nature

https://doi.org/10.1038/s41586-021-03553-9

Accelerated Article Preview

High-dimensional characterization of post-acute sequalae of COVID-19

Received: 18 January 2021

Accepted: 14 April 2021

Accelerated Article Preview Published online 22 April 2021

Cite this article as: Al-Aly, Z. et al. High-dimensional characterization of post-acute sequalae of COVID-19. *Nature* https://doi.org/10.1038/s41586-021-03553-9 (2021).

Ziyad Al-Aly, Yan Xie & Benjamin Bowe

Article

High-dimensional characterization of post-acute sequalae of COVID-19

https://doi.org/10.1038/s41586-021-03553-9

Received: 18 January 2021

Accepted: 14 April 2021

Published online: 22 April 2021

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

The acute clinical manifestations of COVID-19 are well characterized[1,2]; however, its post-acute sequalae have not been comprehensively described. Here, we use the national healthcare databases of the US Department of Veterans Affairs to systematically and comprehensively identify 6-month incident sequalae including diagnoses, medication use, and laboratory abnormalities in 30-day survivors of COVID-19. We show that beyond the first 30 days of illness, people with COVID-19 exhibit higher risk of death and health resource utilization. Our high dimensional approach identifies incident sequalae in the respiratory system and several others including nervous system and neurocognitive disorders, mental health disorders, metabolic disorders, cardiovascular disorders, gastrointestinal disorders, malaise, fatigue, musculoskeletal pain, and anemia. We show increased incident use of several therapeutics including pain medications (opioids and non-opioids), antidepressants, anxiolytics, antihypertensives, and oral hypoglycemics and evidence of laboratory abnormalities in multiple organ systems. Analysis of an array of pre-specified outcomes reveals a risk gradient that increased across severity of the acute COVID-19 infection (non-hospitalized, hospitalized, admitted to intensive care). The findings show that beyond the acute illness, substantial burden of health loss – spanning pulmonary and several extrapulmonary organ systems – is experienced by COVID-19 survivors. The results provide a roadmap to inform health system planning and development of multidisciplinary care strategies to reduce chronic health loss among COVID-19 survivors.

The coronavirus disease 2019 (COVID-19) is a viral illness caused by the severe acute respiratory syndrome coronavirus. The acute clinical manifestations of COVID-19 are well characterized and involve both pulmonary and extrapulmonary systemic manifestations[1,2]. Emerging reports of "long haulers" suggest that beyond the acute setting, some patients with COVID-19 may experience persistent long-lasting clinical manifestations. However, the post-acute sequelae of COVID-19 are not yet clear.

Here we leveraged the breadth and depth of the US Department of Veterans Affairs electronic health databases to undertake a high dimensional approach to comprehensively identify the 6-months outcomes of incident diagnoses (from 379 diagnostic categories), incident medication use (from 380 medication classes), and incident laboratory abnormalities (from 62 laboratory tests) in people who survived the first 30 days of COVID-19.

High dimensional analysis of non-hospitalized COVID-19 vs VHA users

The cohort included 73,435 users of the Veteran Health Administration (VHA) with COVID-19 who survived at least the first 30 days after COVID-19 diagnosis and were not hospitalized, and 4,990,835 VHA users who did not have COVID-19 and were not hospitalized (Supplementary Fig. 1a and b). The median follow-up and interquartile range were 126 (81, 203) and 130 (82, 205) days in the COVID-19 and VHA user groups (Extended data table 1a). Examination of a panel of negative outcome controls including neoplasms and accidental injuries yielded results consistent with a priori expectations (HR 1.03 (0.94, 1.12), and HR 1.03 (0.95, 1.12), respectively) (results of all the negative outcome controls are provided in extended data table 2a). Examination of standardized differences of all high dimensional variables across all outcome specific cohorts (including those selected and those that were not selected in the models) showed that more than 99.99% of standardized differences were <0.1 after adjustment (Supplementary Fig. 2a and b), resulting in similar distributions of baseline characteristics in each group after adjustment (Supplementary table 1).

Beyond the first 30 days of illness, COVID-19 survivors had increased risk of death (HR 1.59 (1.46. 1.73)). We also estimated the adjusted excess burden due to COVID-19 per 1000 persons at 6-months based on the difference of estimated incidence rate between COVID-19 and all users of VHA. The excess death was estimated at 8.39 (7.09. 9.58) per 1000 COVID-19 patients at 6-months. Those with COVID-19 had a higher risk

[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]Nephrology Section, Medicine Service, VA Saint Louis Health Care System, 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]Department of Epidemiology and Biostatistics, College for Public Health and Social Justice, Saint Louis University, Saint Louis, MO, USA. ✉e-mail: zalaly@gmail.com

of outpatient care encounter (HR 1.20 (1.19, 1.21); excess burden 33.22 (30.89, 35.58) and at a greater frequency (0.47 (0.44, 0.49) additional encounter every 30 days) (Extended data table 2b and c).

We evaluated the risk of incident 379 diagnoses (categorized from ICD-10 codes based on Clinical Classifications Software Refined (CCSR)), 380 medication classes, and 62 laboratory tests beyond the first 30 days. For each outcome examined, we built a cohort free of the related outcome at baseline to identify the risk of incident outcome during follow up. Several conditions in almost every organ system exhibited an adjusted hazard ratio greater than 1, and a p value lower than 6.57×10^{-5} (significance level adjusting for multiple comparisons). The adjusted hazard ratio and burden for all outcomes are presented in figure 1a-c and supplementary table 2-4. The result for outcomes that were positively associated with COVID-19 are presented in in figure 2a-c, extended data Fig. 1a-c, supplementary table 5 and are detailed below:

Respiratory conditions

At six months following a COVID-19 infection that did not result in a hospitalization in the first 30-days, excess burden of respiratory conditions was most common and included respiratory signs and symptoms (28.51 (26.40, 30.50) per 1000 COVID-19 patients at 6-months), respiratory failure, insufficiency, arrest (3.37 (2.71, 3.92)), and lower respiratory disease (4.67 (3.96, 5.28)). There was also evidence of high burden of incident use of bronchodilators (22.23 (20.68, 23.67)), antitussives and expectorants (12.83 (11.61, 13.95)), anti-asthmatics (8.87 (7.65, 9.97)), and glucocorticoids (7.65 (5.67, 9.50)).

Diseases of the nervous system

Excess burden of nervous system disorders was evident including nervous system signs and symptoms (excess burden 14.32 (12.16, 16.36)) per 1000 COVID-19 patients at 6-months), neurocognitive disorders (3.17 (2.24, 3.98)), and nervous system disorders (4.85 (3.65, 5.93)) and headache (4.10 (2.49, 5.58)).

Mental health burden

The results showed excess burden of sleep wake disorders (14.53 (11.53, 17.36)) per 1000 COVID-19 patients at 6-months), anxiety and fear-related disorders (5.42 (3.42, 7.29)), and trauma and stress related disorders (8.93 (6.62, 11.09)). These findings were coupled with evidence of excess burden of incident use of non-opioid analgesics (19.97 (17.41, 22.40)), opioid analgesics (9.39 (7.21, 11.43)), antidepressants (7.83 (5.19, 10.30)) and benzodiazepines sedatives and anxiolytics (22.23 (20.68, 23.67)).

Metabolic disorders

Excess burden of several metabolic disorders was evident including disorders of lipid metabolism (12.32 (8.18, 16.24)) per 1000 COVID-19 patients at 6-months), diabetes mellitus (8.23 (6.36, 9.95)), and obesity (9.53 (7.55, 11.37)). These was also evidence of excess burden of incident use of antilipemic agents (11.56 (8.73, 14.19)), oral hypoglycemics (5.39 (3.99, 6.64)), and insulin (4.95 (3.87, 5.90)), and excess burden of elevated low density lipoprotein cholesterol (9.48 (7.02, 11.81)), total cholesterol (9.94 (6.61, 13.11)), triglycerides (9.40 (6.63, 12.03)), and hemoglobin A1C (10.66 (6.77, 14.35)).

Poor general wellbeing

Survivors of COVID-19 exhibited excess burden of poor general wellbeing including malaise and fatigue (12.64 (11.24, 13.93) per 1000 COVID-19 patients at 6-months), muscle disorders (5.73 (4.60, 6.74)), musculoskeletal pain (13.89 (9.89, 17.71)) and anemia (4.79 (3.53, 5.93)). These diagnoses were coupled with laboratory evidence of excess burden of anemia (decreased hemoglobin (31.03 (28.16, 33.76)), decreased hematocrit levels (30.73 (27.64, 33.67)), and low serum albumin (6.44 (4.84, 7.92)).

Cardiovascular conditions

Excess burden of cardiovascular conditions included hypertension (15.18 (11.53, 18.62)) per 1000 COVID-19 patients at 6-months), cardiac dysrhythmias (8.41 (7.18, 9.53)), circulatory signs and symptoms (6.65 (5.18, 8.01)), chest pain (10.08 (8.63, 11.42)), coronary atherosclerosis (4.38 (2.96, 5.67)), and heart failure (3.94 (2.97, 4.80)). There was also evidence of excess burden of incident use of beta blockers (9.74 (8.06, 11.27)), calcium channel blockers (7.18 (5.61, 8.61)), loop diuretics (4.72 (3.59, 5.72)), thiazide diuretics (2.52 (1.37, 3.54)), and antiarrhythmics (1.28 (0.79, 1.67)).

Gastrointestinal system

There was evidence of excess burden of the following conditions: esophageal disorders (6.90 (4.58, 9.07)) per 1000 COVID-19 patients at 6-months), gastrointestinal disorders (3.58 (2.15, 4.88)), dysphagia (2.83 (1.79, 3.76)), abdominal pain (5.73 (3.7, 7.62)). These were coupled with evidence of increased use of laxatives (9.22 (6.99, 11.31)), antiemetics (9.22 (6.99, 11.31)), histamine antagonists (4.83 (3.63, 5.91)), other antacids (1.07 (0.62, 1.42)), and antidiarrheal agents (2.87 (1.70, 3.91)). Laboratory abnormalities included increased risk of incident high levels of alanine aminotransferase (7.62 (5.20, 9.90)).

Other sequalae

There was also evidence of excess burden in incident acute pulmonary embolism (2.63 (2.25, 2.92) per 1000 COVID-19 patients at 6-months) and use of anticoagulants (16.43 (14.85, 17.89)). Other conditions included excess burden of skin disorders (7.52 (5.17, 9.73)), arthralgia and arthritis (5.16 (3.18, 7.01)), and infections (including urinary tract infections (2.99 (1.94, 3.93)) (Figure 2a-c and supplementary table 2-5).

High-dimensional analysis of hospitalized COVID-19 vs seasonal influenza

To gain a better understanding of the spectrum of clinical manifestations in survivors of COVID-19 who got hospitalized, we undertook a comparative evaluation in a cohort of hospitalized individuals with COVID-19 vs. those hospitalized with seasonal influenza (a well-known, well characterized respiratory viral illness).

The hospitalized cohort included 13,654 people with COVID-19 and 13,997 people with influenza who survived at least 30 days after hospital admission (Supplementary Fig. 2a and b). The median follow-up and interquartile range were 150 (84, 217) and 157 (87, 220) days in the COVID-19 and influenza groups (Extended data table 1). Testing of a panel of negative outcome controls including neoplasms and accidental injuries yielded results consistent with a priori expectations (HR 0.98 (0.83, 1.16), and HR 1.02 (0.90, 1.15), respectively) (results of all the negative outcome controls are provided in extended data table 2a). Examination of standardized differences of all high dimensional variables (including those selected and those that were not selected in the models) in all outcome specific cohorts showed that more than 99.75% of standardized differences were <0.1 after adjustment (Supplementary Fig. 4a and b), resulting in similar distributions of baseline characteristics in each group after adjustment (Supplementary table 6).

Beyond the first 30 days of illness, COVID-19 survivors who had been hospitalized for COVID-19 had increased risk of death (HR=1.51 (1.30, 1.76)); excess death was estimated at 28.79 (19.52, 36.85) per 1000 persons at 6-months. Those with COVID-19 exhibited a higher

risk of outpatient care encounter (HR 1.12 (1.08, 1.17), excess burden 6.37 (4.01, 9.03)) and with greater frequency (1.45 (1.28, 1.63) additional encounters every 30 days) (Extended data table 2b and c).

Compared to those hospitalized with seasonal influenza, and beyond the first 30 days of illness, COVID-19 survivors who had been hospitalized for COVID-19 had a higher burden of a broad array of pulmonary and extrapulmonary systemic manifestations including neurologic disorders (nervous system disorders (19.78 (12.58, 26.19) per 1000 hospitalized COVID-19 patients) and neurocognitive disorders (16.16 (10.40, 21.19)), mental health disorders (e.g. mental and substance use conditions 7.75 (4.72, 10.10)), metabolic disorders (e.g. disorders of lipid metabolism 43.53 (28.71, 57.08)), cardiovascular disorders (e.g. circulatory signs and symptoms (17.92 (10.73, 24.35)), gastrointestinal disorders (e.g. dysphagia (19.28 (12.75, 25.13)), coagulation disorders (14.31 (10.08, 17.89)), pulmonary embolism (18.31 (15.83, 20.25)), and other disorders including malaise and fatigue (36.49 (28.13, 44.15)), and anemia (19.08 (10.58, 26.81)) (Extended data Fig. 2a-f, Extended data Fig. 3a-c, and supplementary table 7-10). Analyses of risk and burden of clinical manifestations which additionally adjusted for severity of the acute infection yielded consistent results in both direction and magnitude of estimates (Extended data Fig. 4a-f, extended data Fig. 5a-c, and supplementary table 11-14). A high dimensional comparative evaluation of 6-months outcomes in a cohort of hospitalized individuals with COVID-19 (N=13,654) vs. those hospitalized for other causes (N=901,516) yielded consistent results (Extended data Fig. 6a-f, extended data Fig. 7a-c, and supplementary table 15-18).

Analysis of risk of incident pre-specified post-acute COVID-19 outcomes

To complement our high dimensional approach, and to gain a deeper understanding of the clinical manifestations of post-acute COVID-19 across the severity of the initial acute disease, we evaluated risks of a panel of pre-specified outcomes across care setting of the acute phase of the disease (non-hospitalized, hospitalized, and admitted to intensive care – a proxy indicator of disease severity) and benchmarked these to a common reference group (the broader population of the Veterans Affairs Health Care System (N=4,990,835)) (Extended data table 1b). Assessment of standardized differences across the four groups showed none were <0.1 after adjustment (Supplementary Fig. 5). Results reveal: a) increased risk of a broad array of specific clinical manifestations including acute coronary disease, arrythmias, acute kidney injury, chronic kidney disease, memory problems, thromboembolic disease, and several others (Figure 3, supplementary table 19 and 20; b) the risks were evident even in those who were not hospitalized with COVID-19; and c) a risk gradient that increased across care setting of the acute COVID-19 infection (non-hospitalized, hospitalized, admitted to intensive care) where people who required intensive care during the acute phase exhibited the highest risk (Figure 3, supplementary table 19 and 20).

To gain a better understanding of whether these post-acute prespecified outcomes are unique to COVID-19 itself or whether they represent a general post-viral syndrome, we then further conducted comparative analyses (adjusted as specified in methods including adjusting for severity of the acute infection) of the prespecified outcomes among people hospitalized with COVID-19 and seasonal influenza (Extended data table 1a, and supplementary table 6). The results show increased risk and excess burden of a broad array of symptoms and multiple organ involvement among people with COVID-19 (Extended data Fig. 8, supplementary table 21).

Negative exposure controls

In addition to testing negative outcome controls (described above and in extended data table 2a), and to further test robustness of our approach, we developed and tested a pair of negative exposure controls. We posited that exposure to influenza vaccination in odd and even months between October 1, 2017 and September 30, 2019 should be associated with similar risks of clinical outcomes. We therefore tested the associations between exposure to influenza vaccine in even (N=762,039) vs. odd months (N= 599,981) and the full complement of 821 high dimensional clinical outcomes considered in this study (including all diagnoses, medications, and laboratory test results). We used the same data sources, cohort building algorithm, variable definitions, analytic approach (including weighting method), outcomes specification, similar length of follow up, and interpretation. The results showed that none of the associations met the threshold of significance ($p < 6.57x10^{-5}$) considered in this study (Supplementary Fig. 6, supplementary table 22-24).

Discussion

In this study, we use a high dimensional approach to identify the spectrum of clinical abnormalities (incident diagnoses, incident medication use, and incident laboratory abnormalities) experienced by COVID-19 survivors beyond the first 30 days of illness. The results suggest that beyond the first 30 days of illness, people with COVID-19 are at higher risk of death, health care resource utilization, and exhibit a broad array of incident pulmonary and extrapulmonary clinical manifestations including nervous system and neurocognitive disorders, mental health disorders, metabolic disorders, cardiovascular disorders, gastrointestinal disorders, and signs and symptoms related to poor general wellbeing including malaise, fatigue, musculoskeletal pain, and anemia. Increased risk of incident use of several medication classes was also observed including pain medications (opioids and non-opioids), antidepressants, anxiolytics, antihypertensives, antihyperlipidemics, oral hypoglycemics, insulin, and other medication classes. Our analyses of pre-specified outcomes complement the high dimensional approach to identify specific post-acute sequalae with greater diagnostic resolution and reveal two key findings: a) the risk and associated burden of post-acute sequalae is evident even among those whose acute disease was not severe enough to necessitate hospitalization – the segment that represents the majority of people with COVID-19, and b) the risk and associated burden increases across the severity spectrum of the acute COVID-19 infection (non-hospitalized, hospitalized, admitted to intensive care). Our comparative approach to examine post-acute sequalae in those hospitalized with COVID-19 vs. seasonal influenza (using a high dimensional approach and through examination of pre-specified outcomes) suggests substantially higher burden of a broad array of post-acute sequelae in those hospitalized with COVID-19 vs. seasonal influenza – providing differentiating features of post-COVID-19 (both in magnitude of risk and breadth of organ involvement) from a post-influenza viral syndrome. The constellation of evidence suggests that 30-day survivors of COVID-19 exhibited increased risk of death and health resource utilization, and substantial burden of health loss (spanning pulmonary and several extrapulmonary organ systems) and highlights the need for a holistic and integrated multidisciplinary long-term care of COVID-19 survivors.

The mechanism(s) which underly the post-acute and chronic manifestations of COVID-19 are not entirely clear. Some of the manifestations may be driven by a direct effect of the viral infection and may be putatively explained by several hypotheses including persistent virus in immune-privileged sites, aberrant immune response, hyperactivation of the immune system, or autoimmunity[3]. Indirect effects including changes in social (e.g. reduced social contact and loneliness), economic (e.g. loss of employment), and behavioral conditions (e.g. changes in diet and exercise) that may be differentially experienced by people with COVID-19 may also shape health outcomes in COVID-19 survivors and may be responsible drivers of some of the clinical manifestations reported here[4–8]. A better delineation of the direct and indirect effects

and a deeper understanding of the underlying biologic mechanisms and epidemiologic drivers of the multifaceted long-term consequences of COVID-19 is needed[9].

To our knowledge, this is the largest post-acute COVID-19 study to date involving 73,435 non-hospitalized patients with COVID-19, and 4,990,835 controls (corresponding to 2,070,615.52 person years of follow-up), and 13,654 hospitalized patients with COVID-19, and 13,997 patients hospitalized with seasonal influenza (corresponding to 12,179.05 person years of follow-up). We leveraged the breadth and depth of the US Department of Veterans Affairs national health care databases – the largest nationally integrated healthcare delivery system in the US – to undertake a comprehensive high-dimensional comparative approach (relative to control groups) to identify the 6-months health outcomes and clinical manifestations in COVID-19 patients who survived the first 30 days of COVID-19 illness. We further examined risk in a pre-specified set of outcomes with higher diagnostic resolution across care settings to enable a deeper understanding of the clinical symptomatology and diagnoses of post-acute COVID-19 across the spectrum of severity of the acute phase of the infection.

This study has several limitations. While our approach identifies the incident post-acute sequalae in COVID-19 survivors, it does not delineate those that may be direct or indirect consequences of COVID-19 infection. Because of the predominantly male composition of the VA population, our findings may not identify clinical features of post-acute COVID-19 that may be differentially much more pronounced in females and either non-expressed or very rare in males. While our approach demonstrated balance for more than 1150 variables across several data domains (diagnoses, medications, and laboratory data), and yielded successful testing of negative exposure and outcome controls, we cannot completely rule out residual confounding. Finally, 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 may also change over time.

In conclusion, the findings show that beyond the first 30 days of illness, substantial burden of health loss – spanning pulmonary and several extrapulmonary organ systems – is experienced by survivors of the acute phase of COVID-19. Our results inform the global discussion on the post-acute manifestations of COVID-19; the findings provide a roadmap to inform health system planning and development of care strategies aimed at reducing chronic and permanent health loss and optimizing wellness among COVID-19 survivors.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03553-9.

1. Wang, D. *et al*. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *Jama*, https://doi.org/10.1001/jama.2020.1585 (2020).
2. 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*, https://doi.org/10.1136/bmj.m4677 (2021).
3. Long-term immunological health consequences of COVID-19. *British Society for Immunology* (2020).
4. Figueroa, J. D. *et al*. Distinguishing between direct and indirect consequences of covid-19. *Bmj* **369**, m2377, https://doi.org/10.1136/bmj.m2377 (2020).
5. Townsend, E. COVID-19 policies in the UK and consequences for mental health. *Lancet Psychiatry* **7**, 1014-1015, https://doi.org/10.1016/S2215-0366(20)30457-0 (2020).
6. Knipe, D., Evans, H., Marchant, A., Gunnell, D. & John, A. Mapping population mental health concerns related to COVID-19 and the consequences of physical distancing: a Google trends analysis. *Wellcome open research* **5**, 82, https://doi.org/10.12688/wellcomeopenres.15870.2 (2020).
7. Raker, E. J., Zacher, M. & Lowe, S. R. Lessons from Hurricane Katrina for predicting the indirect health consequences of the COVID-19 pandemic. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 12595-12597, https://doi.org/10.1073/pnas.2006706117 (2020).
8. Mahase, E. Covid-19: Mental health consequences of pandemic need urgent research, paper advises. *Bmj* **369**, m1515, https://doi.org/10.1136/bmj.m1515 (2020).
9. Del Rio, C., Collins, L. F. & Malani, P. Long-term Health Consequences of COVID-19. *Jama*, https://doi.org/10.1001/jama.2020.19719 (2020).

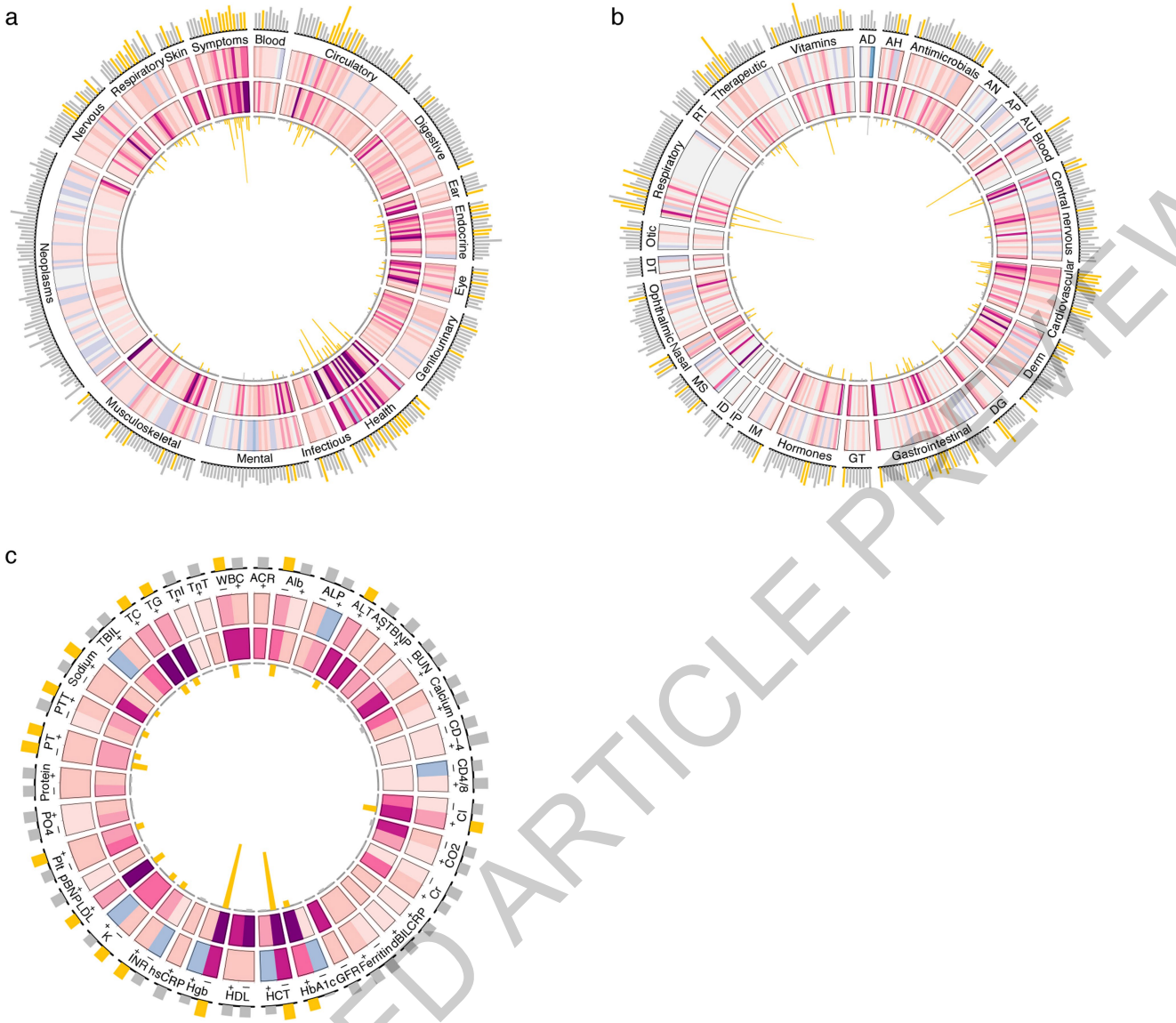


Fig. 1 | High dimensional identification of the incident post-acute sequalae of COVID-19. a) Incident diagnoses, (b) Incident medication use and (c) Incident laboratory abnormalities. All VHA users served as the referent category. Post-acute sequalae were ascertained from 30 days after infection until end of follow-up. Beginning from the outside ring, the first ring represents hazard ratios for the post-acute sequalae of COVID-19. A higher bar indicates larger hazard ratio. Hazard ratios with point estimate larger than one and statistically significant were colored in yellow. The second ring represents excess burden per 1000 COVID-19 patients at 6-months. Color of the cell indicates value of the excess burden, where deeper shades of red indicate higher excess burden and deeper shades of blue indicate greater reduced burden. The third ring represents the baseline incident rate in the control group, where deeper shades of red indicate higher incident rate. The fourth ring represents negative log of the P value, where a higher bar indicates smaller P value and yellow bar indicate statistically significant. ACR, Albumin/ Creatinine Ratio; AD, Antidotes; AH, Antihistamines; ALB, Albumin; ALP, Alkaline phosphatase; ALT, Alanine aminotransferase; AN, Antineoplastics; AP, Antiparasitics; AST, Aspartate aminotransferase; AU, Autonomic; BNP, Brain natriuretic peptide; BUN, Blood urea nitrogen; CD-4, CD-4-cell count; CD4/8, CD4/CD8 Ratio; CL, Cholesterol; CL, Chloride; CO2, Carbon dioxide; Cr, Creatinine; CRP, C-reactive protein; dBIL, Direct bilirubin; Derm, Dermatological; DG, Diagnostic; DT, Dental; GFR, glomerular filtration rate; GT, Genitourinary; HbA1c, Hemoglobin-A1c; HCT, Hematocrit; HDL, High-density lipoprotein cholesterol; Hgb, Hemoglobin; hsCRP, High-sensitivity C-reactive protein; ID, Irrigation/Dialysis; IM, Immunological; INR, International normalized ratio; IP, Intrapleural; K, Potassium; LDL, Low-density lipoprotein cholesterol; MS, Musculoskeletal; pBNP, Pro-B natriuretic peptide; Plt, Platelet; PO4, Phosphate; Protein, Total protein; PT, Prothrombin time; PTT, Partial thromboplastin time; RT, Rectal; TBIL, Total bilirubin; TG, Triglycerides; TnI, Troponin-I; TnT, Troponin-T; WBC, White blood cell.

a

Respiratory signs and symptoms	28.51 (26.40, 30.50)
Hypertension	15.18 (11.53, 18.62)
Sleep wake disorders	14.53 (11.53, 17.36)
Nervous system signs and symptoms	14.32 (12.16, 16.36)
Musculoskeletal pain, not low back pain	13.89 (9.89, 17.71)
Malaise and fatigue	12.64 (11.24, 13.93)
Disorders of lipid metabolism	12.32 (8.18, 16.24)
Chest pain	10.08 (8.63, 11.42)
Obesity	9.53 (7.55, 11.37)
Trauma– and stressor–related disorders	8.93 (6.62, 11.09)
Cardiac dysrhythmias	8.41 (7.18, 9.53)
Diabetes mellitus	8.23 (6.36, 9.95)
Skin disorders (Itch, rash, other)	7.52 (5.17, 9.73)
Esophageal disorders	6.90 (4.58, 9.07)
Circulatory signs and symptoms	6.65 (5.18, 8.01)
Abdominal pain	5.73 (3.70, 7.62)
Muscle disorders	5.73 (4.60, 6.74)
Anxiety and fear–related disorders	5.42 (3.42, 7.29)
Arthralgia and arthritis	5.16 (3.18, 7.01)
Nervous system disorders	4.85 (3.65, 5.93)
Anemia	4.79 (3.53, 5.93)
Lower respiratory disease	4.67 (3.96, 5.28)
Chronic obstructive pulmonary disease	4.44 (3.16, 5.59)
Genitourinary signs and symptoms	4.39 (2.98, 5.68)
Coronary atherosclerosis and other heart disease	4.38 (2.96, 5.67)
Headache	4.10 (2.49, 5.58)
Heart failure	3.94 (2.97, 4.80)
Gastrointestinal disorders	3.58 (2.15, 4.88)
Respiratory failure; insufficiency; arrest	3.37 (2.71, 3.92)
Neurocognitive disorders	3.17 (2.24, 3.98)
Acute phlebitis; thrombophlebitis; thromboembolism	3.05 (2.51, 3.49)
Urinary tract infections	2.99 (1.94, 3.93)
Dysphagia	2.83 (1.79, 3.76)
Asthma	2.82 (1.92, 3.61)
Acute pulmonary embolism	2.63 (2.25, 2.92)
Bacterial infections	2.38 (1.52, 3.13)
Pressure ulcer of skin	2.05 (1.40, 2.59)
Pleurisy, pleural effusion	1.52 (0.95, 1.98)

b

Bronchodilators, sympathomimetic, inhalation	22.23 (20.68, 23.67)
Non–opioid analgesics	19.97 (17.41, 22.40)
Anticoagulants	16.43 (14.85, 17.89)
Non–opioid–containing antitussives/expectorants	12.83 (11.61, 13.95)
Antilipemic agents	11.56 (8.73, 14.19)
NSAIDs	10.94 (8.04, 13.67)
Beta blockers	9.74 (8.06, 11.27)
Anti–inflammatory, topical	9.63 (6.74, 12.37)
Opioid analgesics	9.39 (7.21, 11.43)
Laxatives	9.22 (6.99, 11.31)
Antiasthmatics	8.87 (7.65, 9.97)
Antidepressants	7.83 (5.19, 10.30)
Vitamin D	7.80 (5.36, 10.09)
Glucocorticoids	7.65 (5.67, 9.50)
Vitamin C	7.23 (6.45, 7.90)
Calcium channel blockers	7.18 (5.61, 8.61)
Anti–inflammatories, nasal	6.33 (4.57, 7.96)
Anticonvulsants	5.78 (3.68, 7.72)
Oral hypoglycemic agents	5.39 (3.99, 6.64)
Antifungal, topical	5.10 (3.37, 6.69)
Insulin	4.95 (3.87, 5.90)
Zinc	4.90 (4.39, 5.32)
Penicillins	4.87 (3.15, 6.44)
Histamine antagonists	4.83 (3.63, 5.91)
Skeletal muscle relaxants	4.78 (2.62, 6.79)
Loop diuretics	4.72 (3.59, 5.72)
Nasal and throat, topical	4.13 (3.09, 5.05)
Potassium	3.72 (2.35, 4.96)
Iron	3.57 (2.46, 4.56)
Magnesium	3.36 (2.27, 4.32)
Antiemetics	3.07 (1.66, 4.36)
Cyanocobalamin	2.98 (1.69, 4.14)
Antidiarrheal agents	2.87 (1.70, 3.91)
Thiazides/related diuretics	2.52 (1.37, 3.54)
Vaccines	2.43 (1.43, 3.31)
Multivitamins	2.31 (1.40, 3.10)
Anti–inflammatories, inhalation	1.37 (0.80, 1.83)
Antiarrhythmics	1.28 (0.79, 1.67)
Magnesium containing antacids	1.07 (0.62, 1.42)

c

Hemoglobin lower than M: 14; F: 12 g/dL	31.03 (28.16, 33.76)
Hematocrit lower than M: 42; F: 37 %	30.73 (27.64, 33.67)
Hemoglobin A1C higher than 5.6%	10.66 (6.77, 14.35)
Triglycerides higher than 150 mg/dL	9.94 (6.61, 13.11)
Low density lipoprotein higher than 130 mg/dL	9.48 (7.02, 11.81)
Total cholesterol higher than 200 mg/dL	9.40 (6.63, 12.03)
Serum chloride higher than 107 mmol/L	9.21 (7.05, 11.24)
Total white blood cell count lower than 4.8 K/cmm	8.45 (6.47, 10.29)
Alanine aminotransferase higher than 40 U/L	7.62 (5.20, 9.90)
Serum albumin lower than 3.5 g/dL	6.44 (4.84, 7.92)
Prothrombin time lower than 11.5 sec	4.81 (3.88, 5.64)
Serum potassium lower than 3.5 mmol/L	4.44 (2.92, 5.85)
Platelet count higher than 400 K/cmm	3.05 (2.10, 3.88)
Prothrombin time higher than 14.7 sec	2.99 (2.00, 3.86)
International normalized ratio higher than 1.2 ratio	2.94 (1.96, 3.80)
Partial thromboplastin time higher than 36.5 SEC	2.66 (1.75, 3.46)
Serum sodium higher than 145 mmol/L	1.64 (0.97, 2.20)

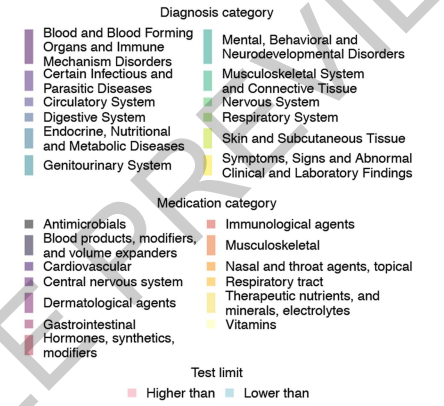


Fig. 2 | Burden of post-acute sequalae of COVID-19 per 1000 persons at 6-month. (a) Incident diagnoses, (b) Incident medication use and (c) Incident laboratory abnormalities. All VHA users served as the referent category. Post-acute sequalae were ascertained from 30 days after infection until end of follow-up. Sequalae were selected based on hazard ratio larger than one, P value less than 6.57×10^{-5}. Excess burdens per 1000 COVID-19 patients at 6-months are presented. Within each domain, outcomes are ranked based on excess burden from high to low. Diagnoses are colored based on diagnosis group, medications are colored based on medication class, and laboratory abnormalities are colored based on higher or lower than normal range.

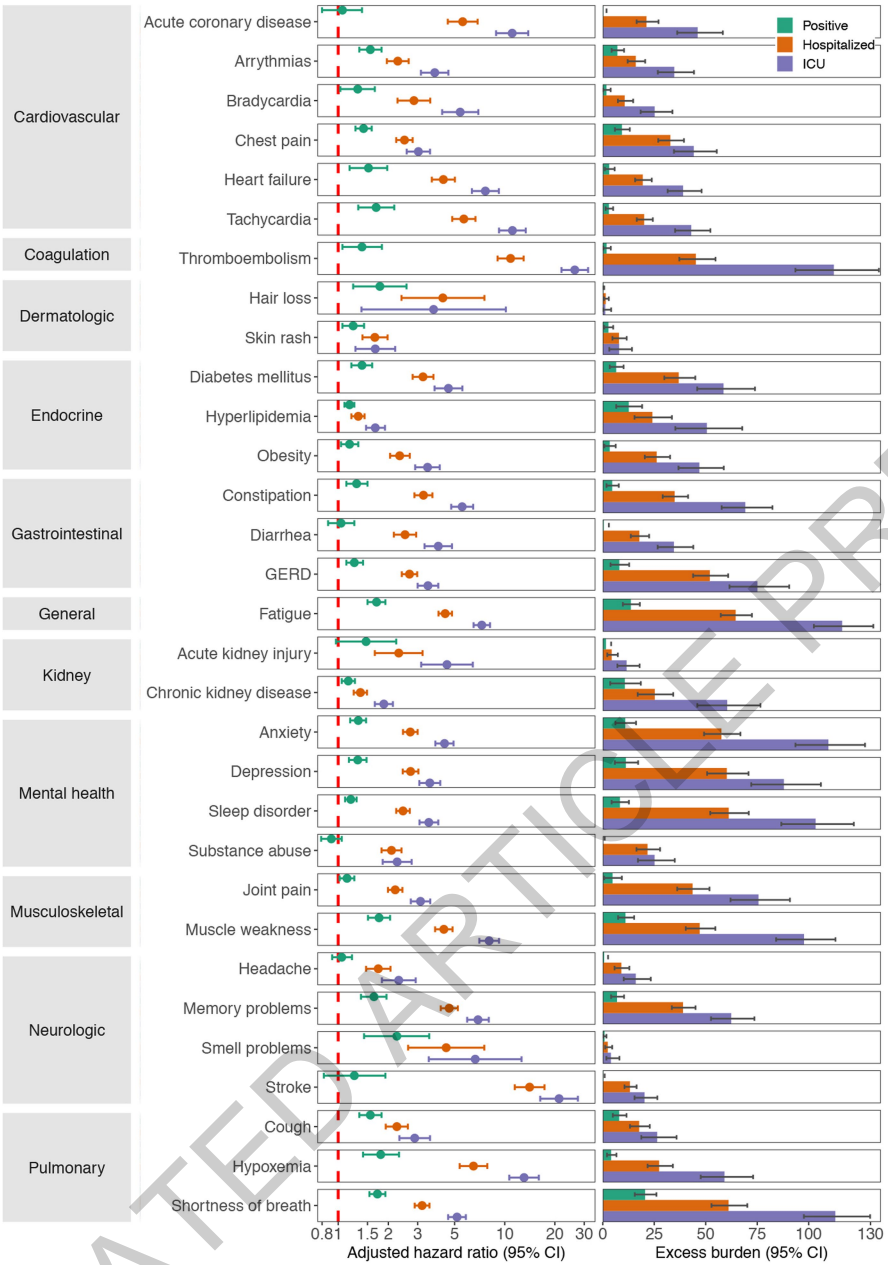


Fig. 3 | Risks and burdens of incident pre-specified high resolution post-acute COVID-19 outcomes at 6 months in mutually exclusive cohorts of people with non-hospitalized COVID-19 (green), people hospitalized for COVID-19 (orange), and people admitted to intensive care for COVID-19 (blue) during the acute phase (first 30 days) of the infection. ; all users of the Veteran Health Administration healthcare system served as the referent category. Outcomes were ascertained from day 30 after COVID-19 diagnosis until end of follow-up. Hazard ratios and 95% confidence intervals and excess burdens per 1000 patients and 95% confidence intervals at 6-months are presented.

Methods

Setting

Users of the US Veteran Health Administration (VHA) were selected from US Department of Veterans Affairs (VA) electronic health care databases. The VHA provides health care to discharged veterans of the US armed forces and operates the largest nationally integrated healthcare system in the United States, with 1,255 health care facilities, including 170 VA Medical Centers and 1,074 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 update daily.

Cohort

The cohort was constructed from 5,808,018 participants who had encountered the VHA between January 01, 2019 and December 31, 2019. Within those alive on March 01, 2020 (N= 5,606,309), a COVID-19 group was selected as those with a COVID-19 positive test between March 01, 2020 and November 30, 2020 (n=98,661). Participants without hospitalization within the first 30 days of their first positive test were further selected (N=76,877). To examine post-acute outcome, we then selected from the COVID-19 group those alive at 30th day after their positive test (COVID-19 participants n=73,435). To generate a comparison group that had a similar distribution of length of follow up, we then matched each COVID-19 participant with 70 VHA users who did not have COVID-19 positive test. In matching, the corresponding 70 VHA users dates of cohort enrollment were matched without replacement with the COVID-19 participant's time of cohort enrollment--the date of testing positive (control group n=5,140,450). In the VHA users group we similarly selected those who were without hospitalization and alive during the first 30 days after the date of enrollment (control group n=4,990,835) (Supplementary Fig. 1a and b). Participants were followed until January 31, 2021.

To compare post-acute outcomes of hospitalized COVID-19 and hospitalized seasonal influenza participants, 15,846 COVID-19 participants who were admitted to a hospital within 30 days after or 5 days before their first positive test were selected from the 98,661 patients with positive COVID-19 test between March 01, 2020 and November 30, 2020. Similarly, 62,909 patients with their first positive seasonal influenza test between October 01, 2016 and February 29, 2020 who encountered the VHA at least once in the calendar year before the test were collected. Within them, 14,948 seasonal influenza patients were admitted to a hospital within 30 days after or 5 days before their first positive influenza test. The hospitalized cohort was further restricted to those alive at 30th day after hospital admission (COVID-19 n=13,654 and seasonal influenza n=14,212). For 215 patients in both hospitalized COVID-19 and seasonal influenza group, only their COVID-19 hospitalizations were used in the analyses (Supplementary Fig. 3a and b). In this cohort, participants were considered enrolled at the time of hospitalization. To balance the duration of follow-up in the hospitalized COVID-19 and seasonal influenza groups, each participant in the seasonal influenza group was independently randomly assigned a duration of follow-up based on the distribution of length of follow up in the hospitalized COVID-19 participants that were followed from date of hospitalization to January 31, 2021.

To examine high resolution pre-specified post-acute COVID-19 outcomes across the severity spectrum of the initial acute disease, we built 4 mutually exclusive cohorts: VHA users without COVID-19 (N=4,990,835), Veterans with COVID-19 (N=73,435), Veterans hospitalized with COVID-19 within the first 30 days of follow up (N=10,068), and Veterans with COVID-19 admitted to the intensive care within the first 30 days of follow up (N=3586). Participants in these cohorts were followed up until January 31, 2021.

Data sources

Electronic health records from VA Corporate Data Warehouse (CDW) were used in this study[10–13]. The CDW Outpatient Encounters domains provided information related to outpatient encounters and Inpatient Encounters domains provided information between hospital admission and discharge[14]. The CDW Outpatient Pharmacy domain and CDW Bar Code Medication Administration domain were used to collect medication data and 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 test and demographic information for COVID-19 patients. In addition, the Area Deprivation Index (ADI), a composite measure of income, education, employment, and housing was obtained from the University of Wisconsin[15].

Post-acute health resource utilization and death

Outcomes which occurred after 30 days of cohort enrollment including death, incident outpatient encounter and frequency of outpatient encounter were examined in both cohorts. Frequency of outpatient encounters was computed based on the number of days with outpatient encounter over days of follow up after 30 days and was reported as number of outpatient encounters per 30 days.

High dimensional post-acute clinical characteristics

Negative outcome and exposure controls. The application of negative controls in clinical epidemiology may help detect both suspected and unsuspected sources of spurious bias and may lessen concerns about unmeasured confounding and other latent biases[16]. In this work, we followed the approach outlined by Lipsitch and colleagues to examine a panel of 8 negative outcome controls (including neoplasms, accidental injuries, scars, fitting or adjustment of orthodontic or dental prosthetic device, fitting or adjustment of hearing device, fitting or adjustment or orthotics, fitting or adjustment of casts, and bandages), where – based on current knowledge – there should be no causal relation between the exposures and risks of the negative outcome controls. We also developed and tested a pair of negative exposure controls defined as exposure to influenza vaccine in odd or even months during the period between October 1, 2017 and September 30, 2019. We posited that there should be no differences in risk of clinical outcomes associated with receipt in influenza vaccine in odd vs even months. The negative exposure controls were tested in all 821 high dimensional outcomes considered in our analyses including diagnoses, medications, and laboratory test results; we used the same data sources, cohort building algorithm, variable definitions, analytic approaches, outcomes specification, similar length of follow up, and interpretation. In the assessment of negative outcome and negative exposure controls, the relation of the exposure-outcome pairs may share the same potential biases with COVID-19 and the outcomes examined in this study including biases in the underlying data, algorithms for construction of cohorts, unmeasured confounders, misspecification of modeling algorithms, outcome ascertainments, analytic considerations, result interpretation, and other latent biases[16,17]. Successful testing of negative controls reduces concerns about both suspected and unsuspected sources of spurious associations, including associations due unmeasured confounding, flaws in the analytic approach, differences in outcome ascertainment, and other sources of bias[16]. In particular, the successful testing of the outcome controls may reduce concerns about biases in outcome ascertainment and unmeasured confounding between the comparison groups (for example, if there was bias in ascertainment of clinical outcomes in one arm vs another, that bias may also extend to ascertainment of neoplasms, accidental injuries or other negative outcome controls tested in this study); the successful testing of the exposure control may reduce concerns about biases in the analytic approach and underlying data (for example, if

there was bias related to the analytic approach, it may also bias the negative exposure control).

Diagnoses

All ICD-10 diagnosis codes from cohort participants from days 30 after COVID-19 diagnosis until end of follow-up were used to define the post-acute diagnosis outcomes. More than 70,000 ICD-10 diagnosis codes were classified into 540 diagnostic categories based on the Clinical Classifications Software Refined (CCSR) version 2021.1, which is developed as part of the Healthcare Cost and Utilization Project sponsored by the Agency for Healthcare Research and Quality[18–20]. We only examined diagnostic categories that may be plausibly be considered post-acute sequalae of COVID-19 in the adult population. Diagnostic categories including external causes of morbidity, injury, poisoning and certain other consequences of external causes, congenital malformations, deformations and chromosomal abnormalities, certain conditions originating in the perinatal period or outcome from pregnancy, childbirth and the puerperium were not examined, yielding 379 diagnostic categories.

Medication use

Cohort participants' prescription records from day 30 after COVID-19 diagnosis until end of follow-up were used to define the post-acute medication use. 3425 medications were classified based on the US Department of Veterans Affairs Drug Classification system into 543 medication classes[21,22]. After removing items in the medication group of investigational agents or prosthetics, supplies and devices, in total 380 different medication outcomes were examined.

Laboratory abnormalities

In total 62 laboratory test abnormalities from 38 laboratory measurements from day 30 after COVID-19 diagnosis until end of follow-up were examined including absolute T cell count, alanine aminotransferase, aspartate aminotransferase, blood urea nitrogen, brain natriuretic peptide, C reactive protein, carbon dioxide, CD4/CD8 ratio, direct bilirubin, estimated glomerular filtration rate, ferritin, hematocrit, hemoglobin, hemoglobin A1c, high density lipoprotein cholesterol, high-sensitivity C-reactive protein, international normalized ratio, low density lipoprotein cholesterol, microalbumin/creatinine ratio, partial thromboplastin time, platelet count, pro B natriuretic peptide, prothrombin time, serum albumin, serum alkaline phosphatase, serum calcium, serum chloride, serum creatinine, serum phosphate, serum potassium, serum sodium, serum total protein, total bilirubin, total cholesterol, total white blood cell count, triglycerides, troponin I and troponin T were identified based on Logical Observation Identifiers Names and Codes (LOINC). Each laboratory test result was classified into abnormally high or abnormally low based on whether results were above the upper normal range or below the lower normal range, in the instance where for a given lab a high, or low, result might be clinically possible. The definition of the abnormality for each laboratory test is presented in supplementary table 4 and 9.

High resolution pre-specified post-acute COVID-19 outcomes

To identify clinical manifestations of post-acute COVID-19 with greater diagnostic resolution, we specified a list of outcomes based on data from the Center of Disease Control and the National Institute of Health workshop on post-acute COVID-19. Outcomes were defined based on definitions validated for use with electronic health records, and integrated information from diagnoses, medications and laboratory measurements when appropriate[23–29]. To gain a deeper understanding of the risks of these outcomes across the severity scale of the acute infection, we examined the risk across the care setting of the acute disease – a proxy indicator of clinical severity in 4 mutually exclusive cohorts (VHA users which served as the referent category, people with COVID-19, people hospitalized for COVID-19, and people admitted to intensive care for COVID-19). In addition, we estimated the risks of these pre-specified outcomes in those hospitalized with COVID-19 and seasonal influenza. The pre-specified high resolution outcomes included acute coronary disease, acute kidney injury, anxiety, arrythmias, bradycardia, chest pain, chronic kidney disease, constipation, cough, depression, diarrhea, type 2 diabetes mellitus, fatigue, gastric esophageal reflux disease, hair loss, headache, heart failure, hyperlipidemia, hypoxemia, joint pain, memory problems, muscle weakness, obesity, shortness of breath, skin rash, sleep disorder, smell disorder, stroke, tachycardia, and thromboembolism. We restricted capture of incident acute coronary disease, stroke and thromboembolism to inpatient diagnoses that were not present on admission. All other prespecified outcomes which may plausibly be encountered in either the outpatient or inpatient setting were accordingly ascertained in the setting in which they first occurred. Among individuals with COVID-19, and for each prespecified outcome, the percentages of outcome which were ascertained from outpatient and inpatient data are presented in supplementary table 19 and 20.

Covariates

Predefined covariates for analyses included demographics such as age, race (white, black, and other), sex, receipt of long-term care; proxies of healthcare utilization such as number of outpatient encounters, number of hospital admissions, number of outpatient prescriptions and number of outpatient eGFR measurements in the year before enrollment. In addition, area deprivation index at patients' residency address as a summary measurement of socio-economic deprivation was included. Sequential Organ Failure Assessment (SOFA) score was employed to adjust for severity of the acute infection in additional high dimensional analyses of the hospitalized COVID-19 vs. hospitalized seasonal influenza cohort[30,31]. To address for the potential non-linear association, all continuous variables were adjusted as restricted cubic spline functions.

To most optimally further adjust the models, we leveraged the multidimensionality of the VA's electronic health care databases to algorithmically identify covariates (potential confounders) spanning multiple domains (diagnoses, pharmacy records, laboratory tests) that showed evidence of difference in prevalence between the comparison groups[32]. In COVID-19 vs. VHA users' cohort, and separately in hospitalized COVID-19 vs. influenza cohort, high dimensional covariates were ascertained within one year before the date of enrollment. Within all diagnoses, medication classes and laboratory tests, we first selected variables that occurs in at least 10 patients in both groups. We then estimated the unadjusted relative risk of each variables with being in the COVID-19 or comparator group. The top 100 high dimensional variables with strongest association with group membership were used along with predefined covariates in the analyses.

To most optimally estimate risk of the set of pre-specified outcomes across the intensity of care needed during the acute infection, in total four sets of high dimensional covariates (corresponding to the four mutually exclusive groups—all VHA users, people with COVID-19, people hospitalized with COVID-19, and people admitted to intensive care with COVID-19) were ascertained based on the unadjusted relative risk of being in each group compared to being in the remaining 3 groups. High dimensional covariates were used along with predefined covariates in the analyses[33].

Statistical analyses

Characteristics of the COVID-19 positive VHA users, VHA users without COVID-19, hospitalized COVID-19 participants and hospitalized seasonal influenza participants were described. The flowcharts of the overall analytic approach are presented in supplementary Fig. 7 and 8.

We estimated the risk of health resource utilization, death, and risk of each diagnosis, medication use and laboratory abnormality between COVID-19 and all VHA users, and separately, between those who had

been hospitalized for COVID-19 and seasonal influenza. To estimate the risk of each incident outcome, we built a cohort of participants without a history of the outcome being examined (for example, risk of insulin use was estimated within a cohort of participants without history of insulin use in the year prior to cohort enrollment). For each outcome specific cohort, propensity scores based on predefined variables and high dimensional algorithmically selected variables were estimated. The propensity scores were then used to compute the overlap weight, which is the probability of membership in the non-observed exposure group (one minus the propensity of in the observed group)[34,35]. We then for all outcome models assessed covariate balance, calculating the standardized difference after application of the overlap weight for all predefined variables, 100 algorithmically selected high dimensional variables, and all high dimensional variables not selected for inclusion in the propensity score models. We present the distribution of these standardized differences for 20 randomly selected outcome specific cohorts, and across all outcomes, and the covariate distributions in overall cohort after adjustment.

Risks of health resource utilizations including outpatient encounter and death between COVID-19 and all VHA users, and between COVID-19 hospitalization and influenza hospitalization were estimated from Cox survival model weighted by overlap weights, where death was considered as a competing risk in the evaluation of health resource utilizations. Frequency of outpatient encounter was modeled based on weighted linear regression. Hazard ratios for each of the outcomes including incident diagnoses, incident medication use, and incident laboratory abnormalities were estimated from cause specific hazard models weighted by overlap weights, where occurrence of death was considered as a competing risk. Event rates per 1000 participants at 6-months (180 days) of follow up in each group, and the adjusted excess burden based on the differences between two groups were estimated. Models were only built for outcomes occurring in at least ten participants from each group. Bonferroni correction was applied in consideration of multiple hypotheses testing for high dimensional outcomes. A P-values of less than $6.57x10^{-5}$ was considered statistically significant. Results are additionally presented with a focus on identified post-acute sequalae of COVID-19, where we selected those with hazard ratio greater than 1, P-values of less than $6.57x10^{-5}$. High dimensional analyses of hospitalized COVID-19 vs. seasonal influenza which additionally adjusted for severity of the acute infection (through inclusion of SOFA scores) were additionally undertaken. In addition, high dimensional analyses were also conducted to evaluate the risk of 6-months clinical outcomes in people hospitalized COVID-19 vs. those hospitalized for other causes. Participants hospitalized for other causes who survived the first 30 days after hospital admission were enrolled between October 01, 2016 and February 29, 2020 (N= 901,516).

We examined the risk of high resolution pre-specified outcomes across care settings of the acute phase of the disease, analyzing differences in risk of clinical manifestations of post-acute COVID-19 between mutually exclusive groups of COVID-19 positive people that were non-hospitalized, hospitalized, and admitted to intensive care, and VHA users who were not COVID-19 positive. Propensity scores for group membership were estimated in outcome specific cohorts free of the related disease at baseline[33]. Standardized differences in the predefined and algorithmically selected high-dimensional covariates are presented after application of overlap weighting[36]. The percentage of outcomes ascertained in the COVID-19 group in an inpatient and outpatient setting are presented. Cox survival models were then constructed to analyze the risk of outcomes using overlap weighting for multiple treatments. Hazard ratios, and event rate differences between each group are reported. We also estimated the risks of pre-specified outcomes among those hospitalized with COVID-19 and seasonal influenza, where SOFA scores were additionally adjusted for.

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 was approved by the Institutional Review Board of the Department of Veterans Affairs St. Louis Health Care System, St. Louis, MO.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data Availability

The data that support the findings of this study are available from the US Department of Veterans Affairs. VA data are made freely available to researchers behind the VA firewall with an approved VA study protocol. For more information, please visit https://www.virec.research.va.gov or contact the VA Information Resource Center (VIReC) at VIReC@va.gov

Code Availability

SAS and R programing codes are available https://github.com/yxie618/HDlongCOVID

10. Xie, Y. *et al*. Proton Pump Inhibitors and Risk of Incident CKD and Progression to ESRD. *Journal of the American Society of Nephrology : JASN* **27**, 3153-3163, https://doi.org/10.1681/ASN.2015121377 (2016).
11. 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, https://doi.org/10.1136/bmjopen-2016-015735 (2017).
12. Xie, Y. *et al*. Long-term kidney outcomes among users of proton pump inhibitors without intervening acute kidney injury. *Kidney international* **91**, 1482-1494, https://doi.org/10.1016/j.kint.2016.12.021 (2017).
13. Xie, Y. *et al*. Higher blood urea nitrogen is associated with increased risk of incident diabetes mellitus. *Kidney international* **93**, 741-752, https://doi.org/10.1016/j.kint.2017.08.033 (2018).
14. 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 Services and Outcomes Research Methodology* **18**, 143-154 (2018).
15. Kind, A. J. & Buckingham, W. R. Making Neighborhood-Disadvantage Metrics Accessible—The Neighborhood Atlas. *New England Journal of Medicine* **378**, 2456-2458 (2018).
16. Lipsitch, M., Tchetgen Tchetgen, E. & Cohen, T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* **21**, 383-388, https://doi.org/10.1097/EDE.0b013e3181d61eeb (2010).
17. Shi, X., Miao, W. & Tchetgen, E. T. A Selective Review of Negative Control Methods in Epidemiology. *Current epidemiology reports* **7**, 190-202, https://doi.org/10.1007/s40471-020-00243-4 (2020).
18. Wei, Y. *et al*. Short term exposure to fine particulate matter and hospital admission risks and costs in the Medicare population: time stratified, case crossover study. *Bmj* **367**, l6258, https://doi.org/10.1136/bmj.l6258 (2019).
19. Aubert, C. E. *et al*. Best Definitions of Multimorbidity to Identify Patients With High Health Care Resource Utilization. *Mayo Clin Proc Innov Qual Outcomes* **4**, 40-49, https://doi.org/10.1016/j.mayocpiqo.2019.09.002 (2020).
20. HCUP CCSR. Healthcare cost and utilization project (HCUP). Agency for Healthcare Research and Quality, Rockville, MD.
21. Olvey, E. L., Clauschee, S. & Malone, D. C. Comparison of critical drug-drug interaction listings: the Department of Veterans Affairs medical system and standard reference compendia. *Clin Pharmacol Ther* **87**, 48-51, https://doi.org/10.1038/clpt.2009.198 (2010).
22. Greene, M., Steinman, M. A., McNicholl, I. R. & Valcour, V. Polypharmacy, drug-drug interactions, and potentially inappropriate medications in older adults with human immunodeficiency virus infection. *Journal of the American Geriatrics Society* **62**, 447-453, https://doi.org/10.1111/jgs.12695 (2014).
23. 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, https://doi.org/10.1136/bmj.l1580 (2019).
24. 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, https://doi.org/10.2337/dc20-1890 (2020).
25. 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, https://doi.org/10.2337/dc20-1231 (2020).
26. Bowe, B. *et al*. Acute Kidney Injury in a National Cohort of Hospitalized US Veterans with COVID-19. *Clinical journal of the American Society of Nephrology : CJASN*, https://doi.org/10.2215/CJN.09610620 (2020).
27. Bowe, B. *et al*. The 2016 global and national burden of diabetes mellitus attributable to PM2.5 air pollution. *The Lancet. Planetary health* **2**, e301-e312, https://doi.org/10.1016/S2542-5196(18)30140-2 (2018).
28. 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 international* **89**, 886-896, https://doi.org/10.1016/j.kint.2015.12.034 (2016).

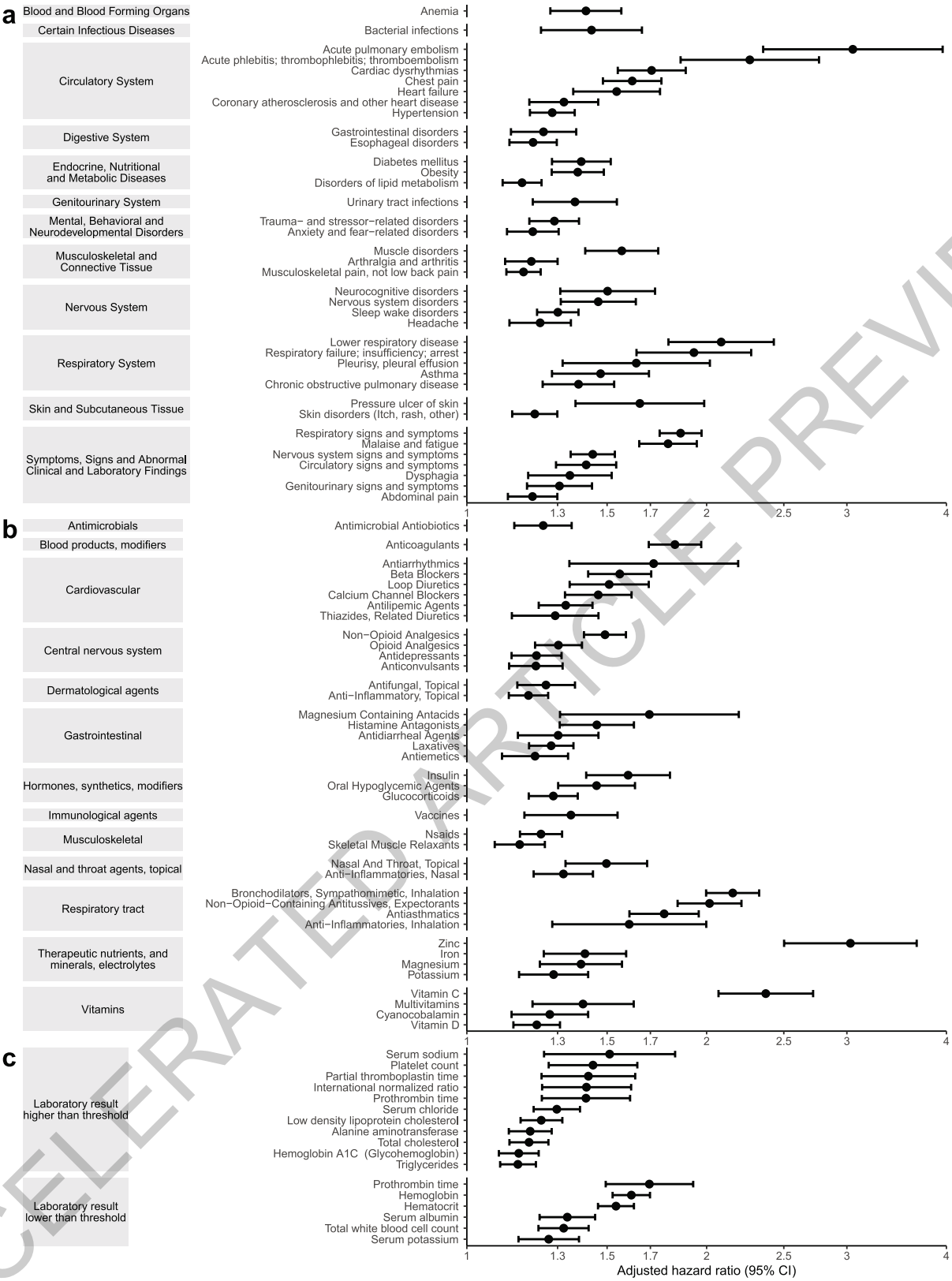
29. Bowe, B. *et al*. High Density Lipoprotein Cholesterol and the Risk of All-Cause Mortality among U.S. Veterans. *Clinical journal of the American Society of Nephrology : CJASN* **11**, 1784-1793, https://doi.org/10.2215/CJN.00730116 (2016).
30. Vincent, J. L. *et al*. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* **22**, 707-710, https://doi.org/10.1007/BF01709751 (1996).
31. Vincent, J. L. *et al*. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Crit Care Med* **26**, 1793-1800, https://doi.org/10.1097/00003246-199811000-00016 (1998).
32. Xie Y *et al*. Comparative Effectiveness of SGLT2 Inhibitors, GLP1 Receptor Agonists, DPP4 Inhibitors and Sulfonylureas on Risk of Kidney Outcomes: emulation of a target trial using healthcare databases. *Diabetes care*, https://doi.org/10.2337/dc20-1890.
33. McCaffrey, D. F. *et al*. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in medicine* **32**, 3388-3414, https://doi.org/10.1002/sim.5753 (2013).
34. Thomas, L. E., Li, F. & Pencina, M. J. Overlap Weighting: A Propensity Score Method That Mimics Attributes of a Randomized Clinical Trial. *Jama* **323**, 2417-2418, https://doi.org/10.1001/jama.2020.7819 (2020).
35. Li, F., Morgan, K. L. & Zaslavsky, A. M. Balancing Covariates via Propensity Score Weighting. *Journal of the American Statistical Association* **113**, 390-400, https://doi.org/10.1080/01621459.2016.1260466 (2018).
36. Li, F. & Li, F. Propensity score weighting for causal inference with multiple treatments. *Ann. Appl. Stat*. **13**, 2389-2415, https://doi.org/10.1214/19-AOAS1282 (2019).

Acknowledgements This study used data from the VA COVID-19 Shared Data Resource. Miao Cai developed the data visualization and Andrew K. Gibson provided technical and editorial assistance. This research was funded by the United States Department of Veterans Affairs and the Institute for Public Health at Washington University in Saint Louis, Missouri, USA (for ZAA), and two American Society of Nephrology and KidneyCure fellowship awards (for YX and BB). The funders of this study had no role in study design; collection, analysis, and interpretation of data; writing the report; and the decision to submit the report for publication. This research project was reviewed and approved by the Institutional Review Board of the Department of Veterans Affairs Saint Louis Health Care System.

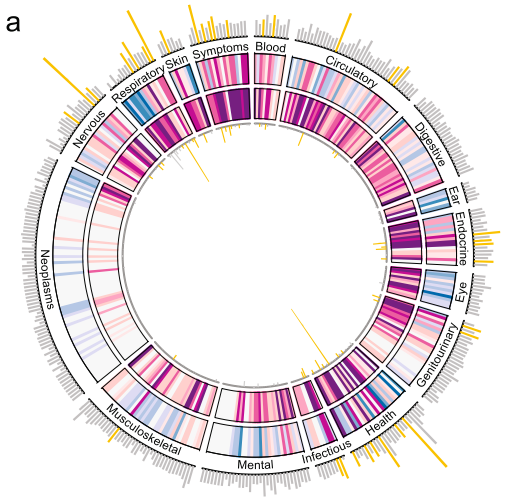
Author contributions ZAA, YX, and BB contributed to the development of the study concept and design. YX and BB contributed to data acquisition. ZAA, YX, and BB contributed to data analysis and interpretation. YX and BB contributed to statistical analysis. ZAA and YX drafted the manuscript. ZAA, YX, and BB contributed to critical revision of the manuscript. ZAA provided administrative, technical, and material support. ZAA provided 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.

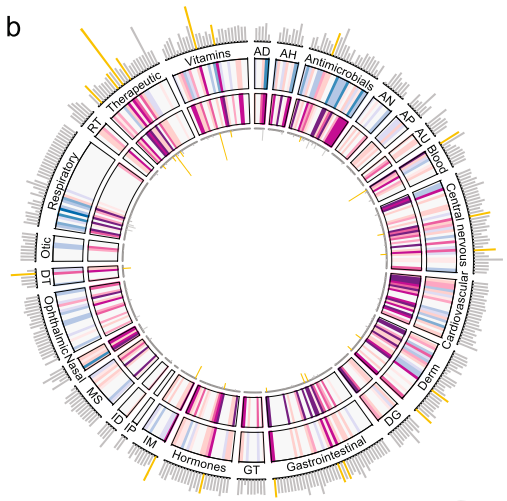
Competing interests The authors declare no competing interests.

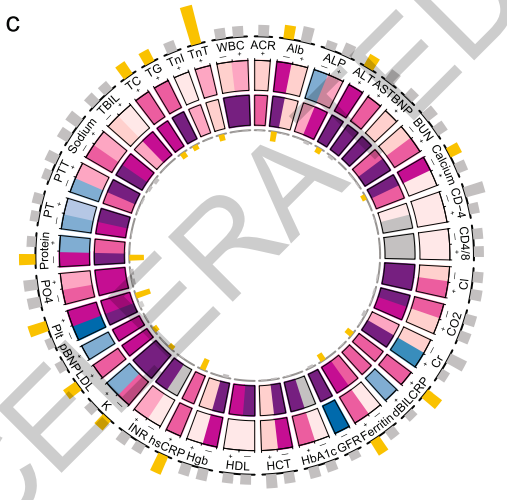
Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03553-9.
Correspondence and requests for materials should be addressed to Z.A.-A.
Peer review information *Nature* thanks Paul Garner, Sachin Yende and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



Extended Data Fig. 1 | Risk of incident post-acute sequalae in COVID-19. (a) Incident diagnoses, (b) Incident medication use and (c) Incident laboratory abnormalities. All users of the Veteran Health Administration healthcare system served as the referent category. Outcomes were ascertained from day 30 after COVID-19 diagnosis until end of follow-up. Adjusted hazard ratios for incident sequalae that are larger than one and P value less than 6.57×10^{-5} are presented. Hazard ratios (dots) and 95% confidence intervals (bars) are presented on log 10 scale.







d

Respiratory failure; insufficiency; arrest	70.38 (64.60, 75.63)
Disorders of lipid metabolism	43.53 (28.71, 57.08)
Fluid and electrolyte disorders	40.23 (31.39, 48.33)
Malaise and fatigue	36.49 (28.13, 44.15)
Acute renal failure	27.93 (20.67, 34.50)
Bacterial infections	27.38 (19.39, 34.68)
Malnutrition	24.08 (19.99, 27.59)
Septicemia	23.63 (18.52, 28.14)
Muscle disorders	22.77 (15.11, 29.72)
Pressure ulcer of skin	20.98 (16.53, 24.79)
Shock	19.99 (17.39, 22.07)
Nutritional and metabolic disorders	19.92 (13.71, 25.41)
Nervous system disorders	19.78 (12.58, 26.19)
Dysphagia	19.28 (12.75, 25.13)
Anemia	19.08 (10.58, 26.81)
Urinary tract infections	18.48 (12.52, 23.74)
Acute pulmonary embolism	18.31 (15.83, 20.25)
Circulatory signs and symptoms	17.92 (10.73, 24.35)
Neurocognitive disorders	16.16 (10.40, 21.19)
Hypotension	15.28 (9.21, 20.64)
Coagulation and hemorrhagic disorders	14.31 (10.08, 17.89)
Lower respiratory disease	12.80 (7.08, 17.81)
Acute phlebitis; thrombophlebitis; thromboembolism	12.76 (9.28, 15.63)
Postprocedural respiratory system complication	10.42 (8.99, 11.42)
Aspiration pneumonitis	7.88 (4.89, 10.23)
Symptoms of mental and substance use conditions	7.75 (4.72, 10.10)
Myopathies	7.39 (6.17, 8.12)

Diagnosis category
- Blood and Blood Forming Organs and Immune Mechanism Disorders
- Certain Infectious and Parasitic Diseases
- Circulatory System
- Endocrine, Nutritional and Metabolic Diseases
- Genitourinary System
- Musculoskeletal System and Connective Tissue
- Nervous System
- Respiratory System
- Skin and Subcutaneous Tissue
- Symptoms, Signs and Abnormal Clinical and Laboratory Findings

e

Anticoagulants	69.22 (56.74, 80.91)
Laxatives	35.29 (24.56, 45.24)
Vitamin C	27.83 (24.82, 30.33)
Vitamin D	25.85 (17.04, 33.74)
Gastric medications	25.82 (14.03, 36.53)
Vaccines	22.42 (17.87, 26.33)
Local anesthetics, topical	22.24 (14.76, 28.93)
Potassium	22.13 (13.38, 30.08)
Zinc	21.71 (20.34, 22.63)
Insulin	20.65 (13.05, 27.42)
Benzodiazepine sedatives/hypnotics anxiolytics	18.26 (12.12, 23.67)
Dermatologicals, topical	14.69 (9.09, 19.58)
Antipsychotics	14.43 (9.38, 18.79)
Mouthwashes	13.84 (9.92, 17.14)
Phosphorus	10.43 (7.03, 13.18)
Cephalosporins	9.47 (5.89, 12.42)

Medication category
- Antimicrobials
- Blood products, modifiers, and volume expanders
- Central nervous system
- Dental and oral agents, topical
- Dermatological agents
- Gastrointestinal
- Hormones, synthetics, modifiers
- Immunological agents
- Therapeutic nutrients, and minerals, electrolytes
- Vitamins

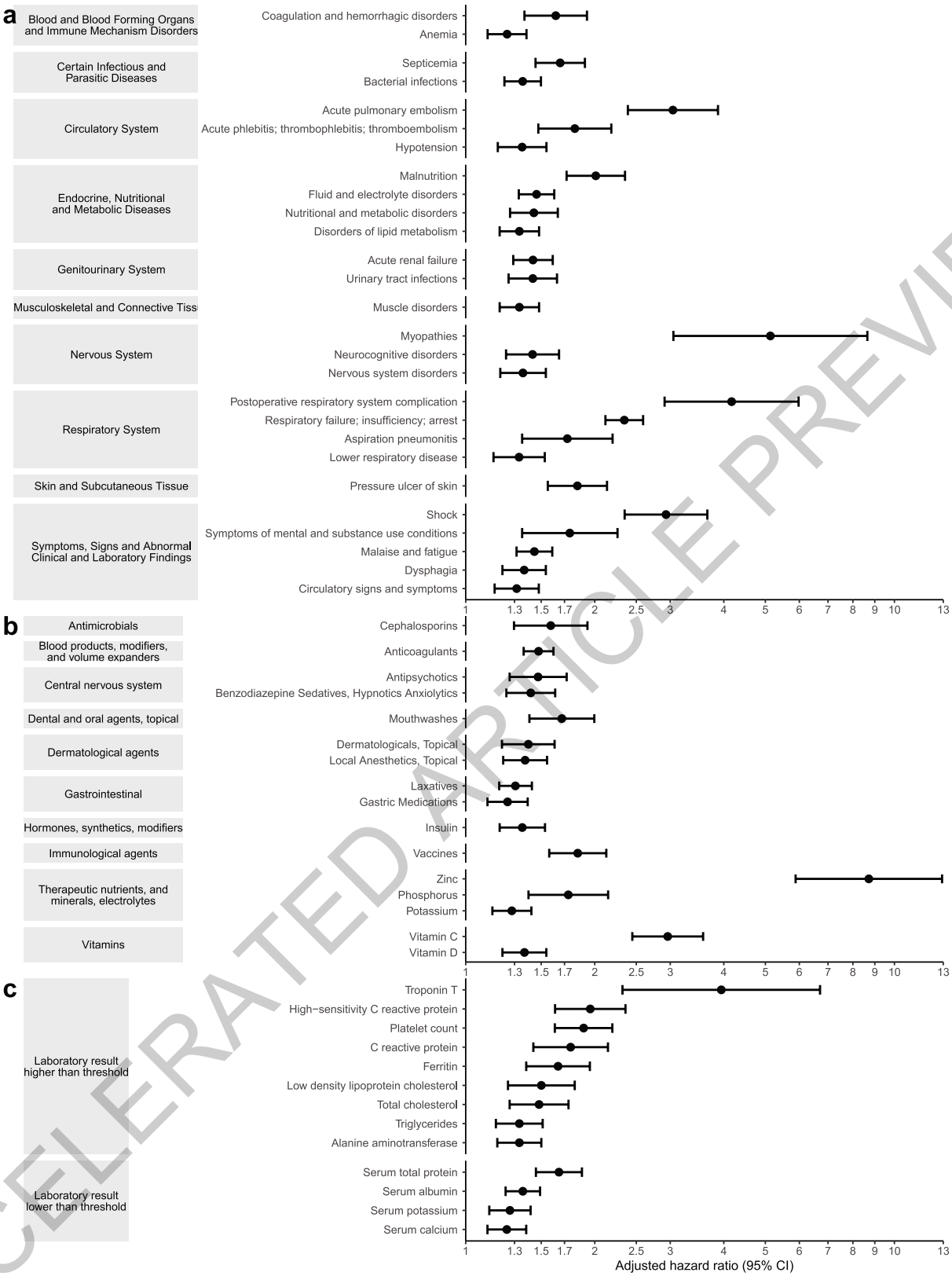
f

Serum albumin lower than 3.5 g/dL	37.55 (27.21, 47.08)
Serum total protein lower than 6 g/dL	27.39 (21.78, 32.38)
Platelet count higher than 400 K/cmm	22.55 (18.24, 26.25)
Serum calcium lower than 8.5 mg/dL	20.31 (11.27, 28.54)
Alanine aminotransferase higher than 40 U/L	20.10 (12.48, 26.92)
Triglycerides higher than 150 mg/dL	19.90 (11.87, 27.04)
Serum potassium lower than 3.5 mmol/L	18.59 (10.38, 25.99)
High−sensitivity C reactive protein higher than 3 mg/L	16.41 (12.80, 19.40)
Total cholesterol higher than 200 mg/dL	14.88 (9.57, 19.42)
Ferritin higher than M: 388; F:252 ng/mL	14.38 (10.18, 17.93)
Low density lipoprotein higher than 130 mg/dL	11.58 (7.02, 15.41)
C reactive protein higher than 4.9 mg/L	11.34 (8.00, 14.08)
Troponin T higher than 0.4 ng/mL	5.28 (4.03, 6.03)

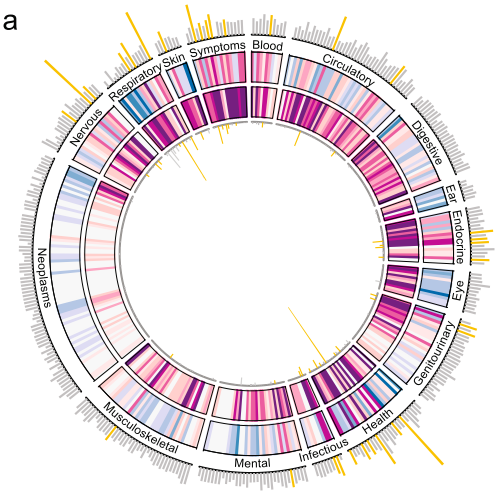
Test limit
- Higher than
- Lower than

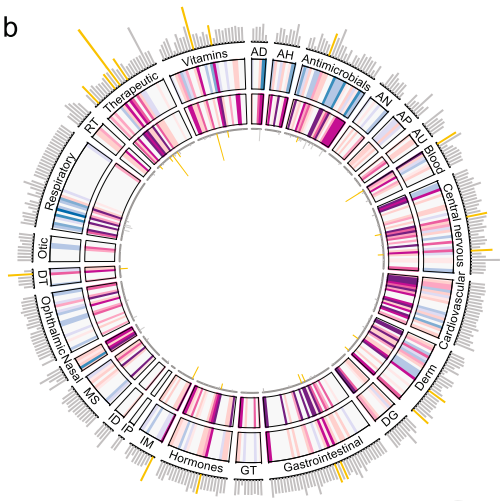
Extended Data Fig. 2 | See next page for caption.

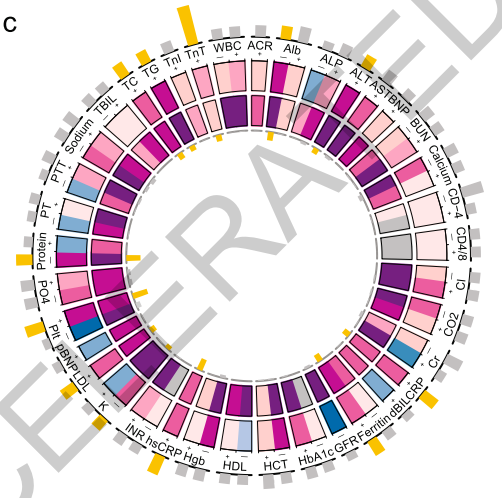
Extended Data Fig. 2 | High dimensional identification of the incident post-acute sequalae of COVID-19 in people who had been hospitalized for COVID-19. (a, d) Incident diagnoses, (b, e) Incident medication use and (c, f) Incident laboratory abnormalities. Hospitalized with seasonal influenza served as the referent category. Post-acute sequalae were ascertained from 30 days after infection until end of follow-up. (a, b, c) Beginning from the outside ring, the first ring represents hazard ratios for the post-acute sequalae of COVID-19. A higher bar indicates larger hazard ratio. Hazard ratios with point estimate larger than one and statistically significant were colored in yellow. The second ring represents excess burden per 1000 COVID-19 patients at 6-months. Color of the cell indicates value of the excess burden, where deeper shades of red indicate higher excess burden and deeper shades of blue indicate greater reduced burden. The third ring represents the baseline incident rate in the control group, where deeper shades of red indicate higher incident rate. The fourth ring represents negative log of the P value, where a higher bar indicates smaller P value and yellow bar indicate statistically significant. (d, e, f) Sequalae were selected based on hazard ratio larger than one and P value less than 6.57×10^{-5}. Excess burdens per 1000 COVID-19 patients at 6-months are presented. Within each domain, outcomes are ranked based on excess burden from high to low. Diagnoses are colored based on diagnosis group, medications are colored based on medication class, and laboratory abnormalities are colored based on higher or lower than normal range.



Extended Data Fig. 3 | Risk of incident post-acute sequalae in COVID-19 patients who had been hospitalized for COVID-19. (a) Incident diagnoses, (b) Incident medication use and (c) Incident laboratory abnormalities. People who had been hospitalized with seasonal influenza served as the referent category. Outcomes were ascertained from day 30 after hospital admission until end of follow-up. Adjusted hazard ratios for incident sequalae that are larger than one and P value less than 6.57×10^{-5} are presented. Hazard ratios (dots) and 95% confidence intervals (bars) are presented on log 10 scale.







d

Respiratory failure; insufficiency; arrest	63.84 (58.39, 68.85)
Disorders of lipid metabolism	33.66 (19.52, 46.72)
Malaise and fatigue	33.56 (25.75, 40.76)
Fluid and electrolyte disorders	31.98 (23.65, 39.70)
Bacterial infections	23.69 (16.07, 30.71)
Acute renal failure	23.29 (16.50, 29.51)
Septicemia	21.15 (16.36, 25.43)
Muscle disorders	19.63 (12.48, 26.18)
Shock	19.40 (17.07, 21.32)
Malnutrition	19.39 (15.42, 22.88)
Nervous system disorders	19.07 (12.68, 24.83)
Pressure ulcer of skin	17.29 (13.08, 20.97)
Nutritional and metabolic disorders	17.22 (11.45, 22.40)
Urinary tract infections	17.09 (11.74, 21.88)
Acute pulmonary embolism	16.54 (14.33, 18.36)
Dysphagia	16.45 (10.32, 21.99)
Circulatory signs and symptoms	14.43 (7.72, 20.52)
Neurocognitive disorders	13.97 (8.32, 18.95)
Anxiety and fear−related disorders	13.82 (7.93, 19.03)
Coagulation and hemorrhagic disorders	12.61 (8.52, 16.14)
Acute phlebitis; thrombophlebitis; thromboembolism	11.54 (8.31, 14.26)
Postoperative respiratory system complication	9.89 (8.76, 10.74)
Symptoms of mental and substance use conditions	7.62 (5.20, 9.56)
Myopathies	7.18 (6.13, 7.87)
Aspiration pneumonitis	6.89 (4.08, 9.16)

Diagnosis category
- Blood and Blood Forming Organs and Immune Mechanism Disorders
- Certain Infectious and Parasitic Diseases
- Circulatory System
- Endocrine, Nutritional and Metabolic Diseases
- Genitourinary System
- Mental, Behavioral and Neurodevelopmental Disorders
- Musculoskeletal System and Connective Tissue
- Nervous System
- Respiratory System
- Skin and Subcutaneous Tissue
- Symptoms, Signs and Abnormal Clinical and Laboratory Findings

e

Anticoagulants	69.38 (56.90, 81.06)
Laxatives	34.29 (23.53, 44.27)
Vitamin C	27.78 (24.76, 30.27)
Vitamin D	25.63 (16.81, 33.53)
Vaccines	22.28 (17.73, 26.20)
Local anesthetics, topical	21.98 (14.49, 28.68)
Potassium	21.79 (13.04, 29.75)
Zinc	21.64 (20.26, 22.57)
Insulin	20.52 (12.92, 27.30)
Benzodiazepine sedatives, hypnotics, anxiolytics	17.65 (11.49, 23.08)
Dermatologicals, topical	14.45 (8.84, 19.35)
Antipsychotics	13.95 (8.88, 18.33)
Mouthwashes	13.63 (9.70, 16.94)
Phosphorus	10.31 (6.91, 13.07)
Cephalosporins	9.42 (5.84, 12.37)

Medication category
- Antimicrobials
- Blood products, modifiers, and volume expanders
- Central nervous system
- Dental and oral agents, topical
- Dermatological agents
- Gastrointestinal
- Hormones, synthetics, modifiers
- Immunological agents
- Therapeutic nutrients, and minerals, electrolytes
- Vitamins

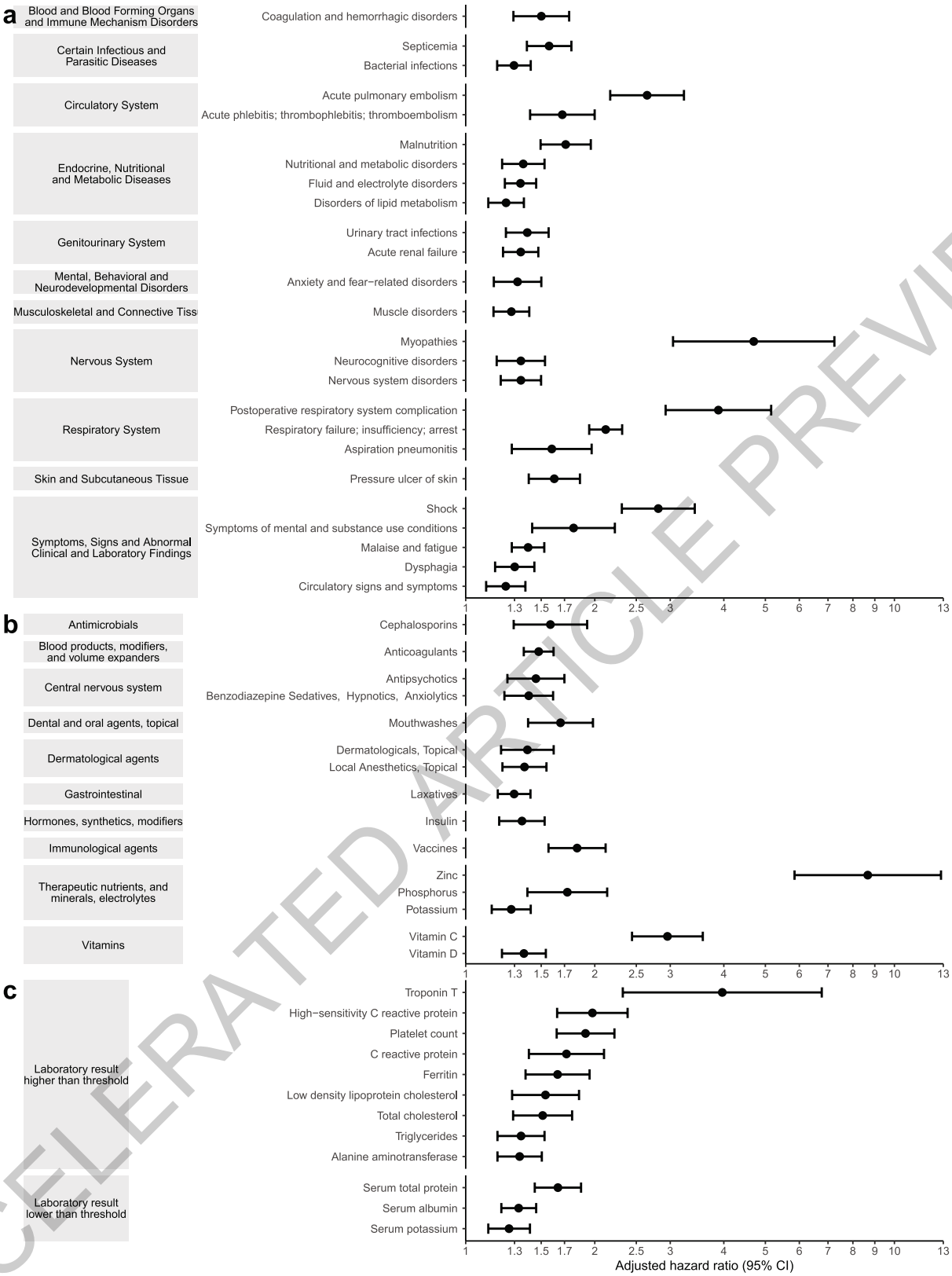
f

Serum albumin lower than 3.5 g/dL	34.70 (24.25, 44.34)
Serum total protein lower than 6 g/dL	26.96 (21.33, 31.96)
Platelet count higher than 400 K/cmm	22.93 (18.63, 26.63)
Triglycerides higher than 150 mg/dL	20.50 (12.47, 27.64)
Alanine aminotransferase higher than 40 U/L	20.19 (12.54, 27.04)
Serum potassium lower than 3.5 mmol/L	18.18 (9.94, 25.61)
High−sensitivity C reactive protein higher than 3 mg/L	16.69 (13.10, 19.67)
Total cholesterol higher than 200 mg/dL	15.81 (10.48, 20.38)
Ferritin higher than M: 388; F:252 ng/mL	14.22 (10.02, 17.78)
Low density lipoprotein higher than 130 mg/dL	12.33 (7.78, 16.16)
C reactive protein higher than 4.9 mg/L	10.85 (7.46, 13.63)
Troponin T higher than 0.4 ng/mL	5.22 (3.97, 5.95)

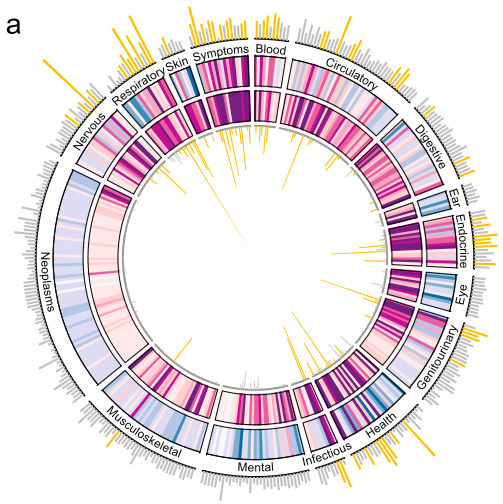
Test limit
- Higher than
- Lower than

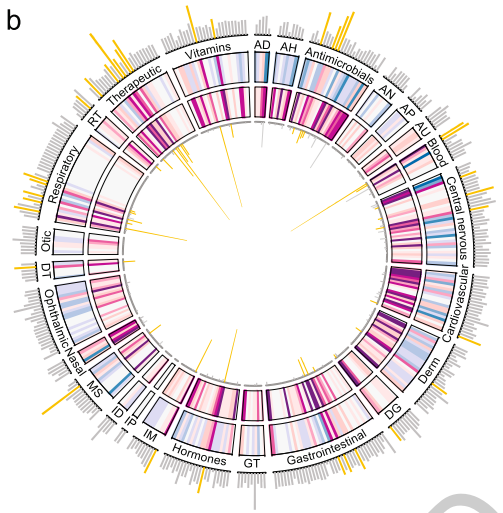
Extended Data Fig. 4 | See next page for caption.

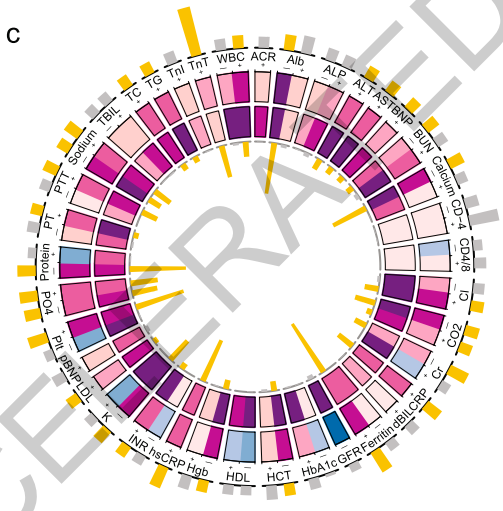
Extended Data Fig. 4 | High dimensional identification of the incident post-acute sequalae of COVID-19 in people who had been hospitalized for COVID-19 after additionally adjusting for severity of the acute infection. (a, d) Incident diagnoses, (b, e) Incident medication use and (c, f) Incident laboratory abnormalities. Hospitalized with seasonal influenza served as the referent category. Post-acute sequalae were ascertained from 30 days after infection until end of follow-up. (a, b, c) Beginning from the outside ring, the first ring represents hazard ratios for the post-acute sequalae of COVID-19. A higher bar indicates larger hazard ratio. Hazard ratios with point estimate larger than one and statistically significant were colored in yellow. The second ring represents excess burden per 1000 COVID-19 patients at 6-months. Color of the cell indicates value of the excess burden, where deeper shades of red indicate higher excess burden and deeper shades of blue indicate greater reduced burden. The third ring represents the baseline incident rate in the control group, where deeper shades of red indicate higher incident rate. The fourth ring represents negative log of the P value, where a higher bar indicates smaller P value and yellow bar indicate statistically significant. (d, e, f) Sequalae were selected based on hazard ratio larger than one and P value less than 6.57×10^{-5}. Excess burdens per 1000 COVID-19 patients at 6-months are presented. Within each domain, outcomes are ranked based on excess burden from high to low. Diagnoses are colored based on diagnosis group, medications are colored based on medication class, and laboratory abnormalities are colored based on higher or lower than normal range.



Extended Data Fig. 5 | Risk of incident post-acute sequalae in COVID-19 patients who had been hospitalized for COVID-19 after additionally adjusting for severity of the acute infection. (a) Incident diagnoses, (b) Incident medication use and (c) Incident laboratory abnormalities. People who had been hospitalized with seasonal influenza served as the referent category. Outcomes were ascertained from day 30 after hospital admission until end of follow-up. Adjusted hazard ratios for incident sequalae that are larger than one and P value less than 6.57×10^{-5} are presented. Hazard ratios (dots) and 95% confidence intervals (bars) are presented on log 10 scale.







d

Respiratory failure; insufficiency; arrest	85.43 (83.47, 87.28)
Respiratory signs and symptoms	58.99 (53.09, 64.59)
Fluid and electrolyte disorders	50.99 (46.70, 55.05)
Malaise and fatigue	45.61 (41.42, 49.57)
Acute renal failure	43.66 (40.41, 46.72)
Bacterial infections	40.76 (36.87, 44.43)
Disorders of lipid metabolism	39.18 (30.04, 47.82)
Septicemia	31.37 (29.08, 33.49)
Muscle disorders	31.31 (27.48, 34.91)
Anemia	25.37 (20.70, 29.77)
Nervous system signs and symptoms	24.79 (18.19, 31.06)
Dysphagia	23.88 (20.43, 27.09)
Nutritional and metabolic disorders	23.20 (19.97, 26.20)
Circulatory signs and symptoms	22.89 (19.03, 26.48)
Lower respiratory disease	22.89 (20.44, 25.13)
Hypertension	22.56 (17.69, 27.13)
Nervous system disorders	20.50 (16.21, 24.51)
Malnutrition	19.91 (17.55, 22.07)
Urinary tract infections	19.70 (16.30, 22.85)
Shock	19.31 (18.13, 20.36)
Hypotension	18.66 (15.53, 21.56)
Acute pulmonary embolism	18.63 (17.48, 19.66)
Cardiac dysrhythmias	17.22 (12.53, 21.61)
Coagulation and hemorrhagic disorders	16.98 (15.03, 18.75)
Sleep wake disorders	16.81 (10.01, 23.22)
Diabetes mellitus	15.71 (11.04, 19.94)
Gastrointestinal disorders	15.48 (11.12, 19.54)
Neurocognitive disorders	14.48 (11.00, 17.68)
Pressure ulcer of skin	14.45 (11.62, 17.05)
Diseases of white blood cells	14.45 (12.08, 16.60)
Esophageal disorders	13.47 (7.64, 18.94)
Acute phlebitis; thrombophlebitis; thromboembolism	13.24 (11.38, 14.90)
Chronic obstructive pulmonary disease	12.35 (7.79, 16.58)
Chronic kidney disease	11.98 (7.94, 15.70)
Obesity	11.98 (7.94, 15.69)
Postprocedural respiratory system complication	10.07 (9.43, 10.61)
Fever	9.30 (7.50, 10.88)
Aspiration pneumonitis	7.85 (6.32, 9.18)
Gastrointestinal hemorrhage	7.27 (4.66, 9.62)
Myopathies	6.99 (6.59, 7.31)
Acute myocardial infarction	6.84 (4.66, 8.76)
Pleurisy, pleural effusion	6.46 (3.66, 8.98)
Symptoms of mental and substance use conditions	6.16 (4.40, 7.69)
Pneumothorax	3.89 (3.29, 4.37)
Cardiac arrest and ventricular fibrillation	3.07 (1.99, 3.94)

Diagnosis category
- Blood and Blood Forming Organs and Immune Mechanism Disorders
- Certain Infectious and Parasitic Diseases
- Circulatory System
- Digestive System
- Endocrine, Nutritional and Metabolic Diseases
- Genitourinary System
- Musculoskeletal System and Connective Tissue
- Nervous System
- Respiratory System
- Skin and Subcutaneous Tissue
- Symptoms, Signs and Abnormal Clinical and Laboratory Findings

e

Anticoagulants	93.50 (87.62, 99.15)
Bronchodilators, sympathomimetic, inhalation	47.49 (43.14, 51.57)
Insulin	36.51 (32.96, 39.82)
Laxatives	32.54 (26.15, 38.63)
Potassium	31.54 (27.21, 35.61)
Vitamin C	24.97 (23.37, 26.44)
Vitamin D	23.67 (18.18, 28.82)
Non-opioid analgesics	22.67 (11.90, 32.99)
Vaccines	22.36 (19.65, 24.85)
Zinc	21.40 (20.95, 21.80)
Histamine antagonists	18.66 (14.64, 22.39)
Magnesium	17.96 (14.54, 21.13)
Non-opioid-containing antitussives/expectorants	14.06 (9.10, 18.69)
Local anesthetics, topical	13.79 (9.08, 18.18)
Benzodiazepine sedatives/hypnotics anxiolytics	12.79 (8.96, 16.34)
Phosphorus	10.62 (8.98, 12.07)
Mouthwashes	10.33 (7.88, 12.54)
Cephalosporins	9.62 (7.58, 11.43)
Thrombolytics	5.04 (4.06, 5.85)
Cardiovascular agents	4.73 (3.33, 5.92)
Bicarbonates	3.91 (3.20, 4.46)
Mucolytics	2.59 (1.64, 3.34)
Extended spectrum penicillins	2.10 (1.49, 2.54)
Other beta-lactams antimicrobials	1.71 (1.27, 2.02)

Medication category
- Antimicrobials
- Blood products, modifiers, and volume expanders
- Cardiovascular
- Central nervous system
- Dental and oral agents, topical
- Dermatological agents
- Gastrointestinal
- Hormones, synthetics, modifiers
- Immunological agents
- Respiratory tract
- Therapeutic nutrients, and minerals, electrolytes
- Vitamins

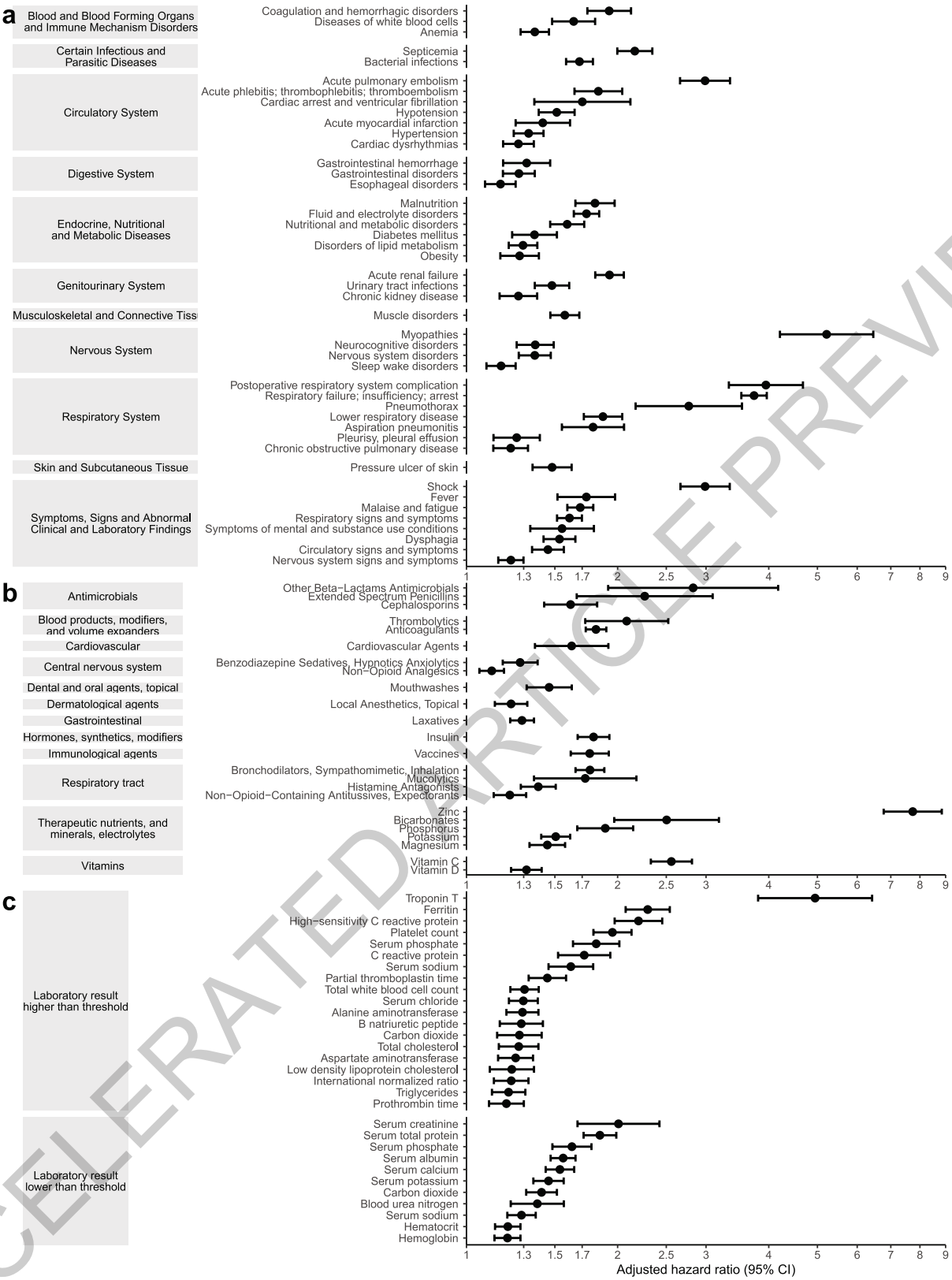
f

Serum albumin lower than 3.5 g/dL	49.25 (43.92, 54.31)
Hematocrit lower than M: 42; F: 37 %	37.76 (26.66, 48.37)
Hemoglobin lower than M: 14; F: 12 g/dL	35.11 (24.50, 45.22)
Serum calcium lower than 8.5 mg/dL	34.40 (29.98, 38.55)
Serum total protein lower than 6 g/dL	29.79 (27.00, 32.39)
Serum potassium lower than 3.5 mmol/L	26.97 (22.65, 31.02)
Serum sodium lower than 136 mmol/L	25.38 (19.36, 31.06)
Carbon dioxide lower than 22 mmol/L	25.02 (20.51, 29.24)
Serum chloride higher than 107 mmol/L	24.44 (18.79, 29.77)
Total white blood cell count higher than 10.8 K/cmm	23.57 (18.32, 28.51)
Platelet count higher than 400 K/cmm	22.84 (20.63, 24.87)
Ferritin higher than M: 388; F:252 ng/mL	20.26 (18.58, 21.78)
Serum phosphate lower than 2.5 mg/dL	17.55 (14.89, 19.99)
Alanine aminotransferase higher than 40 U/L	17.39 (12.83, 21.66)
High-sensitivity C reactive protein higher than 3 mg/L	17.30 (15.62, 18.80)
Partial thromboplastin time higher than 36.5 SEC	15.93 (12.73, 18.87)
Serum phosphate higher than 4.9 mg/dL	14.43 (12.43, 16.23)
Triglycerides higher than 150 mg/dL	13.66 (8.54, 18.43)
Serum sodium higher than 145 mmol/L	12.83 (10.57, 14.88)
Aspartate aminotransferase higher than 40 U/L	12.73 (8.52, 16.64)
International normalized ratio higher than 1.2 ratio	11.34 (7.22, 15.15)
C reactive protein higher than 4.9 mg/L	10.68 (8.77, 12.37)
Total cholesterol higher than 200 mg/dL	10.48 (6.79, 13.86)
Prothrombin time higher than 14.7 sec	10.07 (5.89, 13.94)
B natriuretic peptide higher than 100 pg/mL	8.51 (5.42, 11.32)
Carbon dioxide higher than 32 mmol/L	7.43 (4.52, 10.06)
Low density lipoprotein higher than 130 mg/dL	7.04 (3.79, 9.98)
Blood urea nitrogen lower than 7 mg/dL	6.71 (4.44, 8.72)
Troponin T higher than 0.4 ng/mL	5.11 (4.72, 5.41)
Serum creatinine lower than 0.5 mg/dL	5.04 (4.00, 5.90)

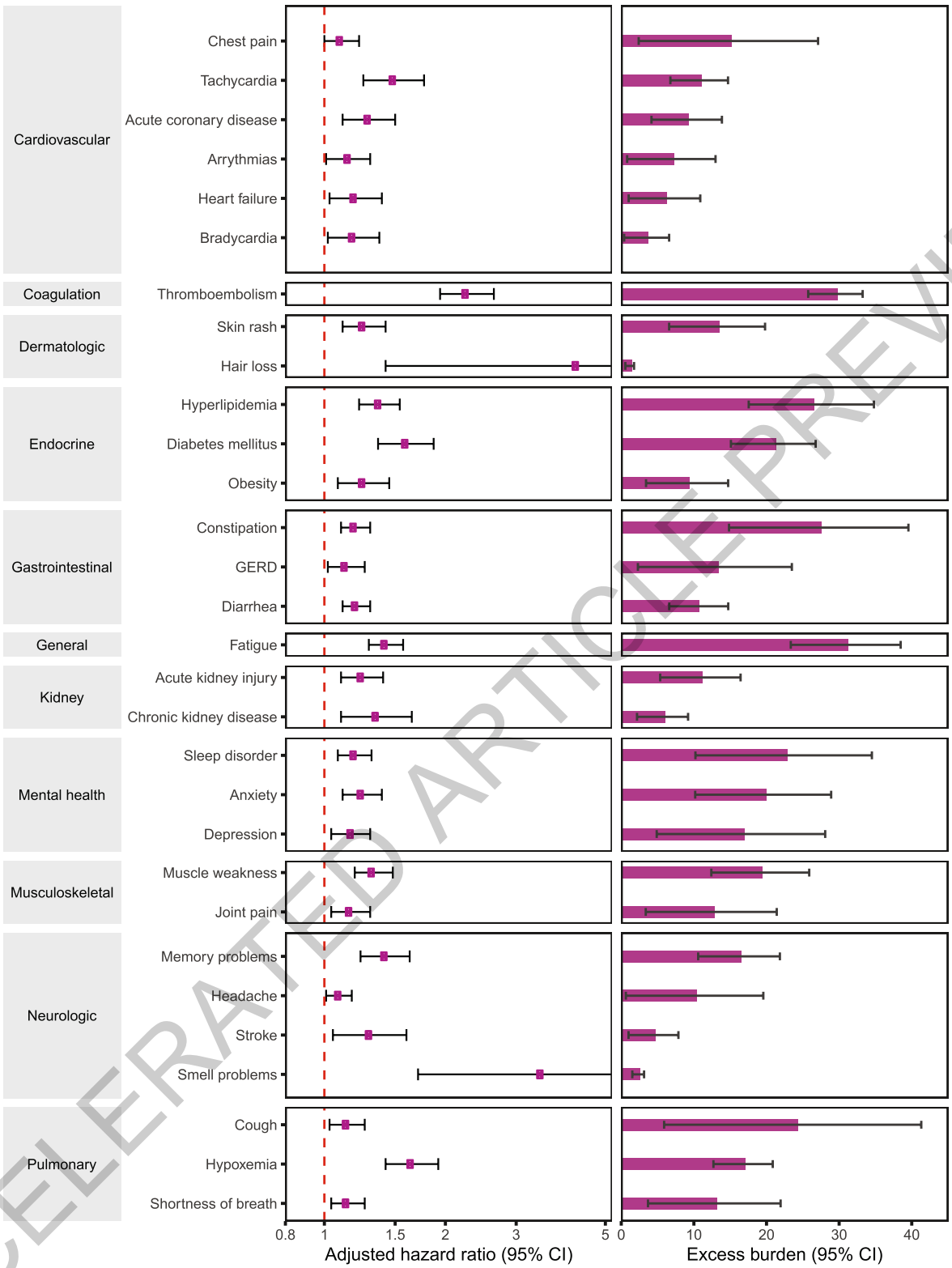
Test limit
- Higher than
- Lower than

Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | High dimensional identification of the incident post-acute sequalae of COVID-19 in people who had been hospitalized for COVID-19. (a, d) Incident diagnoses, (b, e) Incident medication use and (c, f) Incident laboratory abnormalities. Hospitalized for other causes served as the referent category. Post-acute sequalae were ascertained from 30 days after infection until end of follow-up. (a, b, c) Beginning from the outside ring, the first ring represents hazard ratios for the post-acute sequalae of COVID-19. A higher bar indicates larger hazard ratio. Hazard ratios with point estimate larger than one and statistically significant were colored in yellow. The second ring represents excess burden per 1000 COVID-19 patients at 6-months. Color of the cell indicates value of the excess burden, where deeper shades of red indicate higher excess burden and deeper shades of blue indicate greater reduced burden. The third ring represents the baseline incident rate in the control group, where deeper shades of red indicate higher incident rate. The fourth ring represents negative log of the P value, where a higher bar indicates smaller P value and yellow bar indicate statistically significant. (d, e, f) Sequalae were selected based on hazard ratio larger than one and P value less than 6.57×10^{-5}. Excess burdens per 1000 COVID-19 patients at 6-months are presented. Within each domain, outcomes are ranked based on excess burden from high to low. Diagnoses are colored based on diagnosis group, medications are colored based on medication class, and laboratory abnormalities are colored based on higher or lower than normal range.



Extended Data Fig. 7 | Risk of incident post-acute sequalae in COVID-19 patients who had been hospitalized for other causes. (a) Incident diagnoses, (b) Incident medication use and (c) Incident laboratory abnormalities. People who had been hospitalized for other causes served as the referent category. Outcomes were ascertained from day 30 after hospital admission until end of follow-up. Adjusted hazard ratios for incident sequalae that are larger than one and P value less than 6.57×10^{-5} are presented. Hazard ratios (dots) and 95% confidence intervals (bars) are presented on log 10 scale.



Extended Data Fig. 8 | Risks and burdens of incident pre-specified high resolution post-acute COVID-19 outcomes at 6 months in hospitalized people with COVID-19 vs. seasonal influenza (the referent category). Hospitalized people with seasonal influenza served as the referent category. Outcomes were ascertained from day 30 after hospital admission until end of follow-up. Hazard ratios and 95% confidence intervals and excess burdens per 1000 patients and 95% confidence intervals at 6-months are presented.

Extended Data Table 1 | Characteristics of study cohorts

a

Characteristics	COVID-19 vs. VHA users		Hospitalized COVID-19 vs. seasonal Influenza	
	COVID-19 N=73,435	VHA users N=4,990,835	Hospitalized COVID-19 N=13,654	Hospitalized seasonal influenza N=13,997
Age (IQR)	60.70 (47.58, 71.59)	66.68 (51.87, 73.91)	70.25 (60.69, 75.72)	70.14 (62.98, 77.04)
Race (%)				
White	51,601 (70.27)	3,826,222 (76.66)	8120 (59.47)	10,235 (73.12)
Black	18,287 (24.90)	930,798 (18.65)	4610 (33.76)	3124 (22.32)
Other	3547 (4.83)	233,815 (4.68)	924 (6.77)	638 (4.56)
Gender (%)				
Male	64,555 (87.91)	4,514,365 (90.45)	12,861 (94.19)	13,207 (94.36)
Female	8880 (12.09)	476,470 (9.55)	793 (5.81)	790 (5.64)
Long term care (%)	2462 (3.35)	31,944 (0.64)	1706 (12.49)	1146 (8.19)
Number of outpatient encounter (IQR) *	3 (2, 5)	2 (1, 4)	7 (5, 11)	9 (5, 13)
Number of hospital admission (IQR) *	0 (0, 0)	0 (0, 0)	0 (0, 1)	0 (0, 1)
Number of prescriptions received (IQR) *	8 (4, 13)	6 (3, 11)	12 (7, 18)	16 (8, 23)
Number of outpatient eGFR measurements (IQR) [a]	1 (1, 2)	1 (0, 2)	4 (2, 8)	5 (2, 10)
Area deprivation index (IQR)	54.31 (43.84, 62.99)	53.71 (41.89, 62.60)	54.05 (42.87, 61.39)	52.82 (40.12. 61.31)
Sequential Organ Failure Assessment Score (IQR)	NA	NA	1 (0, 2)	1 (0, 2)
Follow up days (IQR)	126 (81, 203)	130 (82, 205)	150 (84, 217)	157 (87, 220)
Total Person-years (Sum)	29,723.73	2,040,891.79	5959.01	6220.04

b

Characteristic	VHA users N=4,990,835	COVID-19 without hospitalization N=73,435	Hospitalized COVID-19 without admit to intensive care N=10,068	Hospitalized COVID-19 admitted to intensive care N=3586
Age (IQR)	66.68 (51.87, 73.91)	60.70 (47.58, 71.59)	70.07 (60.32, 75.83)	70.40 (61.87, 75.53)
Race (%)				
White	3,826,222 (76.66)	51,601 (70.27)	5969 (59.29)	2151 (59.98)
Black	930,798 (18.65)	18,287 (24.90)	3417 (33.94)	1193 (33.27)
Other	233,815 (4.68)	3547 (4.83)	682 (6.77)	242 (6.75)
Gender (%)				
Male	4,514,365 (90.45)	64,555 (87.91)	9470 (94.06)	3391 (94.56)
Female	476,470 (9.55)	8880 (12.09)	598 (5.94)	195 (5.44)
Long term care (%)	31,944 (0.64)	2462 (3.35)	1305 (12.96)	401 (11.18)
Number of outpatient encounter (IQR) *	2 (1, 4)	3 (2, 5)	7 (5, 11)	7 (5, 12)
Number of hospital admission (IQR) *	0 (0, 0)	0 (0, 0)	0 (0, 1)	0 (0, 1)
Number of prescriptions received (IQR) *	6 (3, 11)	8 (4, 13)	12 (7, 18)	12(7, 18)
Number of outpatient eGFR measurements (IQR) *	1 (0, 2)	1 (1, 2)	4 (2, 8)	4 (2, 9)
Area deprivation index (IQR)	53.71 (41.89, 62.60)	54.31 (43.84, 62.99)	53.69 (42.87, 61.31)	54.53 (42.87, 61.99)
Follow up days (IQR)	130 (82, 205)	126 (81, 203)	151 (83, 217)	145 (85, 217)
Total Person-years (Sum)	2,040,891.79	29,723.73	4409.01	1550.00

(a) Characteristics of (1) people with COVID-19 and users of the Veterans Health Administration (VHA), and (2) people hospitalized with COVID-19 and people hospitalized seasonal influenza. (b) Characteristics of four mutually exclusive groups: (1) users of Veterans Health Administration (VHA), (2) people with COVID-19 without hospitalization, (3) people hospitalized with COVID-19 but not admitted to intensive care, and (4) people with COVID-19 and admitted to intensive care. *. Data collected within one year before the cohort enrollment.

Extended Data Table 2 | Results of negative controls, and evidence of high risk of death and health resource utilization.

a

Negative outcome control	COVID-19 vs. VHA users Hazard Ratio (95% confidence interval) *, ‡	Hospitalized COVID-19 vs. Hospitalized Seasonal Influenza Hazard Ratio (95% confidence interval) †, ‡
Accidental injuries	1.03 (0.95, 1.12)	1.02 (0.90, 1.15)
Neoplasms	1.03 (0.94, 1.12)	0.98 (0.83, 1.16)
Scars	0.98 (0.67, 1.44)	1.06 (0.38, 2.95)
Fitting or adjustment of other devices (orthodontic or dental prosthetic device)	1.04 (0.95, 1.14)	0.96 (0.74, 1.24)
Fitting or adjustment of hearing aids	1.04 (0.94, 1.11)	1.02 (0.76, 1.37)
Fitting or adjustment of orthotics	1.13 (0.93, 1.38)	1.06 (0.62, 1.82)
Fitting or adjustment of casts	0.97 (0.82, 1.14)	0.90 (0.62, 1.30)
Bandages	0.99 (0.86, 1.13)	1.08 (0.91, 1.28)

b

Cohort	Outcomes	Hazard ratio (95% confidence interval) ‡	Incident rate per 1000 at 6-months in COVID-19 group (95% confidence interval) ‡	Incident rate per 1000 at 6-months in comparison group (95% confidence interval) ‡	Excess burden per 1000 at 6-months (95% confidence interval) ‡
COVID-19 vs. VHA users *	Death	1.59 (1.46, 1.73)	22.77 (20.90, 24.81)	14.38 (13.19, 15.68)	8.39 (7.09, 9.58)
	Outpatient encounter	1.20 (1.19, 1.21)	946.69 (944.94, 948.41)	913.48 (911.11, 915.81)	33.22 (30.89, 35.58)
Hospitalized COVID-19 vs. hospitalized seasonal influenza †	Death	1.51 (1.30, 1.76)	87.92 (76.11, 101.47)	59.13 (51.07, 68.41)	28.79 (19.52, 36.85)
	Outpatient encounter	1.12 (1.08, 1.17)	990.32 (988.50, 991.90)	983.95 (981.28, 986.30)	6.37 (4.01, 9.03)

c

Cohort	Number of outpatient encounters per 30 days In COVID-19 group (95% confidence interval) ‡	In comparison group (VHA users or seasonal influenza) (95% confidence interval) ‡	Excess encounters (95% confidence interval) ‡
COVID-19 vs. VHA users *	3.03 (3.00, 3.05)	2.56 (2.56, 2.57)	0.47 (0.44, 0.49)
Hospitalized COVID-19 vs. hospitalized seasonal influenza †	7.51 (7.37, 7.65)	6.06 (5.95, 6.16)	1.45 (1.28, 1.63)

(a) Results of testing negative outcome controls in people with COVID-19 compared to users of the Veterans Health Administration (VHA), and in people who had been hospitalized with COVID-19 compared to people who had been hospitalized with seasonal influenza. (b) Risk of death and health resource utilization in people with COVID-19 compared to users of the Veterans Health Administration (VHA), and in people who had been hospitalized with COVID-19 compared to people who had been hospitalized with seasonal influenza. (c) Health resource utilization in people with COVID-19 compared to users of the Veterans Health Administration (VHA), and in people who had been hospitalized with COVID-19 compared to people who had been hospitalized with seasonal influenza. *. For analyses of people with COVID-19 vs. VHA users, outcomes were ascertained from 30 days after COVID-19 diagnosis and VHA users served as the referent category. †. For analyses of people who had been hospitalized with COVID-19 vs. people who had been hospitalized with seasonal influenza, outcomes were ascertained from 30 days after hospital admission and people who had been hospitalized with seasonal influenza served as the referent category. ‡. Results based on models adjusted through overlap weighting.

nature research

Corresponding author(s): Ziyad Al-Aly

Last updated by author(s): Apr 8, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	SAS Enterprise Guide version 7.1 was used to collect data for the study. ICD-10 diagnosis codes were classified into 540 diagnostic categories based on the Clinical Classifications Software Refined (CCSR) version 2021.1 (Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality).
Data analysis	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).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data are available from the US Department of Veterans Affairs

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	To achieve better precision of the study results, we enrolled all users of the US Veterans Health Administration and followed them until January 31, 2021. To our knowledge, this is the largest and most comprehensive post-acute COVID-19 study to date involving 73,435 non-hospitalized patients with COVID-19, and 4,990,835 controls (2,070,615.52 person years of follow-up), and 13,654 hospitalized patients with COVID-19, and 13,997 patients hospitalized with seasonal influenza (12,179.05 person years of follow-up).
Data exclusions	To examine the risk of post-acute outcomes beyond the first 30 days of illness, we predefined our exclusion criteria and excluded participants who did not survive the first 30 days of COVID-19 illness.
Replication	To provide results with better accuracy and precision, we included all users of the US Veterans Health Administration to estimate the study results. The finding was not replicated because no other internal or external datasets with large sample size and high dimensional information is currently available to us.
Randomization	Exposure allocation was not random. To adjust for potential confounders, we applied overlap weighting based on propensity scores constructed from a) predefined variables including demographic, healthcare utilization and contextual factors, and b) from algorithmically selected high dimensional variables from domains including diagnoses, pharmacy records, and laboratory tests.
Blinding	We conducted an observational study. Blinding was not possible.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Study participants are users of the US Veteran Health Administration. The overall study population had a median age of 67, 87% were male, and 77% of cohort participants were of white race. 73,435 people with COVID-19 and 4,990,835 controls were evaluated. Additionally, 13,654 people hospitalized with COVID-19 and 13,997 people hospitalized with seasonal influenza were also evaluated. The hospitalized group comprised of 10,068 patients who were hospitalized but did not require intensive care, and 3,586 patients who were admitted to intensive care.
Recruitment	Participants were recruited if they had at least 1 encounter with the US Veteran Health Administration in the year prior to cohort enrollment. Non users of the VA health care system were not included. The characteristics of the study population may be different from the general population (US or global population). Other biases due to recruitment including self-selection bias are unlikely to bias the results of this study.
Ethics oversight	The study was approved by the Institutional Review Board of the Veterans Affairs St. Louis Health Care System, St. Louis, MO, US

Note that full information on the approval of the study protocol must also be provided in the manuscript.