Ethnic Composition and Outcomes of COVID-19 Patients

Ahmed M. Alaa[1], Zhaozhi Qian[2], Jem Rashbass[3], Keith Gomes Pinto[3], Jonathan Benger[3], and Mihaela van der Schaar[2]

[1]University of California, Los Angeles, CA 90095, USA
[2]University of Cambridge Centre for Mathematical Sciences, Wilberforce Rd, Cambridge CB3 0WA, UK
[3]NHS Digital

The rate of SARS-CoV-2 infections, COVID-19-related hospitalizations and deaths are hypothesized to be disproportionately high among members of the Black, Asian, Mixed-race and Ethnic minorities (BAME) community. In this report, we analyse the characteristics and outcomes of COVID-19 patients from the BAME population in comparison with the White population. Our goal is to analyse the ethnic composition of the current COVID-19 population in order to verify whether BAME patients are particularly vulnerable to COVID-19, and identify potential factors that could explain any increased susceptibility to the disease.

Data

In our analysis, we linked five different data sets in order to construct a comprehensive record of the different stages of COVID-19 patients' health trajectories. The data sets involved in our analysis are described below.

SGSS data. The first data set involved testing data obtained from the Second Generation Surveillance System (SGSS), covering the period to 16th April 2020. The data set contains 78,443 people with positive SARS-CoV-2 lab tests. This data set comprises patients at the first stage of their interaction with the NHS (diagnosis).

CHESS data. The next stage is hospitalization and comprises patients who are more severely ill. Data for COVID-19 hospitalizations was obtained from the COVID-19 Hospitalization in England Surveillance System (CHESS) established by Public Health England (PHE). This data covered the period from February 8th to April 14th 2020 (7,714 hospital admissions, including 3,092 ICU admissions from 94 NHS hospital trusts across England).

PDS data. We utilized the Personal Demographics Service (PDS) data from NHS Digital, which records 15,090 deaths due to COVID-19 up to April 20th, 2020.

HES data. We used data from Hospital Episode Statistics (HES) in order to obtain information regarding each patient's ethnicity. If a patient had multiple HES records with conflicting ethnicity, we chose the one with highest frequency.

Primary care data. Finally, we used primary care prescription medicine data between July 2019 and January 2020 to extract information on each patient's likely pre-existing medical conditions. Comorbidities were inferred on the basis of the British National Formulary (BNF) chapter code for all medications prescribed to each patient.

Using the data sets above, we constructed one comprehensive data set for 72,358 COVID-19 patients, including information on whether each patient was hospitalized, whether they were admitted to the ICU, their comorbidities, in addition to basic demographic information (age, sex and geographical region), and their outcome on April 20th, 2020. The data extraction process is shown in Figure 1.

Methods

We divided the COVID-19 population in England into 4 ethnic groups: White, Asian, Black and Other ethnic background. The definition of these ethnic groups were based on the standard HES ethnicity index. Descriptive statistics for age and the ethnic groups of interest were collected from the data sets above. In addition, a multivariate mixed-effects Cox regression model was fit to adjust for the effect of age and comorbidities on outcome, we also fit, with the fixed effects as the patient-level clinical predictors, and the random effects as the patient's geographical location.

Results

Of 78,443 patients with a positive COVID-19 test in the SGSS testing data, 72,358 (92%) had information on ethnicity and formed the population of interest (Figure 1). By comparing the representation of patients from different ethnic groups at the different stages of COVID-19 disease (Figure 2), we found that BAME patients constitute a disproportionately large fraction of hospitalizations and

ICU admissions, especially in younger age groups. Conversely, in the group of patients aged 70 years and older, most of the patients were from the White population. Table 1 provides further summary statistics for each ethnic group. The proportion of patients who died in the Black and Other ethnic groups exceeded the proportion of these groups in the general population as reported in the 2011 census, whilst the proportion of deaths in the White ethnic group was lower. However, direct comparisons with the whole population in England in 2011 should be treated with caution because COVID-19 cases have occurred disproportionately in London, the West Midlands and other urban areas where the population tends to be more ethnically diverse, and the underlying population may have altered to some extent since 2011 (see below).

Age disparities among ethnic groups. Figure 3 provides a detailed break down of the age distribution of COVID-19 patients within each ethnic group. A significant difference in the age distributions of the White and BAME populations is apparent (Panel A). In particular, we found that the median age of diagnosed patients in the Asian ethnic group was 51 years, significantly younger than the White patient population, for which the median age was 69 years. In Panels C and B of Figure 3, we compare the representation of each ethnic group within specific age groups, and compare these with the ethnic composition of the baseline demographic based on the 2011 census in England and Wales. Compared to the baseline demographic, we observed an excess in the number of patients from the Asian population in the age group of less than 40 years old (15%) and 40-50 years old (19%) (Figure 3, Panel B). On the other hand, the elderly population (more than 70 years old) was dominated by the white population (84%).

The marked difference in the age distribution between the white and BAME groups explains the lower rates of ICU admission in the white population, and higher rates of ICU admission in all other ethnic groups. The age gap between the white and BAME populations may be explained by: (i) the difference in the underlying age distribution within each ethnic group in the population, (ii) differences in social deprivation which is consistently higher in BAME populations (Figure 5), or (iii) differences in the age for the first onset of chronic diseases and co-morbidities that themselves influence outcome (Figure 6).

Adjusting for baseline demographics. While BAME patients appear to be over-represented in the COVID-19 diagnoses with respect to their representation in the overall demographics of England (Figure 3, panels B and C), it is important to note that the COVID-19 population is more concentrated in the more ethnically diverse regions of England. That is, while the population of greater London comprises only around 14% of the overall population of England, more than 24% of COVID-19 cases in our dataset were associated with the greater London area. By comparing the rates of diagnosis in each ethnic group with their corresponding regional demographics (Figure 4), we found that the diagnosis of COVID-19 generally reflects the underlying ethnic distribution in each region, however White and Asian populations are slightly less likely to be diagnosed with COVID-19, compared to their regional representation, while Black and Mixed populations are more likely to be diagnosed.

Prevalence of comorbidities. The vulnerability of younger patients in ethnic minorities to COVID-19 may also be explained by the prevalence of comorbidities in these populations. As we can see in Figure 6, younger BAME patients displayed higher prevalence cardiovascular disease, endocrine disease and hypertension compared to the white population.

Based on the observations above, the prevalence of comorbidities at a younger age, combined with a younger demographic, higher deprivation and the fact that ethnically diverse regions in England exhibited higher infection rates may explain the disproportionate rates of hospitalizations and ICU admissions among young BAME patients. In the rest of this report, we examine the potential effect of ethnicity as a risk factor when adjusting for age and comorbidities through a multi-variate regression model.

Ethnicity and risk elevation. Figure 7 shows the coefficients for all the clinical predictors under consideration in the fitted multi-variate Cox model for mortality. We can see that an Asian ethnic background significantly elevates mortality risk even when adjusting for age and comorbidities. This is explored further in Figure 8 in which this analysis is repeated, separating the Asian ethnic group into Indian, Pakistani and Bangladeshi in comparison to other Asian ethnicity, and demonstrating that the Indian, Pakistani and Bangladeshi ethnic group has a higher mortality risk than other Asian individuals. To further illustrate the elevated risk within the Asian subgroup, Figure 9 shows the Kaplan-Meier estimates for the probability of death plotted against the number of days since diagnosis stratified by various age groups. We plot these curves for both patients in the overall population and patients who have CHESS records (hospitalized). As we can see, Asian patients display an elevated risk in older age groups (more than 70 years old) for both the hospitalized and non-hospitalized subgroups, and for the hospitalized age group between 60 and 70 years old.

Discussion

There is a marked difference in the age distribution of BAME and White patients diagnosed with COVID-19. While younger patients in the BAME population exhibited disproportionate diagnosis and ITU admission, this apparent over-diagnosis may be explained by the wider spread of COVID-19 in ethnically diverse regions in England, the younger median age of the BAME population compared to the white population, higher rates of deprivation in the BAME population and the higher prevalence of other medical conditions within this population.

While the disproportionate diagnosis rate in BAME can be explained through demographic factors, we still observed significantly higher rates of hospitalization within younger BAME patients. Moreover, an Asian ethnic background, particularly Indian, Pakistani or Bangladeshi, was found

	Population (2011)	Diagnosis	Hospitalization	ICU Admission	Deaths
White	48,209,395 (86%)	53,820 (74.6%)	5,149 (79.6%)	1,629 (68.7%)	11,286 (79.4.6%)
Asian	4,213,531 (7.5%)	7,100 (9.8%)	581 (9.0%)	346 (14.6%)	989 (6.9%)
Black	1,864,890 (3.3%)	4,778 (6.6%)	323 (5.0%)	155 (6.5%)	811 (5.7%)
Other	1788096 (3.2%)	6,515 (9.0%)	414 (6.4%)	239 (10.0%)	1,127 (7.9%)

Table 1: Ethnic composition of patients at each stage of the COVID-19 trajectory.

to be a significant independent risk factor for mortality even after adjusting for age and co-morbidities. Of note, endocrine diseases (the vast majority of which will be diabetes) are also a strong independent predictor of mortality, and more prevalent in the Asian ethnic group. The elevated risk within the Asian ethnic group may be attributed to other confounding clinical, social or genetic factors, however we are unable to explore these further in this analysis.

	Respiratory	Cardiovasc.	Endocrine	Malignant & Immunosup.
Asian	21.7%	64.9%	47.1%	1.7%
Black	13.7%	69.8%	43.5%	3.8%
OtherEth	15.7%	53.9%	30.9%	1.7%
White	29.2%	71.2%	34.2%	3.3%

Table 2: Prevalence of comorbidities in different ethnic groups.

N Comorb. / Ethnicity	0	1	2	>=3
Asian	25%	26%	38%	12%
Black	25%	27%	39%	9%
OtherEth	40%	24%	30%	6%
White	20%	35%	30%	15%

Table 3: Number of comorbidities in different ethnic groups.

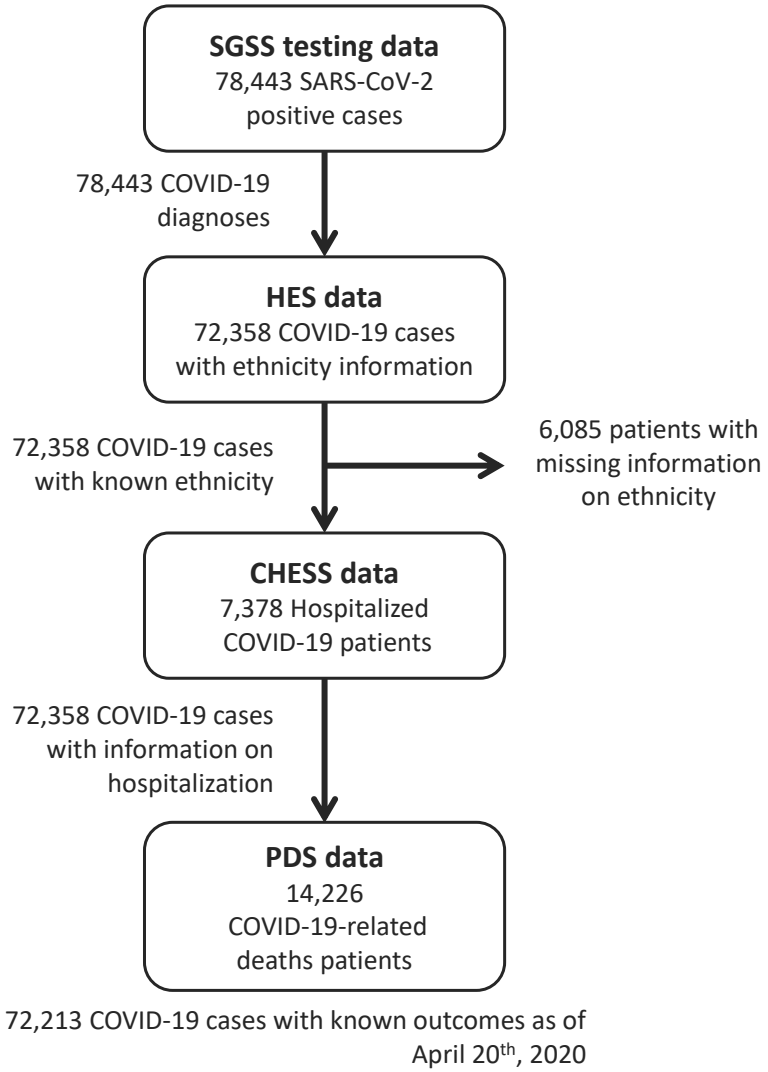

Figure 1: Data linkage and patient inclusion.

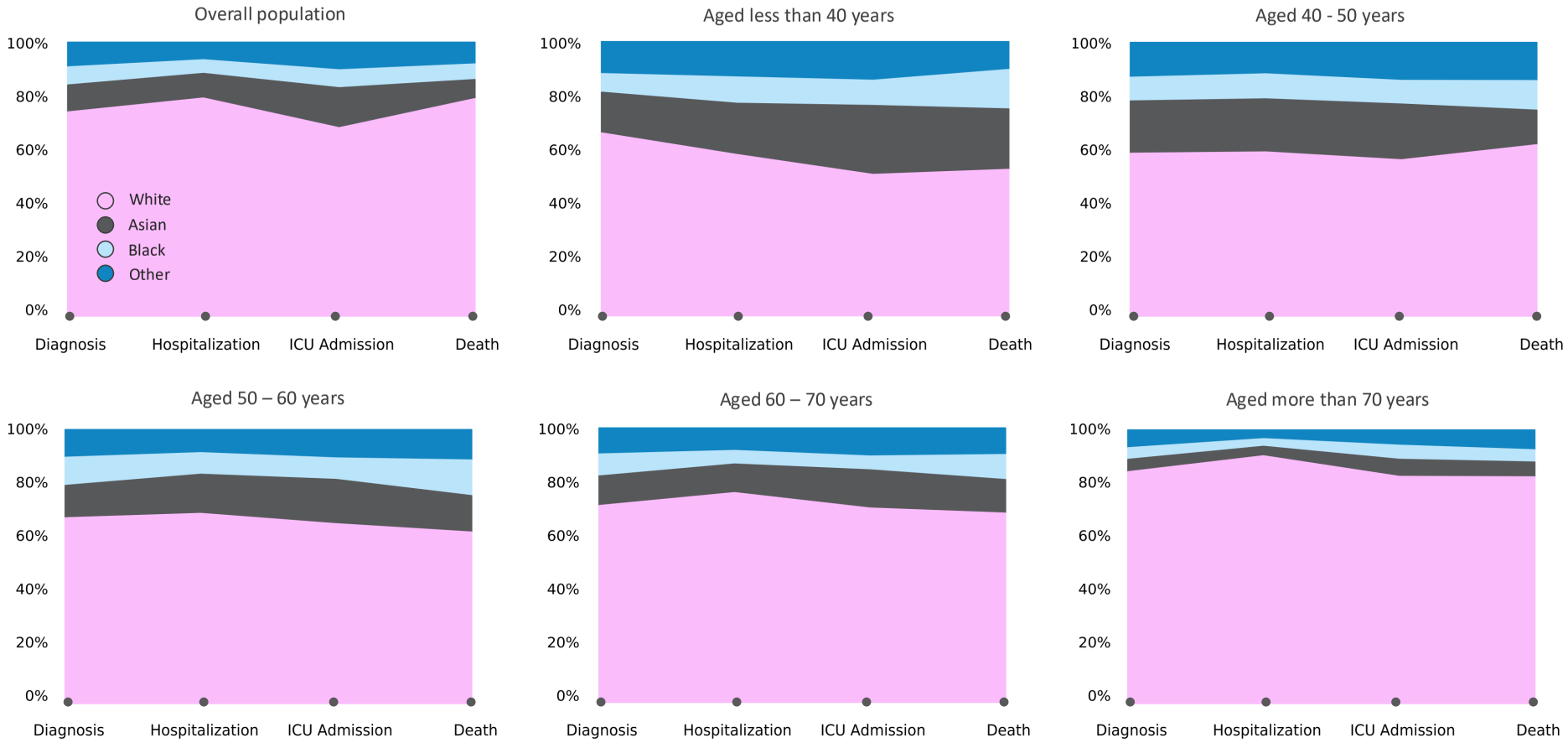

Figure 2: Representation of different ethnic groups for various age ranges across different stages of the COVID-19 trajectory.

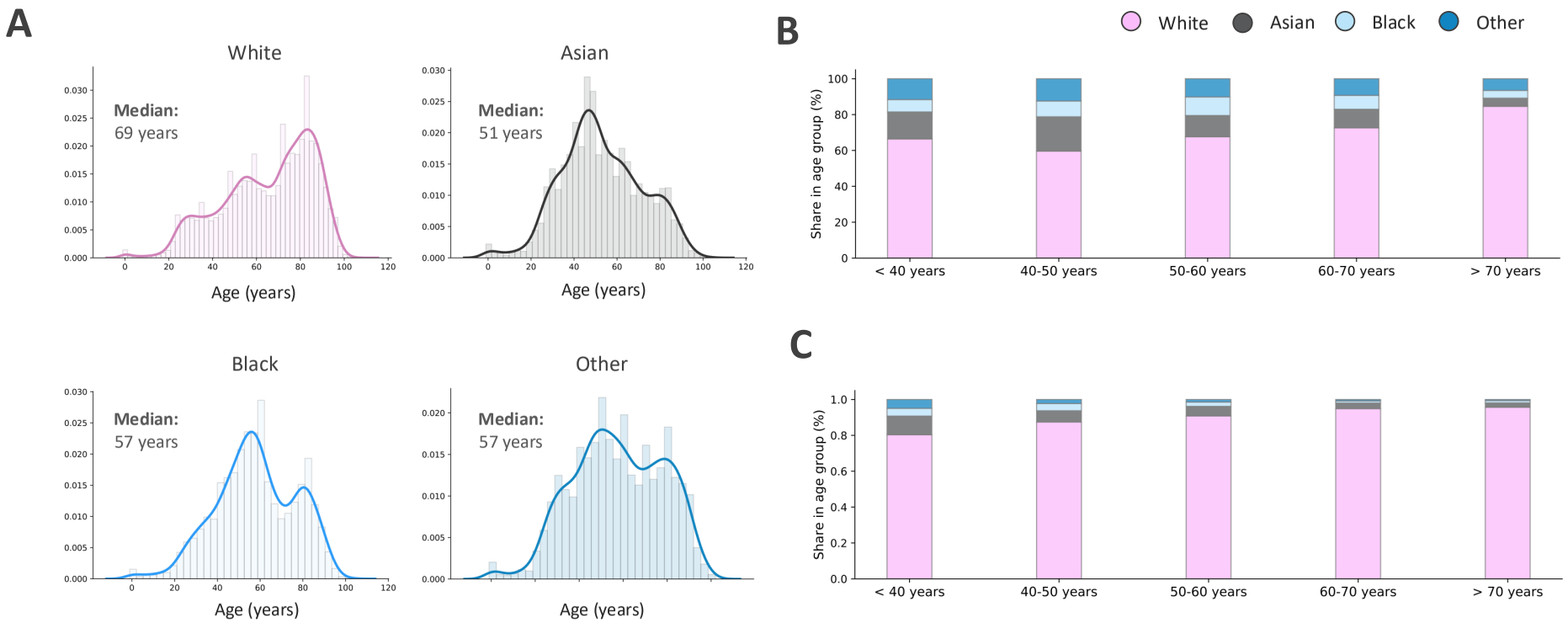

Figure 3: **(A)** Age distribution for COVID-19 patients within the different ethnic groups. **(B)** Representation of the different ethnic groups across various age ranges within the COVID-19 population. **(C)** Representation of the different ethnic groups across various age ranges in the population of England based on the 2011 census.

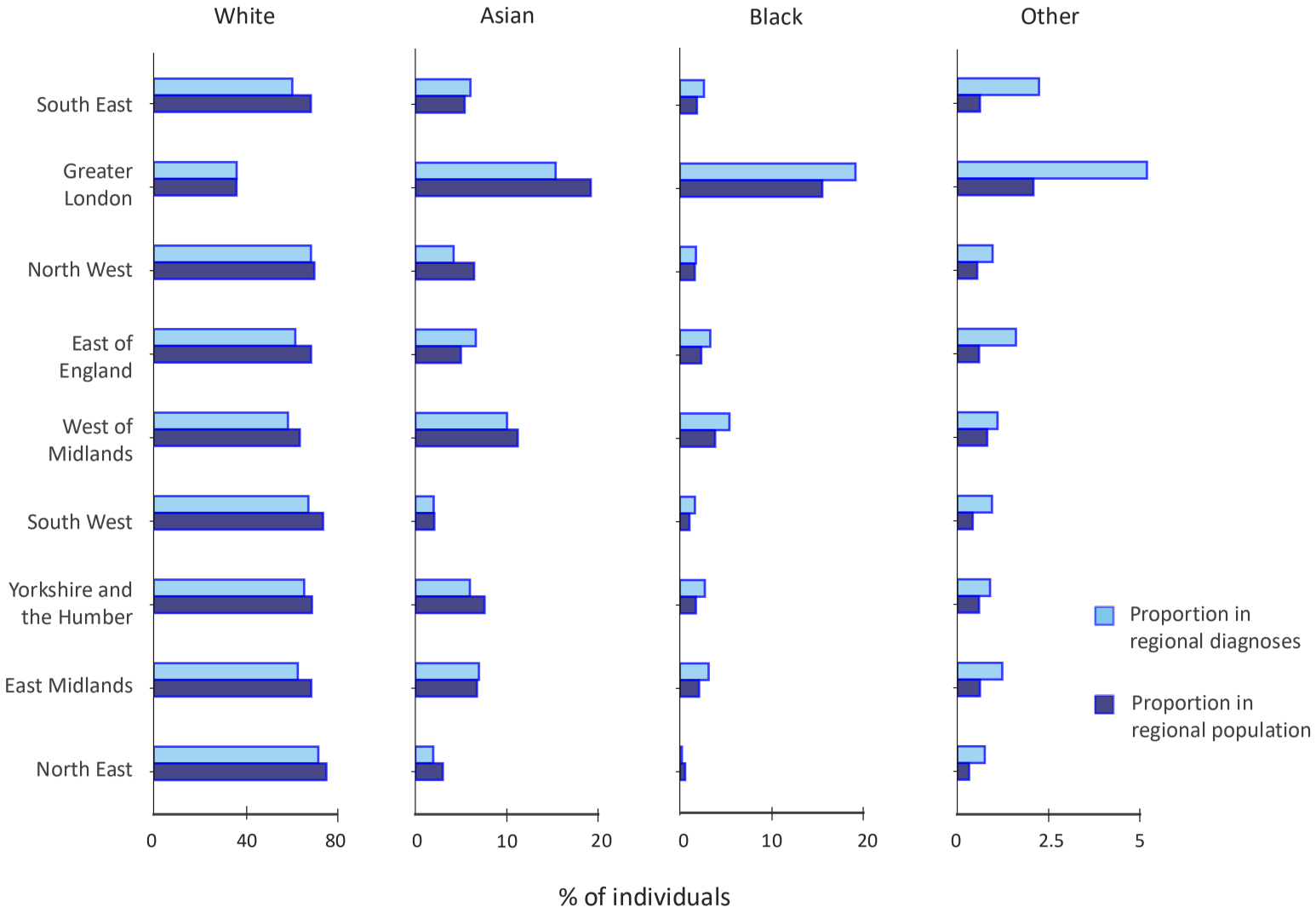

Figure 4: Regional proportion of COVID-19 diagnoses within the different ethnic groups compared to the regional demographics based on the 2011 census for England.

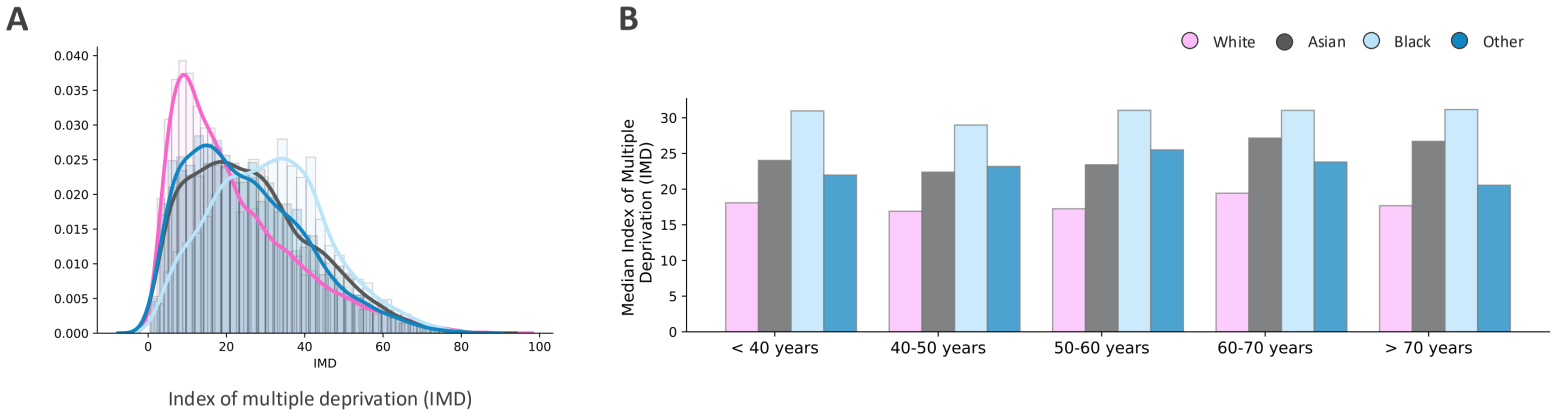

Figure 5: **(A)** Histogram for the index of multiple deprivation (IMD) among patients in various ethnic groups. **(B)** Median IMD in various age ranges within the different ethnic groups.

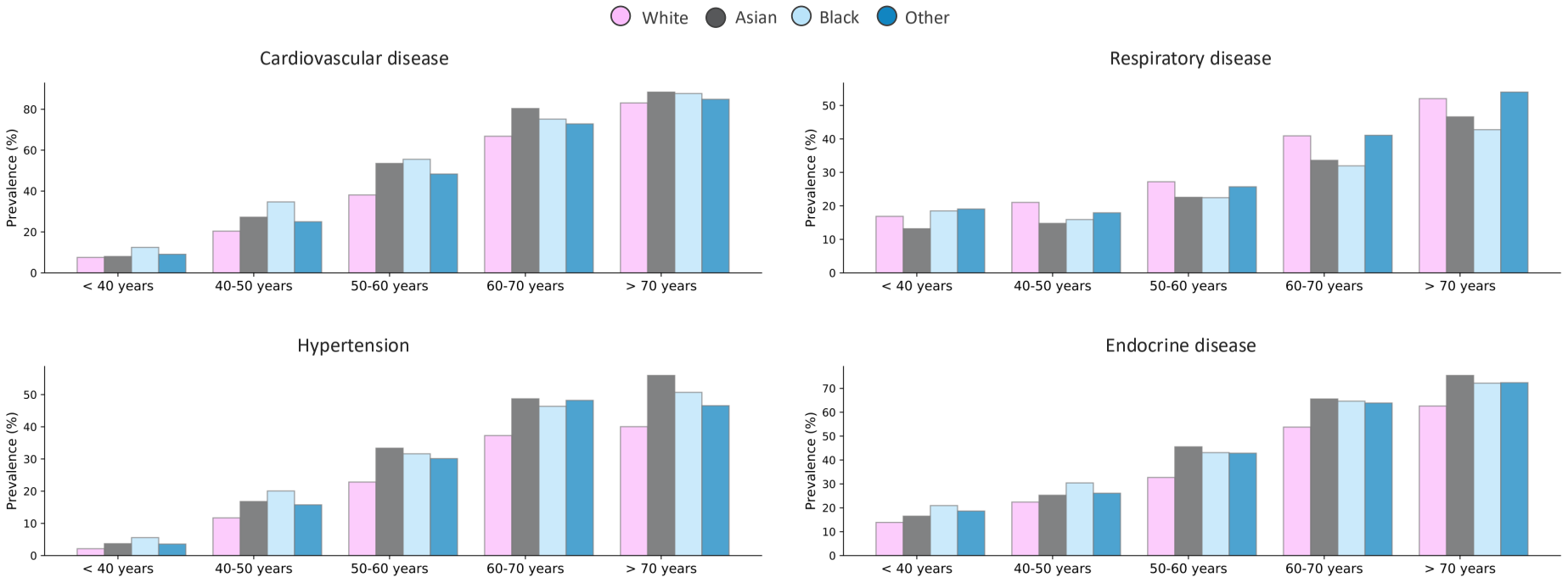

Figure 6: Prevalence of various comorbidities among different ethnic and age groups.

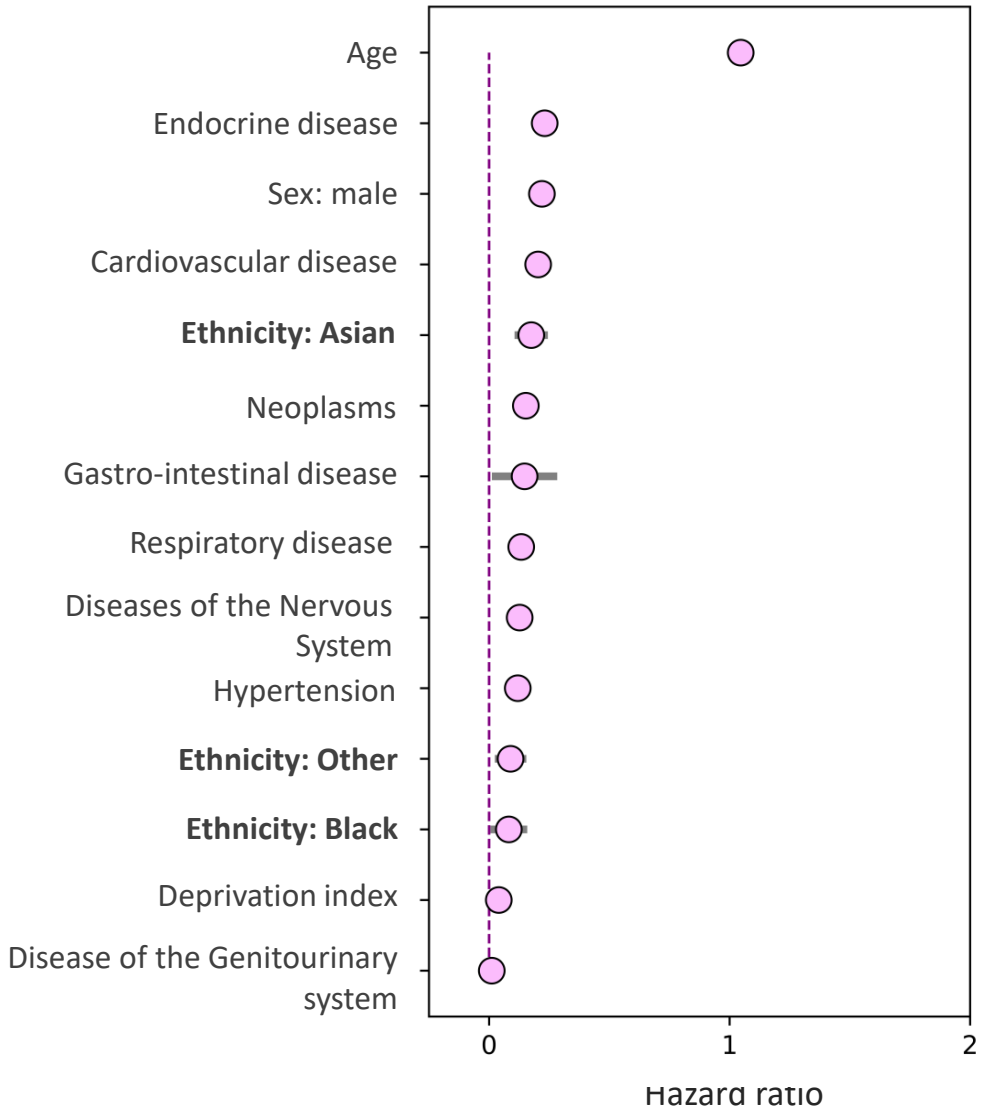

Figure 7: Coefficients of the Cox mixed-effects model for mortality.

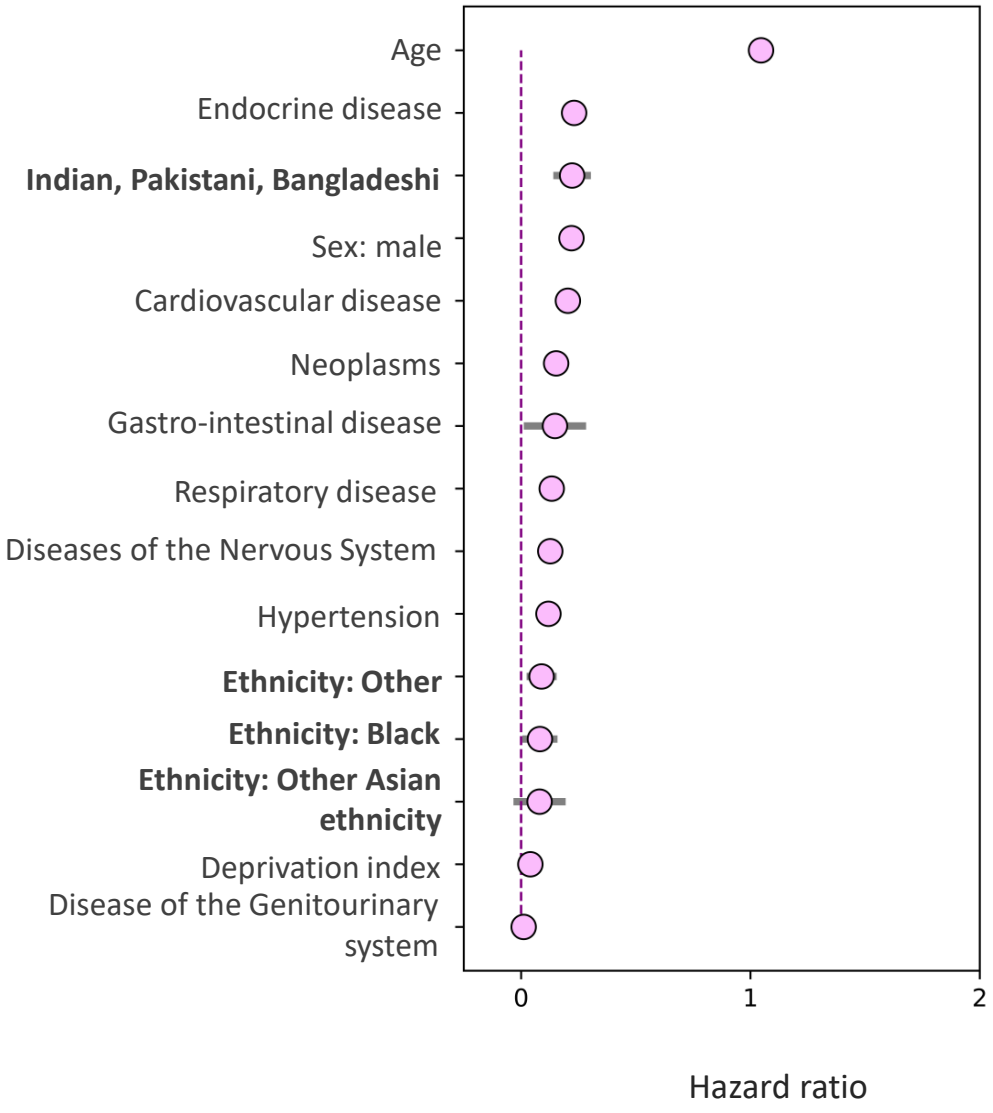

Figure 8: Coefficients of the Cox mixed-effects model for mortality with separate modeling of the Indian subcontinent ethnicity.

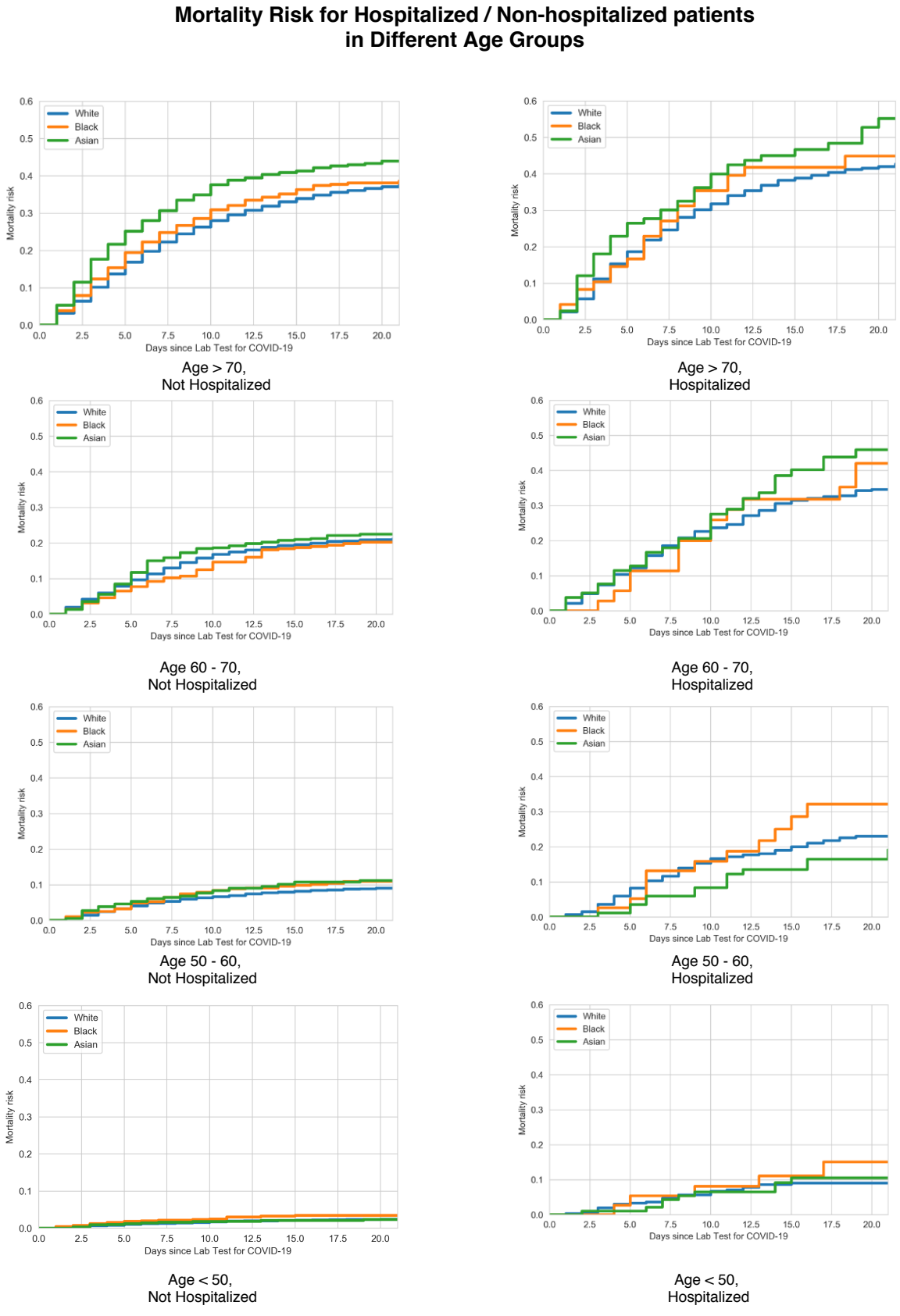

Figure 9: Death probabilities over time.