

1 Characteristics of 24,516 Patients Diagnosed with COVID-19 Illness in a National
2 Clinical Research Network: Results from PCORnet®

3 Short Title: Characteristics of 24,516 Patients Diagnosed with COVID-19 Illness

4 Jason P. Block, MD, MPH¹, Keith A. Marsolo, PhD², Kshema Nagavedu, MPH¹, L. Charles
5 Bailey, MD, PhD³, Henry Cruz, BA⁴, Christopher B. Forrest, MD, PhD³, Kevin Haynes,
6 PharmD, MSCE⁵, Adrian F. Hernandez, MD, MHS⁶, Rainu Kaushal, MD, MPH⁷, Abel Kho,
7 MD⁸, Kathleen M. McTigue, MD, MPH, MS⁹, Vinit P. Nair, BPharm, MS¹⁰, Richard Platt, MD,
8 MSc¹, Jon Puro, MPA/HA¹¹, Russell L. Rothman, MD, MPP¹², Elizabeth Shenkman, PhD¹³,
9 Lemuel Russell Waitman, PhD¹⁴, Mark G. Weiner, MD⁷, Neely Williams, EdD¹⁵, Thomas W.
10 Carton, PhD, MS¹⁶ on behalf of PCORnet® Network Partners*

11 ¹Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical
12 School, Boston, Massachusetts; ²Department of Population Health Sciences, Duke University,
13 Durham, North Carolina; ³Applied Clinical Research Center, Department of Pediatrics,
14 Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania; ⁴ PCORnet®, National Patient-
15 Centered Clinical Research Network, Washington, D.C.; ⁵HealthCore, Wilmington, Delaware;
16 ⁶Duke Clinical Research Institute, Durham, North Carolina; ⁷Department of Population Health
17 Sciences Weill Cornell Medicine, New York, New York; ⁸Center for Health Information
18 Partnerships, Northwestern University, Chicago, Illinois; ⁹Department of Medicine, University
19 of Pittsburgh, Pittsburgh, Pennsylvania; ¹⁰PRACnet, Sharon, MA; ¹¹OCHIN, Inc., Portland, OR;
20 ¹²Center for Health Services Research, Vanderbilt University Medical Center, Nashville,
21 Tennessee; ¹³Health Outcomes & Biomedical Informatics, College of Medicine, University
22 of Florida, Gainesville, Florida; ¹⁴Department of Internal Medicine, University of Kansas
23 Medical Center, Kansas City, Kansas; ¹⁵Community Partners Network Inc, Nashville, Tennessee;
24 ¹⁶Louisiana Public Health Institute, New Orleans, LA

25

26

27 *Collaborative authors are listed in an Appendix

28

29 **Address correspondence to:** Jason Block, MD, MPH, Division of Chronic Disease Research
30 Across the Lifecourse, Department of Population Medicine, Harvard Pilgrim Health Care
31 Institute, Harvard Medical School, 401 Park Drive, Boston, MA 02215, 617-867-4841,
32 jblock1@partners.org

33

34

35 Abbreviations: EHR – electronic health record, CDM – Common Data Model, ETL – extract-
36 transfer-load procedure to move data from a clinical data warehouse to the PCORnet Common
37 Data Model, HTN – hypertension, Diabetes – Type 2 Diabetes Mellitus, Pulm – pulmonary
38 disease, CKD – chronic kidney disease, CAD – coronary artery disease, CHF – congestive heart
39 failure, Coag – Coagulopathy, Cancer – Cancer other than non-melanoma skin cancer,

40 Immunosupp – use of immunosuppressive medications; Any – any care setting, Amb –
41 ambulatory care setting, IP/ED – inpatient or emergency department, Vent – inpatient or
42 emergency department with procedure codes indicating use of mechanical ventilation, PNA -
43 pneumonia

44
45 Manuscript Word Count: 3909

46 Abstract Word Count: 313

47 **Abstract**

48

49 **Background:** National data from diverse institutions across the United States are critical for
50 guiding policymakers as well as clinical and public health leaders. This study characterized a
51 large national cohort of patients diagnosed with COVID-19 in the U.S., compared to patients
52 diagnosed with viral pneumonia and influenza.

53 **Methods and Findings:** We captured cross-sectional information from 36 large healthcare
54 systems in 29 U.S. states, participating in PCORnet[®], the National Patient-Centered Clinical
55 Research Network. Patients included were those diagnosed with COVID-19, viral pneumonia
56 and influenza in any care setting, starting from January 1, 2020. Using distributed queries
57 executed at each participating institution, we acquired information for patients on care setting
58 (any, ambulatory, inpatient or emergency department, mechanical ventilator), age, sex, race,
59 state, comorbidities (assessed with diagnostic codes), and medications used for treatment of
60 COVID-19 (hydroxychloroquine with or without azithromycin; corticosteroids, anti-interleukin-
61 6 agents).

62 During this time period, 24,516 patients were diagnosed with COVID-19, with 42% in an
63 emergency department or inpatient hospital setting; 79,639 were diagnosed with viral pneumonia
64 (53% inpatient/ED) and 163,984 with influenza (41% inpatient/ED). Among COVID-19
65 patients, 68% were 20 to <65 years of age, with more of the hospitalized/ED patients in older age
66 ranges (23% 65+ years vs. 12% for COVID-19 patients in the ambulatory setting). Patients with
67 viral pneumonia were of a similar age, and patients with influenza were much younger.
68 Comorbidities were common, especially for patients with COVID-19 and viral pneumonia, with
69 hypertension (32% for COVID-19 and 46% for viral pneumonia), arrhythmias (20% and 35%),
70 and pulmonary disease (19% and 40%) the most common. Hydroxychloroquine was used in
71 treatment for 33% and tocilizumab for 11% of COVID-19 patients on mechanical ventilators
72 (25% received azithromycin as well).

73 **Conclusion and Relevance:** PCORnet leverages existing data to capture information on one of
74 the largest U.S. cohorts to date of patients diagnosed with COVID-19 compared to patients
75 diagnosed with viral pneumonia and influenza.

76 The novel coronavirus SARS-CoV-2 has resulted in unprecedented challenges to the global
77 medical and public health sectors. The virus has spread rapidly across the world, leading
78 policymakers, clinicians, first responders, and public health professionals to coordinate responses
79 with limited information about effects on diverse populations and potential treatments. Over the
80 first few months of the pandemic, data on patients with COVID-19 came from varied sources,
81 often as single-institution or regional case series or through case reports to public health
82 authorities, typically based on positive lab tests (1-5). Case reports to public health authorities
83 and popular dashboards that rely on this information often lack detailed information about
84 patient characteristics, use of treatments, or course of illness (2, 6-9).

85

86 A rapidly evolving pandemic presents tremendous uncertainty. The pattern of disease is typically
87 unknown initially, and knowledge can flow in asymmetric ways, often driven by anecdotes, press
88 reports, and social media. Throughout much of the COVID-19 pandemic, testing has been
89 insufficient, limiting knowledge about the scope of the pandemic and characteristics of patients
90 infected. Treatments have become mainstays before being fully vetted (10-12).

91

92 Leaders, institutions, researchers, and patients need extensive and accurate information about
93 characteristics of patients infected, utilization of treatments and disease course, ideally from
94 diverse settings across large geographic regions. Research networks that can capture this
95 information across many healthcare systems have an important role during a pandemic response,
96 especially because they can provide more generalizable, real-world information on a diverse
97 population. In this article, we describe characteristics of one of the largest national cohorts of
98 patients to date diagnosed with COVID-19 and compare this cohort to patients diagnosed with

99 other viral illnesses in institutions participating in PCORnet[®], the National Patient-Centered
100 Clinical Research Network.

101

102 **Methods**

103 Developed with funding by the Patient-Centered Outcomes Research Institute[®] (PCORI[®]),
104 PCORnet[®] is a network-of-networks developed to facilitate comparative effectiveness research
105 using electronic health record (EHR) and other healthcare data (13, 14). The Network currently
106 has participation from 251 medical institutions, 337 hospitals, 3,574 primary care practices, and
107 1,024 community clinics embedded within nine Clinical Research Networks (i.e., consortia of
108 health systems called PCORnet Network Partners) and 61 data contributing sites (i.e., individual
109 health systems); the Network also has participation from two Health Plan Research Networks.
110 PCORnet Network Partners (NPs) are well suited to develop and analyze cohorts of patients
111 across many of the largest health systems in the United States. While set up to accommodate
112 comparative effectiveness and epidemiologic research, the network has features that allow for a
113 transition to rapid querying of data in a medical and public health crisis. In this study on COVID-
114 19 patients, 37 of the 61 data contributing sites participated.

115

116 *PCORnet Common Data Model*

117 PCORnet operates as a distributed research network, where data are held locally behind
118 institutional firewalls, with queries sent to institutions from a central PCORnet Coordinating
119 Center and results returned in the form of aggregate counts or summary statistics. NPs also have
120 the capacity to generate patient-level data and to link electronic health records to claims data for
121 specific projects. The NPs utilize the PCORnet Common Data Model, in which each data

122 contributing site maps its clinical and/or administrative claims data into a standardized
123 representation (15). The most recent version of the PCORnet Common Data Model, v5.1,
124 provides standards for information on demographics, diagnoses, vital signs, laboratories,
125 prescriptions and dispensed medications, medications administered (multiple settings),
126 procedures, and facility and patient geographic information, among other information (15).
127 These data are normally updated quarterly by sites, and the PCORnet Coordinating Center
128 curates the data to ensure conformance with the Common Data Model, data plausibility, and data
129 completeness (16). The Coordinating Center provides metrics based on these data quality
130 domains to sites, allowing them to troubleshoot any data problems; some deficiencies require
131 mitigation before the site can participate in PCORnet queries.

132

133 Prompted by the COVID-19 pandemic, PCORnet NPs agreed to abruptly change the operation of
134 the network, allowing for rapid capture of information on patients infected with COVID-19. To
135 facilitate this, starting in early April, NPs began updating their data weekly instead of quarterly.
136 Rather than doing this for all patients (which can involve millions of patients and require
137 considerable time), PCORnet sites focused the refresh on patients that had at least one
138 International Classification of Diseases, 10th revision, Clinical Modification (ICD-10-CM)
139 diagnostic code for a respiratory illness or had a lab order for a SARS-CoV-2 test recorded
140 beginning January 1, 2020 (**Fig 1**) (17). This respiratory illness filter created a COVID-19-
141 focused Common Data Model that included all information available in the typical “full
142 population” Common Data Model but for many fewer patients. The Common Data Model
143 respiratory illness filter included ICD-10-CM codes for coronavirus (COVID-19 and other),
144 pneumonia, bronchitis, influenza, upper respiratory infections, fever, cough, respiratory failure or

145 acute respiratory distress syndrome, and shortness of breath (**S Methods Note**). The codes for
146 SARS-CoV-2 laboratory orders were Special Use terms identified via the Logical Observation
147 Identifiers Names and Codes (LOINC) Prerelease Page, as well as guidance from the Centers for
148 Medicare and Medicaid Services (CMS) on SARS-CoV-2-related Healthcare Common
149 Procedure Coding System (HCPCS) codes (18). Sites first began populating their COVID-19
150 Common Data Model during the week of March 27th. The timeline for implementation varied by
151 site based on ability to change data infrastructure and alter governance to allow for frequent
152 updates and queries.

153

154 **Fig 1. PCORnet Common Data Models, including standard and COVID-19 Common Data**
155 **Model.**

156 The PCORnet Common Data Model includes information collected through various clinical data
157 systems, including electronic health records (EHR) and health insurance claims or pharmacy
158 dispensing records. Once data are entered by the user, systems transfer data to a data warehouse
159 that is unique to their own health system and system vendor (e.g., Epic, Cerner). The typical
160 process for populating the PCORnet Common Data Model is through an extraction-transfer-
161 loading (ETL) procedure that allows for transformation of raw information into a form that is
162 compliant with the Common Data Model specifications. This transformation allows for
163 consistency across health systems and centralized querying of data, and the PCORnet ETL and
164 Common Data Model data refresh occurs quarterly. The PCORnet COVID-19 Common Data
165 Model utilizes a similar process but rather than including information on all patients, it includes
166 information only on patients who have had at least one respiratory illness diagnostic code or a
167 lab test ordered for SARS-CoV-2 since January 1, 2020. This process occurs weekly and
168 includes data with a lag time between 1 day to 7 days, depending on the site.

169

170 *Query Development Process and Cohort Characterization*

171 The PCORnet Coordinating Center has developed standardized statistical programs, called
172 modular programs, that can capture aggregate descriptive information on defined cohorts. These
173 programs define a cohort or multiple cohorts and then capture broad descriptive information,
174 including demographics, comorbidities (and symptoms based on billing codes), prescriptions or
175 hospital-administered medications, presence of laboratory results, and geographic information.

176 NPs return data as counts and frequencies for variables specified; the PCORnet Coordinating
177 Center combines results across NPs, producing a summary report of Network-wide aggregate
178 results. Submission and use of these aggregate data are covered by a Network-wide Data
179 Security Agreement.

180

181 Starting from the broader group of patients whose information was in the COVID-19 Common
182 Data Model, we defined a cohort of patients with ICD-10-CM diagnostic codes for COVID-19 or
183 other coronaviruses (our case definition for COVID-19 for this study), viral pneumonia, or
184 influenza (**S1 Table**). The case definition did not use laboratory data. We assessed patients with
185 these three infections to enable some characterization of patients with other infections relevant to
186 COVID-19 as comparison cohorts. The viral pneumonia and influenza cohorts excluded patients
187 who had COVID-19 (or other coronavirus) ICD-10-CM codes. We stratified our cohort based on
188 these three infections as well as the setting of diagnosis: 1) any setting, 2) ambulatory, 3)
189 inpatient or emergency department, or 4) inpatient or emergency department with procedure
190 codes identifying use of mechanical ventilation. Patients could be included in multiple care
191 settings, if patients received diagnostic codes in more than one. We identified mechanical
192 ventilation usage through ICD-10-CM, HCPCS and Current Procedural Terminology (CPT)
193 codes. All codes for case definitions, comorbidities, treatments and labs were posted on GitHub
194 ([https://github.com/PCORnet-DRN-OC/Query-
195 Details/tree/master/COVID_weekly_query/2020_04_29](https://github.com/PCORnet-DRN-OC/Query-Details/tree/master/COVID_weekly_query/2020_04_29)).

196

197 *Comorbidities, Treatments and Symptoms*

198 We captured aggregate data across NPs for demographic characteristics of patients, including
199 age group, sex, and race, using value sets defined in the PCORnet Common Data Model v5.1.
200 Sex categories allowed for Male, Female, and Other. For race, several categories, including
201 Native Hawaiian/Other Pacific Islander and Alaskan Native/American Indian, were less common
202 and typically <1% for all cohorts.

203

204 We included a detailed assessment of comorbidities. We chose these based on discussions with
205 experts regarding the comorbidities that might affect outcomes with COVID-19, as well as
206 information presented in early publications regarding COVID-19 outcomes. Patients were
207 considered to have a disease/condition if they had at least one ICD-10-CM code for the
208 disease/condition recorded any time over a three-year period prior to the date of their index
209 diagnosis. We required only one code to account for the possibility that some patients had their
210 first encounter with the healthcare system for their admission for COVID-19 and may only have
211 one code listed for a comorbidity. Using RxNorm and National Drug Codes (NDC) lists, we
212 identified use of immunosuppressive medications over the same three-year period (19, 20).
213 Similarly, we captured information on medications being used for treatment, with an assessment
214 of prescriptions given in the 14-day period before and after the index diagnosis. These included
215 hydroxychloroquine and chloroquine, alone or in combination with azithromycin, as well as
216 corticosteroids and anti-IL-6 agents (siltuximab, sarilumab, tocilizumab) used to treat the
217 cytokine storm that can occur with severe COVID-19. We also captured use of critical care
218 procedure/billing CPTs (99291, 99292) during this period and the presence of laboratory tests,
219 identified with LOINC codes, that are commonly used in the assessment of COVID-19. Lastly,
220 we categorized weight status for each cohort, using the most recent height and weight available,

221 up to one year prior to the index diagnosis date. We calculated measures separately by age group,
222 using body mass index (BMI) for adults and Centers for Disease Control and Prevention macros
223 for calculation of BMI z-scores and weight status for children < 20 years of age (21).

224

225 **Results**

226 The query captured aggregate data on 24,516 patients with a COVID-19 diagnosis from 36 data
227 contributing sites, across 29 states; 37 sites responded to the query, but 1 site had no cases
228 reported (**S2 Table**). Across the 36 sites, percent contributions to the overall COVID-19
229 population ranged from <1% to 14% with 11 sites each contributing $\geq 4\%$; these 11 sites
230 included 70% of all COVID-19 patients. The most commonly represented states were Illinois
231 (18%), Wisconsin (11%), and Florida (9%), with 18% of patients missing geographic
232 information. Among patients diagnosed with COVID-19, 42% were diagnosed in an emergency
233 department or inpatient hospital setting, and the most commonly used codes for the diagnosis of
234 COVID-19 were B34.2 (19%), B97.29 (33%), and U07.1 (62%). A lab test for SARS-CoV-2
235 was present for 68% of patients diagnosed (**S1 Figure**). For comparison conditions, the query
236 identified 79,639 diagnosed with viral pneumonia (53% in hospital/ED) and 163,984 with
237 influenza (41% in hospital/ED). Labs commonly used in the assessment of patients with COVID-
238 19 were variably available in the PCORnet Common Data Model, with creatinine, alanine
239 aminotransferase and aspartate aminotransferase, and lymphocyte count the most common.

240

241 Most COVID-19 patients were diagnosed in March and April 2020 (25% and 53% respectively),
242 in contrast to viral pneumonia and influenza patients, who were most often diagnosed in January
243 and February (64% and 84% respectively). Only 0.4% of influenza cases were diagnosed in

244 April/May 2020 (**Table 1**). Among COVID-19 patients, 68% were 20 to <65 years of age, with
245 more of the hospitalized/ED patients in older age ranges (23% 65+ years compared to 12% for
246 COVID-19 patients in the ambulatory setting) (**Table 1**). Patients with influenza were younger
247 than patients with COVID-19 or viral pneumonia (63% hospitalized/ED with influenza were <20
248 years of age compared to 16% for COVID-19 and 21% for viral pneumonia). More female
249 patients (55%) received a COVID-19 diagnosis overall, with a predominance of men among
250 those on a mechanical ventilator (57%).

251

252 Patients with COVID-19 were more likely to have black or African American race (25%) than
253 patients with influenza (18%) or viral pneumonia (15%). Among hospitalized/ED patients, the
254 proportion of patients with black or African American race was similar for COVID-19 and
255 influenza patients (29% and 28%) but higher than for patients with viral pneumonia (18%).
256 However, COVID-19 (16%) and influenza (15%) patients were more commonly categorized as
257 having “missing” or “other” race than those with viral pneumonia (8%).

258

Table 1: Characteristics of patients diagnosed with COVID-19, viral pneumonia and influenza, January 1 – May 18, 2020, 36 PCORnet Data Contributing Institutions

	COVID Diagnosis Any Care Setting		COVID Diagnosis in an Ambulatory Setting		COVID Diagnosis in an Inpatient or Emergency Department Setting		COVID Diagnosis in an Inpatient or Emergency Department Setting AND Mechanical Ventilation		Viral Pneumonia Diagnosis AND NO COVID Diagnosis in an Inpatient or Emergency Department Setting		Influenza Diagnosis and No COVID Diagnosis in an Inpatient or Emergency Department Setting	
	24,516		9,305		10,351		921		42,100		66,558	
	N	%	N	%	N	%	N	%	N	%	N	%
Age												
<i>0-<2</i>	1,087	4%	196	2%	770	7%	60	7%	1,968	5%	9,018	14%
<i>2-<10</i>	976	4%	309	3%	544	5%	44	5%	4,742	11%	24,207	36%
<i>10-<20</i>	1,105	5%	494	5%	457	4%	33	4%	2,003	5%	8,647	13%
<i>20-<45</i>	8,890	36%	3,960	43%	2,864	28%	125	14%	6,650	16%	13,173	20%
<i>45-<65</i>	7,918	32%	3,101	33%	3,249	31%	309	34%	10,966	26%	7,063	11%

65-<75	2,530	10%	778	8%	1,287	12%	197	21%	7,088	17%	2,517	4%
75-<85	1,329	5%	313	3%	764	7%	109	12%	5,544	13%	1,362	2%
85+	644	3%	107	1%	410	4%	22	2%	3,103	7%	540	1%
Sex												
<i>Female</i>	13,447	55%	5,255	56%	5,231	51%	381	41%	20,828	49%	34,126	51%
<i>Male</i>	10,965	45%	3,970	43%	5,108	49%	525	57%	21,219	50%	32,318	49%
<i>Other¹</i>	104	0%	80	1%	12	0%	0	0%	52	0%	114	0%
Race												
<i>Native Hawaiian or Other Pacific Islander</i>	133	1%	55	1%	70	1%	0	0%	192	0%	438	1%
<i>American Indian or Alaska Native</i>	110	0%	35	0%	62	1%	0	0%	176	0%	215	0%
<i>Asian</i>	668	3%	234	3%	315	3%	27	3%	829	2%	2,061	3%
<i>Black or African American</i>	6,148	25%	1,729	19%	3,002	29%	236	26%	7,690	18%	18,911	28%

<i>White</i>	12,523	51%	5,053	54%	5,159	50%	492	53%	29,174	69%	33,465	50%
<i>Multiple Race</i>	291	1%	104	1%	107	1%	0	0%	506	1%	1,747	3%
<i>Other¹/Missing</i>	4,621	19%	2,071	22%	1,607	16%	119	13%	3,488	8%	9,706	15%
Month²												
<i>January 2020</i>	1,193	5%	153	2%	937	9%	90	10%	13,490	32%	26,717	40%
<i>February 2020</i>	828	3%	105	1%	664	6%	65	7%	11,553	27%	28,087	42%
<i>March 2020</i>	6,240	25%	1,449	16%	2,925	28%	358	39%	10,937	26%	11,542	17%
<i>April 2020</i>	13,070	53%	5,796	62%	4,735	46%	365	40%	5,363	13%	179	0%
<i>May 2020</i>	3,162	13%	1,783	19%	1,082	10%	21	2%	757	2%	19	0%

*Columns may not add to the total N of patients due to some missing data that is not categorized separately as missing. Further, small cell sizes with <11 patients are suppressed, leading to some differences in totals and percent values.

¹ "Other" includes all remaining PCORnet CDM values not specified by this request.

² For by year-month counts, the reported period represents the month and year in which the health event of interest first occurred within the query period. Data for May 2020 only includes information through the middle of the; the query to acquire this data was circulated on May 7th and ended on May 18th.

259
260
261 Comorbidities were common among all infections. Among COVID-19 patients, hypertension
262 was the most common comorbidity (32% for all patients) followed by arrhythmias (20%),
263 pulmonary disease (19%), type 2 diabetes (17%), anemia (14%), and asthma (12%). These
264 diagnoses were more common among COVID-19 patients being treated in the inpatient or
265 emergency department setting (**Fig 2A**). Compared to patients with viral pneumonia treated in
266 the inpatient/ED setting, COVID-19 patients in this same care setting had fewer comorbidities
267 (arrhythmia: 33% vs. 45%; pulmonary disease: 24% vs. 43%; congestive heart failure: 13% vs.
268 26%; coronary artery disease: 13% vs. 24%); patients with influenza in the inpatient/ED setting
269 had much lower prevalence of comorbidities, likely, in large part, because patients were younger
270 (arrhythmia: 15%; pulmonary disease: 21%; congestive heart failure 4%; coronary artery
271 disease: 4%) (**Fig 2B**). Asthma rates were similar across all three disease groups (14% to 19%).
272 Obesity was more common among COVID-19 patients <65 years old hospitalized or in ED (40%
273 among 20-44-year-olds; 45% among 45-64-year-olds; 36% among 65-74-year-olds) compared to
274 patients with viral pneumonia (32%, 38%, 36%, respectively) and influenza (26%, 37%, 35%,
275 respectively) (**S3 Table**).

276
277 **Fig 2. Comorbidities for Patients with COVID-19 Across Care Settings and Compared to**
278 **Patients with Viral Pneumonia and Influenza in the Inpatient/Emergency Department**
279 **Setting.**

280 These figures show prevalence of various important comorbidities for COVID-19 infections
281 across care settings (A), including all patients and those diagnosed in ambulatory settings,
282 inpatient/emergency department, or using mechanical ventilation. In (B), we compared patients
283 with COVID-19, viral pneumonia, and influenza among those who were hospitalized or in the
284 emergency department. Care settings are not mutually exclusive. Patients could be in all care
285 settings, such as if a patient receives a COVID-19 diagnosis in each of the ambulatory and
286 inpatient/emergency department settings.

287

288 Among patients hospitalized or in the ED, critical care procedure/billing codes appeared for 13%
289 of patients diagnosed with COVID-19 compared to 16% of those with viral pneumonia and 2%
290 with influenza. Of treatments assessed, corticosteroids and hydroxychloroquine were used
291 commonly for COVID-19 patients in the inpatient/emergency department setting (12% and 15%,
292 respectively); 10% of patients in this setting used the combination of hydroxychloroquine and
293 azithromycin (**Table 2**). Chloroquine use was very infrequent. Tocilizumab was used most often
294 among patients on mechanical ventilators (11%) as was hydroxychloroquine (33%), with rare
295 use of other anti-interleukin-6 agents. Hydroxychloroquine or the combination of it with
296 azithromycin was used for 1% or fewer of the patients hospitalized or in the ED with viral
297 pneumonia or influenza.

298

299

Table 2: Treatments used for COVID-19 and Other Infections

	COVID Any		COVID Ambulatory		COVID Emergency Department /Inpatient		COVID Ventilator		Viral Pneumonia Emergency Department /Inpatient		Influenza Emergency Department /Inpatient	
Hydroxychloroquine	1,915	8%	257	3%	1,514	15%	302	33%	618	1%	94	0%
Hydroxychloroquine +Azithromycin*	1,250	5%	125	1%	993	10%	233	25%	326	1%	26	0%
Corticosteroids	2,270	9%	354	4%	1,202	12%	198	21%	10,259	24%	5,074	8%
Tocilizumab	346	1%	12	0.001%	284	3%	101	11%	16	0.0004%	0	0%

*Not mutually exclusive; patients on hydroxychloroquine and azithromycin are also counted in the hydroxychloroquine row.

300 **Discussion**

301 In this article, we describe one of the largest cohorts reported to date, using information on
302 24,516 patients diagnosed with COVID-19 across the U.S. Unlike most other studies, we
303 compare patients across care settings and describe patients diagnosed with viral pneumonia
304 (N=79,639) and influenza (N=163,984) during the COVID-19 pandemic. Comorbidities were
305 common among patients diagnosed with COVID-19, especially those cared for in the emergency
306 department and inpatient setting; prevalence of comorbidities was more higher among patients
307 diagnosed with viral pneumonia and lower among those with influenza. Treatments, such as
308 hydroxychloroquine were commonly used among patients diagnosed with COVID-19. Among
309 patients on mechanical ventilators, 1/3 of patients were using hydroxychloroquine and 1/4 using it
310 in combination azithromycin. Obtaining real-world, comprehensive information on a diverse,
311 national population of patients with COVID-19 is critical to help guide policymakers as well as
312 clinical and public health leaders regarding planning for future COVID-19 response. These
313 results provide information on the high burden of chronic disease among COVID-19 patients,
314 compared to those with other similar infections, especially among patients requiring ED visits or
315 hospitalization. The depth and breadth of the PCORnet data infrastructure will facilitate wide-
316 ranging observational and interventional research, with strong generalizability.

317

318 Several prior studies have assembled and characterized cohorts of COVID-19 patients mostly
319 using lab test results and extracts from medical records or case reports. These studies originated
320 in China and have emerged from various hotspots across the world as the pandemic has spread.
321 One of the first reports included data on 1099 hospitalized patients across 30 Chinese provinces,
322 representing 14% of patients who had been hospitalized in 552 hospitals at the time. In the

323 cohort, mean age was 47 years, and 24% of patients had an underlying health condition, of which
324 hypertension (15%) and diabetes (7%) were the most common. Patients with more severe disease
325 were more likely to have an underlying health condition (39%) (22). Underscoring the need for
326 large cohorts over diverse geographic regions, subsequent reports from China have found wide
327 variation in characteristics of patients infected with COVID-19 (23). For example, in one study
328 of hospitalized patients across China, patients in Hubei province, the epicenter of the outbreak,
329 were older and had a higher burden of comorbidities (33%) compared to patients outside of
330 Hubei (20%) (24). Diabetes rates were 12% vs. 5%, and hypertension was 24% vs. 12%.

331

332 Studies in the United States began with single case reports or case series in Washington state (1,
333 25) and have grown to include large collections of patients, typically within single healthcare
334 systems or from case reports to public health authorities. In the Northwell healthcare system in
335 New York State, investigators captured data on 5700 inpatients, mostly hospitalized in March
336 2020, from their EHR data warehouse. Median age of patients was 63 years, 23% were black or
337 African American, and 40% were female; 57% had hypertension, 42% obesity (severe obesity
338 19%), and 34% diabetes. Among patients, 14% received treatment in an intensive care unit, and
339 12% had received mechanical ventilation (26).

340

341 Research and surveillance networks have begun reporting on COVID-19. The Observational
342 Health Data Sciences and Informatics (OHDSI) group, which manages the Observational
343 Medical Outcomes Partnership (OMOP) Common Data Model, has published data characterizing
344 patients with COVID-19 compared to influenza. OHDSI uses a similar process to PCORnet, with
345 queries that do not require sharing of patient-level data (27). OHDSI network partners

346 contributed information on COVID-19 patients hospitalized in South Korea (captured from a
347 national health insurance claims database) and in three US healthcare systems. Of 6806 patients
348 (75% of them from South Korea) identified based on positive SARS-CoV-2 lab tests, the most
349 common chronic conditions were hypertension (22% to 70% across sites), anemia (14% to 30%),
350 and diabetes (18% to 43%), with the highest burden of chronic disease by far among the 577
351 patients cared for in Department of Veterans Affairs facilities. In contrast to our study, patients
352 in this OHDSI cohort hospitalized with influenza from 2014 to 2019 were older and had a higher
353 burden of comorbidities than those with COVID-19.

354

355 Data on children with COVID-19 are sparse. Among 2572 children under 18 years reported to
356 the Centers for Disease Control and Prevention, 345 children had information on underlying
357 health conditions; 23% of these children had documentation of at least one, most of which were
358 asthma or cardiovascular disease (8). Our study has information available for 3168 children
359 under 20 years of age, 56% of whom were hospitalized; in future iterations of our work, we will
360 have the capacity to report separately on the specific comorbidity burden among children
361 diagnosed with COVID-19 compared to other viral illnesses.

362

363 COVID-19 surveillance in PCORnet provides several unique opportunities. The network
364 infrastructure and large samples can allow for creation of research-ready cohorts and in-depth
365 assessment of patients, based on demographic or clinical characteristics. Analyses can be done in
366 a distributed fashion without sharing of patient-level data, including predictive modeling, and
367 aggregation across diverse sites can increase generalizability. Large sample sizes with granular
368 clinical information will be increasingly important to capture information on patient subgroups,

369 defined by demographics or less common presentations or outcomes of COVID-19, such as
370 emerging reports of neurologic complications or rare complications in children (28). PCORnet
371 will also facilitate observational research and clinical trials on various treatments; even in this
372 early phase of data capture, we have identified thousands of patients receiving treatments, such
373 as hydroxychloroquine. By continuously updating and refining COVID-19 data, PCORnet will
374 allow for contemporaneous comparisons to patients diagnosed with other similar illnesses, such
375 as viral pneumonia and influenza in this study as well as concurrent trials; soon, we will compare
376 patients who have “positive” and “negative” SARS-CoV-2 tests. The investment in changing the
377 processes for data management – moving to a weekly update for patients with respiratory illness
378 from a quarterly update – set up the network to make ongoing contributions to the investigation
379 of COVID-19.

380
381 In contrast, PCORnet does not allow for representative estimates of overall COVID-19 cases
382 across the U.S. Because PCORnet has a fixed number of institutions and is not designed to be
383 representative of the U.S. population, the prevalence of cases across states in PCORnet is not
384 equivalent to the actual state prevalence of COVID-19 patients. Also, in this early stage of
385 exploring COVID-19 in PCORnet, the case definition depended on diagnostic codes. Patients
386 actively diagnosed with COVID-19, which in our case included codes for other coronaviruses,
387 likely differ from those who test positive for SARS-CoV-2. Diagnostic codes are less sensitive
388 than lab test results though perhaps more specific, considering that many patients with positive
389 lab tests are asymptomatic and do not have the syndrome of COVID-19. Sensitivity of diagnostic
390 codes for COVID-19 was likely much lower for patients who were diagnosed before an ICD-10-
391 CM code was available for COVID-19 (April 1, 2020) (29). Prior to that time, the Centers for

392 Disease Control and Prevention provide interim guidance on use of existing coronavirus codes
393 that were not specific to COVID-19 (30). The presence of the U07.1 code already has led to
394 more case identification, with most of the cases identified in April and May using this code.
395 Because of the uncertainty of diagnostic codes in the early phase of the pandemic, some patients
396 with other coronavirus infections would have been misclassified as having COVID-19 in our
397 study.

398

399 *Limitations*

400 Several limitations are inherent to working with structured healthcare data. First, electronic
401 health records may not have complete data on some patients, especially patients diagnosed in
402 virtual visits or who were diagnosed and treated entirely outside of a PCORnet health system,
403 including in skilled nursing facilities or nursing homes. In such cases, the PCORnet Common
404 Data Model may not have any information about a COVID-19 infection or inadequate
405 information on chronic diseases and medication use. We separately captured information on
406 patients who were seen in the healthcare system prior to their hospitalization or emergency
407 department visit for COVID-19; among these patients, prevalence of comorbidities was slightly
408 higher when compared to all patients seen in the emergency department or hospitalized. Missing
409 data can also be an issue for variables such as race. Some medications being used in research
410 protocols also were likely missing because they are not captured in routine medication
411 administration tables. Second, because the research network was set up to accommodate
412 quarterly refreshes of clinical data, we quickly changed protocols to allow for weekly updates.
413 The standard quarterly data refreshes undergo detailed curation to ensure that data are compliant
414 with standards of the Common Data Model, and such efforts are not possible with weekly

415 refreshes. Erroneous records may exist that will only be identified later through data curation.
416 The PCORnet Coordinating Center will soon implement regular data curation for the COVID-19
417 Common Data Model. Third, during these initial queries, we categorized weight status by age
418 group, but we did not fully separate children or adults for the determination of comorbidities and
419 treatments; this led to lower prevalence of comorbidities for patients with influenza. Future
420 queries will separate categorize adults and children. Fourth, overlap likely existed between
421 disease groups. Patients diagnosed with viral pneumonia or influenza might have had COVID-19
422 but either did not receive a test or were misclassified based on diagnostic coding. Only 12% of
423 patients with viral pneumonia and 1% with influenza had a recorded SARS-CoV-2 test,
424 compared to 68% for patients with COVID-19. These SARS-CoV-2 test results are likely an
425 undercount; some sites have not yet fully loaded their test results into PCORnet data models, and
426 others might not have information on tests done outside of the health system, such as in drive-
427 thru testing facilities or private labs.

428

429 *Conclusion and Next Steps in Development*

430 PCORnet has captured information on one of the largest cohorts to date of patients diagnosed
431 with COVID-19 and other respiratory illnesses. Patients diagnosed with COVID-19 have a high
432 burden of chronic disease, and treatments are common. We have updated data infrastructure to
433 allow for weekly updates and queries of patients, with COVID-19 case definitions initially based
434 on diagnostic codes. We will soon identify and characterize cohorts based on lab results, along
435 with additional improvements such as identifying patients based on institutional registries,
436 improved definitions of care settings, and detailed stratifications by patient characteristics.

437

438 **Collaborative Authors: PCORnet[®] Network Partners**

439 H. Timothy Bunnell, PhD, Nemours Children’s Health System; Ryan Carnahan, PharmD, MS,
440 The University of Iowa College of Public Health, Department of Epidemiology; Alanna M.
441 Chamberlain, PhD, MPH, Department of Health Sciences Research, Mayo Clinic; Dimitri A
442 Christakis, MD MPH, Center for Child Health, Behavior & Development, Seattle Childrens
443 Research Institute, University of Washington; Lindsay G. Cowell, MS, PHD, UT Southwestern
444 Medical Center; Janis L. Curtis, MSPH, MA, Duke University School of Medicine; Julia A.
445 Fearrington, Harvard Pilgrim Health Care Institute; Soledad A. Fernandez, PhD, The Ohio State
446 University; David A Hanauer, MD, MS, University of Michigan; Kristin Herman, Tallahassee
447 Memorial Healthcare, Inc; Rachel Hess, MD, MS, Departments of Population Health Sciences
448 and Internal Medicine, University of Utah School of Medicine; Philip Giordano, MD, Orlando
449 Health, Inc.; Wenke Hwang, PhD, Penn State University College of Medicine; Adam Lee, MBA,
450 NC TraCS Institute, University of North Carolina, Chapel Hill; Harold Lehmann MD
451 PhD, Health Science Informatics, Johns Hopkins School of Medicine; Simon Lin, MD, MBA,
452 Nationwide Children's Hospital; Scott Mackey, DO, LCMC Health; Narayana Mazumber, MS,
453 Allina Health; James McClay, MD, MS, FACEP, FAMIA, University of Nebraska Medical
454 Center; Abu Saleh Mohammad Mosa, PhD, MS, FAMIA, University of Missouri School of
455 Medicine; Jihad S. Obeid, MD, FAMIA, Biomedical Informatics Center, Department of Public
456 Health Sciences, Medical University of South Carolina; Brian Ostasiewski, B.S., Wake Forest
457 School of Medicine; Nathan M. Pajor, MD, MS, Division of Pulmonary Medicine, Cincinnati
458 Children’s Hospital Medical Center, and Departments of Pediatrics, University of Cincinnati;
459 Aruradha Paranjape, MD, MPH, Lewis Katz School of Medicine at Temple University; Suchitra
460 Rao, MBBS, MSCS, University of Colorado School of Medicine and Children's Hospital
461 Colorado; Pedro Rivera, OCHIN; Patricia S. Robinson, PhD, APRN, NE-BC, AdventHealth;
462 Marc Roseman, Ann & Robert H. Lurie Children’s Hospital of Chicago; Steffani Roush,
463 Marshfield Clinic Health System; Jonathan C. Silverstein, MD, MS, Department of Biomedical
464 Informatics, University of Pittsburgh School of Medicine; Alexander Stoddard, MS,CTSI,
465 Medical College of Wisconsin; William E. Trick, Center for Health Equity & Innovation, Cook
466 County Health; Katherine J Worley, MLIS, Vanderbilt University Medical Center
467

468 **Acknowledgement**

469 The Clinical Research Networks, Health Plan Research Networks, and Coordinating Center
470 reported in this publication are Network Partners in PCORnet, the National Patient-Centered
471 Clinical Research Network. PCORnet has been developed with funding from the Patient-
472 Centered Outcomes Research Institute (PCORI). The Network Partners’ participation in
473 PCORnet is funded through the following PCORI Awards: Coordinating Center (PCORI-CC2-
474 Duke-2016); ADVANCE (RI-CRN-2020-001); CaPriCORN (RI-CRN-2020-002); GPC (RI-
475 CRN-2020-003); HealthCore (HP-1510-32545); Insight (RI-CRN-2020-004); OneFlorida (RI-
476 CRN-2020-005); PaTH (RI-CRN-2020-006); PEDSnet (RI-CRN-2020-007); PRACnet (HP-
477 1511-32713); REACHnet (RI-CRN-2020-008); and, STAR (RI-CRN-2020-009). Authors report
478 no competing interests associated with this publication.

479

480 **References**

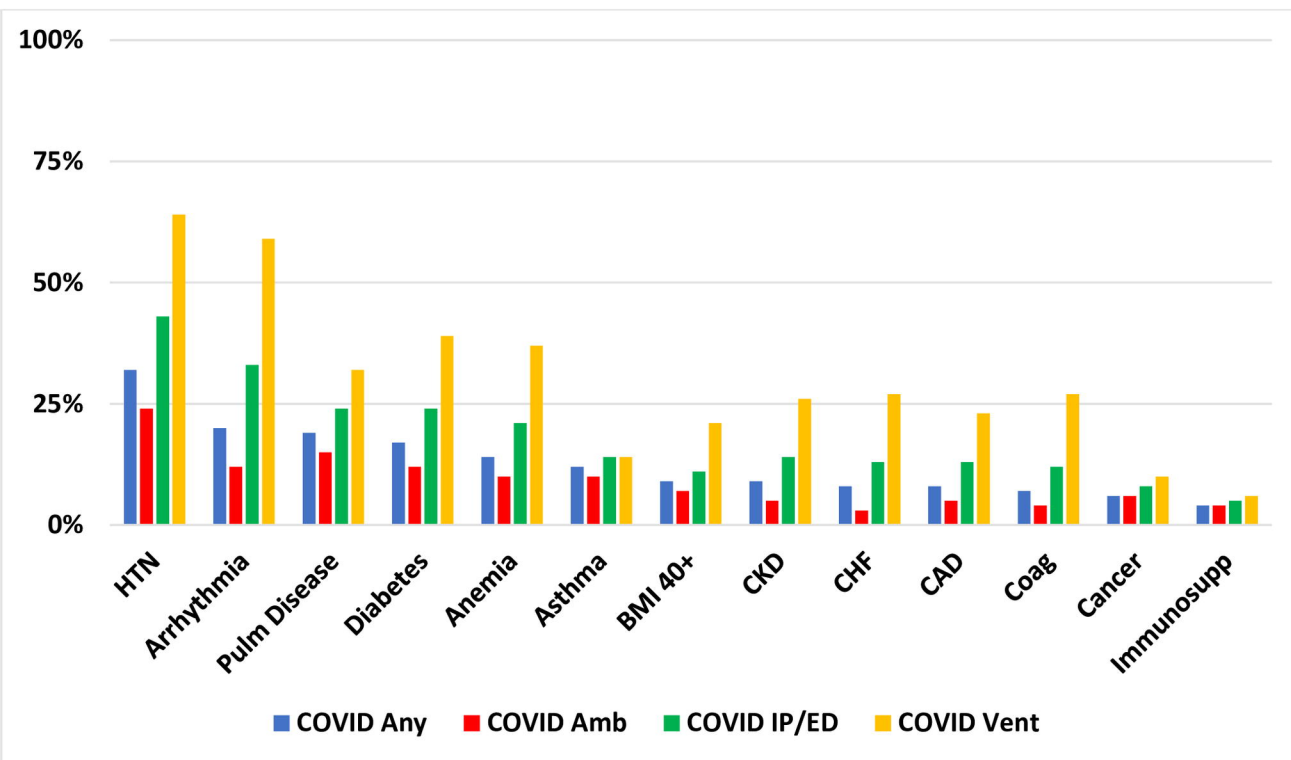
481

- 482 1. Bhatraju PK, Ghassemieh BJ, Nichols M, Kim R, Jerome KR, Nalla AK, et al. Covid-19
483 in Critically Ill Patients in the Seattle Region - Case Series. *The New England journal of*
484 *medicine*. 2020.
- 485 2. Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al.
486 Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed
487 Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. *MMWR Morbidity and*
488 *mortality weekly report*. 2020;69(15):458-64.
- 489 3. Goyal P, Choi JJ, Pinheiro LC, Schenck EJ, Chen R, Jabri A, et al. Clinical
490 Characteristics of Covid-19 in New York City. *The New England journal of medicine*. 2020.
- 491 4. Myers LC, Parodi SM, Escobar GJ, Liu VX. Characteristics of Hospitalized Adults With
492 COVID-19 in an Integrated Health Care System in California. *Jama*. 2020.
- 493 5. Nikpouraghdam M, Jalali Farahani A, Alishiri G, Heydari S, Ebrahimnia M, Samadinia
494 H, et al. Epidemiological characteristics of coronavirus disease 2019 (COVID-19) patients in
495 IRAN: A single center study. *Journal of clinical virology : the official publication of the Pan*
496 *American Society for Clinical Virology*. 2020;127:104378.
- 497 6. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United
498 States, February 12-March 16, 2020. *MMWR Morbidity and mortality weekly report*.
499 2020;69(12):343-6.
- 500 7. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions
501 Among Patients with Coronavirus Disease 2019 - United States, February 12-March 28, 2020.
502 *MMWR Morbidity and mortality weekly report*. 2020;69(13):382-6.
- 503 8. Coronavirus Disease 2019 in Children - United States, February 12-April 2, 2020.
504 *MMWR Morbidity and mortality weekly report*. 2020;69(14):422-6.
- 505 9. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in
506 real time. *The Lancet Infectious diseases*. 2020;20(5):533-4.
- 507 10. DeJong C, Wachter RM. The Risks of Prescribing Hydroxychloroquine for Treatment of
508 COVID-19-First, Do No Harm. *JAMA internal medicine*. 2020.
- 509 11. Mercurio NJ, Yen CF, Shim DJ, Maher TR, McCoy CM, Zimetbaum PJ, et al. Risk of QT
510 Interval Prolongation Associated With Use of Hydroxychloroquine With or Without
511 Concomitant Azithromycin Among Hospitalized Patients Testing Positive for Coronavirus
512 Disease 2019 (COVID-19). *JAMA cardiology*. 2020.
- 513 12. Borba MGS, Val FFA, Sampaio VS, Alexandre MAA, Melo GC, Brito M, et al. Effect of
514 High vs Low Doses of Chloroquine Diphosphate as Adjunctive Therapy for Patients
515 Hospitalized With Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection:
516 A Randomized Clinical Trial. *JAMA network open*. 2020;3(4):e208857.
- 517 13. Collins FS, Hudson KL, Briggs JP, Lauer MS. PCORnet: turning a dream into reality.
518 *Journal of the American Medical Informatics Association : JAMIA*. 2014;21(4):576-7.
- 519 14. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet,
520 a national patient-centered clinical research network. *Journal of the American Medical*
521 *Informatics Association : JAMIA*. 2014;21(4):578-82.
- 522 15. The National Patient-Centered Clinical Research Network (PCORnet), Data-Driven
523 Common Model. 2020 [Available from: <https://pcornet.org/data-driven-common-model/>].

- 524 16. Qualls LG, Phillips TA, Hammill BG, Topping J, Louzao DM, Brown JS, et al.
525 Evaluating Foundational Data Quality in the National Patient-Centered Clinical Research
526 Network (PCORnet(R)). EGEMS (Washington, DC). 2018;6(1):3.
- 527 17. International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-
528 CM) 2020 [Available from: <https://www.cdc.gov/nchs/icd/icd10cm.htm>].
- 529 18. LOINC prerelease terms 2020 [Available from: <https://loinc.org/prerelease/>].
- 530 19. RxMix: creating applications from NLM Drug APIs 2020 [Available from:
531 <https://mor.nlm.nih.gov/RxMix/>].
- 532 20. First Databank. Functional classification system. 2020 [Available from:
533 <http://www.fdbhealth.com/>].
- 534 21. Kuczumski RJ, Ogden CL, Grummer-Strawn LM, Flegal KM, Guo SS, Wei R, et al.
535 CDC growth charts: United States. Advance Data. 2000(314):1-27.
- 536 22. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of
537 Coronavirus Disease 2019 in China. The New England journal of medicine. 2020;382(18):1708-
538 20.
- 539 23. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, et al. Clinical characteristics of
540 coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. The
541 Journal of infection. 2020.
- 542 24. Liang WH, Guan WJ, Li CC, Li YM, Liang HR, Zhao Y, et al. Clinical characteristics
543 and outcomes of hospitalised patients with COVID-19 treated in Hubei (epicenter) and outside
544 Hubei (non-epicenter): A Nationwide Analysis of China. The European respiratory journal.
545 2020.
- 546 25. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of
547 2019 Novel Coronavirus in the United States. The New England journal of medicine.
548 2020;382(10):929-36.
- 549 26. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al.
550 Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized
551 With COVID-19 in the New York City Area. Jama. 2020.
- 552 27. Burn E, You SC, Sena AG, Kostka K, Abedtash H, Abrahao MTF, et al. An international
553 characterisation of patients hospitalised with COVID-19 and a comparison with those previously
554 hospitalised with influenza. Pre-print on medRxiv. 2020 [Available from:
555 <https://www.medrxiv.org/content/10.1101/2020.04.22.20074336v1.full.pdf>].
- 556 28. Helms J, Kremer S, Merdji H, Clere-Jehl R, Schenck M, Kummerlen C, et al. Neurologic
557 Features in Severe SARS-CoV-2 Infection. The New England journal of medicine. 2020.
- 558 29. ICD-10-CM official coding and reporting guidelines April 1, 2020 through September
559 30, 2020 [Available from: <https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf>].
- 560 30. ICD-10-CM official coding guidelines - supplement coding encounters related to
561 COVID-19 coronavirus outbreak 2020 [Available from: [https://www.cdc.gov/nchs/data/icd/ICD-
562 10-CM-Official-Coding-Gudance-Interim-Advice-coronavirus-feb-20-2020.pdf](https://www.cdc.gov/nchs/data/icd/ICD-10-CM-Official-Coding-Gudance-Interim-Advice-coronavirus-feb-20-2020.pdf)].

563

A)



B)

