



**FAST TRACK**

# Risk of clinical sequelae after the acute phase of SARS-CoV-2 infection: retrospective cohort study

Sarah E Daugherty,<sup>1</sup> Yinglong Guo,<sup>1</sup> Kevin Heath,<sup>2</sup> Micah C Dasmariñas,<sup>1</sup> Karol Giuseppe Jubilo,<sup>1</sup> Jirapat Samranvedhya,<sup>1</sup> Marc Lipsitch,<sup>3</sup> Ken Cohen<sup>2,4</sup>

<sup>1</sup>OptumLabs at UnitedHealth Group, Minneapolis, MN, USA

<sup>2</sup>OptumCare, Minneapolis, MN, USA

<sup>3</sup>Harvard T H Chan School of Public Health, Boston, MA, USA

<sup>4</sup>OptumLabs at UnitedHealth Group, Golden, CO, USA

Correspondence to:

S E Daugherty  
SarahDaugherty@uhg.com  
(ORCID 0000-0002-4148-2881)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2021;373:n1098  
<http://dx.doi.org/10.1136/bmj.n1098>

Accepted: 26 April 2021

## ABSTRACT

### OBJECTIVE

To evaluate the excess risk and relative hazards for developing incident clinical sequelae after the acute phase of SARS-CoV-2 infection in adults aged 18-65.

### DESIGN

Retrospective cohort study.

### SETTING

Three merged data sources from a large United States health plan: a large national administrative claims database, an outpatient laboratory testing database, and an inpatient hospital admissions database.

### PARTICIPANTS

Individuals aged 18-65 with continuous enrollment in the health plan from January 2019 to the date of a diagnosis of SARS-CoV-2 infection. Three comparator groups, matched by propensity score, to individuals infected with SARS-CoV-2: a 2020 comparator group, an historical 2019 comparator group, and an historical comparator group with viral lower respiratory tract illness.

### MAIN OUTCOME MEASURES

More than 50 clinical sequelae after the acute phase of SARS-CoV-2 infection (defined as the date of first SARS-CoV-2 diagnosis (index date) plus 21 days) were identified using ICD-10 (international classification of diseases, 10th revision) codes. Excess risk in the four months after acute infection and hazard ratios with Bonferroni corrected 95% confidence intervals were calculated.

## RESULTS

14% of adults aged  $\leq 65$  who were infected with SARS-CoV-2 (27 074 of 193 113) had at least one new type of clinical sequelae that required medical care after the acute phase of the illness, which was 4.95% higher than in the 2020 comparator group. The risk for specific new sequelae attributable to SARS-CoV-2 infection after the acute phase, including chronic respiratory failure, cardiac arrhythmia, hypercoagulability, encephalopathy, peripheral neuropathy, amnesia (memory difficulty), diabetes, liver test abnormalities, myocarditis, anxiety, and fatigue, was significantly greater than in the three comparator groups (2020, 2019, and viral lower respiratory tract illness groups) (all  $P < 0.001$ ). Significant risk differences because of SARS-CoV-2 infection ranged from 0.02 to 2.26 per 100 people (all  $P < 0.001$ ), and hazard ratios ranged from 1.24 to 25.65 compared with the 2020 comparator group.

## CONCLUSIONS

The results indicate the excess risk of developing new clinical sequelae after the acute phase of SARS-CoV-2 infection, including specific types of sequelae less commonly seen in other viral illnesses. Although individuals who were older, had pre-existing conditions, and were admitted to hospital because of covid-19 were at greatest excess risk, younger adults (aged  $\leq 50$ ), those with no pre-existing conditions, or those not admitted to hospital for covid-19 also had an increased risk of developing new clinical sequelae. The greater risk for incident sequelae after the acute phase of SARS-CoV-2 infection is relevant for healthcare planning.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Small observational studies of patients admitted to hospital have shown that some covid-19 survivors had short term and long term clinical sequelae. Few studies have characterized the excess risk of clinical sequelae attributable to SARS-CoV-2 after the acute infection in adults aged  $\leq 65$  in a large generalizable sample.

## WHAT THIS STUDY ADDS

14% of individuals aged  $\leq 65$  who were infected with SARS-CoV-2 (27 074 of 193 113) developed at least one new type of clinical sequelae that required medical care after the acute phase of the illness, which was 4.95% higher than the 2020 comparator group.

An increased risk of specific clinical sequelae after the acute infection was noted across a range of organ systems, including cardiovascular, neurologic, kidney, respiratory, and mental health complications.

The risk for incident sequelae increased with age, pre-existing conditions, and admission to hospital for covid-19, but in adults aged  $\leq 50$  and those with no pre-existing conditions or not admitted to hospital for covid-19, the risk for some clinical sequelae was still elevated.

## Introduction

Emerging data suggest that the sequelae of infection with SARS-CoV-2 and the disease it causes, covid-19, could vary in presentation and extend beyond the typical postviral recovery period. Hence epidemiologic interest in morbidity after the acute infection in survivors is growing. Some survivors experience serious complications during the acute phase of the illness, affecting pulmonary, cardiovascular, hepatic, renal, cognitive, and neurologic function.<sup>1-4</sup> Survivors also report a range of persistent symptoms adversely affecting physical, mental, and social wellbeing.<sup>5-8</sup> At least some of these complications might occur independent of the severity of covid-19.<sup>9</sup> Although individuals admitted to hospital with community acquired (non-covid-19) pneumonia or influenza are at risk of cardiovascular, cerebrovascular, and neurological complications,<sup>10 11</sup> the degree of increased risk resulting from SARS-CoV-2 infection

is unclear. Longitudinal studies on survivors of other coronaviruses (Middle East respiratory syndrome (MERS) and severe acute respiratory syndrome (SARS)) suggest long term physical and mental sequelae are not uncommon.<sup>12-14</sup>

Most published studies so far have been small and mainly focused on clinical sequelae in patients admitted to hospital for covid-19.<sup>15</sup> Hence many of these studies might not apply to the larger population of individuals infected with SARS-CoV-2. Little is known of the incidence of clinical sequelae caused by SARS-CoV-2 infection after the acute phase of the illness in adults aged 18-65 who might be considered to have a lower risk of severe covid-19. Also, few studies have been powered to evaluate whether factors such as age, sex, pre-existing conditions, and admission to hospital because of covid-19 modify the risk of clinical sequelae after the acute infection.

We estimated the excess risk and hazard ratios of new clinical sequelae attributable to SARS-CoV-2 in adults aged 18-65 after the acute phase of covid-19. Our analysis included a large generalizable sample of commercially insured adults; determined objective outcomes with the valid ICD-10 (international classification of diseases, 10th revision) codes in claims; and had the power to detect rare diagnoses and evaluate associations across subgroups.

## Methods

### Data source

We conducted a retrospective cohort analysis with three data sources within the UnitedHealth Group Clinical Discovery Database: de-identified administrative outpatient and inpatient claims, outpatient laboratory results for SARS-CoV-2, and a hospital admissions database updated daily with patients admitted with a primary, secondary, or tertiary diagnosis of covid-19. Quality control efforts for data were applied (supplementary appendix A). All data were for commercially insured patients enrolled with one large national health insurance provider in the United States.

### Study population

#### *Individuals diagnosed as having SARS-CoV-2*

This group consisted of individuals aged 18-65 with continuous enrollment in the health plan from 1 January 2019 to the index date. We defined the index date as the date of the first occurrence of any of these events: diagnosis of a primary, secondary, or tertiary diagnosis of covid-19, identified by administrative claims with ICD-10 code U07.1 or either B34.2 or B97.29 before 1 April 2020; or documentation of a positive polymerase chain reaction (PCR) test in an outpatient laboratory dataset; or admitted to hospital for covid-19, identified from a hospital admissions database with a diagnosis code on admission or a primary, secondary, or tertiary diagnosis code of U07.1 or U07.2.

We included ICD-10 codes B34.2 and B97.29 because many physicians used these codes early in

the pandemic to clinically diagnose infection before the US Centers for Disease Control and Prevention recommended U07.1 as the primary code for clinical diagnosis on 1 April 2020. The U07.1 code accounted for 98.2% of all clinically diagnosed patients in this study.

For individuals with a positive PCR test identified in our database (n=77 273), 45% also had a clinical diagnosis code of U07.1. We assumed that the remaining individuals with a positive PCR test without diagnostic codes were asymptomatic or had mild symptoms and did not seek medical care.

For patients admitted to hospital for covid-19 (n=21 746), over half (51.3%) were coded as U07.1 (covid-19, virus identified) and the other half were coded with the World Health Organization emergency code for clinically diagnosed covid-19 with no laboratory confirmed infection, U07.2 (covid-19, virus not identified). Although U07.2 was not officially adopted by the US, many clinicians used this code to identify patients with suspected covid-19.

We excluded individuals with a positive SARS-CoV-2 antibody serology without documented infection (n=28 810) because index dating of the illness was not possible. We also excluded individuals with a diagnosis code of B34.2 or B97.29 on or after 1 April 2020 (n=24 865) and individuals admitted to hospital for suspected covid-19 but missing diagnosis code U07.1 or U07.2 in the primary, secondary, or tertiary position (n=1247). Figure 1 shows a flowchart with details of population sampling.

#### *2020 comparator group*

The 2020 comparator group was individuals aged 18-65 who did not have a clinical diagnosis related to covid-19, a positive PCR test, or were admitted to hospital for covid-19 in 2020. Continuous enrollment in the health plan was required from 1 January 2019 to a randomly assigned index date drawn from the SARS-CoV-2 infection group.

#### *2019 comparator group*

We created this historical comparison group to account for possible ascertainment bias because of reduced use of healthcare services during the 2020 pandemic. Individuals aged 18-65 were required to have continuous enrollment in the health plan from 1 January 2018 to a randomly assigned month and day in 2019, drawn from the SARS-CoV-2 infection group.

#### *Viral lower respiratory tract illness comparator group*

We created this historical comparison group to evaluate the clinical sequelae specific to SARS-CoV-2 infection because many serious viral illnesses have a risk of morbidity after the acute illness. The viral lower respiratory tract illness group included individuals aged 18-65 who developed influenza (J09, J10, J11), non-bacterial pneumonia (J12, J18.9), acute bronchitis (J20), acute lower respiratory infection (J22), or chronic obstructive pulmonary disease with acute lower respiratory infection (J44.0), between

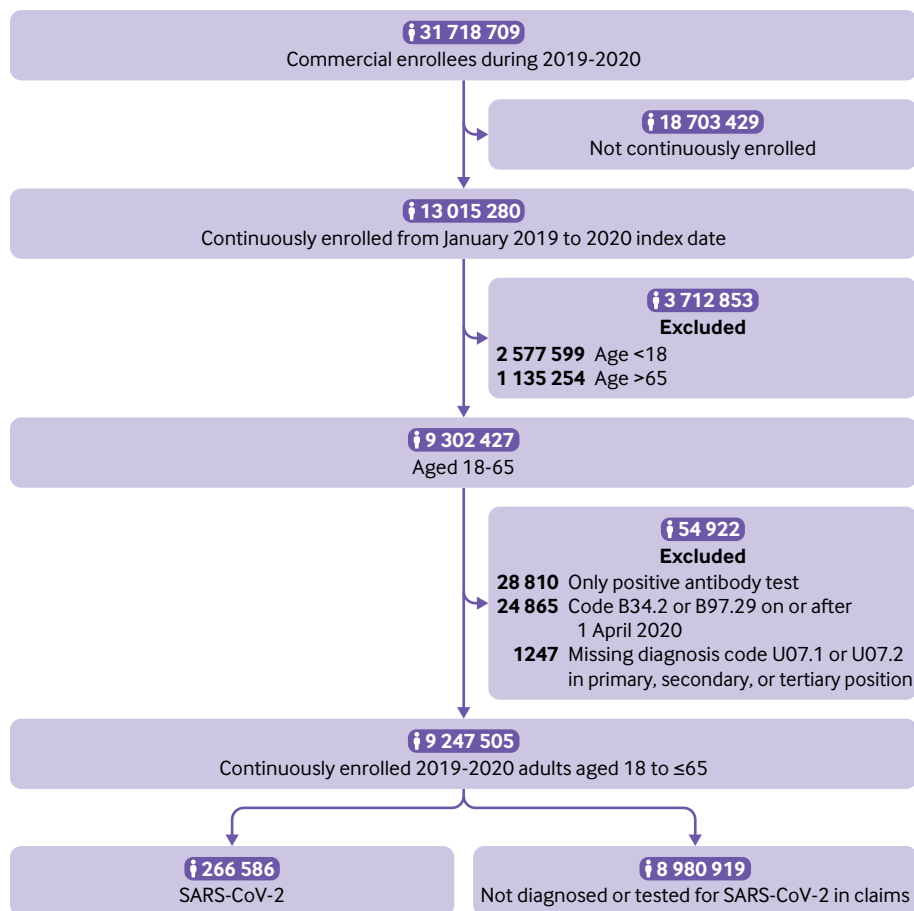


Fig 1 | Flowchart with details of population sampling of the 2020 cohort sample

1 January 2017 and 31 October 2017, 1 January 2018 and 31 October 2018, or 1 January 2019 and 31 October 2019. We included chronic obstructive pulmonary disease exacerbation because instances are typically induced by a virus and identified only with code J44.0. We defined the index date as the date of the first diagnosis of viral lower respiratory tract illness in the corresponding year of the cohort. Individuals were required to have continuous enrollment in the health plan from 1 January 2016, 1 January 2017, or 1 January 2018 to the index date, respectively.

#### Main outcomes

We used ICD-10 codes to identify new clinical diagnoses from the administrative claims data between 1 January 2020 and 31 October 2020. eTable 1 shows the ICD-10 classification details. We created domain clusters based on clinically similar diagnoses, and we included atopic dermatitis as a negative control.<sup>16</sup>

#### Study variables

We used administrative claims between 1 January 2019 and 30 days before the index date to determine length of hospital stay, previous clinical conditions, history of comorbidities derived from the Charlson Index and Elixhauser score, and previous visits to a

primary care physician, cardiologist, or nephrologist. Demographic, clinical, and testing data were obtained from administrative claims between 1 January 2020 and 31 October 2020. We derived socioeconomic status scores specific to zip codes and proportions of white, black, and Hispanic populations by zip code.<sup>17</sup> Missing values for zip code derived variables (socioeconomic status scores n=11 713; race n=444 080) were imputed with the median of non-missing values for these variables. Because all other variables were derived from administrative claims and the continuous enrollment criteria were applied, clinical events without a claim were considered to have not occurred and were given a value of zero.

#### Follow-up

Follow-up periods for the primary analysis for individuals infected with SARS-CoV-2 and comparator groups started at the index date plus 21 days and continued up to a diagnostic event, disenrollment from the insurance plan (because of death or withdrawal), or the end of the study period (31 October 2020 or 31 October of the corresponding year for the historical comparator groups), whichever occurred first. We performed a secondary analysis evaluating rates by month of follow-up starting 30 days before the index date and ending after six months of follow-up.

## Statistical methods

### *Propensity score matching*

We used matching by propensity score to create three cohorts with similar baseline characteristics and relevant confounders.<sup>18</sup> We constructed a propensity score for every individual based on 108 variables (age, number; sex, male yes/no; socioeconomic status scores specific to zip codes, number; proportion of white, black, and Hispanic populations in a zip code, number; state, 10 binary variables in the top 10 states of SARS-CoV-2 diagnosis; index month, January-March, April, May, June, July, August, and September yes/no; pre-existing comorbidities, yes/no; total length of stay as an inpatient in the previous year, number in days; previous number of visits to a primary care physician, cardiologist, or nephrologist, number; and previous clinical conditions, yes/no) by logistic regression with ridge penalty.<sup>19 20</sup>

Because of the large study population (about 10 million), we did not perform conventional one-to-one nearest neighbor matching. We generated 40 000 bins based on the propensity scores of the SARS-CoV-2 infected group. In each small bin, we created equal sized sets of individuals with SARS-CoV-2 and in the comparator group. We matched within the bins and then combined all sets to form the propensity matched groups. This process was repeated for each of the comparator groups (2020, 2019, and viral lower respiratory tract illness groups). We achieved three balanced cohort groups with this approach (eTable 1a and eFigure 1a-c).

### *Data analysis*

We evaluated demographic and clinical factors with the *t* test and the Pearson  $\chi^2$  test for the unmatched population to compare numeric and categorical data, respectively. We used a paired *t* test and the McNemar test for evaluating numeric and categorical factors in the matched populations.

We calculated the proportion of individuals in the matched populations with a follow-up time of at least 21 days after the index date who did not develop new clinical sequelae, or developed one or more than one new type of clinical sequelae after the acute infection. No events that occurred before or during the acute phase were counted. Symptoms such as fatigue, myalgia, and anosmia, measured by ICD-10 codes, were included in this assessment of clinical sequelae after the acute infection.

When we calculated the risk for a specific incident outcome after the acute infection, individuals with the diagnosis of interest before the index date (in the SARS-CoV-2 and respective comparator group) or individuals (and their comparator match) who experienced the diagnosis of interest during the acute phase were removed from the post-acute calculation. Individuals with less than 21 days of follow up after the index date and their matched pair were also excluded from the analysis. Therefore, the denominator for the incidence of each outcome included only those at risk for the diagnosis of interest at the index date plus 21

days. We reported the risk difference as the difference between the cumulative incidence calculated by the Kaplan-Meier estimator at day 120 from the time origin (index date plus 21 days) in the SARS-CoV-2 group and the comparator group, multiplied by 100. We calculated the 95% confidence intervals and P values with a pairwise bootstrap method, with a one sided test. We tested the proportional hazards assumption with Schoenfeld residuals.<sup>21</sup> Hazard ratios and 95% confidence intervals were estimated by fitting Cox proportional hazards models using robust variance estimator with clustering for matched pairs. We used the Wald test to evaluate significance at the 0.05 level. Individuals were censored based on the outcome of interest, disenrollment from the insurance plan, or the end of the study period (31 October 2020).

In a secondary analysis, we created seven one month time intervals (one month before the index date to six months after the index date). Within each time interval, we only included the matched pairs who were both at risk for the diagnosis of interest at the beginning of each time interval. We estimated the hazard ratios and confidence intervals by fitting Cox proportional hazard models using robust variance estimator with clustering for matched pairs. Individuals were censored based on the diagnosis of interest, disenrollment from the insurance plan, or the end of the time interval.

Because of the large number of comparisons, we applied the Bonferroni correction for all P values and confidence intervals, by multiplying P values by N and estimating  $(1-(0.05 \div N)) \times 100\%$  confidence intervals, where N=51 and is the number of clinical sequelae we tested. For all data analyses, we used Python with scikit-learn, statsmodels, lifelines, and scipy libraries and R with survminer,<sup>22</sup> survival,<sup>23</sup> glmnet,<sup>24</sup> and stats libraries.<sup>25</sup>

### *Subgroup analysis*

We performed four stratified analyses (age (18 to  $\leq 34$ ,  $>34$  to  $\leq 50$ , and  $>50$ ), sex (male/female), any pre-existing clinical comorbidity (yes/no), and admitted to hospital for covid-19 (yes/no) with the SARS-CoV-2 and 2020 comparator group. We constructed a propensity score, and performed matching by propensity score within each subgroup for age, sex, and pre-existing condition status before calculating the risk difference or hazard ratio within each stratum. The SARS-CoV-2 and 2020 comparator groups, matched for propensity score, were stratified based on whether individuals with SARS-CoV-2 were admitted to hospital for covid-19. To test for interaction, we evaluated the significance of the fitted coefficient for the interaction term in a model including the main effect variables, with a Bonferroni correction ( $N=51 \times$  number of levels in each subgroup analysis).

### *Sensitivity analysis*

We evaluated the robustness of the definition of the post-acute phase by varying the cut-off points from the index date at 14 and 28 days. We also considered a period effect because of changes in the availability

of testing and advancements in management and treatment over time. We evaluated differences by two periods (January-June and July-October). Finally, we evaluated the number of visits to a primary care physician after the acute infection by type of diagnostic method (PCR positive, clinical diagnosis, and admitted to hospital) and among the three comparator groups.

### Patient and public involvement

This retrospective analysis did not directly involve patients or the public in the development of the research question or conduct of the analysis because of funding and training restrictions for retrospective analyses. A patient reviewer provided insightful comments and contributed to the expansion of the outcomes reported in this manuscript.

### Results

Among 9 247 505 individuals meeting the study criteria in 2020, we identified 266 586 (2.9%) with SARS-CoV-2 infection. After matching by propensity score, we identified 266 586 matched pairs for the primary (2020) and secondary (2019) comparison groups (100% of SARS-CoV-2 individuals matched). We identified 244 276 matched pairs for the viral lower respiratory tract illness comparison group (91.6% of SARS-CoV-2 individuals matched).

Individuals with SARS-CoV-2 infection were more likely than their unmatched 2020 and 2019 comparators to be younger, to be women, have a lower socioeconomic status score index, live in a zip code with a higher proportion of black or Hispanic individuals, have a pre-existing comorbidity or at risk condition, been admitted to hospital with a longer length of stay during the preceding year, visited a primary care physician or other specialists more often, and live in the northeast or southern US (areas of high incidence during the study period; all  $P < 0.05$ ) (table 1). Among those individuals with SARS-CoV-2, 8.2% were admitted to hospital and 1.1% were admitted to the intensive care unit. We found different patterns between individuals with SARS-CoV-2 and the viral lower respiratory tract illness comparison group who were more likely to be older, to be women, have a comorbidity, smoke, visit a primary care physician or cardiologist more often, and live in the southern or midwestern US. After matching by propensity score, most of these differences were resolved, although some small differences between the matched SARS-CoV-2 infected group and the viral lower respiratory tract illness comparator were significant (all  $P < 0.05$ ) (eTable 1a). Despite these minor differences, balance was achieved overall, as shown by the standardized mean differences between matched groups by key variables at less than 0.10 (eFig 1a-c).

Among the matched individuals infected with SARS-CoV-2 identified in our study with a follow-up time of at least 21 days from the index date ( $n = 193\ 113$ ), 85.98% had no new clinical sequelae after the acute infection that required medical care during their follow-up, 10.01% had one type of new sequelae that required

medical care, and 4.01% had more than one type of new sequelae (table 2). The proportion of individuals diagnosed as having any new clinical sequelae after the acute phase was higher in the SARS-CoV-2 infected group than in the three comparator groups, although the differences were smallest compared with the viral lower respiratory tract illness group.

Estimates of risk difference by type of new clinical sequelae were calculated for individuals infected with SARS-CoV-2 who were still at risk 21 days after the index date ( $n = 193\ 113$  for 2020 comparison). Follow-up time was considered up to day 141 after the index date (120 days from the start of the post-acute period and 73.5 centile of the follow-up distribution, with a median of 87 days (interquartile range 45-124 days)). Figure 2 summarizes the most common new clinical outcomes in the SARS-CoV-2 group (incidence  $\geq 0.1\%$ ) and eTable 2a the less common outcomes. Symptoms are included separately in eTable 2b and are not represented in figure 2. Overall, the excess risk attributable to SARS-CoV-2 infection was low for incident diagnoses (0.02-2.26 per 100 individuals) four months after the acute phase (fig 2 and eTable 2a-b). The increased risk, however, was consistently seen for many outcomes across all three comparison groups (2020, 2019, and viral lower respiratory tract illness groups; eTable 2a).

Despite the small absolute risk attributable to SARS-CoV-2 infection, the hazard ratios for individuals infected with SARS-CoV-2 and the 2020 comparator group after the acute infection were large (significant hazard ratios of 1.24-25.65 (all  $P < 0.001$  except for atopic dermatitis ( $P = 0.033$ ); eTable 2c-d and fig 2). When we evaluated rates over time, hazard ratios were highest in the first month of the index date but were substantially elevated up to six months for some common events, such as hypertension (hazard ratio 1.81 (95% confidence interval 1.10 to 2.96)), diabetes (2.47 (1.14 to 5.38)), sleep apnea (2.31 (1.23 to 4.32)), and fatigue (2.20 (1.48 to 3.27)), suggesting the hazard for some new clinical sequelae was sustained months after the initial SARS-CoV-2 infection (eTable 3). An increased risk 30 days before the index date was likely because of a delay in testing or documentation of a confirmed diagnosis in symptomatic individuals. Figure 3 shows select graphs of the cumulative hazards for the most common or most severe clinical sequelae.

Excess risk for developing many new outcomes after the acute infection increased significantly (most  $P < 0.001$ ) with age (fig 4 and eTable 4a). The risk for clinical sequelae was greatest in individuals aged  $> 50$  but the absolute risk in young adults aged 18-34 was significantly elevated, albeit modestly so, for some conditions including, but not limited to, hypertension, arrhythmia, hypercoagulability, amnesia, diabetes, and fatigue (all  $P < 0.001$ ). The risk of developing any mental health outcome was significantly increased regardless of age ( $P_{\text{interaction}} = 0.35$ ).

Excess risk for new clinical sequelae after acute covid-19 rarely differed between men and women, apart from fatigue and anosmia (more commonly diagnosed

**Table 1 | Demographics, comorbidities, and clinical factors in unmatched adults aged 18 to ≤65, UnitedHealth Group Clinical Discovery Database up to 31 October 2020**

| Characteristic  | Total 2020 population | 2020 SARS-CoV-2 infected group | 2020 Comparator group* | 2019 Comparator group* | vLRTI group*      |
|---|-----------------------|--------------------------------|------------------------|------------------------|-------------------|
| Total No  | 9 247 505             | 266 586                        | 8 980 919              | 9 722 381              | 1 655 907         |
| Follow-up from index date (days; median (IQR))            | —                     | 95 (42-135)                    | —                      | —                      | 226 (136-268)*    |
| Age (mean (SD))   | 42.4 (13.6)           | 41.7 (13.9)                    | 42.4 (13.6)*           | 42.5 (13.6)*           | 44.5 (13.4)*      |
| Age (No (%))  |                       |                                |                        |                        |                   |
| ≥18 to ≤34  | 2 903 571 (31.4)      | 90 497 (33.9)                  | 2 813 074 (31.3)       | 3 038 968 (31.3)       | 417 742 (25.2)    |
| >34 to ≤50  | 3 261 246 (35.3)      | 89 171 (33.4)                  | 3 172 075 (35.3)       | 3 426 770 (35.2)       | 587 892 (35.5)    |
| >50   | 3 082 688 (33.3)      | 86 918 (32.6)                  | 2 995 770 (33.4)       | 3 256 643 (33.5)       | 650 273 (39.3)    |
| Sex   |                       |                                |                        |                        |                   |
| Male  | 4 640 393 (50.2)      | 126 980 (47.6)                 | 4 513 413 (50.3)*      | 4 861 274 (50.0)*      | 707 084 (42.7)*   |
| SES index (median (IQR))                                  | 53.2 (51.6-55.2)      | 52.9 (51.2-55.1)               | 53.2 (51.6-55.2)*      | 53.1 (51.6-55.2)*      | 52.9 (51.2-55.2)* |
| Proportion of black individuals (mean (SD))               | 0.10 (0.15)           | 0.13 (0.17)                    | 0.10 (0.15) *          | 0.10 (0.15) *          | 0.11 (0.15) *     |
| Proportion of black individuals (No (%))                  |                       |                                |                        |                        |                   |
| ≤0.25   | 8 218 593 (88.9)      | 226 443 (84.9)                 | 7 992 150 (89.0)       | 8 618 587 (88.6)       | 1 462 821 (88.3)  |
| >0.25 to ≤0.50  | 666 921 (7.2)         | 25 327 (9.5)                   | 641 594 (7.1)          | 712 056 (7.3)          | 131 297 (7.9)     |
| >0.50 to ≤0.75  | 252 382 (2.7)         | 10 241 (3.8)                   | 242 141 (2.7)          | 272 726 (2.8)          | 44 280 (2.7)      |
| >0.75   | 109 609 (1.2)         | 4 575 (1.7)                    | 105 034 (1.2)          | 119 012 (1.2)          | 17 509 (1.1)      |
| Proportion of Hispanic individuals (mean (SD))            | 0.15 (0.17)           | 0.19 (0.21)                    | 0.14 (0.17)*           | 0.15 (0.17)*           | 0.15 (0.17)*      |
| Proportion of Hispanic individuals (No (%))               |                       |                                |                        |                        |                   |
| ≤0.25   | 7 705 168 (83.3)      | 201 097 (75.4)                 | 7 504 071 (83.6)       | 8 053 655 (82.8)       | 1 359 968 (82.1)  |
| >0.25 to ≤0.50  | 1 040 572 (11.3)      | 39 727 (14.9)                  | 1 000 845 (11.1)       | 1 120 296 (11.5)       | 202 486 (12.2)    |
| >0.50 to ≤0.75  | 352 428 (3.8)         | 16 934 (6.4)                   | 335 494 (3.7)          | 380 593 (3.9)          | 64 624 (3.9)      |
| >0.75   | 149 337 (1.6)         | 8 828 (3.3)                    | 140 509 (1.6)          | 167 837 (1.7)          | 28 829 (1.7)      |
| Comorbidity (No (%))                                      |                       |                                |                        |                        |                   |
| Any   | 3 311 964 (35.8)      | 120 749 (45.3)                 | 3 191 215 (35.5)*      | 3 584 107 (36.9) *     | 996 015 (60.1)*   |
| Previous conditions (No (%))                              |                       |                                |                        |                        |                   |
| Alzheimer dementia  | 6760 (0.1)            | 385 (0.1)                      | 6375 (0.1)*            | 8283 (0.1)*            | 3203 (0.2)*       |
| Asthma  | 355 183 (3.8)         | 15112 (5.7)                    | 340 071 (3.8)*         | 382 552 (3.9)*         | 162 684 (9.8)*    |
| Cystic fibrosis   | 1284 (0.0)            | 64 (0.0)                       | 1220 (0.0)*            | 1349 (0.0)*            | 818 (0.0)*        |
| Immunodeficiency  | 57 139 (0.6)          | 3101 (1.2)                     | 54 038 (0.6)*          | 60 545 (0.6)*          | 22 040 (1.3)*     |
| Pulmonary fibrosis  | 9436 (0.1)            | 488 (0.2)                      | 8948 (0.1)*            | 10566 (0.1)*           | 5444 (0.3)*       |
| Sickle cell disease                                       | 4008 (0.0)            | 273 (0.1)                      | 3735 (0.0)*            | 4258 (0.0)*            | 1211 (0.1)*       |
| Smoking   | 289 418 (3.1)         | 8113 (3.0)                     | 281 305 (3.1)*         | 321 713 (3.3)*         | 104 321 (6.3)*    |
| Type 1 diabetes   | 46 947 (0.5)          | 2157 (0.8)                     | 44 790 (0.5)*          | 52 806 (0.5)*          | 18 342 (1.1)*     |
| Type 2 diabetes   | 474 752 (5.1)         | 22619 (8.5)                    | 452 133 (5.0)*         | 521 746 (5.4)*         | 171 703 (10.4)*   |
| Thalassemia   | 6055 (0.1)            | 226 (0.1)                      | 5829 (0.1)*            | 6467 (0.1)*            | 1358 (0.1)        |
| Previous length of stay in hospital (days; mean (SD))     | 0.3 (2.9)             | 0.5 (5.9)                      | 0.3 (2.7)*             | 0.3 (2.9)*             | 0.5 (5.1)         |
| Previous PCP visit (No (%))                               | 5 142 739 (55.6)      | 186 313 (70.0)                 | 4 956 426 (55.2)       | 5 617 767 (57.8)       | 107 686 (75.1)    |
| No of PCP visits (mean (SD))                              | 1.7 (2.7)             | 2.5 (3.5)                      | 1.7 (2.6)*             | 1.8 (2.8)*             | 2.9 (3.7)*        |
| No of cardiology visits (mean (SD))                       | 0.2 (1.0)             | 0.3 (1.7)                      | 0.2 (1.0)*             | 0.2 (1.1)*             | 0.4 (1.7)*        |
| No of nephrology visits (mean (SD))                       | 0.0 (0.7)             | 0.1 (1.6)                      | 0.0 (0.6)*             | 0.0 (0.7)*             | 0.1 (1.5)         |
| Region (No (%))   |                       |                                |                        |                        |                   |
| South   | 3 924 142 (42.4)      | 117 292 (44.0)                 | 3 806 850 (42.4)*      | 4 039 716 (41.6)*      | 799 967 (48.3)*   |
| Midwest   | 2 171 722 (23.5)      | 54 494 (20.4)                  | 2 117 228 (23.6)*      | 2 312 541 (23.8)*      | 363 017 (21.9)*   |
| Northeast   | 1 590 979 (17.2)      | 56 071 (21.0)                  | 1 534 908 (17.1)*      | 1 714 661 (17.6)*      | 281 385 (17.0)*   |
| West  | 1 538 866 (16.6)      | 38 300 (14.4)                  | 1 500 566 (16.7)*      | 1 631 945 (16.8)*      | 208 353 (12.6)*   |
| Unknown   | 21796 (0.2)           | 429 (0.2)                      | 21 367 (0.2)*          | 23 518 (0.2)*          | 3185 (0.2)*       |
| Clinical characteristics (No (%))                         |                       |                                |                        |                        |                   |
| Covid-ICU visit flag                                      | —                     | 2786 (1.1)                     | —                      | —                      | —                 |
| Admission to hospital with covid-19                       | —                     | 21 746 (8.2)                   | —                      | —                      | —                 |
| Diagnostic method (No (%))†                               |                       |                                |                        |                        |                   |
| PCR test positive   | —                     | 77 273 (29.0)                  | —                      | —                      | —                 |
| Clinical diagnosis (not PCR test)                         | —                     | 178 642 (67.0)                 | —                      | —                      | —                 |
| Admitted to hospital with covid-19 (not PCR or diagnosis) | —                     | 10 671 (4.0)                   | —                      | —                      | —                 |

vLRTI=viral lower respiratory tract illness; IQR=interquartile range; SD=standard deviation; SES=socioeconomic status; PCP=primary care physician; PCR=polymerase chain reaction; ICU=intensive care unit.

\*All P values ≤0.05 comparing SARS-CoV-2 individuals with three comparison groups.

†Presented as hierarchical although many individuals had more than one diagnostic method recorded.

in women) and myocarditis, hypercoagulability, deep vein thrombosis, kidney injury, and sleep apnea (more commonly diagnosed in men) (fig 4 and eTable 4b). With a few exceptions, individuals with pre-existing conditions (fig 4 and eTable 4c) and individuals admitted to hospital with covid-19 (fig 4 and eTable

4d) had a greater excess risk of developing new clinical sequelae because of SARS-CoV-2 infection.

We did not find a significant period effect for most outcomes (eTable 4e). In our sensitivity analysis, the risk differences increased when the post-acute phase was shortened to an index date plus 14 days (eTable

**Table 2 | Proportion of adults aged 18 to ≤65 with new clinical sequelae of SARS-CoV-2 infection after the acute phase compared with three groups matched by propensity score, UnitedHealth Group Clinical Discovery Database up to 31 October 2020**

| No of sequelae* | No (%) SARS-CoV-2 infected (n=193 113) | No (%) 2020 comparator (n=193 113) | Difference (%) | No (%) SARS-CoV-2 infected (n=193 264) | No (%) 2019 comparator (n=193 264) | Difference (%) | No (%) SARS-CoV-2 infected (n=181 613) | No (%) vLRTI comparator (n=181 613) | Difference (%) |
|-----------------|--|------------------------------------|----------------|--|------------------------------------|----------------|--|-------------------------------------|----------------|
| None            | 166 039 (85.98)                        | 175 593 (90.93)                    | —              | 166 094 (85.94)                        | 174 963 (90.53)                    | —              | 155 083 (85.39)                        | 158 078 (87.04)                     | —              |
| 1               | 19 328 (10.01)                         | 13 671 (7.08)                      | 2.93†          | 19 408 (10.04)                         | 14 224 (7.36)                      | 2.68†          | 18 858 (10.38)                         | 17 627 (9.71)                       | 0.68†          |
| ≥2              | 7746 (4.01)                            | 3849 (1.99)                        | 2.02†          | 7762 (4.02)                            | 4077 (2.11)                        | 1.91†          | 7672 (4.22)                            | 5908 (3.25)                         | 0.97†          |

vLRTI=viral lower respiratory tract illness.

\*Sequelae are new clinical diagnoses determined after the index date plus 21 days.

†McNemar test was performed and all P values &lt;0.01.

5a). Similar risk differences were seen for the index date plus 21 days and the index date plus 28 days (eTable 5b), suggesting an index date plus 21 days is a reasonable start to the post-acute phase of the illness.

## Discussion

### Principal findings

Our retrospective study conducted in a large administrative database evaluated the excess risk of developing a wide range of clinical sequelae after the acute phase of SARS-CoV-2 infection in commercially insured adults aged 18-65. We found that 14% of individuals aged ≤65 who were infected with SARS-CoV-2 developed at least one new type of clinical sequelae that required medical care after the acute phase of SARS-CoV-2 infection, which was 4.95% higher than the 2020 comparator group and 1.65% higher than individuals diagnosed as having viral lower respiratory tract illness. This finding suggests that the SARS-CoV-2 virus is not unique in causing clinical sequelae after the acute infection. Our results confirmed an excess risk for specific types of sequelae in the four months after the acute phase (index date plus 21 days). Our analysis also showed that although the risk increased with age, pre-existing conditions, and admission to hospital for covid-19, younger adults (aged ≤50), those with no pre-existing conditions, and individuals infected with SARS-CoV-2 not admitted to hospital were also at risk for new clinical sequelae after the acute infection. Finally, our results suggested that the risk for some clinical sequelae, such as mental health diagnoses, were increased regardless of age and pre-existing condition.

### Comparison with other studies

When we considered the risk attributable to SARS-CoV-2 infection, several clinical sequelae were increased in survivors after the acute infection regardless of comparison group (2020, 2019, or viral lower respiratory tract illness group). These outcomes included chronic respiratory failure, cardiac irregularities, such as tachycardia and arrhythmia, hypercoagulability in the form of pulmonary embolism and deep vein thrombosis, anxiety, encephalopathy, peripheral neuropathy, amnesia, diabetes, liver test abnormalities, myocarditis, and fatigue. Many

of these outcomes have been previously reported in case studies or observational studies during the acute phase of covid-19 (including tachycardia,<sup>26</sup> hypercoagulability,<sup>27</sup> mental health outcomes,<sup>28</sup> encephalopathy,<sup>29,30</sup> diabetes,<sup>31</sup> and amnesia<sup>32</sup>). A few studies have also highlighted persistent symptoms or new clinical diagnoses after the acute infection,<sup>32-38</sup> although few have reported on a full range of new clinical diagnoses across multiple organ systems in such a large population.

Studies of individuals infected with other coronaviruses have shown arrhythmia and tachycardia sequelae in survivors of SARS,<sup>39,40</sup> and central and peripheral nervous system sequelae in survivors of SARS and MERS.<sup>41</sup> The proportion of individuals infected with SARS-CoV-2 with a new diagnosis of encephalopathy (0.23) and peripheral neuropathy (0.31) in our study (encephalopathy risk difference 0.09-0.19 and peripheral neuropathy risk difference 0.12-0.16) was closer to the higher rates of central and peripheral nervous system sequelae seen with MERS.<sup>41</sup> In another study, anxiety was reported to be the most common type of mental health sequelae in individuals with a diagnosis of covid-19 (hazard ratio 1.59-2.62) at 14-90 days after SARS-CoV-2 infection.<sup>42</sup> Our results on mental health outcomes (anxiety; hazard ratio 1.24-1.54) were similar, although absolute risks were lower in our study population.

We also identified excess risk for clinical sequelae that were not unique to SARS-CoV-2 and are commonly seen with other serious viral infections. The magnitude of the relative risk for these incident sequelae (eg, hypertension, stroke, and kidney injury), however, was nearly twice that typically seen in the general population in a normal year. Because of the scale of the SARS-CoV-2 pandemic in the US, these findings suggest more planning for healthcare resources is needed to look at the health complications in survivors.

### Strengths and limitations

Our study had the power to quantify the small, but not trivial, risk among younger, healthier adults. In our study population of adults aged ≤65, more than 90% of individuals recovered at home. Younger individuals might experience complications caused by covid-19, especially if they are admitted to hospital and have pre-existing conditions,<sup>43</sup> but few studies have reported

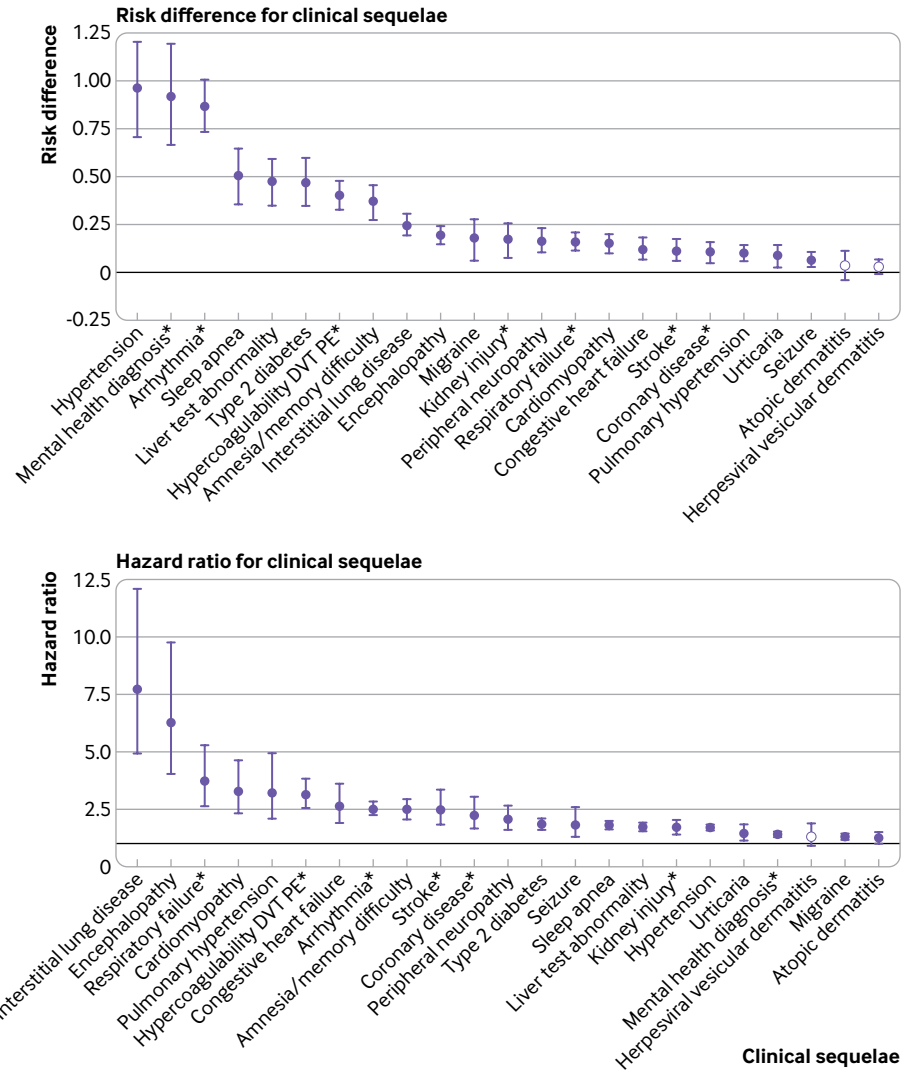


Fig 2 | Risk difference (per 100 individuals) and hazard ratios for the most common clinical sequelae in the SARS-CoV-2 versus the 2020 comparator group, UnitedHealth Group Clinical Discovery Database up to 31 October 2020. Clinical sequelae are diagnoses with an incidence  $\geq 0.1$  in the SARS-CoV-2 group during the first four months after the acute infection (index date plus 21 days) and highest in hierarchy if an aggregate diagnosis. We adopted this rule to avoid confidence intervals that were too wide to display. Symptoms are not displayed. All associations for each of the 51 outcomes are listed in eTable 2a-2d. Filled symbols indicate significant risk difference or hazard ratio (Bonferroni corrected P value  $\leq 0.05$ ). Atopic dermatitis is present in all plots as a negative control. \*Aggregate diagnosis includes all subdiagnoses listed in eTable 1. DVT=deep vein thrombosis; PE=pulmonary embolism

new outcomes after the acute phase of the illness in individuals with milder symptoms.

Fatigue is often the most reported symptom after the acute infection, with self-reported estimates in surveys ranging from 13.6%<sup>36</sup> to 77.7%<sup>44</sup> depending on whether individuals were admitted to hospital and length of follow-up. Although fatigue was also the most common diagnosis after the acute infection in this cohort (4.64%), our estimates only reflected fatigue that was reported to and noted by the physician. Most symptoms, when associated with a primary viral illness such as covid-19, are frequently not coded by the clinician because they are presumed to be part of the infectious process. Also, many individuals with covid-19 might not seek medical care unless the symptoms are unusually long lasting or severe. ICD-10

codes have been shown to be valid for many clinical diagnoses<sup>45</sup> but are inaccurate and unreliable for determining symptoms.<sup>46</sup> Therefore, we intentionally did not evaluate a broad list of symptoms in this study as we expect that the true incidence of symptoms was not accurately reflected when determined by ICD-10 codes and would be best established through patient reported surveys.

We could not determine race or ethnicity at the individual level in the commercial claims. More research is needed to better understand how race and ethnicity modify the risk for long term clinical sequelae. Moreover, our population required continuous enrollment from January 2019 until the index date, and so individuals who might not have been insured during this period were not included in our sample.



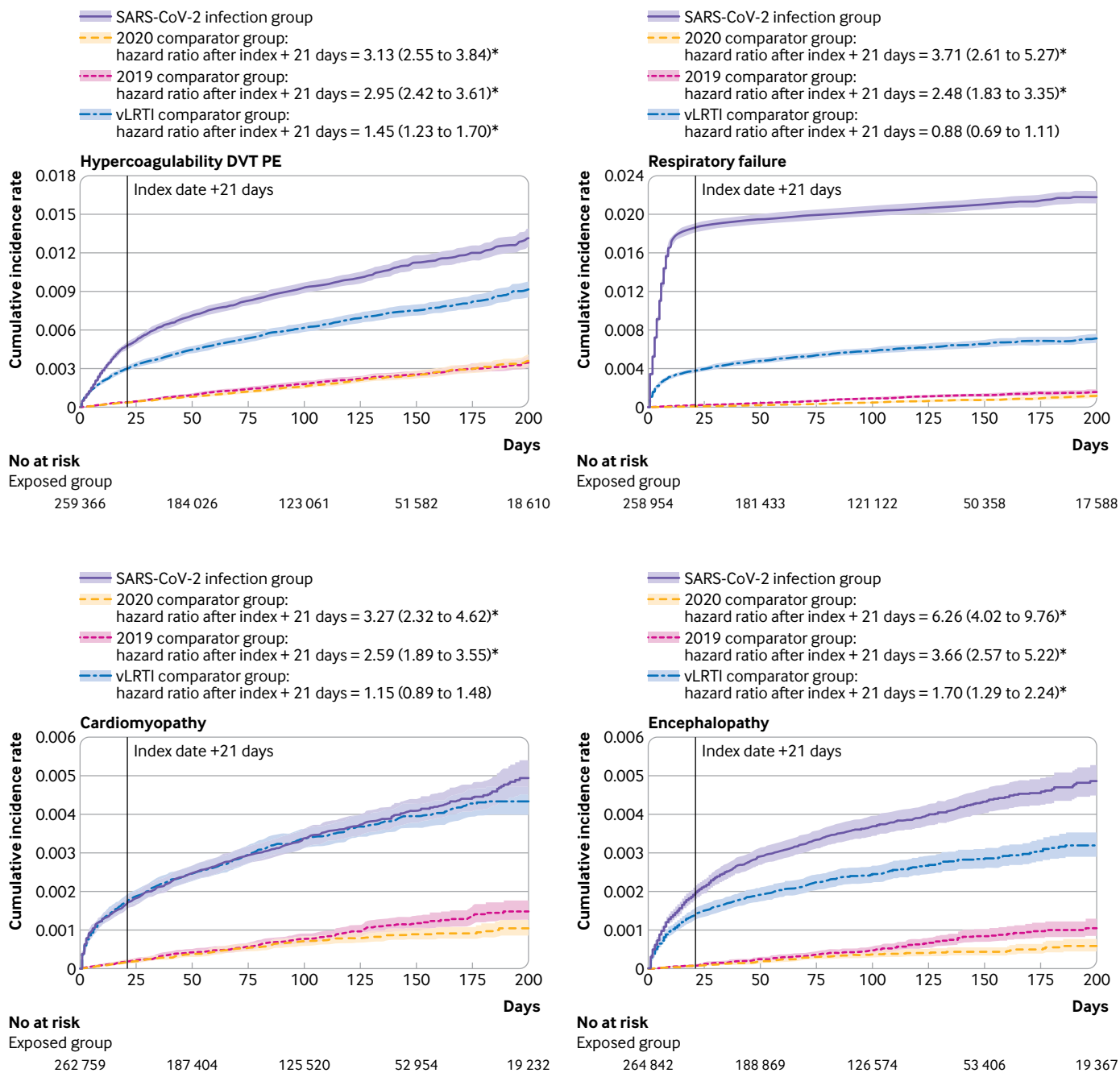
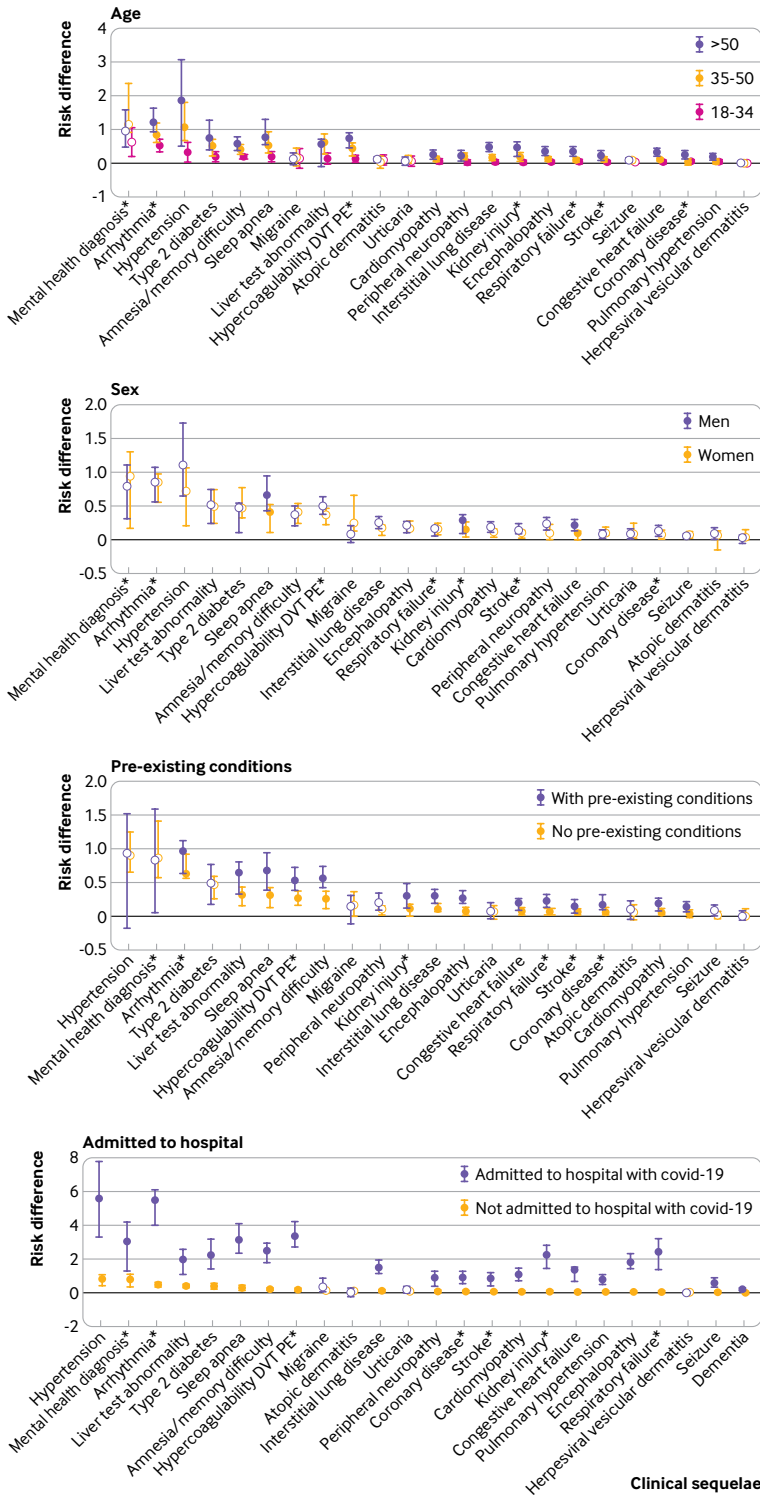


Fig 3 | Cumulative incidence rate for new clinical sequelae and symptoms after SARS-CoV-2 infection with three groups matched by propensity score, UnitedHealth Group Clinical Discovery Database up to 31 October 2020. vLRTI=viral lower respiratory tract illness. \*P<0.05

Death might be an important competing risk. We did not perform a competing risk analysis, however, for two reasons. First, we could not identify mortality as an outcome in our commercial claims database. Disenrollment from the insurance plan might be used as a proxy but the reason for disenrollment was not available. Therefore, we could not distinguish between withdrawal because of loss or change in employment and withdrawal because of death. Second, less than 5% of people disenrolled from the insurance plan during the follow-up period in our matched cohorts. Also, the difference in the number of disenrollments between individuals infected with SARS-CoV-2 and the

three comparison groups was minimal (range 0.03-0.43%), suggesting that disenrollment was not an informative censoring event.

We might have misclassified individuals because of the retrospective nature of our study and the inherent limitations of using claims to define variables. First, we relied on a database of laboratory results to identify individuals with a positive PCR test. Individuals infected with SARS-CoV-2 might have been misclassified into the 2020 comparator group if they tested positive outside of our network and did not have enough symptoms to require medical care. This misclassification would likely have biased our findings toward the null.



**Fig 4 | Risk difference (per 100 individuals) for the most common clinical sequelae by subpopulation, UnitedHealth Group Clinical Discovery Database up to 31 October 2020.** Clinical sequelae are diagnoses with an incidence  $\geq 0.1$  in the SARS-CoV-2 group for any subgroups during the first four months after the acute infection (index date plus 21 days) and highest in hierarchy if an aggregate diagnosis. We adopted this rule to avoid confidence intervals that were too wide to display. Symptoms are not displayed. All associations for each of the 51 outcomes are listed in eTable 4a-4d. Filled symbols indicate significant interaction term (Bonferroni corrected P value  $\leq 0.05$ ). Atopic dermatitis is present in all plots as a negative control. \*Aggregate diagnosis includes all subdiagnoses listed in eTable 1. DVT=deep vein thrombosis; PE=pulmonary embolism

Second, because we required continuous enrollment for study eligibility, we captured clinical diagnoses and comorbidities that occurred in the previous year and in the year of the index date up until 30 days before the index date. An incident or new diagnosis in the post-acute phase for a specific condition might have been an exacerbation of a pre-existing condition that did not receive medical care within the 13-22 month window created by our continuous enrollment criteria. Third, in the group with no pre-existing comorbidities, we might have misclassified some individuals, potentially inflating the incidence in that subgroup. This misclassification is likely minimal, however, because most medical conditions and comorbidities require individuals to participate in at least annual check-ups with primary care physicians or specialists. Fourth, we might not have captured a comprehensive list of all ICD-10 codes for each outcome, although the most common ones were included. Finally, because SARS-CoV-2 is a new virus, we believe that physicians might have underestimated the clinical significance of some outcomes, especially early in the pandemic. Our sensitivity analysis, however, does not support this temporal lack of documentation because we found no significant difference in risk by period.

Our small excess risks might be a result of increased medical care after SARS-CoV-2 infection, such that the ascertainment (diagnosis) rather than the incidence of a new diagnosis was triggered by SARS-CoV-2 infection. This ascertainment bias is unlikely to fully explain our results, however, because individuals with viral lower respiratory tract illness should receive similar medical care after the acute illness but have fewer visits to primary care physicians than individuals infected with SARS-CoV-2, regardless of the method of diagnosis (eFigure 2).

This retrospective analysis was strengthened by its large sample size that powered the assessment of multiple and rare outcomes simultaneously overall and across multiple subgroups. We used valid ICD-10 codes to determine outcomes, and included a broad definition of SARS-CoV-2 infection (positive PCR results or clinical diagnosis), producing a generalizable sample of individuals.

**Policy implications**

With almost 70 million individuals infected with SARS-CoV-2 worldwide and rising, the number of survivors with potential sequelae after covid-19 will continue to grow. To manage these patients effectively, understanding the incidence and natural history of these sequelae is important. Our results provide clinicians with a comprehensive understanding of the excess risk for over 50 clinical morbidities across multiple organ systems affecting adults aged  $\leq 65$  after the acute phase of SARS-Cov-2 infection. Knowing the magnitude of risk for rare and common clinical sequelae might improve the diagnosis and management of individuals infected with SARS-CoV-2. Also, our results could help providers and other key stakeholders anticipate the scale of future health

complications and improve planning on the use of healthcare resources.

### Conclusion

We found that 14% of individuals with SARS-CoV-2 infection developed a new type of clinical sequelae that required medical care after the acute phase of the illness, which was 4.95% higher than the 2020 comparator group. An increased and sustained risk for clinical sequelae was seen during the four months after the acute illness, particularly, but not exclusively, in individuals with pre-existing conditions or admitted to hospital for covid-19. More follow-up is needed to determine resolution of risk over time.

We thank Andrea Rotnitzky and Jamie Robins for reviewing and providing feedback on our methods.

**Contributors:** SED, YG, KH, JS, MCD, KGJ, ML, and KC played an active role in all aspects of the development of the research, including design, conduct, and interpretation of the data; preparation, review, and approval of the manuscript; and decision to submit the manuscript for publication. A description of the data source is included in supplementary appendix A. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. SED and YG serve as joint guarantors who accept full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

**Funding:** The study was funded by OptumLabs, the research and development arm of UnitedHealth Group, and the authors SED, KC, KH, YG, MCD, KGJ, and JS are full time employees at UnitedHealth Group. UnitedHealth Group, the organization, was not involved in determining the details of the study design, analysis plan, interpretation of data, or writing of the report.

**Competing interests:** All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare: support from OptumLabs for the submitted work and no other organization or company provided support for this work; SED, YG, KH, JS, MCD, KGJ declare that they have had no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted. KC declares consulting from Pfizer. ML declares honoraria/consulting from Merck, Sanofi-Pasteur, Bristol Myers-Squibb, and Antigen Discovery; research funding (institutional) from Pfizer, and unpaid scientific advice to Janssen, Astra-Zeneca, One Day Sooner, and Covaxx (United Biomedical).

**Ethical approval:** This research was determined to be exempt from human research regulations by the UnitedHealth Group Office of Human Research Affairs (action ID: 2020-0076-01).

**Data sharing:** The data are proprietary and are not available for public use but, under certain conditions, might be made available to editors and their approved auditors under a data use agreement to confirm the findings of this study.

The lead author, SED, affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

**Dissemination to participants and related patient and public communities:** The content of this paper will be summarized and made publicly available on the United In Research ([unitedinresearch.com](http://unitedinresearch.com)) platform sponsored by OptumLabs.

**Provenance and peer review:** Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- 1 Leung TYM, Chan AYL, Chan EW, et al. Short- and potential long-term adverse health outcomes of COVID-19: a rapid review. *Emerg Microbes Infect* 2020;9:2190-9. doi:10.1080/22221751.2020.1825914
- 2 Bonow RO, O'Gara PT, Yancy CW. Cardiology and COVID-19. *JAMA* 2020;324:1131-2. doi:10.1001/jama.2020.15088
- 3 Winkelmayer WC, Khairallah P, Charytan DM. Nephrology and COVID-19. *JAMA* 2020;324:1137-8. doi:10.1001/jama.2020.16779
- 4 Josephson SA, Kamel H. Neurology and COVID-19. *JAMA* 2020;324:1139-40. doi:10.1001/jama.2020.14254
- 5 Sudre CH, Murray B, Varsavsky T, et al. Attributes and predictors of Long-COVID: analysis of COVID cases and their symptoms collected by the Covid Symptoms Study App. *medRxiv* [Preprint] 2020;2020.10.19.20214494.
- 6 Carfi A, Bernabei R, Landi F, Gemelli Against COVID-19 Post-Acute Care Study Group. Persistent Symptoms in Patients After Acute COVID-19. *JAMA* 2020;324:603-5. doi:10.1001/jama.2020.12603
- 7 McGinty EE, Presskreischer R, Han H, Barry CL. Psychological Distress and Loneliness Reported by US Adults in 2018 and April 2020. *JAMA* 2020;324:93-4. doi:10.1001/jama.2020.9740
- 8 Tenforde MW, Kim SS, Lindsell CJ, et al. IVY Network Investigators. CDC COVID-19 Response Team, IVY Network Investigators. Symptom Duration and Risk Factors for Delayed Return to Usual Health Among Outpatients with COVID-19 in a Multistate Health Care Systems Network - United States, March-June 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:993-8. doi:10.15585/mmwr.mm6930e1
- 9 van den Borst B, Peters JB, Brink M, et al. Comprehensive health assessment three months after recovery from acute COVID-19. *Clin Infect Dis* 2020. <https://doi.org/10.1093/cid/ciaa1750>.
- 10 Chow EJ, Rolfes MA, O'Halloran A, et al. Acute Cardiovascular Events Associated With Influenza in Hospitalized Adults: A Cross-sectional Study. *Ann Intern Med* 2020;173:605-13. doi:10.7326/M20-1509
- 11 Sellers SA, Hagan RS, Hayden FG, Fischer WA 2nd. The hidden burden of influenza: A review of the extra-pulmonary complications of influenza infection. *Influenza Other Respir Viruses* 2017;11:372-93. doi:10.1111/irv.12470
- 12 Das KM, Lee EY, Singh R, et al. Follow-up chest radiographic findings in patients with MERS-CoV after recovery. *Indian J Radiol Imaging* 2017;27:342-9. doi:10.4103/ijri.IJRI\_469\_16
- 13 Rogers JP, Chesney E, Oliver D, et al. Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the COVID-19 pandemic. *Lancet Psychiatry* 2020;7:611-27. doi:10.1016/S2215-0366(20)30203-0
- 14 Wu Q, Zhou L, Sun X, et al. Altered Lipid Metabolism in Recovered SARS Patients Twelve Years after Infection. *Sci Rep* 2017;7:9110. doi:10.1038/s41598-017-09536-z
- 15 Docherty AB, Harrison EM, Green CA, et al. ISARIC4C investigators. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 2020;369:m1985. doi:10.1136/bmj.m1985
- 16 Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010;21:383-8. doi:10.1097/EDE.0b013e3181d61eeb
- 17 Agency for Healthcare Research and Quality. Creation of New Race-Ethnicity Codes and SES Indicators for Medicare Beneficiaries - Chapter 3. 2008. <https://archive.ahrq.gov/research/findings/final-reports/medicareindicators/medicareindicators3.html>
- 18 Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* 2011;46:399-424. doi:10.1080/00273171.2011.568786
- 19 le Cessie S, van Houwelingen JC. Ridge Estimators in Logistic Regression. *J R Stat Soc Ser C Appl Stat* 1992;41:191-201.
- 20 Schaefer RL, Roi LD, Wolfe RA. A ridge logistic estimator. *Commun Stat Theory Methods* 1984;13:99-113. doi:10.1080/03610928408828664
- 21 Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika* 1982;69:239-41. doi:10.1093/biomet/69.1.239
- 22 Kassambara A, Kosinski M, Biecek P, Fabian S. survminer: Drawing Survival Curves using "ggplot2". 2020. <https://CRAN.R-project.org/package=survminer>
- 23 Therneau TM, Lumley T, Atkinson E, Crowson C. survival: Survival Analysis. 2020. <https://CRAN.R-project.org/package=survival>
- 24 Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010;33:1-22. doi:10.18637/jss.v033.i01
- 25 R: The R Project for Statistical Computing. <https://www.r-project.org/>
- 26 Bhatta A, Mayer MM, Adusumalli S, et al. COVID-19 and cardiac arrhythmias. *Heart Rhythm* 2020;17:1439-44. doi:10.1016/j.hrthm.2020.06.016

- 27 Malas MB, Naazie IN, Elsayed N, Mathlouthi A, Marmor R, Clary B. Thromboembolism risk of COVID-19 is high and associated with a higher risk of mortality: A systematic review and meta-analysis. *EClinicalMedicine* 2020;29:100639. doi:10.1016/j.eclinm.2020.100639
- 28 Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. *Brain Behav Immun* 2020;89:531-42. doi:10.1016/j.bbi.2020.05.048
- 29 Ye M, Ren Y, Lv T. Encephalitis as a clinical manifestation of COVID-19. *Brain Behav Immun* 2020;88:945-6. doi:10.1016/j.bbi.2020.04.017
- 30 Helms J, Kremer S, Merdji H, et al. Neurologic Features in Severe SARS-CoV-2 Infection. *N Engl J Med* 2020;382:2268-70. doi:10.1056/NEJMc2008597
- 31 Sathish T, Kapoor N, Cao Y, Tapp RJ, Zimmet P. Proportion of newly diagnosed diabetes in COVID-19 patients: A systematic review and meta-analysis. *Diabetes Obes Metab* 2021;23:870-4. doi:10.1111/dom.14269
- 32 Garrigues E, Janvier P, Kherabi Y, et al. Post-discharge persistent symptoms and health-related quality of life after hospitalization for COVID-19. *J Infect* 2020;81:e4-6. doi:10.1016/j.jinf.2020.08.029
- 33 Moreno-Pérez O, Merino E, Leon-Ramirez J-M, et al. COVID-19-ALC research group. Post-acute COVID-19 syndrome. Incidence and risk factors: A Mediterranean cohort study. *J Infect* 2021;82:378-83. doi:10.1016/j.jinf.2021.01.004
- 34 Hill JB, Garcia D, Crowther M, et al. Frequency of venous thromboembolism in 6513 patients with COVID-19: a retrospective study. *Blood Adv* 2020;4:5373-7. doi:10.1182/bloodadvances.2020003083
- 35 Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet* 2021;397:220-32. doi:10.1016/S0140-6736(20)32656-8
- 36 Logue JK, Franko NM, McCulloch DJ, et al. Sequelae in Adults at 6 Months After COVID-19 Infection. *JAMA Netw Open* 2021;4:e210830. doi:10.1001/jamanetworkopen.2021.0830
- 37 Chopra V, Flanders SA, O'Malley M, Malani AN, Prescott HC. Sixty-Day Outcomes Among Patients Hospitalized With COVID-19. *Ann Intern Med* 2021;174:576-8. doi:10.7326/M20-5661
- 38 Dennis A, Wamil M, Kapur S, et al. Multi-organ impairment in low-risk individuals with long COVID. *medRxiv* [Preprint] 2020. <https://doi.org/10.1101/2020.10.14.20212555>
- 39 Lau S-T, Yu W-C, Mok N-S, Tsui P-T, Tong W-L, Cheng SW. Tachycardia amongst subjects recovering from severe acute respiratory syndrome (SARS). *Int J Cardiol* 2005;100:167-9. doi:10.1016/j.ijcard.2004.06.022
- 40 Yu C-M, Wong RS-M, Wu EB, et al. Cardiovascular complications of severe acute respiratory syndrome. *Postgrad Med J* 2006;82:140-4. doi:10.1136/pgmj.2005.037515
- 41 Ellul MA, Benjamin L, Singh B, et al. Neurological associations of COVID-19. *Lancet Neurol* 2020;19:767-83. doi:10.1016/S1474-4422(20)30221-0
- 42 Taquet M, Luciano S, Geddes JR, Harrison PJ. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. *Lancet Psychiatry* 2021;8:130-40.
- 43 Cunningham JW, Vaduganathan M, Claggett BL, et al. Clinical Outcomes in Young US Adults Hospitalized With COVID-19. *JAMA Intern Med* 2020;181:379-81. doi:10.1001/jamainternmed.2020.5313
- 44 Davis HE, Assaf GS, McCorkell L, et al. Characterizing Long COVID in an International Cohort: 7 Months of Symptoms and Their Impact. *medRxiv* [Preprint] 2021. <https://doi.org/10.1101/2020.12.24.20248802>
- 45 Quan H, Li B, Saunders LD, et al. IMECCHI Investigators. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res* 2008;43:1424-41. doi:10.1111/j.1475-6773.2007.00822.x
- 46 Crabb BT, Lyons A, Bale M, et al. Comparison of International Classification of Diseases and Related Health Problems, Tenth Revision Codes with Electronic Medical Records Among Patients With Symptoms of Coronavirus Disease 2019. *JAMA Netw Open* 2020;3:e2017703. [https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list\\_uids=32797176&dopt=Abstract](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=32797176&dopt=Abstract) doi:10.1001/jamanetworkopen.2020.17703

**Web appendix:** Supplemental tables and figures  
**Web appendix:** Appendix A - UnitedHealth Group  
 Clinical Research Database