Baseline Demographics and Clinical Characteristics Among 3471 US Patients Hospitalized with COVID-19 and Pulmonary Involvement: A Retrospective Study

Jennie H. Best · Shalini V. Mohan · Amanda M. Kong ·

Krish Patel · John M. Pagel · Boris Ivanov · Otis W. Brawley ·

Krutika Jariwala-Parikh · James L. Zazzali · John Pauk

ABSTRACT

Introduction: Coronavirus disease 2019 (COVID-19) can present as a range of symptoms, from mild to critical; lower pulmonary involvement, including pneumonia, is often associated with severe and critical cases. Understanding the baseline characteristics of patients hospitalized with COVID-19 illness is essential for effectively targeting clinical care and allocating resources. This study aimed to describe baseline demographics and clinical characteristics of US patients hospitalized with COVID-19 and pulmonary involvement.

Methods: US patients with COVID-19 and pulmonary involvement during an inpatient admission from December 1, 2019, to May 20, 2020, were identified using the IBM Explorys® electronic health records database. Baseline (up to 12 months prior to first COVID-19 hospitalization) demographics and clinical characteristics and preadmission (14 days to 1 day prior to admission) pulmonary diagnoses were assessed. Patients were stratified by sex, age, race, and geographic region.

Results: Overall, 3471 US patients hospitalized with COVID-19 and pulmonary involvement were included. The mean (SD) age was 63.5 (16.3) years; 51.2% of patients were female, 55.0% African American, 81.6% from the South, and 16.8% from the Midwest. The most common comorbidities included hypertension (27.7%), diabetes (17.3%), hyperlipidemia (16.3%), and obesity (9.7%). Cough (27.3%) and dyspnea (15.2%) were the most common preadmission pulmonary symptoms. African American patients were younger (mean [SD], 62.5 [15.4] vs. 67.8 [6.2]) with higher mean (SD) body mass index (33.66 [9.46] vs. 30.42 [7.86]) and prevalence of diabetes (19.8% vs. 16.7%) and lower prevalence of chronic obstructive pulmonary disease (5.6% vs. 8.2%) and smoking/tobacco use (28.1% vs. 37.2%) than White patients.

Conclusions: Among US patients primarily from the South and Midwest hospitalized with

J. H. Best (✉) · S. V. Mohan · J. L. Zazzali
Genentech, Inc., South San Francisco, CA, USA
e-mail: best.jennie@gene.com

A. M. Kong · B. Ivanov · K. Jariwala-Parikh
IBM Watson Health, Cambridge, MA, USA

K. Patel · J. M. Pagel
Swedish Cancer Institute, Seattle, WA, USA

O. W. Brawley
Johns Hopkins Medicine, Baltimore, MD, USA

J. Pauk
Swedish Medical Center, Seattle, WA, USA

COVID-19 and pulmonary involvement, the most common comorbidities were hypertension, diabetes, hyperlipidemia, and obesity. Differences observed between African American and White patients should be considered in the context of the complex factors underlying racial disparities in COVID-19.

Keywords: Baseline characteristics; Coronavirus; COVID-19; Demographics; Hospitalized; Infectious disease; Pulmonary involvement

Key Summary Points

Why carry out this study?

Understanding the baseline characteristics of patients hospitalized with COVID-19 across a variety of US geographic regions is essential for targeting clinical care and allocating resources.

This study examined baseline demographics and clinical characteristics of patients hospitalized with COVID-19 and pulmonary involvement from a US electronic health records database that includes patients primarily from the South and Midwest; patients were stratified by geographic region, race, sex, and age.

What was learned from this study?

Among US patients primarily from the South and Midwest hospitalized with COVID-19 and pulmonary involvement, the most common comorbidities were hypertension, diabetes, hyperlipidemia, and obesity.

Compared with White patients, African American patients were younger, with higher mean BMI, higher prevalence of concurrent diabetes, and lower prevalence of COPD and smoking/tobacco use; it is important to view these data in light of underlying reasons for racial disparities in COVID-19.

This study provides real-world data from geographic regions with limited published COVID-19 data; such information can be used to help target clinical care and allocate resources.

DIGITAL FEATURES

This article is published with digital features to facilitate understanding of the article. To view digital features for this article go to https://doi.org/10.6084/m9.figshare.12961658.

INTRODUCTION

The global coronavirus disease 2019 (COVID-19) pandemic began in Wuhan, China, and quickly spread internationally [1]. As of July 28, 2020, there were over 4.3 million confirmed cases of COVID-19 in the US and over 148,000 deaths [2]. COVID-19 can present as a range of symptoms, from mild to critical; lower pulmonary involvement, including severe pneumonia, dyspnea, or blood oxygen saturation $\leq 93\%$, among other potential symptoms, is often associated with severe and critical cases [3].

Understanding the baseline characteristics of patients hospitalized with COVID-19 illness is essential for effectively targeting clinical care and allocating resources. Two large studies of patients hospitalized with COVID-19 illness in New York observed that the majority of patients were male and the most common comorbidities were hypertension, obesity, and diabetes [4, 5]. In comparison, a study of patients hospitalized with COVID-19 in Louisiana reported that the majority of patients were female and 70% were Black non-Hispanic; in that study, increased odds of hospital admission were associated with Black race, increasing age, higher Charlson Comorbidity Index (CCI), public insurance, residence in a low-income area, and obesity [6]. These studies demonstrate that patients from one state may not be representative of patients

throughout the US and that studies across a variety of geographic regions are needed to specifically tailor clinical responses to COVID-19.

The objective of this study was to examine baseline demographics and clinical characteristics of US patients hospitalized with COVID-19 and pulmonary involvement. This study utilized the IBM Explorys® US electronic health records (EHR) database, which includes a racially diverse group of patients primarily from the South and Midwest. Patients in the database with COVID-19 and pulmonary involvement from December 1, 2019, to May 20, 2020, were included; patients were examined overall and stratified by sex, age, race, and geographic region.

METHODS

Study Design and Patient Population

This retrospective study used EHR data from the US-based IBM Explorys® dataset from December 1, 2019, to May 20, 2020. This database is a convenience sample of health systems and contains data on approximately 70 million patients (since 2000) from 39 large health systems that include approximately 400 hospitals and 400,000 providers. Currently, the Explorys population comprises patients from the Midwest (39%), South (33%), West (18%), and Northeast (5%). Explorys includes EHR, outgoing billing, and adjudicated claims from both commercial and public payers. The venue of care is longitudinal, including inpatient, ambulatory, emergency, and post-acute care data. Data include diagnoses, procedures, laboratory results, vital signs, patient-reported outcomes, encounter-level data, providers, and other clinical, financial, and operational data.

This was a retrospective, observational study, and all study data were deidentified and fully compliant with Health Insurance Portability and Accountability Act Regulations (45 CFR 164.514e); therefore, approval from an institutional review board was not required, and informed consent was not obtained. Variables were measured using International Classification of Diseases, Clinical Modification (ICD-9-CM and/or ICD-10-CM) diagnosis, Systemized Nomenclature of Medicine, and procedure codes, Healthcare Common Procedure Coding System codes, and National Drug Codes, as appropriate.

Patients with a COVID-19 diagnosis code and a diagnosis code for pulmonary involvement based on the Centers for Disease Control and Prevention coding guidelines were included (Table S1 in the electronic supplementary material) [7]. Pulmonary involvement included pneumonia, acute bronchitis, bronchitis not otherwise specified, lower respiratory, associated with a respiratory infection not otherwise specified, and acute respiratory distress syndrome. If a patient had more than one hospitalization with COVID-19 and pulmonary involvement, only the first hospitalization was included in this analysis.

Outcomes

Patient demographics, baseline clinical characteristics, and prescribed medications were assessed using available data from up to 12 months through 14 days before hospital admission (baseline period) for patients who had at least one encounter during this period. Symptoms during the 14 days to 1 day prior to hospitalization were assessed separately. Demographics included age, sex, patient-reported race/ethnicity, geographic region, smoking status, body mass index (BMI), and month of admission. Self-identified race data were included as study variables to characterize hospitalized patients. Comorbid conditions included 17 CCI [8] conditions and several select comorbid conditions, including hypertension, diabetes, coronary artery disease, and respiratory conditions (chronic obstructive pulmonary disease, asthma, sleep apnea, and pulmonary fibrosis).

Statistical Analysis

Results are presented for the overall patient sample and were stratified by geographic region, race, sex, and age category (< 60 years

and ≥ 60 years). To compare patient characteristics between cohorts in the stratified data, t tests and χ^2 tests were used. P values < 0.05 were considered statistically significant.

RESULTS

Demographics and Clinical Characteristics

Between December 1, 2019, and May 20, 2020, 91,380 patients had a pulmonary diagnosis; after applying exclusion criteria, 3471 patients had diagnoses of COVID-19 and pulmonary involvement during an inpatient admission and were included (Fig. 1; Table S1 in the electronic supplementary material). The most common pulmonary diagnoses were pneumonia (56.1%) and lower respiratory infection (43.4%).

The mean (SD) and median age were 63.5 (16.3) and 64.0 years, respectively; 1776 patients (51.2%) were female and 1910 (55.0%) were African American (Table 1). Overall, 2834 patients (81.6%) were from the South and 583 (16.8%) were from the Midwest; the most highly represented states included Louisiana (1677 patients [48%]), Ohio (559 [16%]), and Maryland (461 [13%]) as well as the District of Columbia (496 [14%]). Patients' insurance status included unknown (57.6%), private insurance (17.7%), Medicare (14.1%), self-pay (6.8%), and Medicaid (3.1%).

Overall, mean (SD) and median CCI were 0.92 (1.89) and 0, respectively, and the most common comorbidities (n [%]) during the baseline period were hypertension (961 [27.7%]), diabetes (600 [17.3%]), hyperlipidemia (566 [16.3%]), and obesity (338 [9.7%]; Table 2). The mean (SD) and median BMIs in patients with a recorded measurement (2460 patients [70.9%]) were 32.05 (8.85) and 30.70, respectively. A history of smoking/tobacco use was self-reported by 994 patients (28.6%). The most common (n [%]) preadmission prescriptions were atorvastatin (332 [9.6%]), amlodipine (312 [9.0%]), and metoprolol (238 [6.9%]). The most common (n [%]) select baseline prescriptions were steroids (314 [9.0%]), angiotensin II receptor blockers (ARBs; 256 [7.4%]), angiotensin-converting enzyme inhibitors
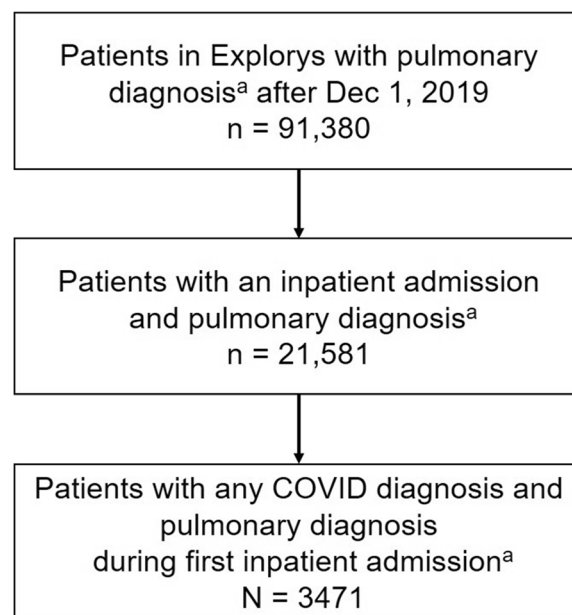

Fig. 1 Patient attrition. [a]Pneumonia, acute bronchitis, bronchitis not otherwise specified, lower respiratory tract infection, other specified respiratory disorder, acute respiratory distress syndrome. *COVID-19* coronavirus disease 2019

(ACEi; 199 [5.7%]), and bronchodilators (143 [4.1%]). There were no baseline prescriptions for Janus kinase inhibitors, dornase alfa, baloxavir marboxil, or intravenous immunoglobulin. Overall, 1621 patients (46.7%) had a healthcare encounter during the 14 days to 1 day prior to their first COVID-19 admission; cough, fever, breathing abnormalities/dyspnea, and hypoxemia were reported in 443 patients (27.3%), 373 (23.0%), 246 (15.2%), and 42 (2.6%), respectively.

Analyses Stratified by Geographic Region

The largest geographic regions represented were the South ($n = 2834$) and Midwest ($n = 583$). Among patients from the South, 1723 (60.8%) were African American and 772 (27.2%) were White, while in the Midwest, 181 (31.0%) were African American and 347 (59.5%) were White. Patients from the South had lower mean (SD) CCI than patients from the Midwest (0.88 [1.87] vs. 1.12 [2.03]; $P = 0.006$); however, the median

Table 1 Demographic characteristics prehospitalization

	All patients N = 3471	Region			Race			Sex		Age, years	
		South n = 2834	Midwest n = 583	Other n = 54	African American n = 1910	White n = 1146	Other n = 415	Male n = 1695	Female n = 1776	< 60 n = 1349	≥ 60 n = 2120
Age, years											
Mean (SD)	63.5 (16.3)	63.6 (16.3)	62.9 (16.0)	63.4 (17.1)	62.5 (15.4)	67.8 (16.2)	56.3 (16.9)	62.2 (15.5)	64.8 (16.9)	47.0 (10.0)	74.0 (9.3)
Median	64.0	65.0	63.0	66.0	63.0	69.0	56.0	63.0	66.0	50.0	73.0
P value	–	Ref	0.320	0.932	Ref	< 0.001	< 0.001	Ref	< 0.001	Ref	< 0.001
Age group (in years), n (%)											
< 18	5 (0.1)	3 (0.1)	2 (0.3)	0	3 (0.2)	1 (0.1)	1 (0.2)	4 (0.2)	1 (0.1)	5 (0.4)	0
18–29	93 (2.7)	76 (2.7)	14 (2.4)	3 (5.6)	47 (2.5)	26 (2.3)	20 (4.8)	38 (2.2)	55 (3.1)	93 (6.9)	0
30–39	195 (5.6)	163 (5.8)	30 (5.1)	2 (3.7)	102 (5.3)	42 (3.7)	51 (12.3)	93 (5.5)	102 (5.7)	195 (14.5)	0
40–49	371 (10.7)	311 (11.0)	55 (9.4)	5 (9.3)	216 (11.3)	82 (7.2)	73 (17.6)	209 (12.3)	162 (9.1)	371 (27.5)	0
50–59	685 (19.7)	539 (19.0)	136 (23.3)	10 (18.5)	422 (22.1)	164 (14.3)	99 (23.9)	369 (21.8)	316 (17.8)	685 (50.8)	0
60–69	809 (23.3)	654 (23.1)	141 (24.2)	14 (25.9)	466 (24.4)	265 (23.1)	78 (18.8)	416 (24.5)	393 (22.1)	0	809 (38.2)
70–79	697 (20.1)	580 (20.5)	108 (18.5)	9 (16.7)	387 (20.3)	255 (22.3)	55 (13.3)	333 (19.6)	364 (20.5)	0	697 (32.9)
80–89	429 (12.4)	351 (12.4)	71 (12.2)	7 (13.0)	198 (10.4)	209 (18.2)	22 (5.3)	177 (10.4)	252 (14.2)	0	429 (20.2)
≥ 90	185 (5.3)	156 (5.5)	25 (4.3)	4 (7.4)	69 (3.6)	100 (8.7)	16 (3.9)	55 (3.2)	130 (7.3)	0	185 (8.7)
Missing	2 (0.1)	1 (0)	1 (0.2)	0	0	2 (0.2)	0	1 (0.1)	1 (0.1)	0	0
P value	–	Ref	0.214	0.959	Ref	< 0.001	< 0.001	Ref	< 0.001	–	–
Sex, n (%)											
Male	1695 (48.8)	1354 (47.8)	312 (53.5)	29 (53.7)	863 (45.2)	578 (50.4)	254 (61.2)	1695 (100)	0	713 (52.9)	981 (46.3)
Female	1776 (51.2)	1480 (52.2)	271 (46.5)	25 (46.3)	1047 (54.8)	568 (49.6)	161 (38.8)	0	1776 (100)	636 (47.1)	1139 (53.7)
P value	–	Ref	0.012	0.388	Ref	0.005	< 0.001	Ref	–	Ref	< 0.001

Table 1 Demographic characteristics prehospitalization

	All patients N = 3471	Region			Race			Sex		Age, years	
		South n = 2834	Midwest n = 583	Other n = 54	African American n = 1910	White n = 1146	Other n = 415	Male n = 1695	Female n = 1776	< 60 n = 1349	≥ 60 n = 2120
Race, n (%)											
Hispanic/Latino	3 (0.1)	0	3 (0.5)	0	0	0	3 (7)	0	3 (2)	2 (0.1)	1 (0)
African American	1910 (55.0)	1723 (60.8)	181 (31.0)	6 (11.1)	1910 (100)	0	0	863 (50.9)	1047 (59.0)	790 (58.6)	1120 (52.8)
White	1146 (33.0)	772 (27.2)	347 (59.5)	27 (50.0)	0	1146 (100)	0	578 (34.1)	568 (32.0)	315 (23.4)	829 (39.1)
Asian	26 (0.7)	18 (0.6)	8 (1.4)	0	0	0	26 (6.3)	15 (0.9)	11 (0.6)	11 (0.8)	15 (0.7)
Other	307 (8.8)	276 (9.7)	13 (2.2)	18 (33.3)	0	0	307 (74.0)	201 (11.9)	106 (6.0)	199 (14.8)	108 (5.1)
Unknown/refused	79 (2.3)	45 (1.6)	31 (5.3)	3 (5.6)	0	0	79 (19.0)	38 (2.2)	41 (2.3)	32 (2.4)	47 (2.2)
P value	–	Ref	< 0.001	< 0.001	–	–	–	Ref	< 0.001	Ref	< 0.001
Geographic region, n (%)											
Northeast	40 (1.2)	0	0	40 (74.1)	4 (0.2)	26 (2.3)	10 (2.4)	23 (1.4)	17 (1.0)	13 (1.0)	27 (1.3)
Midwest	583 (16.8)	0	583 (100)	0	181 (9.5)	347 (30.3)	55 (13.3)	312 (18.4)	271 (15.3)	237 (17.6)	345 (16.3)
South	2834 (81.6)	2834 (100)	0	0	1723 (90.2)	772 (67.4)	339 (81.7)	1354 (79.9)	1480 (83.3)	1092 (80.9)	1741 (82.1)
West	12 (0.3)	0	0	12 (22.2)	2 (0.1)	1 (0.1)	9 (2.2)	5 (0.3)	7 (0.4)	6 (0.4)	6 (0.3)
Unknown/missing	2 (0.1)	0	0	2 (3.7)	0	0	2 (0.5)	1 (0.1)	1 (0.1)	1 (0.1)	1 (0)
P value	–	–	–	–	Ref	< 0.001	< 0.001	Ref	0.098	Ref	0.667
Insurance plan type, n (%)											
Private	616 (17.7)	301 (10.6)	303 (52.0)	12 (22.2)	275 (14.4)	268 (23.4)	73 (17.6)	330 (19.5)	286 (16.1)	336 (24.9)	280 (13.2)
Medicare	490 (14.1)	306 (10.8)	168 (28.8)	16 (29.6)	251 (13.1)	201 (17.5)	38 (9.2)	224 (13.2)	266 (15.0)	38 (2.8)	452 (21.3)
Medicaid	107 (3.1)	45 (1.6)	60 (10.3)	2 (3.7)	63 (3.3)	23 (2.0)	21 (5.1)	51 (3.0)	56 (3.2)	66 (4.9)	40 (1.9)
Self-pay	236 (6.8)	212 (7.5)	9 (1.5)	15 (27.8)	98 (5.1)	25 (2.2)	113 (27.2)	152 (9.0)	84 (4.7)	149 (11.0)	87 (4.1)
Other	20 (0.6)	7 (0.2)	12 (2.1)	1 (1.9)	10 (0.5)	7 (0.6)	3 (0.7)	11 (0.6)	9 (0.5)	11 (0.8)	9 (0.4)
Unknown	2001 (57.6)	1962 (69.2)	31 (5.3)	8 (14.8)	1213 (63.5)	622 (54.3)	166 (40.0)	926 (54.6)	1075 (60.5)	749 (55.5)	1251 (59.0)
P value	–	Ref	< 0.001	< 0.001	Ref	< 0.001	< 0.001	Ref	< 0.001	Ref	< 0.001

Table 1 continued

	All patients	Region			Race			Sex		Age, years	
	N = 3471	South n = 2834	Midwest n = 583	Other n = 54	African American n = 1910	White n = 1146	Other n = 415	Male n = 1695	Female n = 1776	< 60 n = 1349	≥ 60 n = 2120
Month of admission, n (%)											
Dec 2019	4 (0.1)	2 (0.1)	2 (0.3)	0	1 (0.1)	3 (0.3)	0	3 (0.2)	1 (0.1)	2 (0.1)	2 (0.1)
Jan 2020	6 (0.2)	2 (0.1)	4 (0.7)	0	0	5 (0.4)	1 (0.2)	2 (0.1)	4 (0.2)	3 (0.2)	2 (0.1)
Feb 2020	8 (0.2)	3 (0.1)	5 (0.9)	0	1 (0.1)	7 (0.6)	0	2 (0.1)	6 (0.3)	3 (0.2)	5 (0.2)
Mar 2020	1164 (33.5)	939 (33.1)	216 (37.0)	9 (16.7)	711 (37.2)	370 (32.3)	83 (20.0)	561 (33.1)	603 (34.0)	480 (35.6)	684 (32.3)
Apr 2020	1931 (55.6)	1596 (56.3)	297 (50.9)	38 (70.4)	1023 (53.6)	639 (55.8)	269 (64.8)	946 (55.8)	985 (55.5)	703 (52.1)	1227 (57.9)
May 2020	358 (10.3)	292 (10.3)	59 (10.1)	7 (13.0)	174 (9.1)	122 (10.6)	62 (14.9)	181 (10.7)	177 (10.0)	158 (11.7)	200 (9.4)
P value	–	Ref	< 0.001	0.242	Ref	< 0.001	< 0.001	Ref	0.531	Ref	0.024

Ref reference

Table 2 Twelve-month baseline clinical characteristics prior to hospitalization

	All patients	Geographic region			Race			Sex		Age, years	
	N = 3471	South n = 2834	Midwest n = 583	Other n = 54	African American n = 1910	White n = 1146	Other n = 415	Male n = 1695	Female n = 1776	< 60 n = 1349	≥ 60 n = 2120
Deyo CCI											
Mean (SD)	0.92 (1.89)	0.88 (1.87)	1.12 (2.03)[a]	0.50 (1.51)	1.03 (1.98)	0.93 (1.89)	0.36 (1.32)[b]	0.91 (1.90)	0.92 (1.89)	0.54 (1.45)	1.16 (2.09)[c]
Median	0	0	0	0	0	0	0	0	0	0	0
Deyo CCI comorbid conditions, n (%)											
Mild to moderate diabetes	398 (11.5)	303 (10.7)	88 (15.1)[a]	7 (13.0)	246 (12.9)	132 (11.5)	20 (4.8)[b]	188 (11.1)	210 (11.8)	120 (8.9)	278 (13.1)[c]
Renal disease	372 (10.7)	307 (10.8)	62 (10.6)	3 (5.6)	238 (12.5)	116 (10.1)	18 (4.3)[b]	183 (10.8)	189 (10.6)	77 (5.7)	295 (13.9)[c]
Diabetes with chronic complications	278 (8.0)	230 (8.1)	48 (8.2)	0[d]	190 (9.9)	77 (6.7)[e]	11 (2.7)[b]	128 (7.6)	150 (8.4)	65 (4.8)	213 (10)[c]
Chronic pulmonary disease	272 (7.8)	189 (6.7)	80 (13.7)[f]	3 (5.6)	155 (8.1)	108 (9.4)	9 (2.2)[b]	105 (6.2)	167 (9.4)[g]	77 (5.7)	195 (9.2)[c]
Congestive heart failure	238 (6.9)	188 (6.6)	48 (8.2)	2 (3.7)	153 (8.0)	77 (6.7)	8 (1.9)[b]	119 (7.0)	119 (6.7)	43 (3.2)	195 (9.2)[c]
Primary cancer	139 (4.0)	105 (3.7)	33 (5.7)[d]	1 (1.9)	81 (4.2)	52 (4.5)	6 (1.4)[e]	76 (4.5)	63 (3.5)	23 (1.7)	116 (5.5)[c]
Cerebrovascular disease	104 (3.0)	82 (2.9)	22 (3.8)	0	59 (3.1)	41 (3.6)	4 (1.0)	47 (2.8)	57 (3.2)	14 (1.0)	90 (4.2)[c]
Peripheral vascular disease	93 (2.7)	73 (2.6)	20 (3.4)	0	53 (2.8)	38 (3.3)	2 (0.5)[e]	61 (3.6)	32 (1.8)[g]	9 (0.7)	84 (4.0)[c]
Dementia	72 (2.1)	53 (1.9)	18 (3.1)	1 (1.9)	28 (1.5)	40 (3.5)[b]	4 (1.0)	37 (2.2)	35 (2.0)	3 (0.2)	69 (3.3)[c]
Myocardial infarction	53 (1.5)	35 (1.2)	18 (3.1)[f]	0	29 (1.5)	19 (1.7)	5 (1.2)	32 (1.9)	21 (1.2)	9 (0.7)	44 (2.1)[c]
Rheumatologic disease	34 (1.0)	28 (1.0)	5 (0.9)	1 (1.9)	24 (1.3)	10 (0.9)	0[h]	7 (0.4)	27 (1.5)[g]	15 (1.1)	19 (0.9)
Metastatic solid tumor	31 (0.9)	25 (0.9)	6 (1.0)	0	20 (1.0)	9 (0.8)	2 (0.5)	14 (0.8)	17 (1.0)	6 (0.4)	25 (1.2)[i]
Mild liver disease	24 (0.7)	16 (0.6)	6 (1.0)	0	12 (0.6)	10 (0.9)	2 (0.5)	10 (0.6)	14 (0.8)	8 (0.6)	16 (0.8)
Peptic ulcer disease	16 (0.5)	10 (0.4)	6 (1.0)[d]	0	6 (0.3)	9 (0.8)	1 (0.2)	8 (0.5)	8 (0.5)	2 (0.1)	14 (0.7)[i]
Hemiplegia and paraplegia	14 (0.4)	13 (0.5)	1 (0.2)	0	9 (0.5)	5 (0.4)	0	7 (0.4)	7 (0.4)	3 (0.2)	11 (0.5)
HIV	10 (0.3)	8 (0.3)	2 (0.3)	0	7 (0.4)	2 (0.2)	1 (0.2)	7 (0.4)	3 (0.2)	7 (0.5)	3 (0.1)[i]
Moderate to severe liver disease	7 (0.2)	4 (0.1)	2 (0.3)	1 (1.9)[a]	1 (0.1)	4 (0.3)[h]	2 (0.5)[h]	4 (0.2)	3 (0.2)	3 (0.2)	4 (0.2)

Table 2 continued

	All patients	Geographic region			Race			Sex		Age, years	
	N = 3471	South n = 2834	Midwest n = 583	Other n = 54	African American n = 1910	White n = 1146	Other n = 415	Male n = 1695	Female n = 1776	< 60 n = 1349	≥ 60 n = 2120
Select comorbidities, n (%)											
Hypertension	961 (27.7)	736 (26.0)	216 (37.0)[f]	9 (16.7)	581 (30.4)	326 (28.4)	54 (13.0)[b]	446 (26.3)	515 (29.0)	246 (18.2)	715 (33.7)[c]
Diabetes	600 (17.3)	467 (16.5)	124 (21.3)[a]	9 (16.7)	378 (19.8)	191 (16.7)[h]	31 (7.5)[b]	274 (16.2)	326 (18.4)	172 (12.8)	428 (20.2)[c]
Hyperlipidemia	566 (16.3)	413 (14.6)	148 (25.4)[f]	5 (9.3)	316 (16.5)	225 (19.6)[h]	25 (6.0)[b]	274 (16.2)	292 (16.4)	131 (9.7)	435 (20.5)[c]
Obesity	338 (9.7)	243 (8.6)	92 (15.8)[f]	3 (5.6)	221 (11.6)	103 (9.0)[h]	14 (3.4)[b]	128 (7.6)	210 (11.8)[g]	144 (10.7)	194 (9.2)
Coronary artery disease	220 (6.3)	150 (5.3)	68 (11.7)[f]	2 (3.7)	103 (5.4)	102 (8.9)[b]	15 (3.6)	134 (7.9)	86 (4.8)[g]	31 (2.3)	189 (8.9)[c]
COPD	206 (5.9)	154 (5.4)	49 (8.4)[a]	3 (5.6)	107 (5.6)	94 (8.2)[c]	5 (1.2)[b]	84 (5.0)	122 (6.9)[j]	41 (3.0)	165 (7.8)[c]
Sleep apnea	180 (5.2)	120 (4.2)	58 (9.9)[f]	2 (3.7)	105 (5.5)	67 (5.8)	8 (1.9)[c]	76 (4.5)	104 (5.9)	61 (4.5)	118 (5.6)
Asthma	114 (3.3)	71 (2.5)	43 (7.4)[f]	0	68 (3.6)	40 (3.5)	6 (1.4)[h]	33 (1.9)	81 (4.6)[g]	46 (3.4)	68 (3.2)
Pulmonary fibrosis	49 (1.4)	40 (1.4)	8 (1.4)	1 (1.9)	29 (1.5)	20 (1.7)	0[h]	26 (1.5)	23 (1.3)	8 (0.6)	41 (1.9)[c]
Organ transplant	61 (1.8)	44 (1.6)	15 (2.6)	2 (3.7)	34 (1.8)	21 (1.8)	6 (1.4)	41 (2.4)	20 (1.1)[k]	27 (2.0)	34 (1.6)
Influenza	46 (1.3)	29 (1.0)	17 (2.9)[f]	0	24 (1.3)	20 (1.7)	2 (0.5)	24 (1.4)	22 (1.2)	19 (1.4)	27 (1.3)
Pregnancy	23 (0.7)	20 (0.7)	3 (0.5)	0	10 (0.5)	9 (0.8)	4 (1.0)	0	23 (1.3)[g]	22 (1.6)	1 (0)[c]
Smoking/tobacco use[i]	994 (28.6)	777 (27.4)	217 (37.2)[f]	0[f]	536 (28.1)	426 (37.2)[b]	32 (7.7)[b]	525 (31.0)	469 (26.4)[g]	254 (18.8)	740 (34.9)[c]
BMI, has measurement	2460 (70.9)	1857 (65.5)	571 (97.9)[f]	32 (59.3)	1297 (67.9)	1030 (89.9)[b]	133 (32.0)[b]	1151 (67.9)	1309 (73.7)[g]	907 (67.2)	1551 (73.2)[c]
BMI, mean (SD)	32.05 (8.85)	32.16 (8.98)	31.77 (8.46)	30.69 (7.71)	33.66 (9.46)	30.42 (7.86)[b]	29.06 (6.53)[b]	30.75 (7.74)	33.19 (9.58)[g]	35.51 (9.68)	30.05 (7.61)[c]
BMI, median	30.70	30.90	30.27	28.22	32.20	29.52	28.30	29.67	32.10	34.30	29.00
Select baseline prescriptions, n (%)											
Steroids	314 (9.0)	186 (6.6)	126 (21.6)[f]	2 (3.7)	152 (8.0)	152 (13.3)[b]	10 (2.4)[b]	135 (8.0)	179 (10.1)[j]	118 (8.7)	196 (9.2)
Methylprednisolone	99 (2.9)	47 (1.7)	52 (8.9)[f]	0	39 (2.0)	55 (4.8)[b]	5 (1.2)	40 (2.4)	59 (3.3)	41 (3.0)	58 (2.7)
ARBs	256 (7.4)	206 (7.3)	49 (8.4)	1 (1.9)	149 (7.8)	100 (8.7)	7 (1.7)[b]	106 (6.3)	150 (8.4)[j]	78 (5.8)	178 (8.4)[m]
ACE inhibitors	199 (5.7)	142 (5.0)	57 (9.8)[f]	0	102 (5.3)	88 (7.7)[h]	9 (2.2)[c]	108 (6.4)	91 (5.1)	66 (4.9)	133 (6.3)
Bronchodilators	143 (4.1)	100 (3.5)	41 (7.0)[f]	2 (3.7)	73 (3.8)	67 (5.8)[h]	3 (0.7)[b]	44 (2.6)	99 (5.6)[g]	32 (2.4)	110 (5.2)[c]
TNF inhibitors	3 (0.1)	3 (0.1)	0	0	2 (0.1)	1 (0.1)	0	3 (0.2)	0	0	3 (0.1)

Table 2 continued

	All patients	Geographic region			Race				Sex		Age, years	
	$N = 3471$	South	Midwest	Other	African American	White	Other	Male	Female		< 60	≥ 60
		$n = 2834$	$n = 583$	$n = 54$	$n = 1910$	$n = 1146$	$n = 415$	$n = 1695$	$n = 1776$		$n = 1349$	$n = 2120$
Antirejection medications	28 (0.8)	20 (0.7)	7 (1.2)	1 (1.9)	17 (0.9)	11 (1.0)	0	20 (1.2)	8 (0.5)[j]		15 (1.1)	13 (0.6)

Diagnoses/prescriptions using available data from up to 12 months through 14 days before hospital admission

Present in ≥ 5% of the sample

ACE angiotensin-converting enzyme, *ARBs* angiotensin II receptor blockers, *BMI* body mass index, *CCI* Charlson Comorbidity Index, *COPD* chronic obstructive pulmonary disease, *TNF* tumor necrosis factor

[a] $P < 0.01$ for comparison vs. South

[b] $P ≤ 0.001$ for comparison vs. African American

[c] $P ≤ 0.001$ for comparison vs. < 60 years of age

[d] $P < 0.05$ for comparison vs. South

[e] $P < 0.01$ for comparison vs. African American

[f] $P ≤ 0.001$ for comparison vs. South

[g] $P ≤ 0.001$ for comparison vs. male

[h] $P < 0.05$ for comparison vs. African American

[i] $P < 0.05$ for comparison vs. < 60 years of age

[j] $P < 0.05$ for comparison vs. male

[k] $P < 0.01$ for comparison vs. male

[l] Self-reported history of smoking/tobacco use

[m] $P < 0.01$ for comparison vs. < 60 years of age

CCI was the same for both groups (0 for both). Lower proportions of patients from the South had hypertension (26.0% vs. 37.0%; $P < 0.001$), diabetes (16.5% vs. 21.3%; $P = 0.005$), hyperlipidemia (14.6% vs. 25.4%; $P < 0.001$), or obesity (8.6% vs. 15.8%; $P < 0.001$) than patients from the Midwest. In addition, a lower proportion of patients from the South had a history of smoking/tobacco use than patients from the Midwest (27.4% vs. 37.2%; $P < 0.001$). Higher proportions of patients from the Midwest had a baseline prescription for steroids, including inhaled steroids (21.6% vs. 6.6%), ACEi (9.8% vs. 5.0%), and bronchodilators (7.0% vs. 3.5%; $P < 0.001$ for all) than patients from the South. A lower proportion of patients from the South had a preadmission healthcare encounter in the 14 days to 1 day prior to their first COVID-19 admission than patients from the Midwest (42.2% vs. 69.6%; $P < 0.001$); there were no significant differences in preadmission pulmonary diagnoses among these groups.

Analyses Stratified by Race

Overall, 1910 patients self-identified as African American and 1146 as White. Among African American patients, 1723 (90.2%) were from the South compared with 772 (67.4%) of the White patients. African American patients were significantly younger than White patients (median, 63.0 vs. 69.0 years; mean [SD], 62.5 [15.4] vs. 67.8 [16.2] years; $P < 0.001$), and a larger proportion were female (54.8% vs. 49.6%; $P = 0.005$). The mean (SD) CCI was higher in African American patients than in White patients (1.03 [1.98] vs. 0.93 [1.89]), although the difference was not statistically significant; the median CCI was 0 in both groups. Higher proportions of African American patients had diabetes (19.8% vs. 16.7%; $P = 0.032$) and obesity (11.6% vs. 9.0%; $P = 0.025$) and a higher mean (SD) BMI (33.66 [9.46] vs. 30.42 [7.86]; $P < 0.001$) and median BMI (32.20 vs. 29.52) compared with White patients. A higher proportion of White patients had hyperlipidemia (19.6% vs. 16.5%; $P = 0.030$), chronic obstructive pulmonary disease (COPD; 8.2% vs. 5.6%; $P = 0.005$), and self-reported a history of smoking/tobacco use (37.2% vs. 28.1%; $P < 0.001$) than African American patients. Compared with White patients, lower proportions of African American patients had a prescription for steroids (13.3% vs. 8.0%; $P < 0.001$), ACEi (7.7% vs. 5.3%; $P = 0.010$), and bronchodilators (5.8% vs. 3.8%; $P = 0.010$). A lower proportion of African American patients had a preadmission healthcare encounter than White patients (44.8% vs. 56.1%; $P < 0.001$), but there were no significant differences in preadmission pulmonary diagnoses between these groups.

Analyses Stratified by Sex

Among all patients, 48.8% were male and 51.2% female; female patients were older than male patients (median, 66.0 vs. 63.0 years; mean [SD], 64.8 [16.9] vs. 62.2 [15.5]; $P < 0.001$). A higher proportion of female patients had COPD (6.9% vs. 5.0%; $P = 0.017$), and a lower proportion had a history of smoking/tobacco use (26.4% vs. 31.0%; $P < 0.001$) compared with male patients.

Analyses Stratified by Age

A higher proportion of patients who were < 60 years of age had a preadmission healthcare encounter than patients who were ≥ 60 years of age (51.9% vs. 43.4%; $P < 0.001$), with cough and breathing abnormalities/dyspnea reported in higher proportions of patients who were < 60 years of age than in those who were ≥ 60 years of age (33.3% vs. 22.8%; $P < 0.001$ and 18.3% vs. 12.7%; $P = 0.002$).

DISCUSSION

This study examined baseline demographics and clinical characteristics among US patients primarily from the South and Midwest—regions with limited data on patients with COVID-19—who were hospitalized with COVID-19 illness and pulmonary involvement. Overall, 81.6% of patients were from the South and 16.8% from

the Midwest. Two highly represented groups were female (51.2%) and African American (55.0%) patients. The mean (SD) age was 63.5 (16.3) years. Hypertension, diabetes, hyperlipidemia, and obesity were the most common comorbidities. Preadmission pulmonary symptoms were also examined, and the most common were cough and dyspnea.

The most prevalent comorbidities observed were the same as those reported after analysis of EHR data of 5700 patients hospitalized with COVID-19 from the New York City area, with the addition of hyperlipidemia, although they were present in a lower proportion of patients in this study vs. the New York study (hypertension, 27.7% vs. 56.6%; diabetes, 17.3% vs. 33.8%; obesity, 9.7% vs. 41.7%, respectively) [4]. A retrospective cohort study of patients diagnosed with COVID-19 in Louisiana, including nonhospitalized and hospitalized patients, reported similar rates of hypertension and diabetes but a higher rate of obesity than the present study [6]. Diagnoses for obesity, defined by the US Centers for Disease Control and Prevention as BMI $\geq$ 30, may be underreported in the EHR database used in this study, as both the mean and median BMI were > 30 for the overall population [9].

Analyses by region demonstrated that, when compared with patients from the South (primarily from Louisiana, Maryland, and the District of Columbia), patients from the Midwest (primarily from Ohio) had higher mean CCI and increased prevalence of hypertension, diabetes, hyperlipidemia, and obesity. In comparison, according to a 2019 report on obesity in the US, Ohio has higher rates of obesity, hypertension, and diabetes than Maryland and the District of Columbia but lower rates than Louisiana [10]. There were differences in insurance plan type between regions; however, no conclusions can be drawn because these data were missing for the majority of patients.

Although data on the role of race in COVID-19 illness and outcomes are limited and complex, several studies have demonstrated that there may be increased risk of infection and worse clinical outcomes among Black patients. The potential underlying reasons for these findings are important to consider and may include socioeconomic and health inequalities, resource deprivation, and pathophysiologic differences in response to infection, but further research is needed [6, 11–14].

In this study, 55.0% of patients hospitalized with COVID-19 were African American. African American patients are not overrepresented in the Explorys database overall: patients in the South are 16.0% African American and 43.0% White, and patients in the Midwest are 10.9% and 62.9%, respectively. This reflects the US population, which according to the latest Census data is 13.4% Black or African American and 60.4% White (not Hispanic or Latinx) [15].

In the present study, among patients from the South, 60.8% were African American and 27.2% were White, and in the Midwest, 31.0% were African American and 59.5% were White. Increased rates of infection observed in the African American population in this study might be due to the tendency of the COVID-19 epidemic to be more severe in certain geographic areas—in this case, in areas with a high prevalence of African Americans. It also may be due to the disease attacking poorer Americans who have essential jobs and cannot isolate or socially distance. In this study, African American patients were younger and a higher proportion were female compared with White patients; these findings are similar to those of a study of 3626 patients in Louisiana, the majority of whom self-identified as Black non-Hispanic [6]. The mean CCI was higher in African American patients than White patients, although the difference was not statistically significant. In the Louisiana study, a higher mean CCI was also observed in Black patients than White patients hospitalized with COVID-19, and racial differences in CCI have been reported among the general US population [6, 16]. The present study also observed that the proportions of African American patients with diabetes and obesity were higher than those of White patients, which is a finding similar to that in the Louisiana cohort, although that was regardless of hospitalization status. Differences in insurance plan type were observed when patients were stratified by race; however, no conclusions can be drawn because, for the majority of patients, these data were missing.

The implications of the differences in baseline characteristics between Black patients and White patients hospitalized with COVID-19 and pulmonary involvement are not yet clear. However, it appears that obesity and diabetes are risk factors for complications in COVID-19 across all races [4–6]. Several large studies have found the prevalence of obesity and diabetes in US patients with COVID-19 to be higher than that of the general US adult population, in which 39.5% of adults are obese and 10.2% have diagnosed diabetes [4–6, 10, 17]. The prevalence of these health conditions varies among demographic groups, and there are many complex factors underlying these differences, including historical, socioeconomic, and policy inequities that exert systemic effects on daily life that play many roles in health [10, 17]. In a study of > 10,000 COVID-19-related deaths in England, a multivariable model that adjusted for underlying health conditions found that greater deprivation—a socioeconomic measure including income, employment, education, health, crime, and living environment—was associated with increased risk of death and suggests that societal factors likely play a key role in COVID-19 outcomes [14]. Importantly, further studies are needed on the underlying reasons for racial disparities in COVID-19 and other health emergencies to more robustly explain these patterns, mitigate any potential misinterpretations of analyses based on race, and improve outcomes for patients [12–14].

This study has several limitations related to retrospective analysis of EHR data. Medications that patients were taking but that were not prescribed in the baseline period or were prescribed by a healthcare provider outside of the systems that contribute to Explorys were not captured. Patients were also presumed to have filled and taken medications as prescribed, but this cannot be confirmed. Diagnoses that were made by providers outside of the systems that contribute to Explorys were not captured, and conditions for which patients did not seek care were not included. Results may not be generalizable to the US population as a whole, as a majority of the patients were from the South and Midwest. Selection for those populations was not intended, as the selection criteria included nationally used ICD codes; these trends reflect the geographic distribution of the Explorys database and may reflect the nature of the COVID-19 outbreak. These large cohorts of Southern and Midwestern patients are also a strength of this study, because currently the data on baseline characteristics of patients with COVID-19 in these regions are limited. Descriptive statistics were used, and additional analyses are needed to adjust for potential confounding factors. Despite these limitations, this study provides important information from a large real-world population of US patients hospitalized with COVID-19 and pulmonary symptoms.

CONCLUSIONS

Among US patients hospitalized with COVID-19 and pulmonary involvement primarily from the South and Midwest, the most common comorbidities were hypertension, diabetes, hyperlipidemia, and obesity. The higher represented demographic groups included patients from the South, African American patients, and female patients. Compared with White patients, African American patients were younger, with higher mean BMI, higher prevalence of concurrent diabetes, and lower prevalence of COPD and smoking/tobacco use. Although this study confirms trends observed elsewhere, it is important to view these data in the context of the complex underlying reasons for racial disparities in COVID-19; further research is needed to understand the implications of these observations.

ACKNOWLEDGEMENTS

Medical Writing Assistance. Support for third-party writing assistance for this manuscript, provided by Claire Stedden, PhD, of Health Interactions, Inc., was provided by Genentech, Inc.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this manuscript, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Authorship Contributions. All authors had access to the data, provided critical feedback, and contributed to the writing of the manuscript. JHB and SVM conceived and designed the study. AMK provided subject matter expertise, and BI and KJ-P conducted programming and data analysis support. JHB, SVM, AMK, JMP, KP, JP, OWB, and JLZ contributed to the interpretation of the results.

Disclosures. Jennie H. Best, Shalini V. Mohan, and James L. Zazzali are employees and shareholders of Genentech, Inc. Amanda M. Kong, Boris Ivanov, and Krutika Jariwala-Parikh are employees of IBM Watson Health. Krish Patel declares consultation honoraria from Genentech, Inc., outside of the submitted work. John M. Pagel, Otis W. Brawley and John Pauk have nothing to disclose.

Compliance with Ethics Guidelines. This was a retrospective, observational study and all study data were deidentified and fully compliant with Health Insurance Portability and Accountability Act Regulations (45 CFR 164.514e); therefore, approval from an institutional review board was not required and informed consent was not obtained.

Data Availability. IBM Explorys is a proprietary database. The database is available from IBM Watson Health, but restrictions apply to the availability of these data and therefore they are not publicly available. Please contact AMK (akong@us.ibm.com) for information on the Explorys data.

REFERENCES

1. Phelan AL, Katz R, Gostin LO. The novel coronavirus originating in Wuhan, China: challenges for global health governance. JAMA. 2020;323:709–10.

2. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://coronavirus.jhu.edu/map.html. Accessed 2 June 2020.

3. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72,314 cases from the Chinese Center for Disease Control and Prevention. JAMA. 2020;323:1239–42.

4. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA. 2020;323:2052–9.

5. Argenziano MG, Bruce SL, Slater CL, et al. Characterization and clinical course of 1000 patients with coronavirus disease 2019 in New York: retrospective case series. BMJ. 2020;369:m1996.

6. 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;382:2534–43.

7. Centers for Disease Control and Prevention. ICD-10-CM Official Coding Guidelines-Supplement. Coding encounters related to COVID-19 Coronavirus Outbreak. Effective: February 20, 2020. https://www.cdc.gov/nchs/data/icd/ICD-10-CM-Official-Coding-Gudance-Interim-Advice-coronavirus-feb-20-2020.pdf. Accessed 3 June 2020.

8. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol. 1992;45:613–9.

9. Centers for Disease Control and Prevention. Defining adult overweight and obesity. 2020. https://www.cdc.gov/obesity/adult/defining.html. Accessed 8 July 2020.

10. Trust for America's Health. The State of Obesity: Better Policies for a Healthier America 2019. https://media.stateofobesity.org/wp-content/uploads/2019/09/16100613/2019ObesityReportFINAL.pdf. Accessed 25 June 2020.

11. Pan D, Sze S, Minhas J, et al. The impact of ethnicity on clinical outcomes in COVID-19: a systematic review. EClinicalMedicine. 2020;23:100404.

12. El-Khatib Z, Jacobs G, Ikomey G, Neogi U. The disproportionate effect of COVID-19 mortality on ethnic minorities: genetics or health inequalities? EClinicalMedicine. 2020;23:100430.

13. Chowkwanyun M, Reed AL Jr. Racial health disparities and COVID-19—caution and context. N Engl J Med. 2020;383:201–3.

14. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020;584:430–6.

15. US Census Bureau. Quick Facts, United States. Population estimates (V2019). https://www.census.gov/quickfacts/fact/table/US/PST045219. Accessed 24 June 2020.

16. Balasubramanian BA, Garcia MP, Corley DA, et al. Racial/ethnic differences in obesity and comorbidities between safety-net- and non safety-net integrated health systems. Medicine (Baltimore). 2017;96:e6326.

17. Centers for Disease Control and Prevention. National Diabetes Statistics Report 2020. https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf. Accessed 12 July 2020.