Burdens of post-acute sequelae of COVID-19 by severity of acute infection, demographics and health status

Yan Xie [1,2,3], Benjamin Bowe[1,2,3] & Ziyad Al-Aly [1,2,4,5,6 ✉]

The Post-Acute Sequelae of SARS-CoV-2 infection (PASC) have been characterized; however, the burden of PASC remains unknown. Here we used the healthcare databases of the US Department of Veterans Affairs to build a cohort of 181,384 people with COVID-19 and 4,397,509 non-infected controls and estimated that burden of PASC—defined as the presence of at least one sequela in excess of non-infected controls—was 73.43 (72.10, 74.72) per 1000 persons at 6 months. Burdens of individual sequelae varied by demographic groups (age, race, and sex) but were consistently higher in people with poorer baseline health and in those with more severe acute infection. In sum, the burden of PASC is substantial; PASC is non-monolithic with sequelae that are differentially expressed in various population groups. Collectively, our results may be useful in informing health systems capacity planning and care strategies of people with PASC.

[1] Clinical Epidemiology Center, Research and Development Service, VA Saint Louis Health Care System, Saint Louis, MO, USA. [2] Veterans Research and Education Foundation of Saint Louis, Saint Louis, MO, USA. [3] Department of Epidemiology and Biostatistics, College for Public Health and Social Justice, Saint Louis University, Saint Louis, MO, USA. [4] Department of Medicine, Washington University School of Medicine, Saint Louis, MO, USA. [5] Institute for Public Health, Washington University in Saint Louis, Saint Louis, MO, USA. [6] Nephrology Section, Medicine Service, VA Saint Louis Health Care System, Saint Louis, MO, USA. ✉email: zalaly@gmail.com

Emerging reports suggest that beyond the acute illness, some COVID-19 survivors experience myriad clinical abnormalities lasting well beyond the first 30 days of infection[1–7]. We recently developed a high-dimensional approach to comprehensively and systematically characterize the post-acute sequelae of SARS-CoV-2 infection (PASC)—also referred to as post-COVID-19 syndrome, or simply long COVID[1]. Our approach identified 33 sequelae in the pulmonary and several extrapulmonary organ systems including nervous system and neurocognitive disorders, mental health disorders, metabolic disorders, cardiovascular disorders, gastrointestinal disorders, and several other clinical manifestations[1].

However, significant knowledge gaps remain. Specifically, what is the burden of PASC—defined as having at least one post-acute sequela? And does the burden of individual sequelae differ by age, race, sex, and baseline health status? Addressing these questions has been highlighted as an urgent research priority by stakeholders including the World Health Organization, the United States National Institute of Health, the United Kingdom National Institute for Health and Care Excellence, and several others[2,8–13]. Addressing these questions will help inform capacity planning and care of people with PASC.

Here, we answer this call for urgent research, where we leverage the breadth and depth of the US Department of Veterans Affairs (VA) electronic healthcare databases, which operates the largest integrated health care delivery system in the US, to undertake comprehensive large-scale analyses of 181,384 people

with COVID-19 who survived the first 30 days of infection and 4,397,509 non-infected controls, and aimed to estimate the burden of PASC in the overall cohort and among non-hospitalized ($n = 155,987$), hospitalized ($n = 19,359$), and those admitted to intensive care ($n = 6038$) and burdens of the 33 individual sequelae that comprise PASC, identified in prior work[1], in population groups (by age, race, sex, and baseline health status).

Results

We enrolled 181,384 veterans who survived the first 30 days of COVID-19 including 155,987, 19,359, and 6038 who were non-hospitalized, hospitalized, and admitted to intensive care during the acute phase (the first 30 days) of the COVID-19 infection. The control group included 4,397,509 users of the Veterans Health Administration (VHA) with no known COVID-19 infection. The baseline demographic and health characteristics are described in Table 1. The unadjusted numbers and percentages of incident clinical manifestations in the control group (VHA users), in the overall COVID-19 cohort, and according to COVID-19 care setting (non-hospitalized, hospitalized, and admitted to intensive care) during the follow-up are presented in Supplementary Table 1.

In adjusted analyses, 30-day survivors of COVID-19 exhibited increased risk of a broad range of incident sequelae including pulmonary system disorders and cardiovascular, coagulation,

Table 1 Characteristics of (1) users of Veterans Health Administration (VHA), (2) non-hospitalized people with COVID-19, (3) hospitalized COVID-19 but not admitted to intensive care, and (4) hospitalized COVID-19 admitted to intensive care.

	VHA users $n = 4,397,509$	COVID-19 without hospitalization $n = 155,987$	Hospitalized COVID-19 without admit to intensive care $n = 19,359$	COVID-19 admitted to intensive care $n = 6038$
Age (IQR)	67.13 (53.12, 74.48)	62.21 (49.63, 72.60)	70.46 (61.03, 76.05)	70.79 (62.26, 75.66)
Race, White (%)	3,369,525 (76.62)	112,637 (72.21)	12,320 (63.64)	3755 (62.19)
Race, Black (%)	821,787 (18.69)	35,945 (23.04)	5972 (30.85)	1930 (31.96)
Race, other (%)	206,197 (4.69)	7405 (4.75)	1067 (5.51)	353 (5.85)
Sex, male (%)	3,978,601 (90.47)	137,767 (88.32)	18,197 (94.00)	5714 (94.63)
Sex, female (%)	418,908 (9.53)	18,220 (11.68)	1162 (6.00)	324 (5.37)
Long-term care (%)	26,246 (0.60)	4090 (2.62)	1900 (9.81)	537 (8.89)
Number of outpatient encounter (IQR)[a]	2 (1, 3)	2 (2, 3)	6 (4, 8)	6 (4, 8)
Number of hospital admission (IQR)[a]	0 (0, 0)	0 (0, 0)	0 (0, 1)	0 (0, 1)
Number of prescriptions received (IQR)[a]	5 (3, 11)	8 (4, 14)	13 (7, 20)	14 (8, 20)
Number of outpatient eGFR measurements (IQR)[a]	1 (0, 2)	1 (1, 2)	3 (1, 6)	3 (1, 7)
Area Deprivation Index (IQR)	53.91 (41.89, 62.66)	54.31 (43.52, 62.97)	53.92 (42.87, 61.31)	54.75 (42.87, 62.20)
Charlson Comorbidity Index (IQR)	0 (0, 1)	1 (0, 2)	2 (1, 4)	2 (1, 5)
Anxiety (%)	376,584 (8.56)	17,523 (11.23)	3711 (19.17)	1113 (18.43)
Cancer (%)	281,992 (6.41)	11,164 (7.16)	2778 (14.35)	885 (14.66)
Cardiovascular disease (%)	577,784 (13.14)	23,145 (14.84)	5779 (29.85)	2025 (33.54)
Chronic kidney disease (%)	594,808 (13.53)	24,522 (15.72)	6344 (32.77)	2140 (35.44)
Chronic lung disease (%)	495,853 (11.28)	20,816 (13.34)	4609 (23.81)	1638 (27.13)
Cerebrovascular disease (%)	197,540 (4.49)	8556 (5.49)	2606 (13.46)	762 (12.62)
Dementia (%)	138,709 (3.15)	7162 (4.59)	2619 (13.53)	700 (11.59)
Depression (%)	305,704 (6.95)	15,671 (10.05)	2919 (15.08)	831 (13.76)
Diabetes mellitus (%)	1,040,084 (23.65)	46,737 (29.96)	8868 (45.81)	3018 (49.98)
Hepatitis C (%)	12,833 (0.29)	599 (0.38)	233 (1.20)	68 (1.13)
Hyperlipidemia (%)	1,157,588 (26.32)	53,080 (34.03)	6435 (33.24)	2108 (34.91)
Hypertension (%)	1,183,516 (26.91)	49,145 (31.51)	7830 (40.45)	2337 (38.70)
Peripheral artery disease (%)	40,250 (0.92)	1800 (1.15)	664 (3.43)	219 (3.63)
Overweight (%)	1,544,583 (35.12)	50,012 (32.06)	5872 (30.33)	1763 (29.20)
Obesity (%)	2,019,446 (45.92)	85,578 (54.86)	9408 (48.60)	3148 (52.14)
Never smoke (%)	2,211,177 (50.28)	84,853 (54.40)	9576 (49.47)	2979 (49.34)
Former smoker (%)	1,151,586 (26.19)	43,231 (27.71)	6020 (31.10)	1900 (31.47)
Current smoker (%)	1,034,746 (23.53)	27,903 (17.89)	3763 (19.44)	1159 (19.20)
Follow-up days (IQR)	150 (115, 221)	149 (114, 209)	155 (117, 262)	173 (124, 282)
Total person-years (sum)	2,112,246.94	73,438.22	9940.89	3336.82

[a]Data collected within 1 year before the cohort enrollment.

dermatologic, endocrine, gastrointestinal, kidney, mental health, musculoskeletal, and neurologic system disorders; risks and associated burdens for each sequela are provided in Fig. 1 and Supplementary Table 2.

Burden of PASC according to care setting of the acute infection. To gain a better understanding of the toll of long covid in 30-day survivors of acute COVID-19, we first estimated the burden of PASC—defined as having at least one sequela in excess of the control group (VHA users without COVID-19). Unadjusted and adjusted burdens are presented in Supplementary Table 3 and Table 2, respectively. The adjusted burden of PASC in the overall cohort was 73.43 (72.10, 74.72) per 1000 persons at 6 months (Table 2), and increased according to care setting of the acute infection and was 44.51 (43.09, 45.85), 217.08 (212.43, 222.23), and 360.16 (350.53, 369.38) per 1000 persons at 6 months among non-hospitalized, hospitalized, and those who were admitted to intensive care during the first 30 days of infection, respectively (Table 2 and Fig. 2).

Burden of individual sequelae by age, race, sex, and health status. To gain a better understanding of whether and to what extent the burden of each sequelae may be differentially expressed across various population groups, we estimated the burden of incident individual post-acute sequelae per 1000 persons at 6 months (the burden in excess of the control group) by age, race, sex, and baseline health status represented by the number of comorbidities at baseline (Fig. 3 and Supplementary Tables 4 and 5). This analysis showed that the burdens were not uniformly expressed across age, race, sex, and baseline health status. Analyses of differences in burden on the basis of age suggest that while most incident sequelae were higher in older adults; burdens of hyperlipidemia, chest pain, sleep disorders, headache, obesity, mood disorders, cough, and smell problems were higher in people younger than 60 years. Smaller differences of burden on the basis of race were observed where burden of acute kidney injury, diabetes mellitus, chest pain, cough, substance abuse, thromboembolism, headache, and tachycardia were slightly higher in Black persons; GERD and smell problems were higher in white persons. Burden of several sequelae in the respiratory system (shortness of
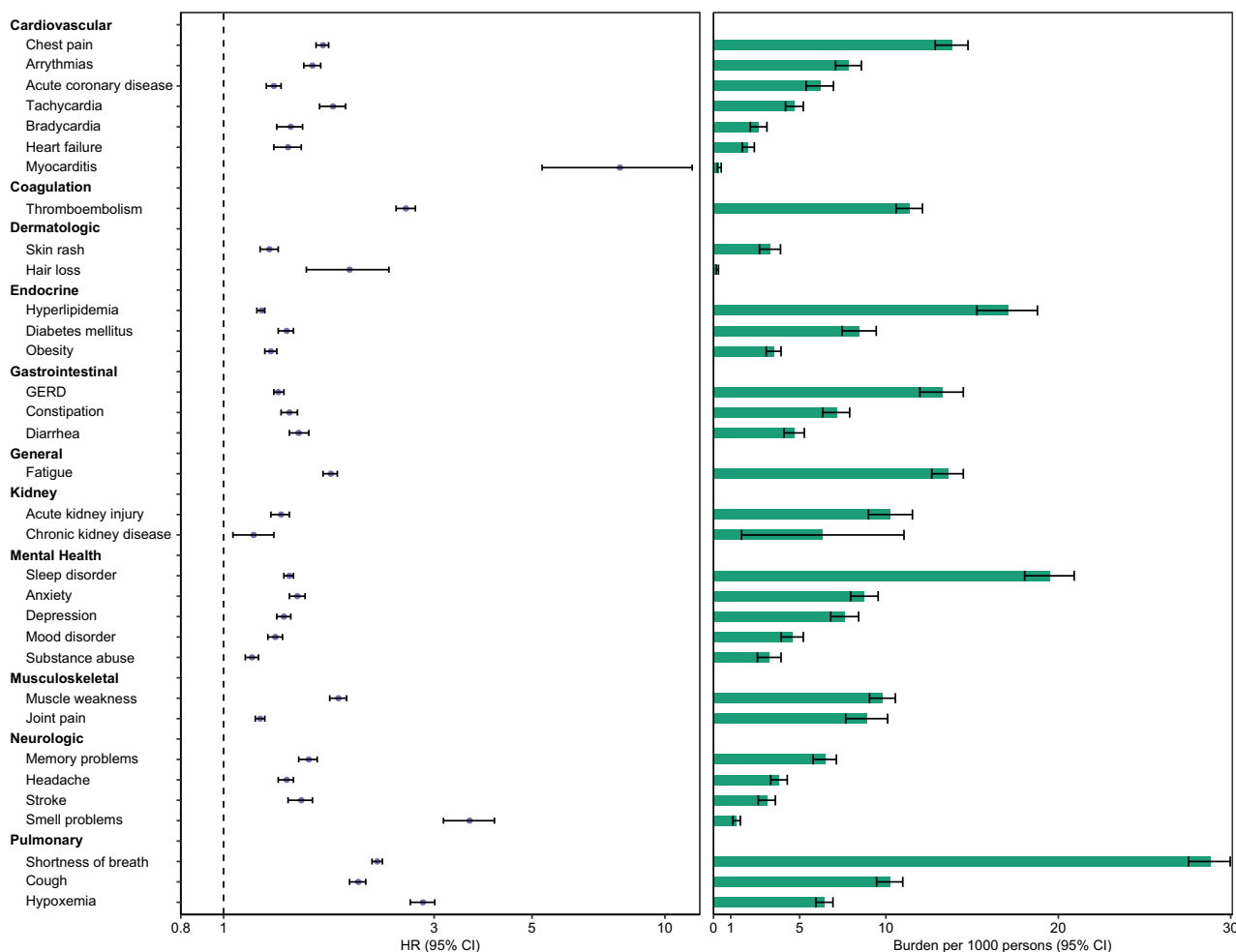

Fig. 1 Risks and burdens of post-acute sequelae of SARS-CoV-2 infection (PASC). Organ systems are presented in bolded font. Post-acute sequelae were ascertained from 30 days after infection until the end of follow-up. Hazard ratio and 95% confidence intervals are presented in the left panel and burdens per 1000 COVID-19 patients at 6 months are presented in the right panel; right limit of the bar represents estimated burden and error bars represent the 95% confidence interval. Models were adjusted for age, race, sex, receipt of long-term care, Area Deprivation Index, number of outpatient encounters, number of hospital admissions, number of outpatient prescriptions, number of outpatient serum creatinine measurements, chronic lung disease, cancer, cardiovascular disease, cerebrovascular disease, dementia, diabetes mellitus, hypertension, hyperlipidemia, depression, anxiety, chronic kidney disease, hepatitis C and peripheral artery disease, overweight, obesity, smoking status, Charlson Comorbidity Index, US geographic region, total number of beds, number of COVID-19 tests administered, COVID-19 positivity rate, and average hospital bed occupancy during the week of participant enrollment.

Table 2 Overall burden of post-acute sequelae of SARS-CoV-2 infection (PASC) per 1000 persons at 6 months in the overall cohort and across care setting (non-hospitalized, hospitalized, and admitted to intensive care during the acute phase of the infection).

| | COVID-19 group | Users of the VHA without COVID-19 | Burden associated with COVID-19[b] |
	Adjusted burden per 1000 persons at 6 months (95% CI)	Adjusted burden per 1000 persons at 6 months (95% CI)	Adjusted burden per 1000 persons at 6 months (95% CI)
Overall	168.01 (166.68, 169.29)	94.56 (94.23, 94.88)	73.43 (72.10, 74.72)
Non-hospitalized COVID-19[a]	139.07 (137.63, 140.44)	94.56 (94.23, 94.88)	44.51 (43.09, 45.85)
Hospitalized COVID-19[a]	311.61 (306.90, 316.82)	94.56 (94.23, 94.88)	217.08 (212.43, 222.23)
COVID-19 admitted to ICU[a]	454.74 (445.01, 463.92)	94.56 (94.23, 94.88)	360.16 (350.53, 369.38)

[a]Care settings during the first 30 days of infection.
[b]Burden defined as having at least one sequela in excess of users of the Veterans Health Administration without COVID-19 after the first 30 days of infection. Burdens were estimated based on Poisson regressions adjusting for age, race, sex, receipt of long-term care, Area Deprivation Index, number of outpatient encounters, number of hospital admissions, number of outpatient prescriptions, number of outpatient serum creatinine measurements, chronic lung disease, cancer, cardiovascular disease, cerebrovascular disease, dementia, diabetes mellitus, hypertension, hyperlipidemia, depression, anxiety, chronic kidney disease, hepatitis C and peripheral artery disease, overweight, obesity, smoking status, Charlson Comorbidity Index, US geographic region, total number of beds, number of COVID-19 tests administered, COVID-19 positivity rate, and average hospital bed occupancy during the week of participant enrollment.
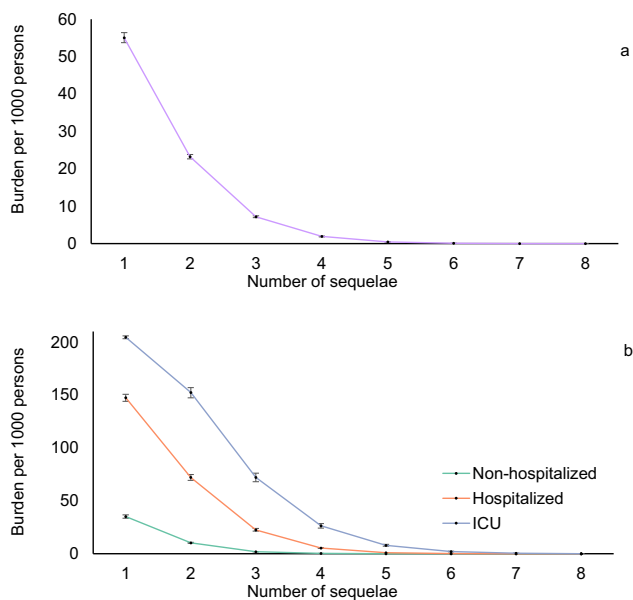

Fig. 2 Burden of post-acute sequelae of COVID-19 as a function of the number of sequelae. a Overall cohort (purple), and **b** by care setting (non-hospitalized (green), hospitalized (orange), and admitted to intensive care (blue) during the acute phase of the infection). Post-acute sequelae were ascertained from 30 days after infection until the end of follow-up. Estimates of burdens per 1000 COVID-19 patients at 6 months are presented; line represent the estimated burden and error bars represent the 95% confidence interval for the corresponding number of sequelae. Models were adjusted for age, race, sex, receipt of long-term care, Area Deprivation Index, number of outpatient encounters, number of hospital admissions, number of outpatient prescriptions, number of outpatient serum creatinine measurements, chronic lung disease, cancer, cardiovascular disease, cerebrovascular disease, dementia, diabetes mellitus, hypertension, hyperlipidemia, depression, anxiety, chronic kidney disease, hepatitis C and peripheral artery disease, overweight, obesity, smoking status, Charlson Comorbidity Index, US geographic region, total number of beds, number of COVID-19 tests administered, COVID-19 positivity rate, and average hospital bed occupancy during the week of participant enrollment.

breath and cough), cardiovascular system (chest pain, and arrhythmia), neurologic system (headache and smell problems), and dermatologic disorders (hair loss, and skin rash) were higher in females. Nearly all sequelae were more pronounced in persons with higher burden of baseline comorbidities (Fig. 4 and Supplementary Table 6).

Burden of individual sequelae by demographic and health characteristics according to care setting of the acute infection. In the overall cohort, burden of individual sequelae increased according to the care setting of the acute infection from non-hospitalized, hospitalized, and those who were admitted to intensive care (Figs. 5–7 and Supplementary Tables 7–9). In each subgroup based on age, race, sex, and baseline health status. The burden for each sequela increased as the intensity of care setting of the acute infection increased (non-hospitalized, hospitalized, admitted to intensive care).

Risk and burden of individual sequelae after 12 weeks of the infection. To test whether the risk of post-acute sequelae persisted beyond the first 12 weeks of the infection, we conducted analyses where we defined post-acute sequelae as incident clinical manifestations in excess of the non-infected controls that occurred after 12 weeks (84 days) from the COVID-19-positive test. In adjusted analyses, the burden of PASC in the overall cohort was 59.36 (57.72, 60.97) per 1000 persons at 6 months (Supplementary Table 10). The burden of PASC increased according to care setting of the acute infection and was 40.60 (38.83, 42.26), 158.31 (152.92, 164.30), and 227.18 (215.65, 238.46) per 1000 persons at 6 months among non-hospitalized, hospitalized, and those who were admitted to intensive care during the first 30 days of infection, respectively (Supplementary Table 10). All individual post-acute sequelae exhibited increased risk and excess burden in COVID-19 compared to the controls. Risks and associated burdens for each sequela are provided in Supplementary Fig. 1 and Supplementary Table 11.

Negative controls. Testing of risks of accidental injuries and neoplasms as negative outcome controls in people with COVID-19 (the overall COVID-19 cohort) yielded non-significant association between COVID-19 exposure and the risks of accidental
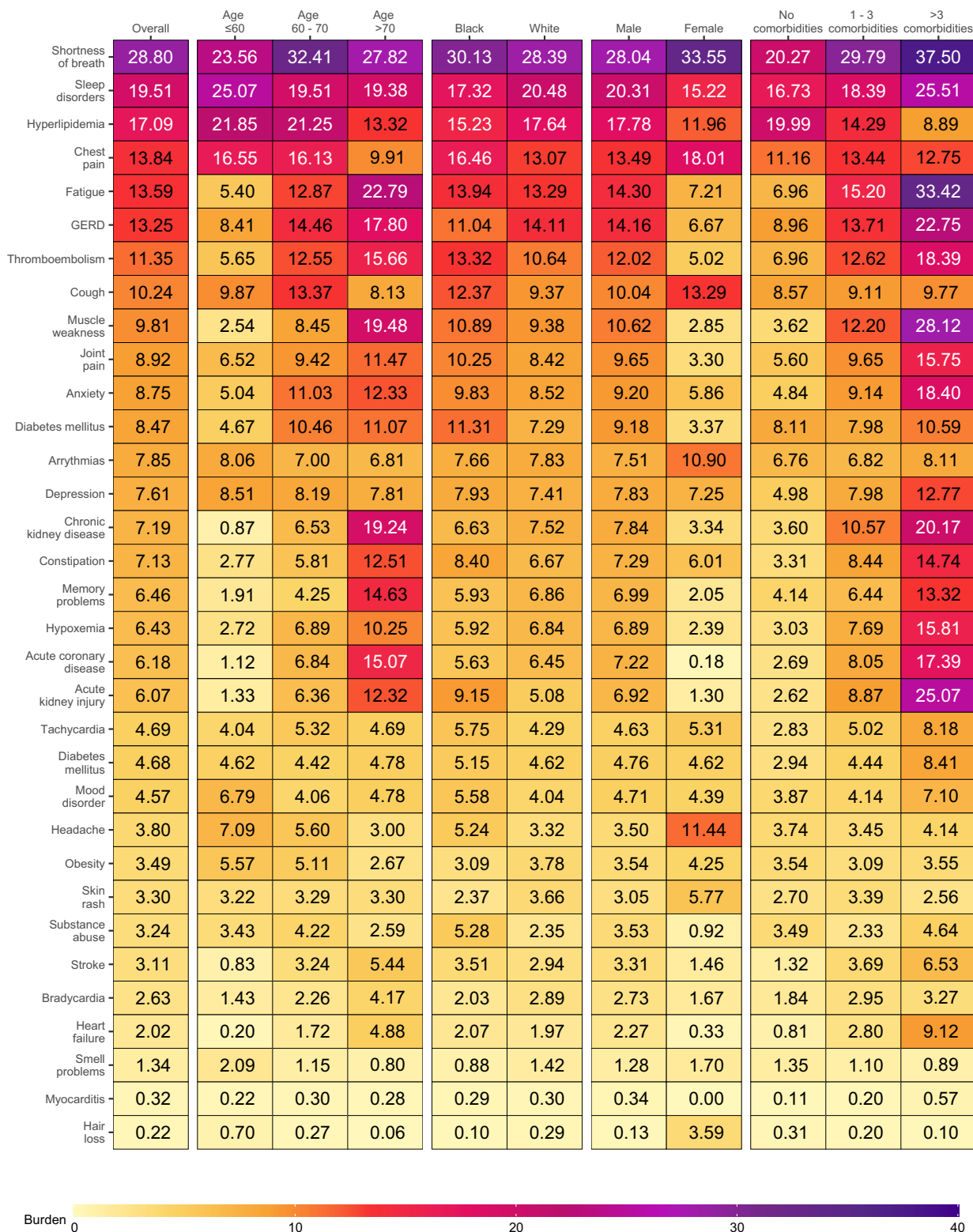
	Overall	Age ≤60	Age 60 - 70	Age >70	Black	White	Male	Female	No comorbidities	1 - 3 comorbidities	>3 comorbidities
Shortness of breath	28.80	23.56	32.41	27.82	30.13	28.39	28.04	33.55	20.27	29.79	37.50
Sleep disorders	19.51	25.07	19.51	19.38	17.32	20.48	20.31	15.22	16.73	18.39	25.51
Hyperlipidemia	17.09	21.85	21.25	13.32	15.23	17.64	17.78	11.96	19.99	14.29	8.89
Chest pain	13.84	16.55	16.13	9.91	16.46	13.07	13.49	18.01	11.16	13.44	12.75
Fatigue	13.59	5.40	12.87	22.79	13.94	13.29	14.30	7.21	6.96	15.20	33.42
GERD	13.25	8.41	14.46	17.80	11.04	14.11	14.16	6.67	8.96	13.71	22.75
Thromboembolism	11.35	5.65	12.55	15.66	13.32	10.64	12.02	5.02	6.96	12.62	18.39
Cough	10.24	9.87	13.37	8.13	12.37	9.37	10.04	13.29	8.57	9.11	9.77
Muscle weakness	9.81	2.54	8.45	19.48	10.89	9.38	10.62	2.85	3.62	12.20	28.12
Joint pain	8.92	6.52	9.42	11.47	10.25	8.42	9.65	3.30	5.60	9.65	15.75
Anxiety	8.75	5.04	11.03	12.33	9.83	8.52	9.20	5.86	4.84	9.14	18.40
Diabetes mellitus	8.47	4.67	10.46	11.07	11.31	7.29	9.18	3.37	8.11	7.98	10.59
Arrythmias	7.85	8.06	7.00	6.81	7.66	7.83	7.51	10.90	6.76	6.82	8.11
Depression	7.61	8.51	8.19	7.81	7.93	7.41	7.83	7.25	4.98	7.98	12.77
Chronic kidney disease	7.19	0.87	6.53	19.24	6.63	7.52	7.84	3.34	3.60	10.57	20.17
Constipation	7.13	2.77	5.81	12.51	8.40	6.67	7.29	6.01	3.31	8.44	14.74
Memory problems	6.46	1.91	4.25	14.63	5.93	6.86	6.99	2.05	4.14	6.44	13.32
Hypoxemia	6.43	2.72	6.89	10.25	5.92	6.84	6.89	2.39	3.03	7.69	15.81
Acute coronary disease	6.18	1.12	6.84	15.07	5.63	6.45	7.22	0.18	2.69	8.05	17.39
Acute kidney injury	6.07	1.33	6.36	12.32	9.15	5.08	6.92	1.30	2.62	8.87	25.07
Tachycardia	4.69	4.04	5.32	4.69	5.75	4.29	4.63	5.31	2.83	5.02	8.18
Diabetes mellitus	4.68	4.62	4.42	4.78	5.15	4.62	4.76	4.62	2.94	4.44	8.41
Mood disorder	4.57	6.79	4.06	4.78	5.58	4.04	4.71	4.39	3.87	4.14	7.10
Headache	3.80	7.09	5.60	3.00	5.24	3.32	3.50	11.44	3.74	3.45	4.14
Obesity	3.49	5.57	5.11	2.67	3.09	3.78	3.54	4.25	3.54	3.09	3.55
Skin rash	3.30	3.22	3.29	3.30	2.37	3.66	3.05	5.77	2.70	3.39	2.56
Substance abuse	3.24	3.43	4.22	2.59	5.28	2.35	3.53	0.92	3.49	2.33	4.64
Stroke	3.11	0.83	3.24	5.44	3.51	2.94	3.31	1.46	1.32	3.69	6.53
Bradycardia	2.63	1.43	2.26	4.17	2.03	2.89	2.73	1.67	1.84	2.95	3.27
Heart failure	2.02	0.20	1.72	4.88	2.07	1.97	2.27	0.33	0.81	2.80	9.12
Smell problems	1.34	2.09	1.15	0.80	0.88	1.42	1.28	1.70	1.35	1.10	0.89
Myocarditis	0.32	0.22	0.30	0.28	0.29	0.30	0.34	0.00	0.11	0.20	0.57
Hair loss	0.22	0.70	0.27	0.06	0.10	0.29	0.13	3.59	0.31	0.20	0.10

Burden 0 ————— 10 ————— 20 ————— 30 ————— 40

Fig. 3 Burden of post-acute sequelae of COVID-19 in the overall cohort and by age, race, sex, and health status. Post-acute sequelae were ascertained from 30 days after infection until the end of follow-up. Estimates of burdens per 1000 COVID-19 patients at 6 months are presented. Models were adjusted for age, race, sex, receipt of long-term care, Area Deprivation Index, number of outpatient encounters, number of hospital admissions, number of outpatient prescriptions, number of outpatient serum creatinine measurements, chronic lung disease, cancer, cardiovascular disease, cerebrovascular disease, dementia, diabetes mellitus, hypertension, hyperlipidemia, depression, anxiety, chronic kidney disease, hepatitis C and peripheral artery disease, overweight, obesity, smoking status, Charlson Comorbidity Index, US geographic region, total number of beds, number of COVID-19 tests administered, COVID-19 positivity rate, and average hospital bed occupancy during the week of participant enrollment when appropriated.
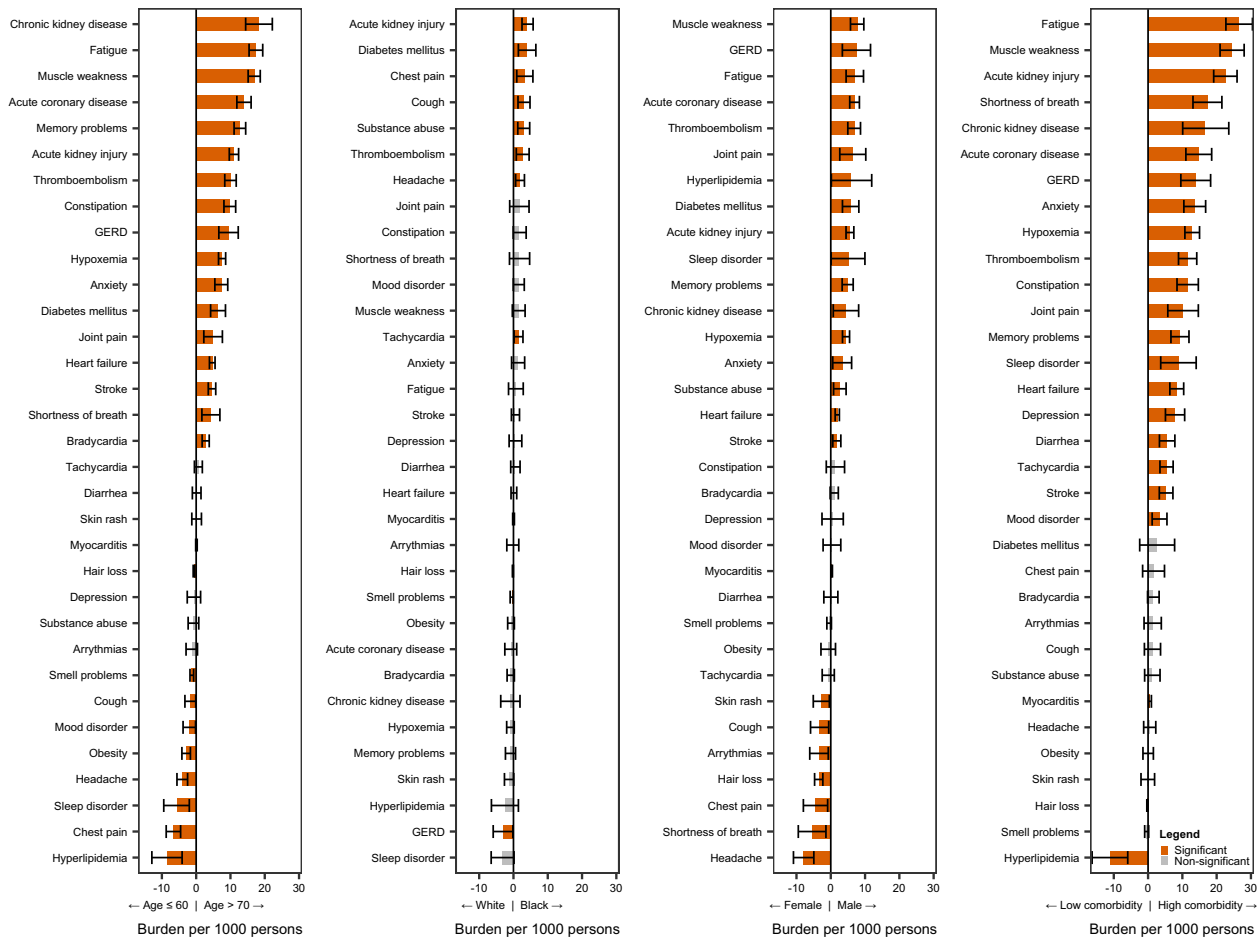
Fig. 4 Differences in burden of individual post-acute sequelae of COVID-19 by age, race, sex, and health status. Differences in burden per 1000 COVID-19 patients at 6 months are presented along with 95% confidence intervals.

injuries and neoplasms (Hazard ratio: 1.04 (0.84, 1.29) and 1.01 (0.88, 1.17), respectively) (Supplementary Table 10). The examination of negative outcome controls in groups according to care setting in non-hospitalized, hospitalized, and admitted to intensive care for COVID-19 also yielded non-significant associations—the results that were consistent with a priori expectations (Supplementary Table 12).

Discussion

In this work, we estimate that PASC—defined as at least one sequela in excess of a non-infected control group—was 73.43 (72.10, 74.72) per 1000 persons at 6 months. The burden of PASC increased as a function of the severity of the acute infection as proxied by the care setting and was 44.51 (43.09, 45.85), 217.08 (212.43, 222.23), and 360.16 (350.53, 369.38) per 1000 persons at 6 months among non-hospitalized, hospitalized, and those who were admitted to intensive care during the first 30 days of infection, respectively. PASC is a multifaceted non-monolithic entity; some of its sequelae were more pronounced in younger individuals, and some were more pronounced in older adults. The same picture was evident in analyses across race groups (white and Black participants), and sex (males and females). The burden of individual sequela was consistently higher in people with poorer baseline health status and increased in a graded fashion according to intensity of the care setting of the acute infection. The constellation of findings shows that among the 30-day survivors of COVID-19, the burden of PASC is substantial (7%); PASC is not a monolithic entity with sequelae that are differentially

expressed in various population groups. In sum, our results provide estimates of the burden for overall PASC; we additionally provide estimates of individual sequelae in various population groups by age, race, sex, and baseline health. Collectively, these results may be useful in informing health systems capacity planning and care strategies of people with PASC.

Here we estimate that the overall burden of PASC (long covid) beyond the first 30 days of illness is around 7%, (4.4%, 21.7%, and 36.5% among non-hospitalized, hospitalized, and those admitted to intensive care, respectively). The estimates were slightly lower but remained substantial using an alternative definition accounting for burden beyond first 12 weeks of illness (6% in overall cohort, and 4.1%, 15.8%, 22.7% in non-hospitalized, hospitalized, and those admitted to intensive care, respectively).

To date, most studies on long covid did not include a control group or simply focused mostly on people who were hospitalized; fewer studies reported burden of long covid in non-hospitalized patients with COVID-19[14–16]. Nevertheless, the evidence from the totality of studies accumulated thus far generally places the burden of long covid at around 4–10%[14–17]. Most importantly, it is now compellingly evident that while burden of PASC (or long covid) is most pronounced in people with poorer baseline health and those who had severe acute infection, the burden of PASC (or long covid) is substantial (and non-trivial) among non-hospitalized individuals (who represent the majority of people with COVID-19)—this potentially may translate into millions of people affected in the US and globally.
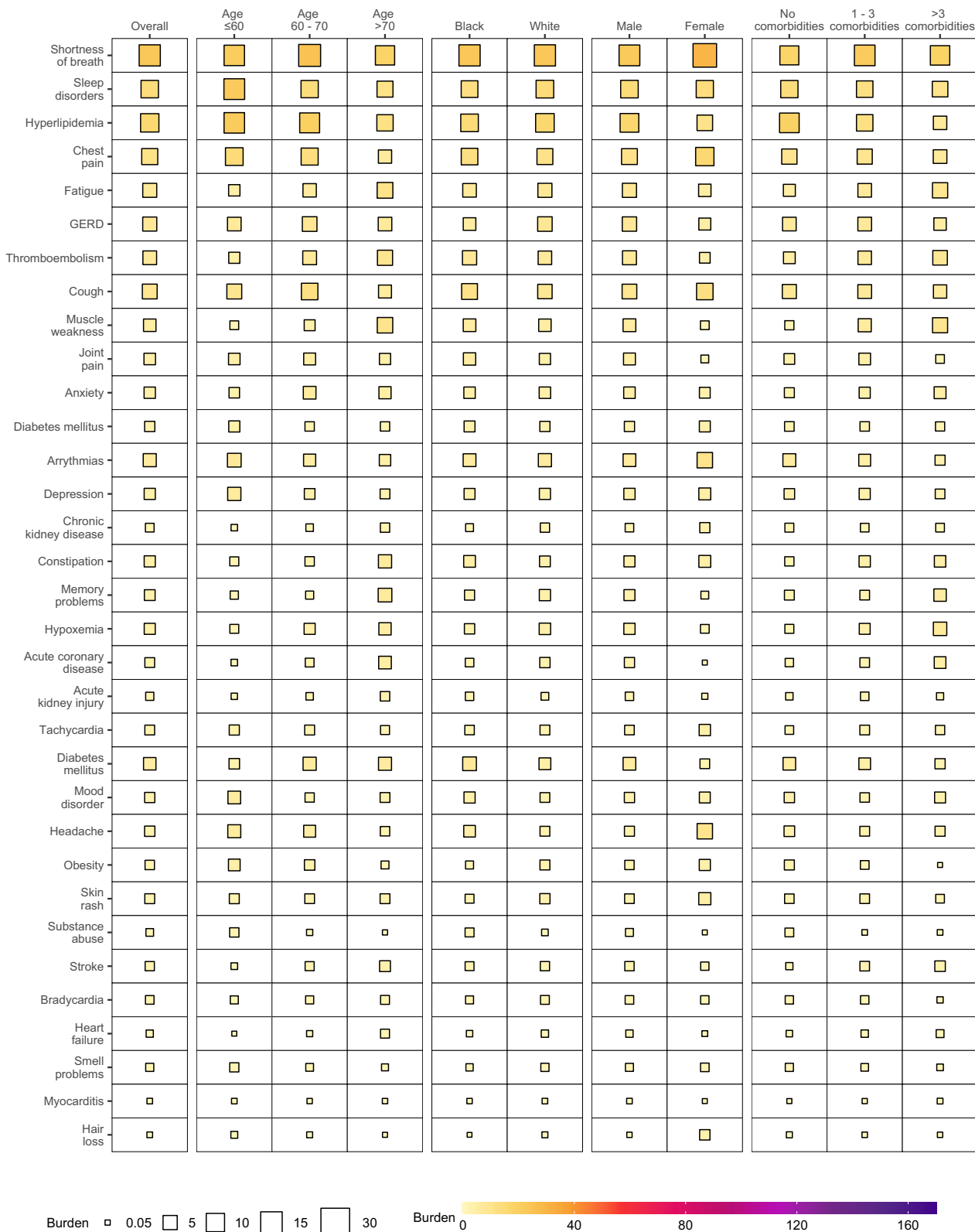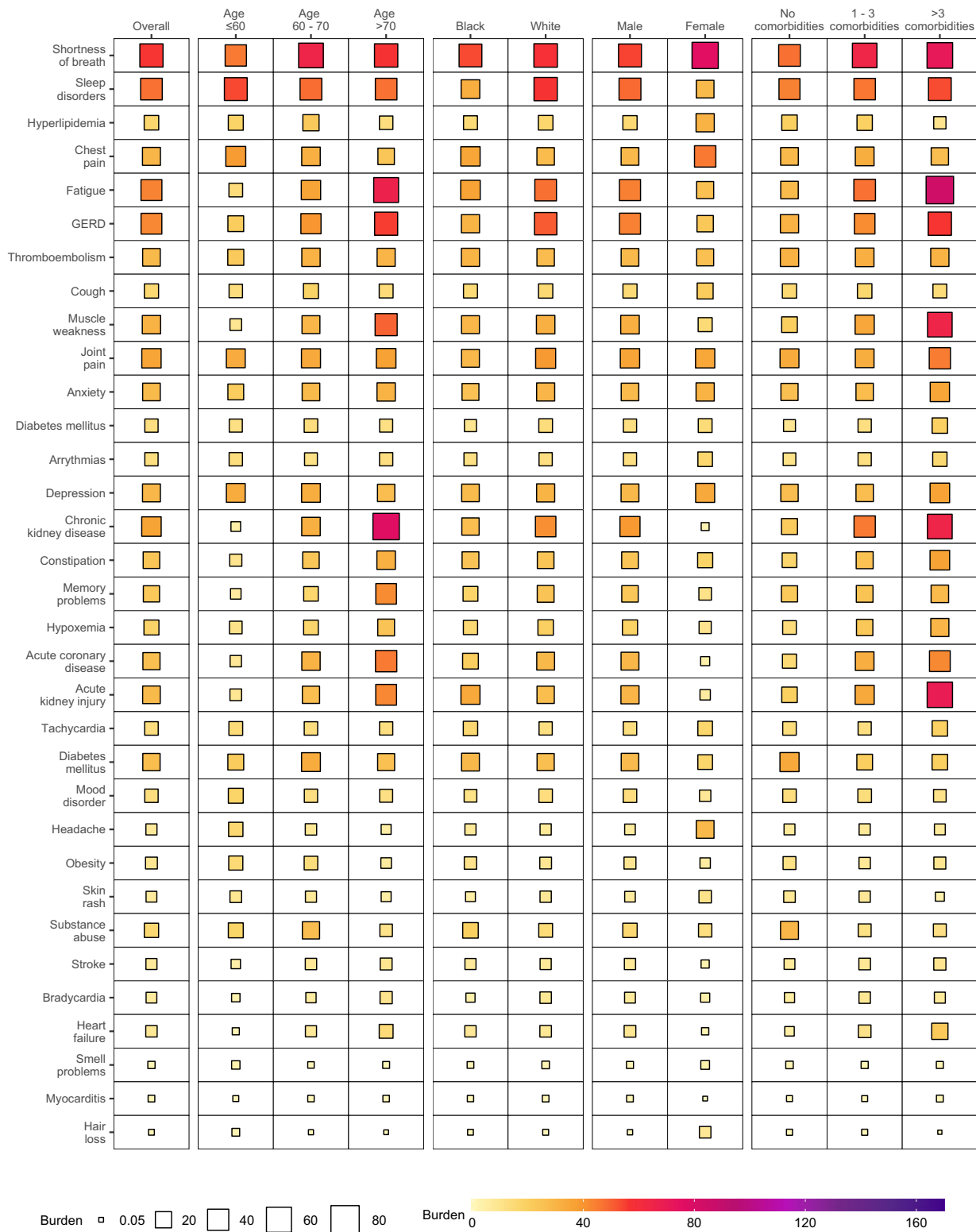
Fig. 5 Burden of post-acute sequelae of COVID-19 in the overall cohort and by age, race, sex, and health status in non-hospitalized COVID-19.
Estimates of burdens per 1000 COVID-19 patients at 6 months are presented. The size of the square represents the burden within each care setting. The intensity of color from light yellow to deep purple represents the range of burdens across care settings. Models adjusted for age, race, sex, receipt of long-term care, Area Deprivation Index, number of outpatient encounters, number of hospital admissions, number of outpatient prescriptions, number of outpatient serum creatinine measurements, chronic lung disease, cancer, cardiovascular disease, cerebrovascular disease, dementia, diabetes mellitus, hypertension, hyperlipidemia, depression, anxiety, chronic kidney disease, hepatitis C and peripheral artery disease, overweight, obesity, smoking status, Charlson Comorbidity Index, US geographic region, total number of beds, number of COVID-19 tests administered, COVID-19 positivity rate, and average hospital bed occupancy during the week of participant enrollment when appropriated.
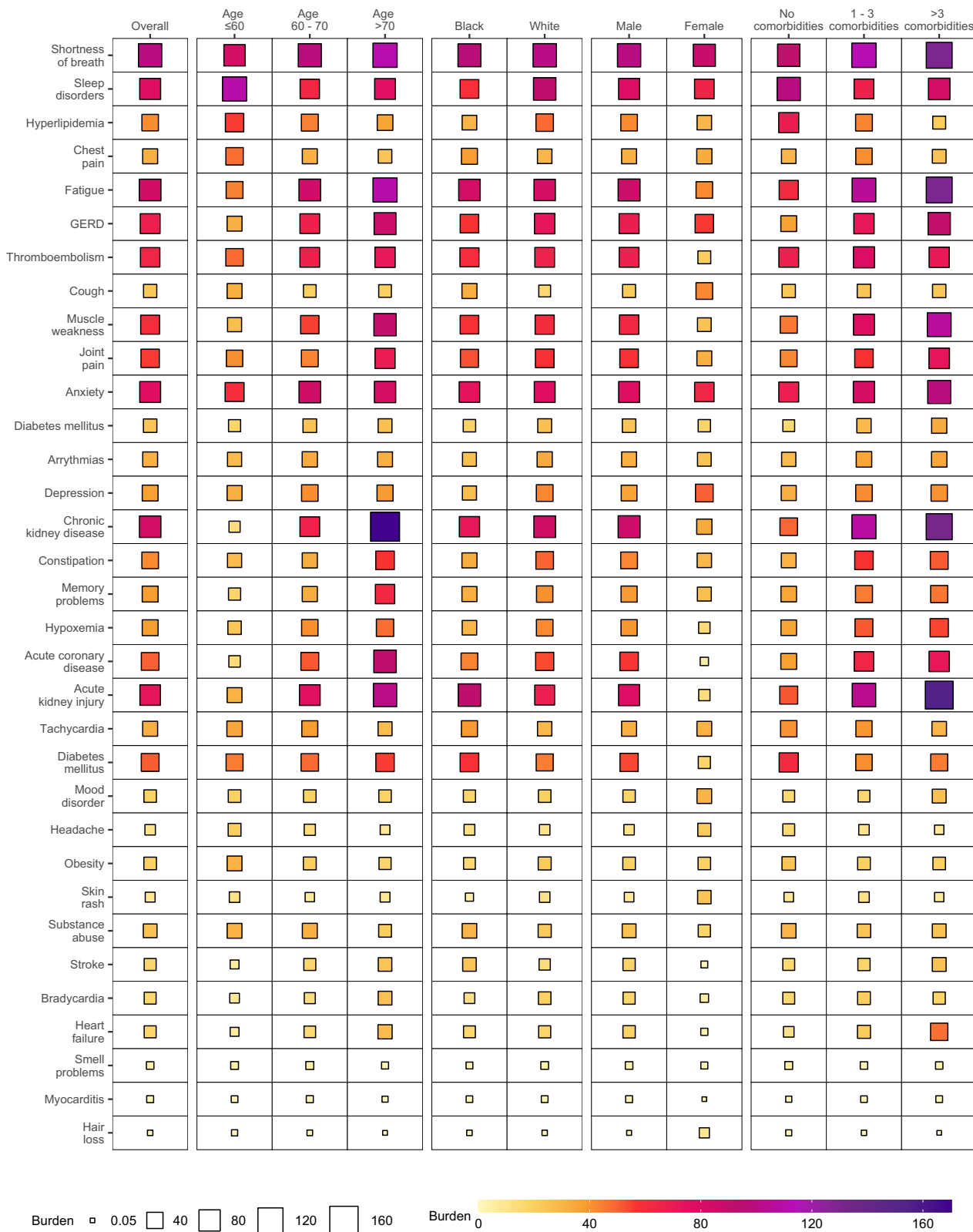
Fig. 6 Burden of post-acute sequelae of COVID-19 in the overall cohort and by age, race, sex, and health status in hospitalized COVID-19. Estimates of burdens per 1000 COVID-19 patients at 6 months are presented. The size of the square represents the burden within each care setting. The intensity of color from light yellow to deep purple represents the range of burdens across care settings. Models adjusted for age, race, sex, receipt of long-term care, Area Deprivation Index, number of outpatient encounters, number of hospital admissions, number of outpatient prescriptions, number of outpatient serum creatinine measurements, chronic lung disease, cancer, cardiovascular disease, cerebrovascular disease, dementia, diabetes mellitus, hypertension, hyperlipidemia, depression, anxiety, chronic kidney disease, hepatitis C and peripheral artery disease, overweight, obesity, smoking status, Charlson Comorbidity Index, US geographic region, total number of beds, number of COVID-19 tests administered, COVID-19 positivity rate, and average hospital bed occupancy during the week of participant enrollment when appropriated.

Fig. 7 Burden of post-acute sequelae of COVID-19 in the overall cohort and by age, race, sex, and health status in COVID-19 patients admitted to intensive care. Estimates of burdens per 1000 COVID-19 patients at 6 months are presented. The size of the square represents the burden within each care setting. The intensity of color from light yellow to deep purple represents the range of burdens across care settings. Models adjusted for age, race, sex, receipt of long-term care, Area Deprivation Index, number of outpatient encounters, number of hospital admissions, number of outpatient prescriptions, number of outpatient serum creatinine measurements, chronic lung disease, cancer, cardiovascular disease, cerebrovascular disease, dementia, diabetes mellitus, hypertension, hyperlipidemia, depression, anxiety, chronic kidney disease, hepatitis C and peripheral artery disease, overweight, obesity, smoking status, Charlson Comorbidity Index, US geographic region, total number of beds, number of COVID-19 tests administered, COVID-19 positivity rate, and average hospital bed occupancy during the week of participant enrollment when appropriated.

The implications of our results are clear. As the number of COVID-19 cases continues to climb across the globe, health systems face the dual challenge of coping with surges in acute infections, and caring for COVID-19 survivors (now accounting for more than 2% of the global population and growing) who will also likely require substantial care to mitigate permanent health loss. This will place additional demand on already strained health systems. Governments and health systems around the globe should be actively devising plans to address the tide of COVID-19 survivors in need of post-COVID-19 care. Our estimates of the burden of PASC (around 7% of infected people) and burden of 33 individual sequelae by age, race, sex, and baseline health status may help inform capacity planning and optimal composition of multidisciplinary post-COVID-19 clinics[11].

Estimates of the burden of individual sequelae by age, race, sex, and baseline health status suggest a more nuanced picture in that the burden of some sequelae was more pronounced in younger adults (e.g., sleep disorders, headache, mood disorders, and smell problems), Black participants (e.g., new onset diabetes mellitus, chest pain, substance abuse, thromboembolism, headache, and tachycardia), and females (e.g., chest pain, arrhythmia, headache, smell problems, hair loss, and skin rash). These estimates provide insights into the long-term health consequences of COVID-19 and suggest that PASC is a complex multifaceted non-monolithic entity that may manifest differently in various population groups. These findings may also have implications on how we define and study the disease. A nuanced definition that considers the heterogeneity of key sequelae by demographic group will be needed. Future work should also focus on understanding whether subtypes or different phenotypes of long covid exist. A better understanding of the putative various phenotypes of long covid, their underlying biologic mechanisms, epidemiologic drivers, and associated clinical outcomes is needed to inform treatment strategies of people with long covid.

Here we use the term "PASC" to denote the consequences of post-covid in excess of what can be ascertained in the control group; while this definition is epidemiologically useful to estimate the burden of disease, a clinical definition is needed to guide the diagnosis and treatment[14]. Other terms used in the literature include "long COVID-19", "post-COVID-19 syndrome", and "post-acute COVID-19 syndrome", and people with symptoms and clinical manifestations beyond the acute phase have been referred to in the lay vernacular as "long haulers"[3,14]. We recognize that this matter is subject to intense research by the scientific community and patient groups with lived experience of post COVID-19 who hold a broad range of views regarding these terminologies[13,14,18,19]. Identification of scientifically accurate definitions and culturally sensitive terms to describe the illness beyond the acute phase will be an important step not only to standardize scientific communications globally but also to support clear and consistent public health messaging about the long-term consequences of COVID-19[14].

The study has several strengths. We used the vast national electronic health care databases of the US Department of Veterans Affairs—the largest nationally integrated healthcare delivery system in the US—to estimate burden of overall PASC and its individual sequelae by age, race, sex, and baseline health status. We comprehensively examined 33 sequelae which were defined based on integrated data from multiple sources including diagnostic codes, medications, and laboratory test results. The simultaneous examination of incident sequelae in the same analytic framework allows the comparative evaluation of risks and burdens of these conditions—providing health care providers, health system planners, public health officials, and the public at large with a priority list of the post-acute clinical conditions encountered in COVID-19 survivors. For each outcome examined, we built a cohort free of the related outcome at baseline to identify the risk of incident outcome during the follow-up—this approach allows the identification of incident clinical manifestations and abnormalities following COVID-19 infection. While we conducted survival analyses to estimate the risk of each outcome examined, we—for each outcome—also estimated the excess burden per 1000 persons due to COVID-19; this measure of risk on the absolute scale also considers the baseline risk and provides a more meaningful estimate of potential harm and can be more effectively communicated to the wider public than measures of relative risks (e.g., hazard ratio).

This study has several limitations. Our approach does not provide mechanistic insights into PASC nor does it delineate the sequelae that are direct or indirect consequences of the SARS-CoV-2 infection. Because of the predominantly male composition of the VA population, our findings may not identify clinical manifestations of post-acute COVID-19 that may be differentially much more pronounced in females and either non-expressed or rare in males. Our study included VHA users, and the results may not be representative of all US veterans. Our study did not include people who may have had COVID-19 but did not have a positive test for it; these individuals may have been included in the control arm, and if they exhibited sequelae related to COVID-19, it would have resulted in underestimation of risk. However, this is unlikely to be present in a large scale since testing constraints were present at the VA for only a brief period in March 2020 during the early phase of the pandemic. Our estimates of risks and associated burdens did not include undiagnosed conditions (i.e., people who may have suffered from some post-covid sequelae but did not seek care and their sequelae were not yet diagnosed). Finally, COVID-19 patients were enrolled in our cohorts from March 01, 2020 to March 15, 2021 and followed until May 01, 2021; as the COVID-19 global pandemic continues to evolve, and as treatment strategies improve, new variants of the virus emerge, and vaccine availability increases, it is likely that the epidemiology, short-term, and long-term outcomes of COVID-19 will likely also change over time[20].

In conclusion, we estimate that the burden of PASC is around 7%; while the burden of PASC increased according to the severity of the care setting of the acute COVID-19 infection, it was not trivial (4.4%) among those who were not hospitalized for acute COVID-19. Our results also show that PASC is not monolithic; the burden of its individual components may be differentially expressed in various population groups. Together, the estimates provided here suggest that the toll of morbidity of COVID-19 extends well beyond the acute phase of the disease. While optimism is rising that—as vaccine availability increases—the pandemic may soon be behind us, the focus on the immediate health effects of COVID-19 allows visibility of the tip of the iceberg. The long-term consequences of COVID-19—as evidenced in our work—are substantial and will reverberate for a long time after the surges in acute infections abate. Long covid (or as we refer to it in this work PASC) is a complex multifaceted post-viral syndrome; it demands greater attention and a coordinated long-term global response strategy.

Methods

Setting. The study utilized the VA electronic health care databases. The VA provides health care to US Veterans and operates the largest national integrated healthcare system in the United States with 1255 health care facilities, including 170 VA medical centers and 1074 outpatient sites located across the United States. Veterans enrolled have access to the Department of Veterans Affairs comprehensive medical benefits package including inpatient hospital care; outpatient services; preventive, primary, and specialty care; prescriptions; mental healthcare; home healthcare; geriatric and extended care; medical equipment; and prosthetics. VA electronic health care databases are updated daily.

Cohort. US Veterans who had at least one outpatient or inpatient encounter with the VHA between January 01, 2019 and December 31, 2019 were selected

($n = 5,808,018$). Among those alive on March 01, 2020 ($n = 5,606,309$), the COVID-19 group consisted of those with a COVID-19-positive test between March 01, 2020 and March 15, 2021 ($n = 191,958$). We then selected those who were alive on the 30th day after their first positive test ($n = 181,384$). COVID-19-positive patients were separated into 3 groups based on the care received during the 30 days after the positive test: (1) non-hospitalized ($n = 155,987$); (2) hospitalized ($n = 19,359$); (3) received intensive care ($n = 6038$). To generate a comparison group without COVID-19, we selected 4,534,600 participants who did not have a COVID-19-positive test, and randomly assigned every 25 of them the same cohort enrollment date as one participant in the COVID-19 group. We then further selected those who were alive 30 days after the enrollment date (VHA group $n = 4,397,509$). Participants were followed until May 01, 2021 (Supplementary Fig. 2).

Data sources. Data used in this study was collected from the VA Corporate Data Warehouse (CDW)[21–26]. The CDW Outpatient Encounters and Inpatient Encounters domains provided information related to diagnoses, procedures, and hospitalization records[27]. The CDW Outpatient Pharmacy domain and CDW Bar Code Medication Administration domain were used to collect prescription information. CDW Patient domain was used to collect demographic information. The CDW Laboratory Results domain was used to collect laboratory test information, and the COVID-19 Shared Data Resource was used to collect COVID-19 tests. The Area Deprivation Index (ADI), a composite measure of income, education, employment, and housing was obtained from the University of Wisconsin[28].

Negative outcome controls. The use of negative controls in observational studies may help detect the presence of both suspected and unsuspected spurious biases; the application of negative controls will test if shared biases in outcome ascertainment, residual confounding, analytic approach, or other latent biases might have influenced the results[29,30]. Here we followed the approach outlined by Lipsitch and collaborators to test accidental injuries and neoplasms as negative outcome controls[29], where, based on current knowledge, we would expect no association between COVID-19 infection and these two negative outcome controls.

Post-acute sequelae of COVID-19. We examined a set of 33 post-acute COVID-19 outcomes; these outcomes were selected based on prior studies[1,31], review of the literature[2,3], and the most recent US National Institute of Health workshop on PASC. Outcomes were defined based on ICD10 codes recorded from inpatient or outpatient encounters, medication records, or laboratory tests when appropriate using definitions validated for use with electronic health records[26,31–41]. Detailed definitions of the outcomes are presented in Supplementary Table 13. Cardiovascular outcomes included acute coronary disease, arrhythmia, bradycardia, chest pain, heart failure, myocarditis, and tachycardia; coagulation outcomes included thromboembolism; dermatologic outcomes included hair loss and skin rash; endocrine outcomes included diabetes mellitus, hyperlipidemia, and obesity; gastrointestinal outcomes included constipation, diarrhea, and GERD; general outcomes included fatigue; kidney outcomes included acute kidney injury and chronic kidney disease; mental health outcomes included anxiety, depression, mood disorder, sleep disorder, and substance abuse; musculoskeletal outcomes included joint pain and muscle weakness; neurologic outcomes included headache, memory problems, smell problems, and stroke; and pulmonary outcomes included cough, hypoxemia, and shortness of breath. Occurrence of incident clinical manifestation was defined as the occurrence of a manifestation that did not occur within past 1 year before cohort enrollment. PASC was defined as the presence of at least one incident clinical manifestation in excess of the non-infected controls.

Covariates. Covariates for analyses included age, race (white, Black, and other; race categorization was based on self-reports), sex, receipt of long-term care, Area Deprivation Index based on residential addresses and proxies of healthcare utilization, such as number of outpatient encounters, number of hospital admissions, number of outpatient prescriptions, and number of outpatient serum creatinine measurements in the year before enrollment. We also included comorbidities, such as chronic lung disease, cancer, cardiovascular disease, cerebrovascular disease, dementia, diabetes mellitus, hypertension, hyperlipidemia, depression, anxiety, chronic kidney disease, and hepatitis C and peripheral artery disease. In addition, covariates included overweight, obesity, smoking status (never, former, and current) and the Charlson Comorbidity Index were also adjusted for. We also adjusted for US geographic region (West, Mid-west, South and Northeast) where the care was received, and additional health system characteristics including total number of beds, number of COVID-19 tests administered, COVID-19 positivity rate, and average hospital bed occupancy during the week of participant enrollment.

Statistical analyses. Characteristics of the VHA users without COVID-19, and those with COVID-19 according to care setting of the acute infection (non-hospitalized, hospitalized, and admitted to intensive care) were described.

Excess burden of PASC, defined as having at least one sequela in excess of VHA users without COVID-19, was estimated using Poisson regression. The excess burdens of having 1, 2, to 33 sequelae at 6 months, as well as the total excess burden of PASC, were estimated in the overall cohort and by care setting of the acute infection.

We then estimated the excess burden of incident individual sequela. For each outcome examined, we built a cohort of participants without a history of the outcome. Cox models adjusting for covariates were used to estimate the hazard ratio of each COVID-19 care setting compared to VHA users, and the survival probability for the 4 groups at 6 months. Cause specific hazard models were used where occurrence of death was considered as competing risk during the analyses. Excess burden per 1000 patients at 6 months were computed as the difference in survival probability between each COVID-19 care setting and the VHA users group. Burden of outcomes in the overall COVID-19 population was computed as the weighted sum of the burden of the three care settings based on the proportion of COVID-19 patients in each care setting. Analyses were also conducted to estimate the excess burden within subgroups by age, race, sex, and baseline health status. Burden differences between subgroups of age ≤60 and >70, Black and white, female and male, and 0 and >3 comorbidity score were then estimated.

To test whether the risk of post-acute sequelae persisted beyond the first 12 weeks of the infection, we conducted additional analyses to estimate the burden of overall PASC and burden of each incident sequela, where we defined post-acute sequelae as incident clinical manifestations in excess of the non-infected controls that occurred after 12 weeks (84 days) from the COVID-19-positive test.

All analyses were done using SAS Enterprise Guide version 7.1 (SAS Institute, Cary, NC). Data visualizations were performed in R 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). The study and informed consent waivers were approved by the Institutional Review Board of the Department of Veterans Affairs St. Louis Health Care System, St. Louis, MO. The study was reported following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for reporting observational studies. This research project was reviewed and approved by the Institutional Review Board of the Department of Veterans Affairs Saint Louis Health Care System.

Data availability
The data that support the findings of this study are available from the VA. VA data are made freely available to researchers behind the VA firewall with an approved VA study protocol. More information is available at https://www.virec.research.va.gov or by contacting the VA Information Resource Center (VIReC) at VIReC@va.gov.

Code availability
SAS codes are available at https://github.com/yxie618/longPASC and at https://doi.org/10.5281/zenodo.5555241[42].

References
1. Al-Aly, Z., Xie, Y. & Bowe, B. High-dimensional characterization of post-acute sequelae of COVID-19. *Nature* **594**, 259–264 (2021).
2. WHO. *In the Wake of the Pandemic: Preparing for Long COVID* (World Health Organization, 2021).
3. Greenhalgh, T., Knight, M., A'Court, C., Buxton, M. & Husain, L. Management of post-acute covid-19 in primary care. *BMJ* **370**, m3026 (2020).
4. Ayoubkhani, D. et al. Post-covid syndrome in individuals admitted to hospital with covid-19: retrospective cohort study. *BMJ* **372**, n693 (2021).
5. Dennis, A. et al. Multiorgan impairment in low-risk individuals with post-COVID-19 syndrome: a prospective, community-based study. *BMJ Open* **11**, e048391 (2021).
6. Daugherty, S. E. et al. Risk of clinical sequelae after the acute phase of SARS-CoV-2 infection: retrospective cohort study. *BMJ* **373**, n1098 (2021).
7. Bowe, B., Xie, Y., Xu, E. & Al-Aly, Z. Kidney outcomes in long COVID. *J. Am. Soc. Nephrol.*, https://doi.org/10.1681/ASN.2021060734 (2021).
8. Collins, F. S. *NIH Launches New Initiative to Study "Long COVID"* (National Institutes of Health, 2021).
9. NHS England. *National Guidance for Post-Covid Syndrome Assessment Clinics* (NHS England, 2020).
10. NICE. *Covid-19 Rapid Guideline: Managing the Long-term Effects of Covid-19* (NICE, 2020).
11. Sivan, M. & Taylor, S. NICE guideline on long covid. *BMJ* **371**, m4938 (2020).
12. Yelin, D. et al. Long-term consequences of COVID-19: research needs. Lancet Infect. Dis. **20**, 1115–1117 (2020).
13. The Lancet. Facing up to long COVID. Lancet **396**, 1861 (2020).
14. Alwan, N. A. & Johnson, L. Defining long COVID: going back to the start. *Med (NY)* **2**, 501–504 (2021).
15. Alwan, N. A. The road to addressing long Covid. *Science* **373**, 491–493 (2021).

16. Rajan, S. et al. *In the Wake of the Pandemic: Preparing for Long COVID (Policy Brief, No. 39.)* (European Observatory on Health Systems and Policies, 2021).

17. Taquet, M. et al. Incidence, co-occurrence, and evolution of long-COVID features: a 6-month retrospective cohort study of 273,618 survivors of COVID-19. *PLoS Med.* **18**, e1003773 (2021).

18. Gorna, R. et al. Long COVID guidelines need to reflect lived experience. *Lancet* **397**, 455–457 (2021).

19. Davis, H. E. et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine* **38**, 101019 (2021).

20. Cai, M., Bowe, B., Xie, Y. & Al-Aly, Z. Temporal trends of COVID-19 mortality and hospitalisation rates: an observational cohort study from the US Department of Veterans Affairs. *BMJ Open* **11**, e047369 (2021).

21. Xie, Y. et al. Proton pump inhibitors and risk of incident CKD and progression to ESRD. *J. Am. Soc. Nephrol.* **27**, 3153–3163 (2016).

22. Xie, Y. et al. Risk of death among users of proton pump inhibitors: a longitudinal observational cohort study of United States veterans. *BMJ Open* **7**, e015735 (2017).

23. Xie, Y. et al. Long-term kidney outcomes among users of proton pump inhibitors without intervening acute kidney injury. *Kidney Int.* **91**, 1482–1494 (2017).

24. Xie, Y. et al. Higher blood urea nitrogen is associated with increased risk of incident diabetes mellitus. *Kidney Int.* **93**, 741–752 (2018).

25. Xie, Y. et al. Clinical implications of estimated glomerular filtration rate dip following sodium-glucose cotransporter-2 inhibitor initiation on cardiovascular and kidney outcomes. *J. Am. Heart Assoc.* **10**, e020237 (2021).

26. Cai, M. et al. Temporal trends in incidence rates of lower extremity amputation and associated risk factors among patients using Veterans Health Administration Services From 2008 to 2018. *JAMA Netw. Open* **4**, e2033953 (2021).

27. Vincent, B. M., Wiitala, W. L., Burns, J. A., Iwashyna, T. J. & Prescott, H. C. Using veterans affairs corporate data warehouse to identify 30-day hospital readmissions. *Health Serv. Outcomes Res. Methodol.* **18**, 143–154 (2018).

28. Kind, A. J. & Buckingham, W. R. Making neighborhood-disadvantage metrics accessible—the neighborhood atlas. *N. Engl. J. Med.* **378**, 2456–2458 (2018).

29. Lipsitch, M., Tchetgen Tchetgen, E. & Cohen, T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* **21**, 383–388 (2010).

30. Shi, X., Miao, W. & Tchetgen, E. T. A selective review of negative control methods in epidemiology. *Curr. Epidemiol. Rep.* **7**, 190–202 (2020).

31. Xie, Y., Bowe, B., Maddukuri, G. & Al-Aly, Z. Comparative evaluation of clinical manifestations and risk of death in patients admitted to hospital with covid-19 and seasonal influenza: cohort study. *BMJ* **371**, m4677 (2020).

32. Xie, Y. et al. Estimates of all cause mortality and cause specific mortality associated with proton pump inhibitors among US veterans: cohort study. *BMJ* **365**, l1580 (2019).

33. Xie, Y. et al. Comparative effectiveness of SGLT2 inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and sulfonylureas on risk of kidney outcomes: emulation of a target trial using health care databases. *Diabetes Care* **43**, 2859–2869 (2020).

34. Xie, Y. et al. Comparative effectiveness of the sodium-glucose cotransporter 2 inhibitor empagliflozin versus other antihyperglycemics on risk of major adverse kidney events. *Diabetes Care* **43**, 2785–2795 (2020).

35. Bowe, B. et al. Acute kidney injury in a national cohort of hospitalized US veterans with COVID-19. *Clin. J. Am. Soc. Nephrol.* **16**, 14–25 (2020).

36. Bowe, B. et al. The 2016 global and national burden of diabetes mellitus attributable to PM2.5 air pollution. *Lancet Planet. Health* **2**, e301–e312 (2018).

37. Bowe, B., Xie, Y., Xian, H., Balasubramanian, S. & Al-Aly, Z. Low levels of high-density lipoprotein cholesterol increase the risk of incident kidney disease and its progression. *Kidney Int.* **89**, 886–896 (2016).

38. Bowe, B. et al. High density lipoprotein cholesterol and the risk of all-cause mortality among U.S. veterans. *Clin. J. Am. Soc. Nephrol.* **11**, 1784–1793 (2016).

39. Bowe, B. et al. Ambient fine particulate matter air pollution and risk of weight gain and obesity in United States veterans: an observational cohort study. *Environ. Health Perspect.* **129**, 047003 (2021).

40. Bowe, B. et al. Ambient fine particulate matter air pollution and the risk of hospitalization among COVID-19 positive individuals: cohort study. *Environ. Int.* **154**, 106564 (2021).

41. Xie, Y. et al. Comparative effectiveness of sodium-glucose cotransporter 2 inhibitors vs sulfonylureas in patients with type 2 diabetes. *JAMA Intern. Med.* **181**, 1043–1053 (2021).

42. Xie, Y. Burden of COVID PASC by populations and manifestations. zenodo https://doi.org/10.5281/zenodo.5555241 (2021).

Author contributions

Z.A.A. and Y.X. contributed to the development of the study concept and design. Z.A.A. and Y.X. contributed to data analysis. Z.A.A. and Y.X. contributed to interpretation of the results. Z.A.A. and Y.X. drafted the manuscript. Z.A.A., Y.X., and B.B. contributed to critical revision of the manuscript. Z.A.A. provided administrative, technical, and material support, as well as supervision and mentorship. Each author contributed important intellectual content during manuscript drafting or revision and accepts accountability for the overall work by ensuring that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved. All authors approved the final version of the report. The corresponding author attests that all the listed authors meet the authorship criteria and that no others meeting the criteria have been omitted.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-26513-3.

Correspondence and requests for materials should be addressed to Ziyad Al-Aly.

Peer review information *Nature Communications* thanks Amitava Banerjee and Nishi Chaturvedi for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.