This highlights national heterogeneity and may, at least in part, be due to either intrinsically lower hospitalisation rates in the North region and/or the disproportionate impact of COVID-19 in populous areas such as São Paulo and Rio de Janeiro (both in the Central-South region).

The disproportionately large percentage of comorbidity-free survivors in the Central-South region (Fig. 3) is remarkable. We may speculate that this may be due to differences in comorbidity ascertainment either due to structural differences in the way data is collected (perhaps comorbidity data was less available from patients who were sicker at the time of presentation) or because less severe patients, perhaps with concerns regarding their comorbid health, presented to hospital preferentially in the Central-South region. It is interesting to observe that *branco* and *pardo* Brazilians have a similar number of comorbidities in these populations. Therefore, it seems unlikely that comorbidities are associated with ethnicity in the group studied, but rather they may be correlated with regional socioeconomic development (education, income, and health).

However, an interplay between ethnic and regional socioeconomic factors is apparent in Fig. 3, with younger *pardo* and *preto* Brazilians seemingly less likely to survive; the difference is more pronounced in the Central-South. For context, the typical life-expectancy in Brazil is 76 years (as of 2017)[23], compared to 80.9 years in Europe.[24] Average life expectancy varies by region, being higher in the Central-South (SC; 79.4) than in the North (MA; 70.9) providing a baseline for the trend shown in Fig. 3.

ICU access may be a factor for regional and ethnic variations in mortality, with *branco* Brazillians more likely to be admitted to ICU once hospitalised. Whilst *branco* Brazilians are more likely to survive overall, by comparing total hospitalisation with deaths after ICU admission, one sees more similar proportions between both ethnicities. Note that the

distribution of comorbidities, symptoms and age does not show strong ethnic variations, especially between pardo and branco (Fig. 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 healthcare for *branco* Brazilians compared to *pardo* Brazilians, as ICU admission policies are known to differ between public and private hospital settings.[25] Note that private healthcare serves only 25% of the population and total spending is comparable to that of public healthcare, implying that, on average, a private-hospital patient costs three times as much as compared to a public-hospital patient.[26] The fact that the proportions of the different ethnicities admitted to ICU with COVID-19 are similar to those in the full 2019 SIVEP-Gripe dataset[16] suggests that this is not a specific feature of COVID-19 (supplementary materials, page 18).

Survival analysis showed that, after age, the most important factor for hospital mortality was being of *pardo* or, to a lesser extent, *preto* ethnicity (HR=1.47(1.33-1.58) and 1.32(1.15-1.52)(95% CI) compared to *branco* baseline). The other risk factors largely replicate worldwide findings, although we see that male sex is perhaps slightly less of a risk factor compared to the findings presented in other series.[20] This ethnic inequity has important social roots and implications: *pardo* and *preto* Brazilians have, on average, less economic security, contagion-prone living conditions, are less likely to be able to stay at home and work remotely, and comprise a significant proportion of health and care workers, making them disproportionately the most vulnerable to COVID-19.[2] In particular, *pardo* and *preto* 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.[27] As a proxy, sewerage cover is also congruent with our findings (supplementary material, page 16). Cox regression results were qualitatively similar for North/Central-South, metropolitan/rural subgroups and public/private predominant healthcare subgroups suggesting robustness to differences in outbreak start and outliers (supplementary material, page 3–11).

We see substantial variation in hazard by region. The states in the North region tend to have higher HRs than those in the Central-South region, concordant with the larger percentage of non-survivors in the North, as shown in Table I. Incorporating the number of ICU beds/ventilators and nurses per 100 million inhabitants for each region as proxies for physical availability of healthcare resources did not qualitatively change our result (supplementary materials, page 3), suggesting a more fundamental difference in healthcare access and trajectory.

Rio de Janeiro, despite high standards of education, income and health, has one of the highest HRs, comparable to those in the Pernambuco and Amazonas states. States in the North have a HR greater than one and a high proportion of *pardo* Brazilians (supplementary materials, page 18). Again, RJ is an exception, exhibiting a much higher HR than the neighbouring state of São Paulo and an ethnicity profile similar to the states of the North region. The HR for the metropolitan area of RJ is twice that of the comparable local rural area (supplementary material, pages 8–9)- similar differences are not seen between other metropolitan areas and their rural neighbouring regions. Furthermore, the disparity in HR between public/private healthcare mortality in Rio de Janeiro was the highest in Brazil suggesting access to high quality healthcare in the metropolitan area together with a large *pardo* community as important drivers of outcome here.

Many *preto* Brazilians may identify themselves as *pardo* Brazilians.[28] For this reason, it is reasonable to consider the *preto* and *pardo* populations together. Indeed, as seen from our analysis, both ethnic groups share higher percentages of non-survivors and higher HRs.

The results of our analysis can then be interpreted according to the interplay of regional and ethnicity effects. We may speculate that the regional effect is due to expected variations in levels of comorbidities (or poorly controlled comorbidities) and general healthcare access, which we might expect to have a notable impact in regions such as the North where socioeconomic levels are lower. We similarly postulate that the ethnicity effect is driven by the greater vulnerability to COVID-19 and reliance on publicly-funded healthcare and limited ICU access in *pardo* and *preto* communities.

For most states, the regional and ethnicity effects are correlated, giving a larger cumulative mortality. Indeed, lower socioeconomic development correlates with a larger *pardo* and *preto* population although RJ is an outlier, with an ethnic composition (ethnicity effect) that is similar to the states in the North region, but high levels of development (regional effect) more akin to those of Central-South states.

Whilst we believe our work is the most comprehensive of its kind to date in Brazil, there are a number of limitations which need discussion. Limitations and possible biases from case ascertainment cannot be ruled out, in common with all observational and database research. Ethnicity is missing in 39% of our data. This is comparable to a recent large dataset from the UK (26%).[8] In Table 2 (second column) we observe that the percentage of hospitalised *branco* Brazilians is lower than the population percentage in the North, and higher in the Central-South, and vice-versa for *pardo* Brazilians. This could indicate that COVID-19 spreads differently through the ethnic groups within the two regions, rather than being an effect of missing ethnicity data. Indeed, in the Central-South the disease spread initially among *branco* Brazilians (especially in populous São Paulo and Rio de Janeiro) who tend to travel internationally more frequently, while in the North more precarious living and working conditions may have been more important (see also the discussion above regarding the Free Economic Zone of Manaus).

We have limited our analysis to patients who were hospitalised, since testing in the community is more likely to be biased according to local factors. Again, however, we cannot be sure that the availability of testing practice is homogeneous even in this population. Indeed, the fact that a large proportion of patients that have tested positive are admitted to hospital clearly shows that testing, at least as far as this dataset is concerned, is performed only when symptoms are severe, indicating in turn that the number of COVID-19 cases in Brazil is likely to be much higher than suggested by available data.[29,30]

Health-seeking behaviour may vary with ethnicity and region; late presentation may be an important determinant of hospital outcome. We are not able to consider this in our analysis, as data for physiological severity at hospital presentation are not available. However a recent UK study did not demonstrate an important effect of physiological severity[17] 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 (supplementary materials, pages 12–15).

Whilst we have focused on hospital mortality, it is important to appreciate that we do not have data on out-of-hospital mortality (which may be substantial) and neither can we robustly address the question of access to hospital services by region, ethnicity or socioeconomic status. As such, a consideration of hospital mortality is likely to substantially underestimate the true impact of COVID-19, and it is plausible to assume that healthcare availability inequities would be further amplified in patients who are not hospitalised. In other words, it is sensible to 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 *preto* Brazilians, and in the country's North region. At the time of writing, the Brazilian federal administration has not supported non-pharmacological interventions such as social distancing. Our results suggest that major metropolitan areas may be particularly affected, and that it is highly plausible that viral spread may be particularly rapid in these settings. Urgent work is needed to understand the impact of R_0 in these areas and testing should be increased. Even without this detail, however, 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 healthcare access, as these may lead to delayed presentation 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 carried out the analysis with descriptive statistics. IB carried out 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 is publicly available from http://plataforma.saude.gov.br/coronavirus/dados-abertos/sivep-gripe. Our analysis code is available from https://github.com/ioanabica/COVID-19-Brazil.

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 and 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. 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 May 10, 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 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 and 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 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).

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.

14 Mehra MR, Desai SS, Kuy S, Henry TD, and Patel AN. Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. *New England Journal of Medicine* 2020; published online May 1. DOI:10.1056/NEJMoa2007621.

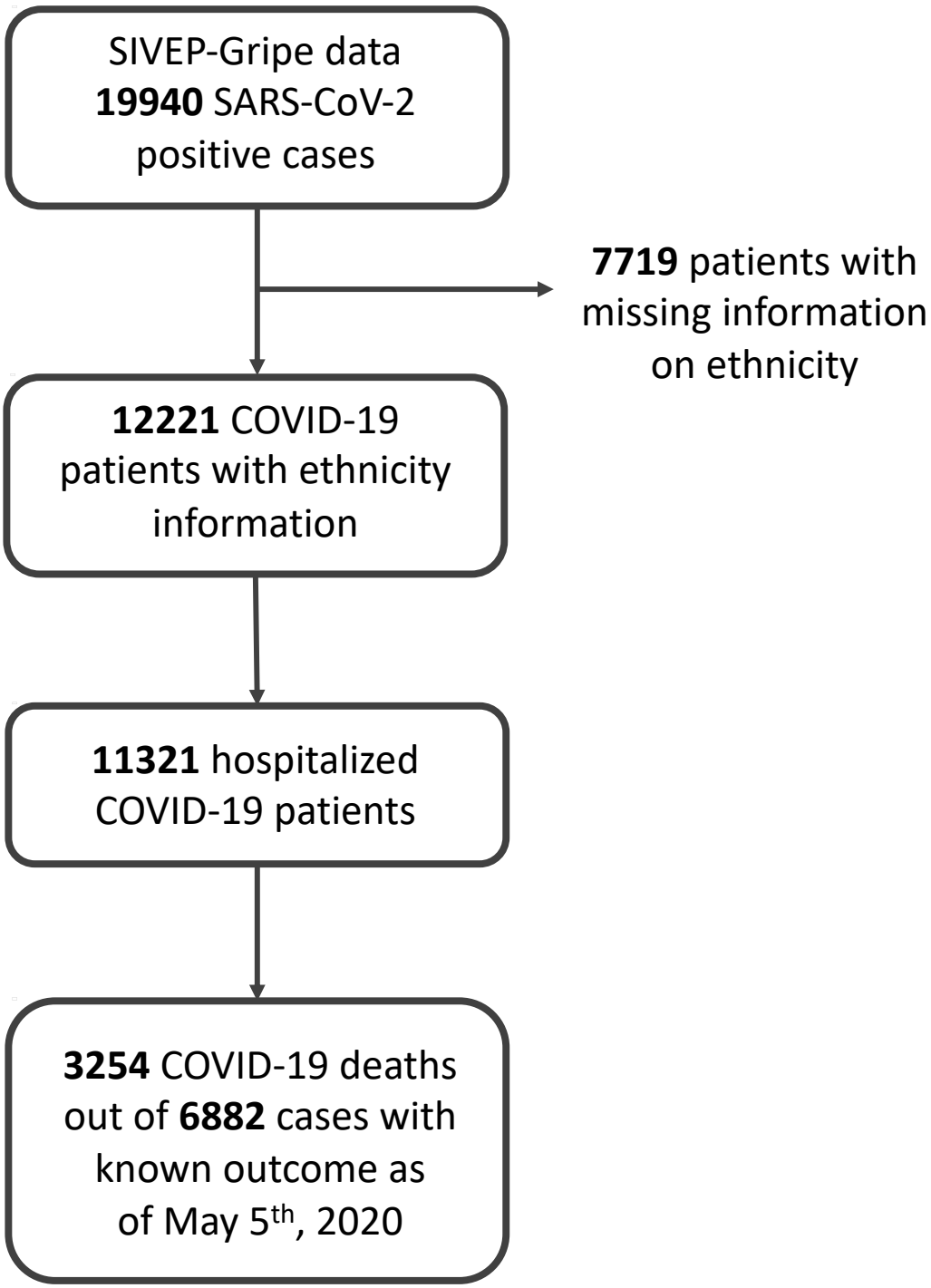
15 Brat GA, Weber GM, Gehlenborg N, et al. International Electronic Health Record-Derived COVID-19 Clinical Course Profiles: The 4CE Consortium. *medRxiv* 2020; published online April 30. DOI: 10.1101/2020.04.13.20059691.

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

17 Qian Z, Alaa AM, van der Schaar M, and Ercole A. Between-centre differences for COVID-19 ICU mortality from early data in England. *Intens Care Med* 2020 (in press).
18 Grambsch PM and Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81**: 515–26.
19 The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) - China, 2020. *China CDC Weekly* 2020; **2**: 113–22.
20 Wenham C, Smith J, and Morgan R. COVID-19: the gendered impacts of the outbreak. *Lancet* 2020; **395**: 846–48.
21 Halpin DMG, Faner R, Sibila O, Badia JR, and Agusti A. Do chronic respiratory diseases or their treatment affect the risk of SARS-CoV-2 infection? *Lancet Respir Med* 2020; **8**: 436–38.
22 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).
23 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).
24 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).
25 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).
26 de Carvalho T. Saúde pública: um panorama do Brasil. 2018. https://www.politize.com.br/panorama-da-saude/ (accessed June 11, 2020).
27 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. *Science of The Total Environment* 2020; **733**: 139357.
28 Bailey SR and Telles EE. Multiracial versus Collective Black Categories: Examining Census Classification Debates in Brazil. *Ethnicities* 2006; **6**: 74–101.
29 Cintra PHP and 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.
30 Alves D, Gaete R, Miyoshi N, Carciofi B, Olveira L, and Sanchez T. Estimativa de Casos de COVID-19. 2020. https://ciis.fmrp.usp.br/covid19-subnotificacao/ (accessed May 10, 2020).

FIGURES

Figure 1: STROBE flowchart of SIVEP-Gripe data used in this study.



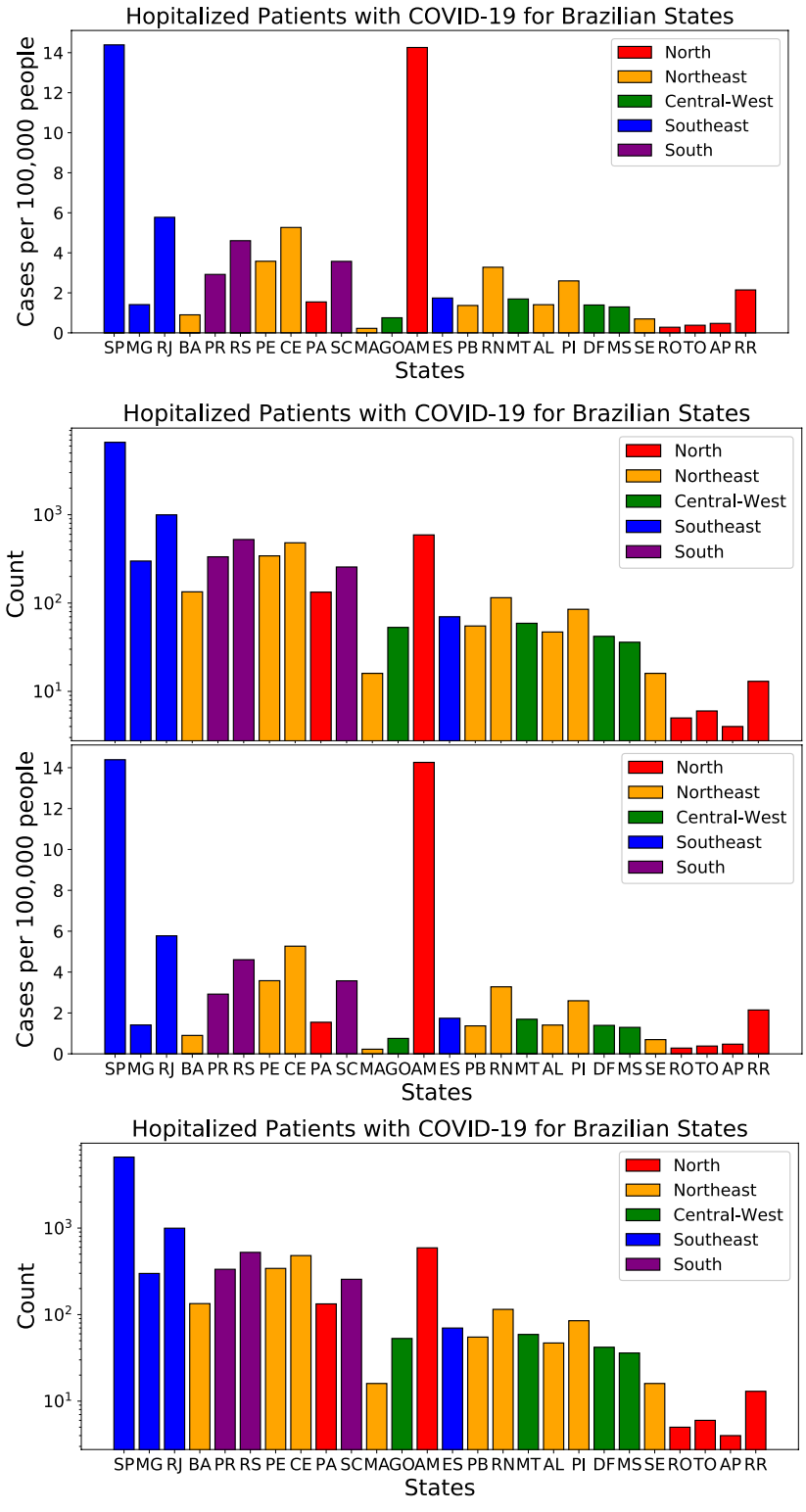
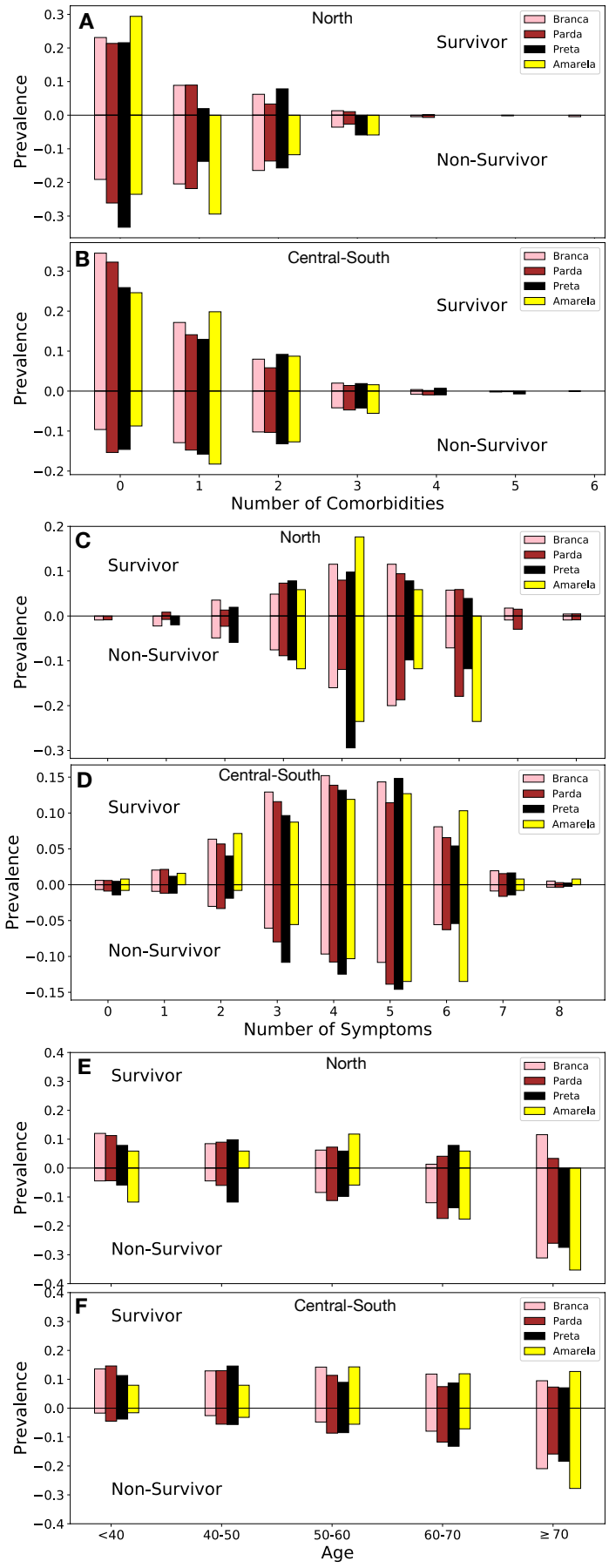
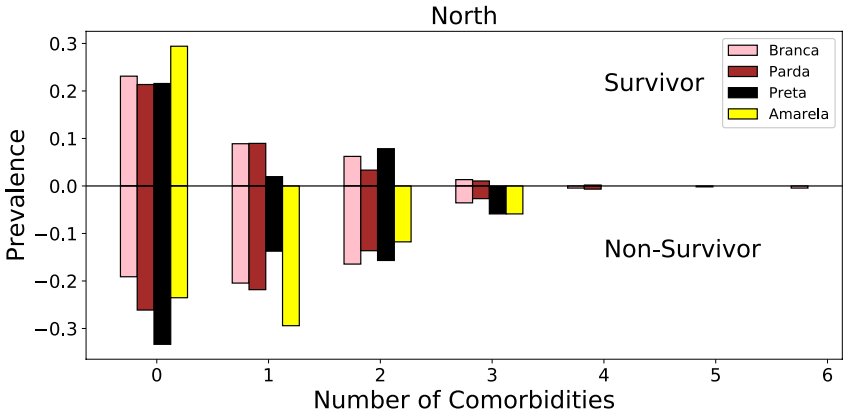
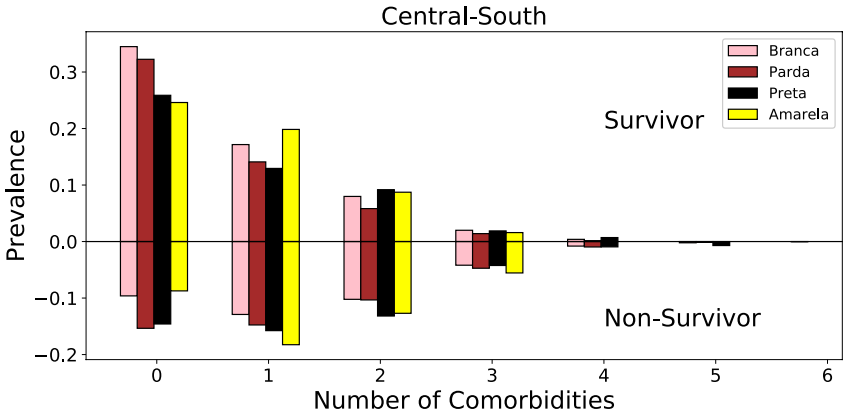
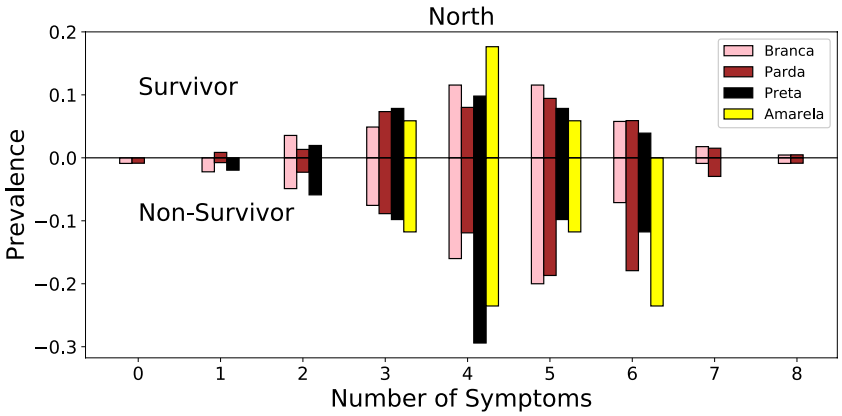


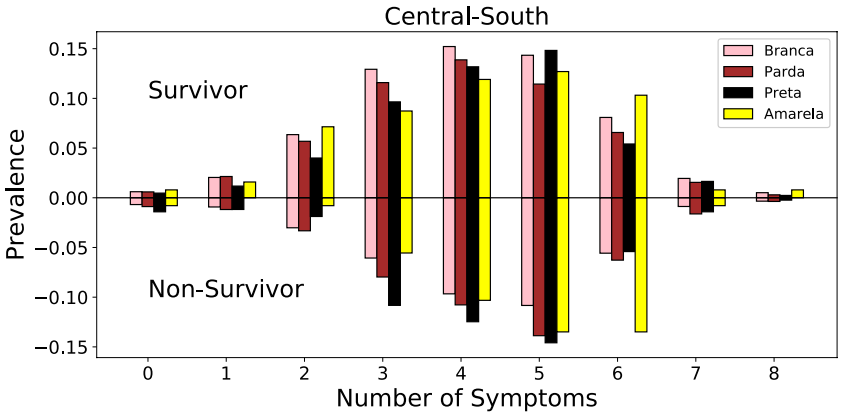
Figure 2: Distribution of the dataset considered in this work (11,321 patients) among the Brazilian states according to absolute (A) and per capita (B) number of cases. States are ordered according to their population, larger on the left.


A
North
Survivor
Non-Survivor
Branca
Parda
Preta
Amarela
Prevalence
B
Central-South
Number of Comorbidities
C
North
D
Central-South
Number of Symptoms
E
North
F
Central-South
<40
40-50
50-60
60-70
≥ 70
Age


North
Branca
Parda
Preta
Amarela
Survivor
Non-Survivor
Prevalence
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
0
1
2
3
4
5
6
Number of Comorbidities


Central-South
Branca
Parda
Preta
Amarela
Survivor
Non-Survivor
Prevalence
0.3
0.2
0.1
0.0
−0.1
−0.2
0
1
2
3
4
5
6
Number of Comorbidities


North
Branca
Parda
Preta
Amarela
Survivor
Non-Survivor
Prevalence
0.2
0.1
0.0
−0.1
−0.2
−0.3
0
1
2
3
4
5
6
7
8
Number of Symptoms


Central-South
Branca
Parda
Preta
Amarela
Survivor
Non-Survivor
Prevalence
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
0
1
2
3
4
5
6
7
8
Number of Symptoms

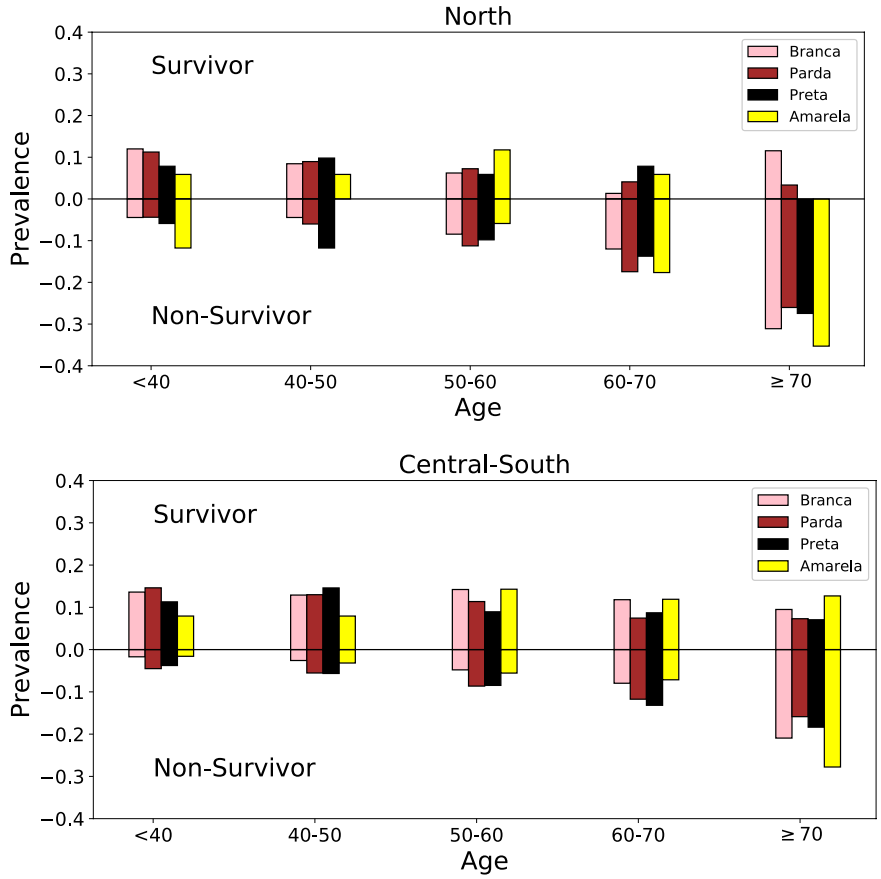


Figure 3: Distributions according to number of comorbidities (A and B), symptoms (C and D) and age (E and F) and ethnicity. The normalisation is such that all the fractions of a given ethnicity add to unity (to adjust for differences in ethnic prevalence). We exclude *indígeno* patients for clarity due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

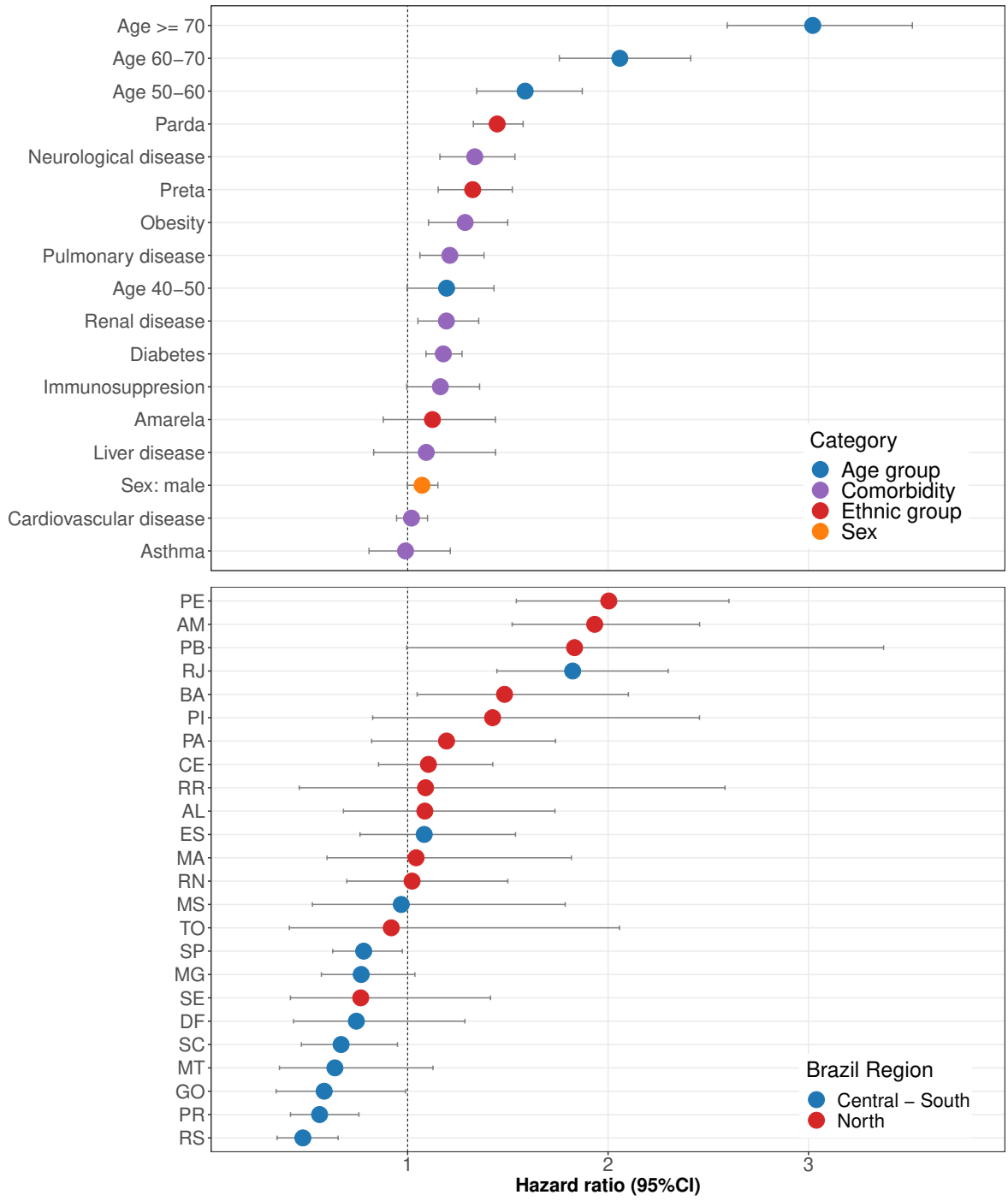


Figure 4: Hazard ratios, with 95% confidence intervals, for all clinical features (fixed effects; A) and for all different states in Brazil (random effects; B) considered in the fitted multivariate mixed-effects Cox model for in-hospital mortality.

TABLES

	survivors	non-survivors
North (2043 data)	**479 (35.5%)**	**871 (64.5%)**
Age [yr]	46.9±19.3	65.3±16.0
Female sex [no. (%)]	218 (39.3%)	337 (60.7%)
Male sex [no. (%)]	261 (32.8%)	534 (67.2%)
Ethnic group [no. (%)]		
branca	89 (39.6%)	136 (60.4%)
parda	366 (34.9%)	683 (65.1%)
preta	16 (31.4%)	35 (68.6%)
amarela	5 (29.4%)	12 (70.6%)
indígena	3 (37.5%)	5 (62.5%)
Comorbidities [no. (%)]		
Cardiovascular disease	95 (22.6%)	325 (77.4%)
Asthma	22 (62.9%)	13 (37.1%)
Diabetes	74 (19.9%)	297 (80.1%)
Pulmonary disease	15 (29.4%)	36 (70.6%)
Obesity	13 (22.4%)	45 (77.6%)
Immunosuppression	28 (57.1%)	21 (42.9%)
Renal disease	13 (18.8%)	56 (81.2%)
Liver disease	4 (23.5%)	13 (76.5%)
Neurological disease	7 (21.2%)	26 (78.8%)
Central-South (9278 data)	**3564 (59.2%)**	**2457 (40.8%)**
Age [yr]	52.2±16.6	67.0±15.8
Female sex [no. (%)]	1525 (60.4%)	1001 (39.6%)
Male sex [no. (%)]	2039 (58.3%)	1456 (41.7%)
Ethnic group [no. (%)]		
branca	2548 (62.0%)	1560 (38.0%)
parda	728 (53.7%)	627 (46.3%)
preta	215 (50.6%)	210 (49.4%)
amarela	69 (54.8%)	57 (45.2%)
indígena	4 (57.1%)	3 (42.9%)
Comorbidities [no. (%)]		
Cardiovascular disease	936 (44.9%)	1147 (55.1%)
Asthma	158 (64.8%)	86 (35.2%)
Diabetes	641 (42.1%)	880 (57.9%)
Pulmonary disease	115 (34.2%)	221 (65.8%)
Obesity	130 (48.9%)	136 (51.1%)
Immunosuppression	104 (40.0%)	156 (60.0%)
Renal disease	87 (27.2%)	233 (72.8%)
Liver disease	25 (37.9%)	41 (62.1%)
Neurological disease	84 (29.7%)	199 (70.3%)

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

North (2043 data)	Brazilian population	Hospitalisation	ICU Admission	Death	Death/ Hospitalisation	Death (not ICU)	Death (ICU)
branca	*28.8%*	342 (16.7%)	127 (19.4%)	136 (15.6%)	39.8%	69 (14.9%)	67 (16.5%)
parda	*61.5%*	1567 (76.7%)	481 (73.8%)	683 (78.4%)	43.6%	368 (79.3%)	315 (77.4%)
preta	*8.8%*	85 (4.2%)	26 (4.0)%)	35 (4.0%)	41.2%	20 (4.3 %)	15 (3.7%)
amarela	*1.2%*	36 (1.8%)	13 (2.0)%)	12 (1.4%)	33.3%	5 (1.1 %)	7 (1.7%)
indígena	*0.7%*	13 (0.6%)	5 (0.8%)	5 (0.6%)	38.5%	2 (0.4%)	3 (0.7%)
Cental-South (9278 data)	**Brazilian population**	**Hospitalisation**	**ICU Admission**	**Death**	**Death/ Hospitalisation**	**Death (not ICU)**	**Death (ICU)**
branca	*58.7%*	6291 (67.8%)	2344 (69.4%)	1560 (63.5%)	24.8%	616 (60.3%)	944 (65.8%)
parda	*33.2%*	2112 (22.8%)	731 (21.6%)	627 (25.5%)	29.7%	278 (27.2%)	349 (24.3%)
preta	*6.8%*	667 (7.2%)	220 (6.5%)	210 (8.6%)	31.5%	108 (10.6%)	102 (7.1%)
amarela	*1.1%*	195 (2.1%)	82 (2.4%)	57 (2.3%)	29.2%	19 (1.9%)	38 (2.7%)
indígena	*0.02%*	13 (0.1%)	2 (0.1%)	3 (0.1%)	23.1%	1 (0.1%)	2 (0.1%)

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