

Association Between Race and COVID-19 Outcomes During the Pre-Vaccination Period in the United States (2020–2021)

Sruthi Adimadhyam¹, Rebecca Hawrusik², Hye Seung Lee³, Caroline J Jjingo⁴, Jane A Gwira⁵, Maria E Kempner², Megan Wiley², Andrew B Petrone², Yueqin Zhao³, Danijela Stojanovic⁶, Efe Eworuke⁷, Adebola Ajao⁶

¹Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA; ²Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA, USA; ³Division of Biometrics VII, Office of Biostatistics, Office of Translational Science, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA; ⁴Division of Anti-Infectives, Office of Infectious Diseases, Office of New Drugs, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA; ⁵Office of Biostatistics and Pharmacovigilance, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA; ⁶Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA; ⁷Epidemiology and Drug Safety, Real World Solutions, IQVIA Inc., Durham, NC, USA

Correspondence: Sruthi Adimadhyam, Email Sruthi_Adimadhyam@populationmedicine.org

Objective: To determine the association between race, a proxy for social determinants of health, and COVID-19 outcomes adjusted for demographic, clinical, and socioeconomic differences.

Methods: We conducted a retrospective cohort study using administrative claims data in the Sentinel Distributed Database. We identified 841,628 individuals diagnosed with COVID-19; and separately, 133,773 individuals hospitalized with COVID-19 between April 1, 2020, and March 31, 2021, in the US. Eligible individuals required at least 6 months of enrollment in a health plan prior to cohort entry. Crude and adjusted associations between self-reported race (Asian, Black or African American, American Indian or Alaska Native [AIAN], Native Hawaiian or Other Pacific Islander [NHOPI], White, or Unknown) and COVID-19 outcomes (hospitalization with COVID-19; critical COVID; 30-day all-cause inpatient mortality) were determined using multivariable logistic regression.

Results: Of the 841,628 individuals with COVID-19, 45.5% were White, 42.1% were of Unknown race, 9.5% Black or African American, 2% Asian, 0.5% NHOPI, and 0.3% AIAN. All subpopulations had increased odds of hospitalization compared to White population (AIAN, adjusted Odds Ratio 1.45 [95% confidence interval 1.11–1.90]; Asian population 1.58 [1.38–1.80]; Black or African American population 1.58 [1.40–1.78]; NHOPI 1.32 [1.22–1.42]). There were 133,773 individuals hospitalized with COVID-19, 56.2% of whom were White, 23.4% of Unknown race, 17.6% Black or African American, 1.5% Asian, 0.9% NHOPI, and 0.4% AIAN. Over half of all hospitalized individuals progressed to critical COVID (58%) and 14.2% died within 30 days. Progression to critical COVID was significantly higher in NHOPI population compared to White population (1.17 [1.03–1.32]). AIAN population (1.52 [1.20–1.91]), Asian population (1.30 [1.13–1.49]), and NHOPI population (1.24 [1.06–1.45]) had significantly higher odds of 30-day inpatient mortality compared to White population.

Conclusion: We identified significant differences in COVID-19 outcomes in different subgroups within a diverse US population.

Keywords: COVID-19, real-world data, multivariable analysis, health outcomes

Introduction

SARS-CoV-2, the causative agent in Coronavirus 19 disease (COVID-19) was first detected in the United States (US) on January 20, 2020.¹ On May 11, 2023, the US federal government declared an end to the COVID-19 Public Health Emergency (PHE).² As of January 2024, there have been over 6.7 million hospitalizations and over 1.1 million deaths from COVID-19 in the US.³

Race is a social construct often used in epidemiologic and health outcomes research across a range of diseases (eg, cardiovascular, metabolic, and infectious diseases such as HIV) as a proxy for social determinants of health, to capture cultural, systemic, and structural factors, including socioeconomic status and access to healthcare, across various subpopulations in the US.^{4–7}

Prior literature estimated the burden of COVID-19 in different subgroups of the US population during the pre-vaccination era.^{8–12} These studies reported a higher burden of preexisting comorbidities such as asthma, chronic obstructive pulmonary disease (COPD), diabetes, hypertension, congestive heart failure (CHF), and chronic kidney disease (CKD),^{10,12,13} at the time of COVID-19 testing,^{11–13} and higher likelihood of progression to COVID-19-related hospitalization^{8–10,14} among Black or African American and/or Hispanic/Latino populations compared to White populations. However, while in-hospital, Black or African American populations were no more likely than White populations to progress to severe COVID-19 (Intensive Care Unit [ICU] care, use of mechanical ventilation or renal replacement therapy) or in hospital death.^{8,12–14} The validity of findings from these preliminary studies may be limited by small sample size, restriction to specific time period or geography, use of broad racial and/or ethnic categories, and lack of control for most socioeconomic factors.

The objective of our study was to assess the association between self-reported race as a proxy for social determinants of health and COVID-19 outcomes including hospitalization, critical COVID-19, and mortality, after controlling for baseline differences in demographic, clinical, and socioeconomic factors at a population-level during the first full year of the COVID-19 pandemic prior to widespread availability of COVID-19 vaccines. We used a large, nationally representative database for this study. The significant size of the database enabled us to retain granular race categories as captured in the database, allowing more accurate characterization of subpopulation-specific differences in the US. Preserving granularity in race categories prevents obscuring variation in disease outcomes and avoids underestimating the burden of disease experienced by various subpopulations that can be a consequence of averaging across heterogeneous groups. Reporting findings by granular categories can help better direct essential resources.

Materials and Methods

We conducted an observational retrospective cohort study using administrative claims data from national and regional insurers in the US to determine the association between race and COVID-19 outcomes from 2020 to 2021. This activity was conducted under the authority of the US Food and Drug Administration (FDA) as part of its Sentinel Initiative, a national system for medical product safety surveillance congressionally mandated by the FDA Act of 2007 Section 905. Consistent with the Health Insurance Portability and Accountability Act of 1996 (HIPAA) (45 C.F.R. §164.512(b)(1)(i)) and the Amended Common Rule, Sentinel activities are deemed to be public health surveillance by a relevant public health authority (here, FDA). Consequently, these activities are permitted under HIPAA without oversight by an Institutional Review Board or consent procedures. Data partners that participate in the Sentinel Initiative rely on FDA's determination that Sentinel activities are public health surveillance and grant permission to access their data to Harvard Pilgrim Health Care Institute, which administers Sentinel activities acting under FDA authorization.^{15–17}

Data Source

We used longitudinal administrative claims data from three insurers, including 2 large national health plans and 1 regional integrated delivery network, that provide medical and drug coverage for their commercially insured beneficiaries in the US. These 3 insurers participate in the FDA's Sentinel Initiative^{18,19} and contribute the most up-to-date claims data as part of the Rapid COVID-19 Sentinel Distributed Database.²⁰ This database includes deidentified data routinely collected as part of administering healthcare to insured beneficiaries including details on diagnoses recorded, procedures performed in the inpatient and outpatient care settings, and medications dispensed in the outpatient setting, billed to insurers. No primary data collection was conducted for this study. The Rapid COVID-19 Sentinel Distributed Database is a distributed network where each participating site retains physical and operational control of their own data that reside behind an institutional firewall to preserve patient privacy. Each site transforms their data into the Sentinel Common Data Model.²¹ These transformed data go through a rigorous quality assurance process prior to being made available for

querying.²² Utilizing a Common Data Model format enables the distribution of data to all participating sites via one common analytic package.²³

Study Cohorts

We created two study cohorts based on the following cohort entry (index) defining events: (1) COVID-19 diagnosis; and (2) hospitalization with COVID-19 occurring between April 1, 2020, and March 31, 2021. Our COVID-19 diagnosed cohort included any eligible individual with evidence of an International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) diagnosis code of U07.1 in any position on the discharge summary, in any care setting, or a positive result for a SARS-COV-2 PCR test. Index date was the date of first qualifying diagnosis or positive test result during the study period. The hospitalized with COVID-19 cohort included any eligible individual with evidence of an inpatient encounter with an ICD-10-CM diagnosis code of U07.1 documented in their discharge summary. Index date was the date of first qualifying hospitalization during the study period. In both cohorts, eligible individuals had to be continuously enrolled in a health plan with medical and drug coverage for at least 183 days prior to the index date so that their baseline health can be characterized. We permitted gaps of up to 45 days when evaluating continuous enrollment as these are typically considered administrative gaps and not lapses in insurance coverage.

Main Independent Variable

Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, or Unknown race) was the main independent variable in the analyses. Race is self-reported by beneficiaries through closed-ended questionnaires administered at the time of enrollment in a health plan. The Sentinel Common Data Model accommodates one value of race per individual. Self-reported ethnicity was not adequately captured (58% missing²⁴) in the Sentinel Distributed Database and therefore not analyzed.

Study Outcomes

Individuals with COVID-19 were evaluated for the occurrence of COVID-19 related hospitalization within 30 days of the index date. COVID-19 related hospitalization was defined as an inpatient encounter with ICD-10-CM diagnosis code of U07.1 on the discharge summary. In the cohort of individuals hospitalized with COVID-19, we evaluated the occurrence of 2 outcomes within 30 days of index date: (1) critical COVID-19; and separately, (2) inpatient mortality. Critical COVID-19 was defined as an inpatient encounter with ICD-10-CM diagnosis code of U07.1 with evidence of admission to the ICU or requiring mechanical ventilation, extracorporeal membrane oxygenation, or renal replacement therapy with an accompanying diagnosis of acute renal failure.²⁵ Inpatient mortality was defined as documentation of a patient discharge disposition of expired during an inpatient encounter. Long-term outcomes were not considered because of the lack of knowledge at the time of study conduct as to what these might be.

Demographic, Clinical, and Socioeconomic Characteristics

We assessed age, sex, census bureau region, urbanicity, and proxies for socioeconomic status namely median household income, median property value and percent unemployment on the index date. Individual-level data on income, housing, and employment are often lacking in administrative claims data. Therefore, proxies for socioeconomic status and urbanicity were measured at the ZIP code level of an individual's most recent primary residence available on the index date and was aggregated across all individuals in each racial population. Socioeconomic markers for each ZIP code tabulation area in the US were gathered from the 2019 American Community Survey (ACS) and mapped to the level of ZIP codes.^{26–30}

We described various clinical characteristics at baseline (defined as 183 days prior to index date) as evidence of at least one healthcare encounter with diagnosis, procedure, or treatments pertaining to the characteristics. These characteristics were deemed relevant to the prognosis and risk of severe illness due to COVID-19 by the US Centers for Disease Control and Prevention³¹ and included metabolic disease and related chronic comorbidities (diabetes, hypertension, obesity, cardiovascular disease, etc.), immunosuppressive conditions (cancer, cystic fibrosis, etc.), use of immunosuppressive medications, and respiratory conditions (asthma, COPD) (Table 1 and [Additional File 1](#)). Additionally, we

Table I Characteristics of Individuals with COVID-19* Between April 1, 2020 to March 31, 2021 in the Sentinel Distributed Database, by Race

	Overall (n = 841,628)	Asian (n = 16,783)	Black or AA** (n = 80,325)	AIAN† (n = 2,853)	NHOPI‡ (n = 4,353)	White (n = 382,674)	Unknown (n = 354,640)
Demographic Characteristics							
Age years							
Mean (SD)	55.2 (19.7)	48.5 (15.4)	63.9 (14.2)	56.2 (15.3)	66.6 (10.2)	63.6 (15.8)	44.4 (21)
0–5 years	8,231 (1%)	137 (0.8%)	85 (0.1%)	7 (0.2%)	5 (0.1%)	971 (0.3%)	7,026 (2%)
6–11 years	12,438 (1.5%)	144 (0.9%)	146 (0.2%)	12 (0.4%)	5 (0.1%)	1,561 (0.4%)	10,570 (3%)
12–18 years	31,580 (3.8%)	351 (2.1%)	448 (0.6%)	59 (2.1%)	20 (0.5%)	5,477 (1.4%)	25,225 (7.1%)
19–29 years	89,419 (10.6%)	2,512 (15%)	3,021 (3.8%)	257 (9%)	151 (3.5%)	22,846 (6%)	60,632 (17.1%)
30–39 years	86,781 (10.3%)	3,428 (20.4%)	5,234 (6.5%)	301 (10.6%)	175 (4%)	25,617 (6.7%)	52,026 (14.7%)
40–49 years	90,333 (10.7%)	2,742 (16.3%)	6,181 (7.7%)	389 (13.6%)	259 (5.9%)	27,779 (7.3%)	52,983 (14.9%)
50–64 years	168,258 (20%)	3,046 (18.1%)	17,084 (21.3%)	715 (25.1%)	580 (13.3%)	68,358 (17.9%)	78,475 (22.1%)
65–74 years	193,277 (23%)	2,939 (17.5%)	29,167 (36.3%)	660 (23.1%)	1,935 (44.5%)	117,733 (30.8%)	40,843 (11.5%)
≥75 years	161,311 (19.2%)	1,484 (8.8%)	18,959 (23.6%)	453 (15.9%)	1,223 (28.1%)	112,332 (29.4%)	26,860 (7.6%)
Female	454,938 (54.1%)	8,559 (51%)	48,828 (60.8%)	1,640 (57.5%)	2,053 (47.2%)	207,119 (54.1%)	186,739 (52.7%)
Hispanic origin							
Yes	52,568 (6.2%)	1,559 (9.3%)	954 (1.2%)	736 (25.8%)	165 (3.8%)	5,247 (1.4%)	43,907 (12.4%)
No	283,161 (33.6%)	4,107 (24.5%)	52,857 (65.8%)	1,341 (47%)	3,622 (83.2%)	220,287 (57.6%)	947 (0.3%)
Unknown	505,899 (60.1%)	11,117 (66.2%)	26,514 (33%)	776 (27.2%)	566 (13%)	157,140 (41.1%)	309,786 (87.4%)
Clinical Characteristics							
Charlson/Elixhauser combined comorbidity score ¹	1.2 (2.3)	0.5 (1.6)	2.1 (3)	1.2 (2.2)	1.7 (2.7)	1.6 (2.6)	0.6 (1.7)
Alcohol or drug abuse	20,122 (2.4%)	119 (0.7%)	3,057 (3.8%)	107 (3.8%)	118 (2.7%)	11,240 (2.9%)	5,481 (1.5%)
Any cancer	51,856 (6.2%)	540 (3.2%)	7,689 (9.6%)	157 (5.5%)	330 (7.6%)	30,996 (8.1%)	12,144 (3.4%)
Asthma	50,991 (6.1%)	745 (4.4%)	7,834 (9.8%)	250 (8.8%)	259 (5.9%)	24,597 (6.4%)	17,306 (4.9%)
Autoimmune conditions	22,663 (2.7%)	220 (1.3%)	4,089 (5.1%)	103 (3.6%)	156 (3.6%)	12,417 (3.2%)	5,678 (1.6%)
Chronic kidney disease	111,502 (13.2%)	1,315 (7.8%)	20,984 (26.1%)	390 (13.7%)	985 (22.6%)	65,784 (17.2%)	22,044 (6.2%)
Congestive heart failure	65,171 (7.7%)	457 (2.7%)	12,851 (16%)	181 (6.3%)	449 (10.3%)	39,974 (10.4%)	11,259 (3.2%)
Chronic obstructive pulmonary disease	78,865 (9.4%)	525 (3.1%)	11,528 (14.4%)	266 (9.3%)	434 (10%)	51,808 (13.5%)	14,304 (4%)
Coronary artery disease	126,840 (15.1%)	1,304 (7.8%)	18,660 (23.2%)	395 (13.8%)	1,021 (23.5%)	80,717 (21.1%)	24,743 (7%)
Cystic fibrosis	118 (0%)	2 (0%)	7 (0%)	0 (0%)	0 (0%)	66 (0%)	43 (0%)
Diabetes mellitus	183,790 (21.8%)	3,141 (18.7%)	33,488 (41.7%)	831 (29.1%)	1,785 (41%)	97,600 (25.5%)	46,945 (13.2%)
HIV	2,605 (0.3%)	38 (0.2%)	840 (1%)	18 (0.6%)	11 (0.3%)	797 (2.0%)	901 (0.3%)
Hypertension	356,107 (42.3%)	4,790 (28.5%)	56,513 (70.4%)	1,238 (43.4%)	2,654 (61%)	201,929 (52.8%)	88,983 (25.1%)
Interstitial lung disease	9,233 (1.1%)	101 (0.6%)	1,267 (1.6%)	29 (1%)	76 (1.7%)	5915 (1.5%)	1,845 (0.5%)
Liver disease	33,090 (3.9%)	601 (3.6%)	4,016 (5%)	151 (5.3%)	288 (6.6%)	17,931 (4.7%)	10,103 (2.8%)
Neurologic conditions	90,725 (10.8%)	661 (3.9%)	12,706 (15.8%)	299 (10.5%)	540 (12.4%)	57,988 (15.2%)	18,531 (5.2%)
Obesity	159,821 (19%)	1,258 (7.5%)	27,103 (33.7%)	574 (20.1%)	883 (20.3%)	80,661 (21.1%)	49,342 (13.9%)
Pulmonary conditions	45,451 (5.4%)	502 (3%)	7,381 (9.2%)	186 (6.5%)	321 (7.4%)	27,082 (7.1%)	9,979 (2.8%)

Sickle cell disease	870 (0.1%)	5 (0%)	564 (0.7%)	2 (0.1%)	4 (0.1%)	45 (0%)	250 (0.1%)
Smoking	97,027 (11.5%)	690 (4.1%)	14,431 (18%)	381 (13.4%)	602 (13.8%)	59,776 (15.6%)	21,147 (6%)
Solid organ or stem cell transplant	644 (0.1%)	16 (0.1%)	125 (0.2%)	0 (0%)	6 (0.1%)	295 (0.1%)	202 (0.1%)
Vascular disease	395,995 (47.1%)	5,283 (31.5%)	59,126 (73.6%)	1,346 (47.2%)	2,878 (66.1%)	225,506 (58.9%)	101,856 (28.7%)
Use of any chemotherapy	12,323 (1.5%)	154 (0.9%)	1,917 (2.4%)	47 (1.6%)	85 (2%)	6,982 (1.8%)	3,138 (0.9%)
Use of any non-systemic glucocorticoids	98,245 (11.7%)	1,716 (10.2%)	12,301 (15.3%)	348 (12.2%)	611 (14%)	52,679 (13.8%)	30,590 (8.6%)
Use of systemic glucocorticoids	192,542 (22.9%)	2,318 (13.8%)	23,140 (28.8%)	724 (25.4%)	972 (22.3%)	104,483 (27.3%)	60,905 (17.2%)
Use of immune modulators	23,000 (2.7%)	325 (1.9%)	3,083 (3.8%)	74 (2.6%)	157 (3.6%)	13,033 (3.4%)	6,328 (1.8%)
Use of immunosuppressants	13,385 (1.6%)	173 (1%)	2,177 (2.7%)	64 (2.2%)	98 (2.3%)	7,176 (1.9%)	3,697 (1%)
Recent use of angiotensin-converting enzyme inhibitors	52,106 (6.2%)	647 (3.9%)	6,324 (7.9%)	236 (8.3%)	364 (8.4%)	29,060 (7.6%)	15,475 (4.4%)
Recent use of angiotensin receptor blockers	47,339 (5.6%)	974 (5.8%)	7,657 (9.5%)	187 (6.6%)	350 (8%)	24,481 (6.4%)	13,690 (3.9%)
Recent use of vasopressors	2,343 (0.3%)	37 (0.2%)	426 (0.5%)	3 (0.1%)	24 (0.6%)	1,307 (0.3%)	546 (0.2%)
Socioeconomic Characteristics							
Average median household income ²	680,00.1 (26,556.9)	83,691.3 (31,133.4)	53,519.9 (20,960.8)	61,081.6 (22,506.9)	67,289.3 (25,098.5)	67,808.3 (24,728.9)	71,070.5 (28,488.2)
Average median home value ²	257,527.4 (197,349.8)	412,125.3 (275,422.3)	192,041 (134,825.9)	222,216 (143,558.1)	269,939.9 (183,190)	245,246.7 (180,485.8)	280,235.3 (219,320.3)
Average percent unemployment ²	5.3 (2.7)	4.8 (1.9)	7.4 (3.9)	6.1 (3.6)	5.5 (2.6)	5 (2.5)	5.1 (2.5)
All socioeconomic indicators available	770,924 (91.6%)	14,519 (86.5%)	76,638 (95.4%)	2,740 (96%)	4,163 (95.6%)	356,084 (93.1%)	316,780 (89.3%)

Notes: * PCR positive for SARS COV-2 or U07.1 diagnosis code in any setting. ¹J Clin Epidemiol. 2011;64(7):749–759. ²These variables were calculated only among those individuals with all socioeconomic indicators available. **African American. †American Indian or Alaska Native. ‡Native Hawaiian or Other Pacific Islander.

described the use of select cardiovascular therapies 30 days prior to index date in line with hypotheses at the time of study conduct regarding increased COVID-19 risks for those taking these medications.³² We described the presence of COVID-19 related symptoms the day prior through 14 days after index date.

Statistical Analysis

The association between race and COVID-19 outcomes was determined at each Data Partner using unadjusted and multivariable adjusted logistic regression. We ran two pre-specified multivariable adjusted models for each cohort. The first model adjusted for demographic and clinical factors associated with illness severity (age, sex, combined comorbidity score,^{33,34} alcohol or drug abuse, asthma, autoimmune conditions, cancer, CKD, COPD, CHF, coronary artery disease (CAD), cystic fibrosis, diabetes mellitus, HIV/AIDS, hypertension, interstitial lung disease (ILD), liver disease, neurological conditions, obesity, pulmonary conditions, sickle cell disease, smoking, solid organ or stem cell transplant, vascular disease, chemotherapy, immunosuppressants, immune modulators, non-systemic glucocorticoids, systemic glucocorticoids, angiotensin converting enzyme inhibitors, angiotensin receptor blockers, and vasopressors). The second model additionally adjusted for socioeconomic factors (median household income, median property value, and percent unemployment in ZIP code of primary residence) and was run in a subset of the cohort with non-missing values for all proxies of socioeconomic status. We used the Firth option, a simplistic solution in the logistic regression procedure in anticipation of sparse data to mitigate bias caused by rare events in a data set.³⁵

Execution of Distributed Analysis

Each site shared aggregated summary data and effect estimates generated from a common analytic package with the coordinating center. Results from each site were reviewed prior to aggregation in a random-effects meta-analysis.³⁶ All analyses were conducted using the Sentinel Routine Querying System³⁷ version 11.3.0 with additional custom programming using SAS 9.4 (SAS Institute, Cary, NC). The analytic package (ie, set of SAS programs) used to query the Sentinel Distributed Database is publicly posted on the Sentinel website for transparency and replicability.³⁸ These programs may be independently executed by the other researchers to replicate findings in similar databases provided the data are formatted in the Sentinel Common Data Model.

Results

Of the 985,837 individuals with a COVID-19 diagnosis claim in the database during the study period, 841,628 (85.4%) met the continuous enrollment eligibility requirement for cohort entry. Similarly, 133,773 (86.8%) of the 154,138 individuals hospitalized for COVID-19 met the continuous enrollment eligibility requirement.

Characteristics of Individuals Diagnosed with COVID-19

A total of 841,628 individuals were diagnosed with COVID-19 between April 1, 2020, and March 31, 2021. Of these, 45.5% were White, 42.1% Unknown, 9.5% Black or African American, 2% Asian, 0.5% Native Hawaiian or Other Pacific Islander, and 0.3% American Indian or Alaska Native. The mean age overall was 55.2 (± 19.7) years, with Asian population being the youngest (48.5 ± 15.4 years) and Native Hawaiian or Other Pacific Islander populations the oldest (66.6 ± 10.2 years). Prevalence of clinical comorbidities varied by race (Table 1). Native Hawaiian or Other Pacific Islander populations had the highest prevalence of CAD, liver disease, and ILD at baseline compared to all other populations. American Indian or Alaska Native populations and Black or African American populations had the highest prevalence of alcohol and drug abuse. Black or African American population had the highest prevalence of hypertension, diabetes, obesity, asthma, COPD, and CKD. All three proxies for socioeconomic status were available for 92% of this cohort. There were no numerical differences in baseline characteristics between the overall population and the subset of the population with known socioeconomic status (Additional File 2). Compared to other populations, Asian population predominantly lived in neighborhoods with the highest household incomes and property values, and the lowest unemployment, whereas Black or African American population predominantly lived in neighborhoods with the lowest household incomes and property values, and the highest unemployment.

Hospitalization with COVID-19 Following Diagnosis with COVID-19

Of the 841,628 individuals diagnosed with COVID-19, 129,802 (15.4%) had the study outcome of hospitalization with COVID-19 in the 30-day follow-up period. Individuals hospitalized with COVID-19 tended to be older (71.6 ± 13.1 vs 52.2 ± 19.2 years) and had a higher burden of comorbidities at baseline compared to those without the outcome at 30 days ([Additional File 3](#)).

Unadjusted analyses showed that Black or African American population diagnosed with COVID-19 had a higher odds of hospitalization with COVID-19 relative to White population (Odds Ratio [OR] 1.55 [95% Confidence Interval (95% CI) 1.27–1.90]). After controlling for baseline demographic and clinical risk factors, all subpopulations had higher odds of hospitalization with COVID-19 relative to White population ([Table 2](#)). The adjusted odds of hospitalization were highest for Black or African American population (OR 1.70 (95% CI: 1.46, 1.97)), followed by American Indian or Alaska Native population, Asian population, and Native Hawaiian or Other Pacific Islander populations. The increased risk for hospitalization was partially explained by baseline differences in socioeconomic factors with the exception of Asian population. Among Asian population, additionally adjusting for socioeconomic factors further increased the risk of hospitalization relative to White population.

Characteristics of Individuals Hospitalized with COVID-19

There were 133,773 individuals hospitalized with COVID-19 between April 1, 2020, and March 31, 2021. Of these, 56.2% were White, 23.4% Unknown, 17.6% Black or African American, 1.5% Asian, 0.9% Native Hawaiian or Other Pacific Islander, and 0.4% American Indian or Alaska Native. The mean age overall was 71.6 (± 13.1) years with Asian population being the youngest (67.5 ± 12.9 years) and White population the oldest (74.5 ± 11.1 years). Among hospitalized individuals, White population had the highest prevalence of COPD, CAD, neurological conditions, and ILD when compared to other racial populations. American Indian or Alaska Native population had the highest prevalence of alcohol or drug abuse, liver disease, autoimmune, and pulmonary conditions, whereas Black or African American population tended to have the highest prevalence of all other baseline chronic comorbidities. Similar to the cohort of individuals diagnosed with COVID-19, Asian population resided in neighborhoods with the highest median incomes and property values and the least unemployment, while Black or African American population resided in neighborhoods with the lowest median incomes and highest unemployment ([Table 3](#) and [Additional File 4](#)).

Critical COVID-19 Following Hospitalization with COVID-19

Of the 133,773 individuals hospitalized with COVID-19, 77,349 (58%) developed critical COVID within 30 days of admission. Individuals with critical COVID were slightly older than those without (72.1 ± 12.3 vs 71.0 ± 14.2 years) and had a higher prevalence of comorbidities at baseline ([Additional File 5](#)).

Crude analyses showed that Native Hawaiian or Other Pacific Islander population was the only population with an increased odds of critical COVID-19 relative to White population (OR 1.23 (95% CI: 1.09–1.39)) ([Table 4](#)). This increased risk persisted after controlling for demographic and clinical factors (OR 1.22 (1.04–1.42)) and socioeconomic factors (OR 1.17 (1.03–1.32)). After controlling for all risk factors, a small but non-significant increased risk of critical COVID-19 was also observed among Asian population (OR 1.13 (0.99–1.29)).

Mortality Following Hospitalization with COVID-19

Of the 133,773 individuals hospitalized with COVID-19, 19,002 (14.2%) had a discharge disposition of expired within 30 days of admission. The deceased were on average older than those discharged alive (76.6 ± 10 vs 70.8 ± 13.4 years) and had a higher prevalence of comorbidities at baseline ([Additional File 6](#)).

Crude analyses did not show any significant difference in 30-day inpatient mortality following hospitalization with COVID-19 in the overall population ([Table 4](#)). After adjusting for demographic and clinical risk factors, American Indian or Alaska Native population (OR 1.57 (1.24–1.97)), Asian population (OR 1.26 (1.10–1.44)), and Native Hawaiian or Other Pacific Islander populations (OR 1.22 (1.04–1.42)) had an increased risk of mortality compared to White population. After adjusting for socioeconomic differences, the increased burden of mortality in these populations

Table 2 Association Between Race and COVID-19 Hospitalization Among Those with COVID-19 Between April 1, 2020, and March 31, 2021, in the Sentinel Distributed Database

	Entire Study Cohort (n = 841,628)				Subset of the Population with Known Socioeconomic Status (n = 770,924)		
	Number of Patients with COVID-19	Number of Patients Hospitalized with COVID-19	Unadjusted Odds Ratio (95% CI)	Odds Ratio Adjusted for Demographic and Clinical Factors (95% CI)	Number of Patients with COVID-19	Number of Patients Hospitalized with COVID-19	Odds Ratio Adjusted for Demographic Clinical and Socioeconomic Factors (95% CI)
American Indian or Alaska Native	2,853	533	1.22 (0.80 1.86)	1.52 (1.20 1.93)	2,740	525	1.45 (1.11 1.90)
Asian	16,783	1,923	0.95 (0.60 1.53)	1.50 (1.30 1.72)	14,519	1,792	1.58 (1.38 1.80)
Black or African American	80,325	23,007	1.55 (1.27 1.90)	1.70 (1.46 1.97)	76,638	22,432	1.58 (1.40 1.78)
Native Hawaiian or Other Pacific Islander	4,353	1,154	1.03 (0.75 1.42)	1.50 (1.11 2.03)	4,163	1,122	1.32 (1.22 1.42)
Unknown	354,640	30,398	0.61 (0.25 1.47)	1.17 (0.84 1.62)	316,780	28,623	1.13 (0.85 1.51)
White	382,674	72,787	Referent group	Referent group	356,084	69,961	Referent group

Table 3 Characteristics of Individuals Hospitalized with COVID-19* Between April 1, 2020 and March 31, 2021 in the Sentinel Distributed Database, by Race

	Overall (n = 133,773)	Asian (n = 1,984)	Black or AA** (n = 23,583)	AIAN† (n = 550)	NHOPI‡ (n = 1,188)	White (n = 75,138)	Unknown (n = 31,330)
Demographic Characteristics							
Age years							
Mean (SD)	71.6 (13.1)	67.5 (12.9)	71.1 (11.6)	68.8 (11.6)	72.9 (8.9)	74.5 (11.1)	65.3 (16.7)
0–5 years	78 (0.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	8 (0%)	70 (0.2%)
6–11 years	43 (0%)	0 (0%)	1 (0%)	0 (0%)	0 (0%)	3 (0%)	39 (0.1%)
12–18 years	187 (0.1%)	0 (0%)	8 (0%)	0 (0%)	0 (0%)	24 (0%)	155 (0.5%)
19–29 years	1,268 (0.9%)	27 (1.4%)	94 (0.4%)	7 (1.3%)	3 (0.3%)	248 (0.3%)	889 (2.8%)
30–39 years	2,723 (2%)	118 (5.9%)	364 (1.5%)	10 (1.8%)	13 (1.1%)	612 (0.8%)	1,606 (5.1%)
40–49 years	5,002 (3.7%)	138 (7%)	752 (3.2%)	31 (5.6%)	25 (2.1%)	1,366 (1.8%)	2,690 (8.6%)
50–64 years	21,935 (16.4%)	325 (16.4%)	4,233 (17.9%)	115 (20.9%)	114 (9.6%)	9,238 (12.3%)	7,910 (25.2%)
65–74 years	44,797 (33.5%)	747 (37.7%)	9,495 (40.3%)	214 (38.9%)	517 (43.5%)	25,291 (33.7%)	8,533 (27.2%)
≥75 years	57,740 (43.2%)	629 (31.7%)	8,636 (36.6%)	173 (31.5%)	516 (43.4%)	38,348 (51%)	9,438 (30.1%)
Female	64,820 (48.5%)	850 (42.8%)	13,179 (55.9%)	293 (53.3%)	444 (37.4%)	35,774 (47.6%)	14,280 (45.6%)
Hispanic origin							
Yes	5,757 (4.3%)	86 (4.3%)	55 (0.2%)	54 (9.8%)	8 (0.7%)	263 (0.4%)	5,291 (16.9%)
No	78,836 (58.9%)	1,057 (53.3%)	18,561 (78.7%)	414 (75.3%)	1,121 (94.4%)	57,617 (76.7%)	66 (0.2%)
Unknown	49,180 (36.8%)	841 (42.4%)	4,967 (21.1%)	82 (14.9%)	59 (5%)	17,258 (23%)	25,973 (82.9%)
Clinical Characteristics							
Charlson/Elixhauser combined comorbidity score ¹	3.0 (3.4)	1.8 (2.8)	3.5 (3.5)	2.4 (3.1)	2.7 (3.3)	3.2 (3.4)	2.3 (3.1)
Alcohol or drug abuse	5,740 (4.3%)	20 (1%)	1,196 (5.1%)	31 (5.6%)	38 (3.2%)	3,351 (4.5%)	1,104 (3.5%)
Any cancer	16,462 (12.3%)	162 (8.2%)	3,166 (13.4%)	55 (10%)	117 (9.8%)	9,855 (13.1%)	3,107 (9.9%)
Asthma	11,819 (8.8%)	151 (7.6%)	2,650 (11.2%)	58 (10.5%)	85 (7.2%)	6,236 (8.3%)	2,639 (8.4%)
Autoimmune conditions	6,218 (4.6%)	42 (2.1%)	1,452 (6.2%)	36 (6.5%)	56 (4.7%)	3,507 (4.7%)	1,125 (3.6%)
Chronic kidney disease	44,858 (33.5%)	535 (27%)	10,032 (42.5%)	148 (26.9%)	431 (36.3%)	25,822 (34.4%)	7,890 (25.2%)
Congestive heart failure	29,504 (22.1%)	216 (10.9%)	6,459 (27.4%)	85 (15.5%)	214 (18%)	17,702 (23.6%)	4,828 (15.4%)
Chronic obstructive pulmonary disease	31,467 (23.5%)	201 (10.1%)	5,300 (22.5%)	105 (19.1%)	193 (16.2%)	20,649 (27.5%)	5,019 (16%)
Coronary artery disease	45,907 (34.3%)	487 (24.5%)	8,243 (35%)	141 (25.6%)	384 (32.3%)	28,719 (38.2%)	7,933 (25.3%)
Cystic fibrosis	25 (0%)	0 (0%)	1 (0%)	0 (0%)	0 (0%)	18 (0%)	6 (0%)
Diabetes mellitus	59,820 (44.7%)	934 (47.1%)	13,284 (56.3%)	287 (52.2%)	634 (53.4%)	32,545 (43.3%)	12,136 (38.7%)
HIV	525 (0.4%)	5 (0.3%)	248 (1.1%)	0 (0%)	2 (0.2%)	146 (0.2%)	124 (0.4%)
Hypertension	100,360 (75%)	1,282 (64.6%)	20,205 (85.7%)	372 (67.6%)	893 (75.2%)	57,923 (77.1%)	19,685 (62.8%)
Interstitial lung disease	4,061 (3%)	41 (2.1%)	656 (2.8%)	16 (2.9%)	35 (2.9%)	2,550 (3.4%)	763 (2.4%)
Liver disease	9,595 (7.2%)	120 (6%)	1,613 (6.8%)	52 (9.5%)	103 (8.7%)	5,558 (7.4%)	2,149 (6.9%)
Neurologic conditions	29,662 (22.2%)	221 (11.1%)	5,289 (22.4%)	96 (17.5%)	208 (17.5%)	18,712 (24.9%)	5,136 (16.4%)
Obesity	41,509 (31%)	285 (14.4%)	9,213 (39.1%)	142 (25.8%)	258 (21.7%)	22,741 (30.3%)	8,870 (28.3%)

(Continued)

Table 3 (Continued).

Pulmonary conditions	30,122 (22.5%)	362 (18.2%)	4,982 (21.1%)	135 (24.5%)	223 (18.8%)	17,213 (22.9%)	7,207 (23%)
Sickle cell disease	324 (0.2%)	2 (0.1%)	234 (1%)	1 (0.2%)	0 (0%)	14 (0%)	73 (0.2%)
Smoking	33,087 (24.7%)	218 (11%)	6,083 (25.8%)	130 (23.6%)	230 (19.4%)	20,931 (27.9%)	5,495 (17.5%)
Solid organ or stem cell transplant	258 (0.2%)	7 (0.4%)	63 (0.3%)	0 (0%)	4 (0.3%)	125 (0.2%)	59 (0.2%)
Vascular disease	108,364 (81%)	1,368 (69%)	20,983 (89%)	407 (74%)	953 (80.2%)	62,977 (83.8%)	21,676 (69.2%)
Use of any chemotherapy	4,333 (3.2%)	46 (2.3%)	818 (3.5%)	14 (2.5%)	37 (3.1%)	2,501 (3.3%)	917 (2.9%)
Use of any non-systemic glucocorticoids	25,166 (18.8%)	355 (17.9%)	4,282 (18.2%)	80 (14.5%)	198 (16.7%)	15,109 (20.1%)	5,142 (16.4%)
Use of systemic glucocorticoids	52,230 (39%)	582 (29.3%)	8,491 (36%)	208 (37.8%)	370 (31.1%)	31,083 (41.4%)	11,496 (36.7%)
Use of immune modulators	7,270 (5.4%)	104 (5.2%)	1,321 (5.6%)	31 (5.6%)	58 (4.9%)	4,107 (5.5%)	1,649 (5.3%)
Use of immunosuppressants	4,005 (3%)	49 (2.5%)	861 (3.7%)	23 (4.2%)	38 (3.2%)	2,194 (2.9%)	840 (2.7%)
Recent use of angiotensin-converting enzyme (ACE) inhibitors	12,029 (9%)	140 (7.1%)	2,031 (8.6%)	60 (10.9%)	106 (8.9%)	6,953 (9.3%)	2,739 (8.7%)
Recent use of angiotensin receptor blockers (ARBs)	11,314 (8.5%)	226 (11.4%)	2,431 (10.3%)	46 (8.4%)	114 (9.6%)	6,024 (8%)	2,473 (7.9%)
Recent use of vasopressors	2,142 (1.6%)	45 (2.3%)	410 (1.7%)	6 (1.1%)	29 (2.4%)	1,180 (1.6%)	472 (1.5%)
Socioeconomic Characteristics							
Average median household income ²	59,622.2 (21,799.8)	74,312.9 (26,438.3)	49,804.8 (18,828.6)	52,975.7 (18,449.9)	64,793.0 (23,512.0)	61,370.9 (21,145.8)	62,002.9 (22,964.5)
Average median home value ²	19,9788.4 (131,753.5)	322,685.3 (196,328.0)	170,609.1 (114,721.9)	170,308.7 (106,561.8)	253,677.7 (175,058.4)	198,423.0 (122,405.3)	216,654.6 (151,086.3)
Average percent unemployment ²	5.9 (3.2)	5.2 (2.0)	7.8 (4.0)	7.2 (4.8)	5.6 (2.6)	5.3 (2.7)	5.6 (3.0)
All socioeconomic indicators available	128,167 (95.8%)	1,844 (92.9%)	22,995 (97.5%)	541 (98.4%)	1,154 (97.1%)	72,172 (96.1%)	29,461 (94%)

Notes: ¹U07.1 diagnosis code in the inpatient care setting. ²Clin Epidemiol. 2011;64(7):749–759. ³These variables were calculated only among those individuals with all socioeconomic indicators available. ⁴**African American. ⁵†American Indian or Alaska Native. ⁶‡Native Hawaiian or Other Pacific Islander.

Table 4 Association Between Race and Critical COVID or Inpatient Mortality Among Those Hospitalized with COVID-19 Between April 1, 2020, and March 31, 2021, in the Sentinel Distributed Database

	Entire Study Cohort (n = 133,773)				Subset of the Population with Known Socioeconomic Status (n = 128,167)		
	Number of Patients Hospitalized with COVID-19	Number of Patients with Critical COVID Outcome	Unadjusted Odds Ratio (95% CI)	Odds Ratio Adjusted for Demographic and Clinical Factors (95% CI)	Number of Patients Hospitalized with COVID-19	Number of Patients with Critical COVID Outcome	Odds Ratio Adjusted for Demographic Clinical and Socioeconomic Factors (95% CI)
American Indian or Alaska Native	550	305	0.92 (0.78 1.09)	1.00 (0.84 1.18)	541	302	0.99 (0.84 1.18)
Asian	1,984	1,146	1.03 (0.91 1.17)	1.10 (1.01 1.21)	1,844	1,070	1.13 (0.99 1.29)
Black or African American	23,583	14,137	0.99 (0.86 1.15)	1.04 (0.92 1.16)	22,995	13,766	1.01 (0.89 1.13)
Native Hawaiian or Other Pacific Islander	1,188	750	1.23 (1.09 1.39)	1.22 (1.09 1.38)	1,154	725	1.17 (1.03 1.32)
Unknown	31,330	17,658	1.02 (0.91 1.14)	1.09 (0.95 1.25)	29,461	16,664	1.10 (0.96 1.26)
White	75,138	43,353	Referent group	Referent group	72,172	41,619	Referent group
	Number of Patients Hospitalized with COVID-19	Number of Patients with Recorded Inpatient Death	Unadjusted Odds Ratio (95% CI)	Odds Ratio Adjusted for Demographic and Clinical Factors (95% CI)	Number of Patients Hospitalized with COVID-19	Number of Patients with Recorded Inpatient Death	Odds Ratio Adjusted for Demographic Clinical and Socioeconomic Factors (95% CI)
American Indian or Alaska Native	550	93	1.15 (0.86 1.52)	1.57 (1.24 1.97)	541	93	1.52 (1.20 1.91)
Asian	1,984	285	0.97 (0.61 1.53)	1.26 (1.10 1.44)	1,844	272	1.30 (1.13 1.49)
Black or African American	23,583	3,356	0.92 (0.88 0.97)	1.08 (0.94 1.24)	22,995	3,291	1.03 (0.87 1.22)
Native Hawaiian or Other Pacific Islander	1,188	207	0.87 (0.37 2.06)	1.22 (1.04 1.42)	1,154	204	1.24 (1.06 1.45)
Unknown	31,330	3,636	0.82 (0.74 0.91)	1.15 (0.97 1.35)	29,461	3,474	1.12 (0.96 1.30)
White	75,138	11,425	Referent group	Referent group	72,172	10,998	Referent group

persisted (American Indian or Alaska Native population (OR 1.52 (1.20–1.91)), Asian population (OR 1.30 (1.13–1.49)), Native Hawaiian or Other Pacific Islander population (OR 1.24 (1.06–1.45)).

Discussion

We evaluated differences in COVID-19 outcomes experienced by various subpopulations within a large US population during the pre-vaccination era (between April 2020 and April 2021) and found that American Indian/Alaskan Native, Asian, and Native Hawaiian or Other Pacific Islander populations had an increased burden of hospitalization with COVID-19 and 30-day inpatient mortality compared to White population. We interpret these non-causal associations as a reflection of underlying social determinants of health in subpopulations. Our results are consistent with other smaller studies that have also shown a significant increased risk of hospitalization in these populations compared to White population after adjusting for baseline demographic, clinical^{8–11} and socioeconomic factors¹¹ over the initial months of the pandemic. We found no increase in COVID-19 related mortality among Black or African American population vs White population as shown in some prior smaller studies.^{9,10} Using a large national and geographically diverse dataset representing the insured US population, our study expands upon existing evidence on the disproportionate burden of COVID-19 pandemic in the pre-vaccination period when effective COVID-19 interventions were minimal by including granular categories of self-reported race and more effectively addressing confounding variables such as clinical and socioeconomic factors.

We saw an increased risk of mortality among AI/AN, Asian, and NHOPI populations compared to White population. Our findings are corroborated in scientific literature reporting excess mortality due to COVID-19 for AI/AN and Asian groups.^{39,40} Life expectancy for AI/AN population before the availability of COVID-19 vaccines in 2021 declined by 6 years compared to estimates from 2019, with differences attributed to poor baseline health and systemic factors such as access to quality healthcare and infrastructure. The Asian population appeared healthier and socioeconomically advantaged in our study at baseline. However, after adjusting for demographic, clinical, and socioeconomic factors, we saw higher odds of hospitalization and inpatient mortality in Asian population compared to White population. This is consistent with a study that examined COVID-19 disparities among Asian Americans.⁴¹ Potential explanations for this could be due to factors like healthcare-related occupation and multigenerational household living that are not captured in administrative claims data but are associated with an increased risk of transmission of SARS-COV-2.⁴² The AI/AN and NHOPI subpopulations have not typically been represented in other studies owing to insufficient sample size or inconsistent data collection.⁴³ Race, conceptualized in our study as a proxy for social determinants of health, can help serve as a proxy for some but not all social experiences, and therefore may not offer a fully representative view of the population. Adequately capturing and accounting for social determinants of health in real-world data might advance our understanding of inequities in health outcomes across subpopulations.

While we observed a higher burden of chronic diseases among various subpopulations in the 183 days prior to COVID-19 diagnosis or hospitalization, it is important to note that such differences in prevalence of comorbidities predate the COVID-19 pandemic in the US. Preexisting differences in burden of chronic diseases, social economic factors, and access to health care that can increase the risk of poor health outcomes were reinforced during the COVID-19 pandemic.⁴ This data demonstrates that targeted health interventions, such as funding for access to and delivery of effective preventive care services, better directed to communities in proportion to their need, are necessary to reduce the burden of comorbidities across the US population and disparities in health outcomes.

Our study findings should be considered in the context of some limitations. The administrative claims database used for this study has a high proportion of unknown race (42%). Unknown race can be a heterogeneous category comprising individuals who choose not to report their race, or, depending on the site, may report more than one race, or whose self-reported race is not available. We addressed this analytically by treating Unknown race as its own population category of self-reported race alongside other subpopulations. We descriptively compared the baseline characteristics of those with known and unknown race and included them alongside other categories of self-reported race in the regression analyses with White population as the referent category. The risk estimates for the Unknown group, while reported, are not interpreted or discussed. We deemed these estimates uninterpretable given expected heterogeneity in the Unknown group. The capture of socioeconomic status was at a neighborhood level as opposed to individual level. Therefore, complete control of confounding by individual socioeconomic status was not feasible. Our study period included a period of rapid transmission of the virus as well as evolution of clinical practices. Therefore, our findings represent differences in

COVID-19 burden in various groups on an average and aggregate level. Our findings may be subject to bias if different subgroups were less likely to be admitted to the hospital due to discriminatory practices which could potentially limit the generalizability of our findings to those with severe presentation of COVID-19. However, we note increased risks even after controlling for measured baseline differences highlighting potential persistent inequities. Lastly, our study population is representative of commercially insured beneficiaries in the US so our data may underestimate the differences among uninsured individuals diagnosed or hospitalized with COVID-19.

As a proxy for social determinants of health, race is frequently linked to differences in health outcomes. There continue to be many challenges stemming from a multitude of factors such as mistrust, language and cultural differences, health literacy, limited access within the health care system that limit access to quality educational, economic, residential, and health opportunities across all populations in the US. The scale of the COVID-19 pandemic in the US has further illuminated preexisting differences in disease burden and health outcomes across US subpopulations.⁴ Health outcomes research frequently utilizes secondary sources of data such as administrative claims data. Our study shows that these databases are imperfect in their ability to capture and account for social determinants of health. Therefore, there is need for greater investment of resources to improve collection of social determinants of health through innovative data linkages⁴⁴ in secondary healthcare databases to improve interpretability of population-based health outcomes research and to inform targeting of healthcare resources to the right populations.

Conclusions

In conclusion, our study, leveraging a large US commercially insured population during the pre-vaccination era (April 2020–March 2021), highlighted significant differences in COVID-19 outcomes across US subpopulations. We found that American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander populations experienced an increased burden of hospitalization with COVID-19 and 30-day inpatient mortality compared to the White population, even after accounting for demographic, clinical, and neighborhood-level socioeconomic factors. The persistence of inequities, despite controlling for numerous baseline factors, points to the need for targeted healthcare interventions that can address the longstanding, pre-existing differences in the burden of chronic diseases across various US subpopulations to mitigate the impact of future public health emergencies. Our study shows that race, used as a proxy for social determinants of health, is imperfectly captured in administrative claims data that are often used for observational research in the US. Improved collection of social determinants of health, including race, in secondary databases will help researchers and organizations identify patient risk factors, design targeted interventions, and advance health equity in the US.

Data Sharing Statement

The data generated in this study are not publicly available. Sentinel uses a distributed data approach in which Data Partners maintain physical and operational control of their own electronic health data after transforming it into a common data model. Sentinel does not save, maintain, or post individual level datasets to preserve patient privacy. The code that was used for querying standardized Sentinel data used for this project and related documentation is posted on the Sentinel website and referenced in Methods.

Ethics Approval and Informed Consent

This project was deemed to be a public health surveillance activity conducted under the authority of the Food and Drug Administration (FDA) and was therefore exempt from review by an Institutional Review Board.

Acknowledgments

Many thanks are due to Data Partners who provided data used in the analysis: CVS Health (Aetna), Blue Bell, PA; Humana Healthcare Research Inc., Louisville, KY; Kaiser Permanente Colorado Institute for Health Research, Aurora, CO.

The data reported in this manuscript were presented at the 39th Annual Meeting of the International Society for Pharmacoepidemiology (ISPE), held August 23–27, 2023, in Halifax, Nova Scotia, Canada. The corresponding conference abstract was published in *Pharmacoepidemiology and Drug Safety* (doi:10.1002/pds.5687). The poster

presentation is available on the Sentinel Initiative website: <https://www.sentinelinitiative.org/news-events/publications-presentations/association-between-race-and-covid-19-outcomes-united-states>

Funding

This publication was supported by the Food and Drug Administration (FDA) Office of Minority Health and Health Equity of the US Department of Health and Human Services (OMHHE/HHS) as part of a financial assistance award [FAIN] totaling \$150,000 with 100% funded by FDA OMHHE/HHS. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by FDA/HHS, or the US Government.

Disclosure

SA, RH, MEK, MW, ABP, and AA are employees of Harvard Pilgrim Health Care Institute, a non-profit organization that conducts work for government and private organizations, including pharmaceutical companies. EE was an employee of the US Food and Drug Administration during the conduct of this study. The authors report no other conflicts of interest in this work.

References

- Centers for Disease Control and Prevention. CDC museum COVID-19 timeline. 2023. Available from: <https://www.cdc.gov/museum/timeline/covid19.html>. Accessed June 20, 2023.
- Centers for Disease Control and Prevention. End of the federal COVID-19 public health emergency (PHE) declaration. 2020. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/your-health/end-of-phe.html>. Accessed June 20, 2023.
- Centers for Disease Control and Prevention. COVID data tracker. 2023. Available from: <https://covid.cdc.gov/covid-data-tracker/#datatracker-home>. Accessed February 16, 2023.
- Alelign YK, Appiah D, Ebong IA. Racial disparities in coronavirus disease 2019 (COVID-19) outcomes. *Curr Opin Cardiol*. 2021;36(3):360–366. doi:10.1097/HCO.0000000000000847
- Cooper RS. Social inequality, ethnicity and cardiovascular disease. *Int J Epidemiol*. 2001;30(suppl_1):S48. doi:10.1093/ije/30.suppl_1.S48
- Hassan S, Gujral UP, Quarells RC, et al. Global Inequity in Diabetes 3. *Lancet Diabetes Endocrinol*. 2023;11(7):509–524. doi:10.1016/S2213-8587(23)00129-8
- Davy-Mendez T, Napravnik S, Eron JJ, et al. Racial, ethnic, and gender disparities in hospitalizations among persons with HIV in the United States and Canada, 2005–2015. *AIDS Lond Engl*. 2021;35(8):1229–1239. doi:10.1097/QAD.0000000000002876
- Adegunsoye A, Ventura IB, Liarski VM. Association of black race with outcomes in COVID-19 disease: a retrospective cohort study. *Ann Am Thorac Soc*. 2020;17(10):1336–1339. doi:10.1513/AnnalsATS.202006-583RL
- Golestaneh L, Neugarten J, Fisher M, et al. The association of race and COVID-19 mortality. *EClinicalMedicine*. 2020;25:100455. doi:10.1016/j.eclinm.2020.100455
- Egede LE, Walker RJ, Garacci E, Raymond JR. Racial/Ethnic differences in COVID-19 screening, hospitalization, and mortality in Southeast Wisconsin. *Health Affairs Project Hope*. 2020;39(11):1926–1934. doi:10.1377/hlthaff.2020.01081
- Muñoz-Price LS, Nattinger AB, Rivera F, et al. Racial disparities in incidence and outcomes among patients with COVID-19. *JAMA Network Open*. 2020;3(9):e2021892. doi:10.1001/jamanetworkopen.2020.21892
- Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and mortality among black patients and white patients with covid-19. *N Engl J Med*. 2020;382(26):2534–2543. doi:10.1056/NEJMsa2011686
- Yehia BR, Winegar A, Fogel R, et al. Association of race with mortality among patients hospitalized with coronavirus disease 2019 (COVID-19) at 92 US hospitals. *JAMA Network Open*. 2020;3(8):e2018039. doi:10.1001/jamanetworkopen.2020.18039
- Gu T, Mack JA, Salvatore M, et al. Characteristics associated with racial/ethnic disparities in COVID-19 outcomes in an academic health care system. *JAMA Network Open*. 2020;3(10):e2025197–e2025197. doi:10.1001/jamanetworkopen.2020.25197
- 45 CFR Part 46 Subpart A – Basic HHS policy for protection of human research subjects. Available from: <https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46/subpart-A>. Accessed October 20, 2022.
- Federal policy for the protection of human subjects. Federal Register. 2017. Available from: <https://www.federalregister.gov/documents/2017/01/19/2017-01058/federal-policy-for-the-protection-of-human-subjects>. Accessed October 20, 2022.
- Rosati K, Jorgensen N, Plc CB. HIPAA and common rule compliance in the sentinel initiative. 49.
- Ball R, Robb M, Anderson SA, Pan GD. The FDA’s sentinel initiative—A comprehensive approach to medical product surveillance. *Clin Pharmacol Ther*. 2016;99(3):265–268. doi:10.1002/cpt.320
- Platt R, Brown JS, Robb M, et al. The FDA sentinel initiative — an evolving national resource. *N Engl J Med*. 2018;379(22):2091–2093. doi:10.1056/NEJMp1809643
- Cocoros NM, Fuller CC, Adimadhyam S, et al. A COVID-19-ready public health surveillance system: the food and drug administration’s sentinel system. *Pharmacoepidemiol Drug Saf*. 2021;30(7):827–837. doi:10.1002/pds.5240
- Sentinel common data model | sentinel initiative. *Sentinel Initiative*. Available from: <https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model>. Accessed May 21, 2022.
- How sentinel gets its data | sentinel initiative. Available from: <https://www.sentinelinitiative.org/about/how-sentinel-gets-its-data>. Accessed November 3, 2022.
- Toh S, Platt R, Steiner JF, Brown JS. Comparative-effectiveness research in distributed health data networks. *Clin Pharmacol Ther*. 2011;90(6):883–887. doi:10.1038/clpt.2011.236

24. Key database statistics | sentinel initiative. Available from: <https://www.sentinelinitiative.org/about/key-database-statistics#ethnicity-distribution-of-members-in-the-sentinel-distributed-database>. Accessed July 26, 2024.
25. World Health Organization. COVID-19 therapeutic trial synopsis. 2020. Available from: https://cdn.who.int/media/docs/default-source/blue-print/covid-19-therapeutic-trial-synopsis.pdf?sfvrsn=44b83344_1&download=true. Accessed March 8, 2026.
26. U.S. Census Bureau. About the American community survey. *Census.gov*. Available from: <https://www.census.gov/programs-surveys/acs/about.html>. Accessed November 3, 2022.
27. U.S. Census Bureau. 2019: ACS 5-year estimates data profiles. Selected economic characteristics. Median household income [CSV data file]. Available from: https://api.census.gov/data/2019/acs/acs5/profile?get=NAME,DP03_0062E&for=zip%20code%20tabulation%20area:*. Accessed November 3, 2022.
28. U.S. Census Bureau. 2019: ACS 5-year estimates subject tables. Employment status. Percent unemployment [CSV data file]. Available from: https://api.census.gov/data/2019/acs/acs5/subject?get=NAME,S2301_C04_001E&for=zip%20code%20tabulation%20area:*. Accessed November 3, 2022.
29. U.S. Census Bureau. 2019: ACS 5-year estimates data profiles. Selected housing characteristics. Median property value [CSV data file]. Available from: https://api.census.gov/data/2019/acs/acs5/profile?get=NAME,DP04_0089E&for=zip%20code%20tabulation%20area:*. Accessed November 3, 2022.
30. ZIP code to ZCTA crosswalk – UDS mapper. Available from: <https://udsmapper.org/zip-code-to-zcta-crosswalk/>. Accessed November 3, 2022.
31. CDC. People with certain medical conditions. Centers for disease control and prevention. 2022. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html>. Accessed November 3, 2022.
32. Sommerstein R, Kochen MM, Messerli FH, Gräni C. Coronavirus disease 2019 (COVID-19): do angiotensin-converting enzyme inhibitors/angiotensin receptor blockers have a biphasic effect? *J Am Heart Assoc*. 2020;9(7):e016509. doi:10.1161/JAHA.120.016509
33. Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol*. 2011;64(7):749–759. doi:10.1016/j.jclinepi.2010.10.004
34. Sun JW, Rogers JR, Her Q, et al. Adaptation and validation of the combined comorbidity score for ICD-10-CM. *Med Care*. 2017;55(12):1046–1051. doi:10.1097/MLR.0000000000000824
35. Karabon P. Rare events or non-convergence with a binary outcome? The power of firth regression in PROC LOGISTIC. *SAS Global Forum 2020*. 2020;2020:9.
36. Hertzmark E, Spiegelman D. The SAS METAANAL Macro. 2017. Available from: https://ysph.yale.edu/cmips/research/software/metaanal_340162_284_47911_v2.pdf. Accessed February 6, 2024.
37. Browse analytic development / qrp - sentinel version control system. Available from: <https://dev.sentinelssystem.org/projects/AD/repos/qrp/browse>. Accessed September 21, 2022.
38. Racial differences in COVID-19 outcomes (2020-2021) | sentinel initiative. Available from: <https://www.sentinelinitiative.org/studies/drugs/individual-drug-analyses/racial-differences-covid-19-outcomes-2020-2021>. Accessed June 22, 2023.
39. Yuan AY, Atanasov V, Barreto N, et al. Understanding racial/ethnic disparities in COVID-19 mortality using a novel metric: COVID excess mortality percentage. *Am J Epidemiol*. 2024;193(6):853–862. doi:10.1093/aje/kwae007
40. Goldman N, Park SS, Beltrán-Sánchez H. Life expectancy among native Americans during the COVID-19 pandemic: estimates, uncertainty, and obstacles. *Am J Epidemiol*. 2023;193(6):846–852. doi:10.1093/aje/kwad244
41. Kalyanaraman Marcello R, Dolle J, Tariq A, et al. Disaggregating Asian race reveals COVID-19 disparities among Asian American patients at New York City’s public hospital system. *Public Health Rep*. 2022;137(2):317–325. doi:10.1177/00333549211061313
42. Chin MK, Doãn LN, Chong SK, Wong JA, Kwon SC, Yi SS. Asian American subgroups and the COVID-19 experience: what we know and still don’t know. *Health Aff Forefr*. doi:10.1377/forefront.20210519.651079
43. Inconsistent data masks the pandemic’s toll on Native Hawaiians and Pacific Islanders. *PBS News*. 2022. Available from: <https://www.pbs.org/newshour/show/inconsistent-data-masks-the-pandemics-toll-on-native-hawaiians-and-pacific-islanders>. Accessed March 8, 2026.
44. Tonnu-Mihara I, Rojas-Lazic C, Mack M, Eshete B, Stephenson JJ, Grabner M. SA42 innovative methods for integrating social determinants of health data with administrative claims to facilitate health equity research. *Value Health*. 2023;26(6):S404–S405. doi:10.1016/j.jval.2023.03.2258

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

Dovepress
Taylor & Francis Group