


BMJ Open Severity of COVID-19 and adverse long-term outcomes: a retrospective cohort study based on a US electronic health record database

Nick Jovanoski,¹ Xin Chen,² Ursula Becker,¹ Kelly Zalocusky,² Devika Chawla,² Larry Tsai,² Michelle Borm,³ Margaret Neighbors,² Vincent Yau ²

To cite: Jovanoski N, Chen X, Becker U, *et al.* Severity of COVID-19 and adverse long-term outcomes: a retrospective cohort study based on a US electronic health record database. *BMJ Open* 2021;**11**:e056284. doi:10.1136/bmjopen-2021-056284

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-056284>).

Received 30 August 2021
Accepted 17 November 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Global Access, F. Hoffmann-La Roche Ltd, Basel, Switzerland
²Product Development, Genentech Inc, South San Francisco, California, USA
³Product Development Medical Affairs, F. Hoffmann-La Roche Ltd, Basel, Switzerland

Correspondence to

Dr Vincent Yau;
yau.vincent@gene.com

ABSTRACT

Objective To identify potential risk factors for adverse long-term outcomes (LTOs) associated with COVID-19, using a large electronic health record (EHR) database.
Design Retrospective cohort study. Patients with COVID-19 were assigned into subcohorts according to most intensive treatment setting experienced. Newly diagnosed conditions were classified as respiratory, cardiovascular or mental health LTOs at >30–≤90 or >90–≤180 days after COVID-19 diagnosis or hospital discharge. Multivariate regression analysis was performed to identify any association of treatment setting (as a proxy for disease severity) with LTO incidence.
Setting Optum deidentified COVID-19 EHR dataset drawn from hospitals and clinics across the USA.
Participants Individuals diagnosed with COVID-19 (N=57 748) from 20 February to 4 July 2020.
Main outcomes Incidence of new clinical conditions after COVID-19 diagnosis or hospital discharge and the association of treatment setting (as a proxy for disease severity) with their risk of occurrence.

Results Patients were assigned into one of six subcohorts: outpatient (n=22 788), emergency room (ER) with same-day COVID-19 diagnosis (n=11 633), ER with COVID-19 diagnosis ≤21 days before ER visit (n=2877), hospitalisation without intensive care unit (ICU; n=16 653), ICU without ventilation (n=1837) and ICU with ventilation (n=1960). Respiratory LTOs were more common than cardiovascular or mental health LTOs across subcohorts and LTO incidence was higher in hospitalised versus non-hospitalised subcohorts. Patients with the most severe disease were at increased risk of respiratory (risk ratio (RR) 1.86, 95% CI 1.56 to 2.21), cardiovascular (RR 2.65, 95% CI 1.49 to 4.43) and mental health outcomes (RR 1.52, 95% CI 1.20 to 1.91) up to 6 months after hospital discharge compared with outpatients.

Conclusions Patients with severe COVID-19 had increased risk of new clinical conditions up to 6 months after hospital discharge. The extent that treatment setting (eg, ICU) contributed to these conditions is unknown, but strategies to prevent COVID-19 progression may nonetheless minimise their occurrence.

INTRODUCTION

The COVID-19 pandemic caused by the novel SARS-CoV-2 has imposed an immense burden of morbidity and mortality worldwide.¹

Strengths and limitations of this study

- This study used a large electronic health record database containing a rich source of patient-level medical and administrative records from hospitals, emergency departments and outpatient centres across the USA.
- Multivariate logistic regression analysis was used to adjust for measured confounders and assess the association of increasing COVID-19 severity (proxied by treatment setting) with the risk of new clinical conditions being diagnosed up to 6 months after COVID-19 diagnosis or hospital discharge.
- A sensitivity analysis assessing the association of increasing COVID-19 severity (proxied by treatment setting) with the risk of a new cancer diagnosis served as a negative control.
- The main limitation of this retrospective study is that we use treatment setting as a proxy for COVID-19 severity and therefore it is difficult to tease out associations specific to the treatment setting (eg, invasive ventilation) from the underlying COVID-19 severity; any differences that exist between cohorts could bias the results and as all potential confounders may not be controlled for, the results do not indicate causality.
- Additional limitations include missing information on smoking status, the lack of a COVID-19-negative control group, the possibility of missing data, being restricted to examining conditions captured by International Classification of Disease-10 codes, the lack of information on COVID-19 treatments received and the lack of laboratory values or other biomarkers to better characterise disease.

Although the majority of patients experience mild or moderate symptoms that resolve within a few weeks of initial infection, increasing evidence suggests that a subset of patients continue to display symptoms beyond 4 weeks after infection.^{2–3} These symptoms are wide ranging and often extend beyond the typical initial symptoms of COVID-19 to include respiratory (eg, dyspnoea, decreased

exercise capacity), cardiovascular (eg, heart palpitations, chest pain) and mental health (eg, confusion, disorientation) disorders.^{4 5} Notably, such outcomes have been observed even in patients with mild acute COVID-19 symptoms.⁶ These prolonged symptoms have collectively been referred to by several names including postacute COVID-19, post-COVID-19 syndrome (PCS), postacute sequelae of SARS-CoV-2 infection and possibly more commonly ‘long COVID-19’.^{7 8} However, due to the overlapping and non-specific range of symptoms experienced, the medical community has not yet converged on precise definitions and it is possible that distinct subsets of patients with long COVID-19 exist. It has also been suggested that long COVID-19 can be further subdivided into subacute COVID-19 (4–12 weeks after initial onset of COVID-19 symptoms) and PCS (beyond 12 weeks).^{4 9} The underlying pathogenic mechanisms of long COVID-19 are not well understood, but multiple causes have been proposed, including immune dysregulation and viral persistence.¹⁰ Additionally, in patients with severe disease requiring treatment in the intensive care unit (ICU), non-specific secondary effects cannot be ruled out, similar to those observed in ‘postintensive care syndrome’.¹¹

High-quality clinical data on respiratory, cardiovascular and neurologic sequelae of SARS-CoV-2 infection are beginning to emerge^{12–14} and several observational studies and patient registries have been established to better understand the long-term outcomes (LTOs) of COVID-19.^{15 16} However, little is known about the potential baseline factors that may predict the development of long COVID-19.

Retrospective cohort studies using electronic health records (EHRs) are uniquely positioned due to their size and convenience to provide insights into factors underlying long COVID-19 development and the range of long COVID-19 conditions that exist. The Optum deidentified COVID-19 EHR dataset contains patient-level medical and administrative records from hospitals, emergency departments, outpatient centres and laboratories across the USA. This dataset has previously been used to

describe key epidemiological features of a large cohort of hospitalised patients with COVID-19¹⁷ and to develop a prognostic model of in-hospital mortality.¹⁸

The current study used the Optum deidentified COVID-19 EHR dataset to better understand the types of LTOs encountered by patients with long COVID-19, to define the factors that predict their diagnosis and to understand the role that treatment setting (as a proxy for COVID-19 severity) plays in the manifestation of these outcomes.

METHODS

Database

Individuals with COVID-19 diagnosed between 20 February 2020 and 4 July 2020 were extracted from the Optum deidentified COVID-19 EHR dataset (569 149 individuals from 3 832 315 in the entire dataset). This dataset contains patient-level medical and administrative records from hospitals, emergency departments, outpatient centres and laboratories across the USA. All data were deidentified according to the Health Insurance Portability and Accountability Act Expert Method and managed according to Optum customer data use agreements. The COVID-19 EHR dataset comprises clinical information sourced from hospital networks that provided data meeting Optum’s internal data quality criteria. Data cleaning methods used were as described previously.¹⁷

Patients and study design

Eligible patients (overall COVID-19 cohort) had ≥ 1 of the following: a COVID-19 diagnosis code (U07.1, U07.2), a positive diagnostic test for SARS-CoV-2 infection (eg, molecular or antigen test) or a B97.29 diagnosis code (other coronavirus as the cause of diseases classified elsewhere) without a negative SARS-CoV-2 molecular test within 14 days. The index date was defined as the date of COVID-19 diagnosis or COVID-19-related hospitalisation (as defined below), whichever occurred first. The

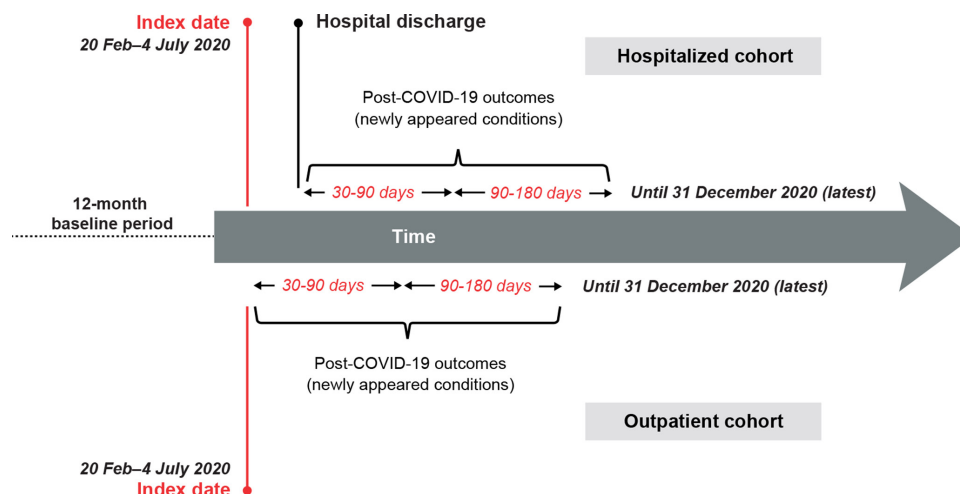


Figure 1 Overall study design.

baseline period was defined as the 12 months prior to the index date and a minimum of 180 days follow-up was required for all patients. The overall study design is shown in [figure 1](#).

Eligible patients were assigned into the following six subcohorts according to treatment setting: (1) Outpatient, patients with a COVID-19 diagnosis and no record of hospitalisation or an emergency room (ER) visit within 21 days of diagnosis; (2) ER on diagnosis, COVID-19 diagnosis on the same day as ER visit; (3) ER, COVID-19 diagnosis prior to ER visit, that is, patients with an ER visit within 21 days after COVID-19 diagnosis (excluding diagnosis date); (4) Hospitalisation without ICU, patients hospitalised with no record of ICU admission; (5) Hospitalised with ICU but no ventilation, patients hospitalised with record of ICU admission but no record of ventilator or extracorporeal membrane oxygenation (ECMO) use during ICU stay and (6) Hospitalised with ICU and ventilation, patients hospitalised with record of ICU admission and ventilator or ECMO use during ICU stay.

Hospitalisation was defined as an inpatient or ER overnight visit with an initial COVID-19 diagnosis made during hospitalisation and within 7 days of admission or an inpatient or ER overnight visit within 21 days of the initial COVID-19 diagnosis, where the hospital had a record of this diagnosis. Contiguous ER and inpatient visits with a gap of up to 1 day were considered a single hospitalisation. If a patient had multiple eligible hospitalisations, only data from the first hospitalisation were considered, as described previously.¹⁷

Modeling and statistical analysis

LTOs occurring >30–≤180 days after hospital discharge or COVID-19 diagnosis were categorised into one of the two time windows (>30–≤90 days or >90–≤180 days) and were further classified as respiratory, cardiovascular or mental health conditions (online supplemental table 1).¹⁹ LTOs were selected to capture a broad range of potential sequelae, even if there was no strong clinical or pathological rationale for their choice, given the absence of sufficient clinical data regarding established complications associated with COVID-19. Multivariate logistic regression analyses were performed to determine the association of disease severity (proxied by treatment setting) with the three LTO classifications. Covariates were intended to encompass the main known risk factors for developing severe COVID-19²⁰ and included demographic information (ie, age, gender, race, ethnicity, diagnosis month, insurance type, obesity status) and baseline health conditions (ie, those included in the Charlson Comorbidity Index (CCI) (online supplemental table 2). CCI was treated as a numeric variable, while all other variables were treated as categorical. Age was binned into <18, 18–29, 30–39, 40–49, 50–65, 65–74, 75–84 and ≥85 years. Date of diagnosis was also binned into months in 2020 (pre April, April, May, June, July; allowing for ≥180 days follow-up until 31 December 2020 at the latest). Patients were excluded from the regression model examining a

specific LTO category if they had a diagnosis in that category in the 12 months prior to the index date (eg, if a patient had an asthma diagnosis 12 months prior to the index date, they would be excluded from the model for respiratory LTOs).

All statistical analysis was performed using R V.3.6.3.²¹ Using the sjstats package, regression was performed using the function 'glm' and the risk ratio (RR) was calculated by converting the OR using the function 'OR to RR'.²² Increased risk of diagnosis of a health condition was implied when the RR and both the low and high 95% CI limits were >1 and decreased risk was implied when the RR and low and high 95% CIs were <1.

Sensitivity analysis

A sensitivity analysis was performed to investigate the potential association of disease severity (proxied by treatment setting) with risk of a new cancer diagnosis, to serve as a negative control. The same set of covariates was used as per the main analysis, but cancer diagnosis was the only LTO examined. Currently, no evidence exists to suggest that COVID-19 severity increases the risk of a new cancer diagnosis. Thus, an association here may indicate that the associations from the main analysis may be driven by other differences between patients across treatment settings.

Patient and public involvement

No patient involved.

RESULTS

Patient population

In total, 57748 patients were eligible for the overall COVID-19 cohort. [Table 1](#) presents descriptive statistics of the patients by subcohort. Mean age tended to be higher in patients in hospitalised subcohorts (53.2–57.7 years) than in those in non-hospitalised subcohorts (41.0–46.8 years). Overall, 53.3% of patients were female. Across all patients, 50.3% were Caucasian, 22.8% were African American, 3.2% were Asian and the remaining 23.6% were missing information on race. Additionally, 67.5% were of non-Hispanic ethnicity, while data on ethnicity were missing for 11.8% of patients. Overall, 19% of patients were obese and the mean weighted CCI Score was 1.20. Information on smoking status was missing for 93.1% of patients ([table 1](#)). Full details of demographics and baseline characteristics are provided in online supplemental table 3.

The proportions of patients with incipient respiratory, cardiovascular and/or mental health conditions that were diagnosed either >30–≤90 days or >90–≤180 days after COVID-19 diagnosis or hospital discharge are provided in [table 2](#). The proportions of patients with new LTOs were generally higher in the subcohorts with more severe disease (ie, the ER subcohort and all hospitalised subcohorts) compared with the outpatient subcohort. In addition, the proportion of patients with respiratory LTOs

Table 1 Baseline characteristics of patients with COVID-19, overall and by subcohort

	Subcohort						
	All patients (N=57 748)	Outpatient (n=22 788)	ER on diagnosis (n=11 633)	ER (n=2877)	Hospitalisation without ICU (n=16 653)	ICU without ventilation (n=1837)	ICU with ventilation (n=1960)
Mean age (SD), years	47.93 (18.76)	46.78 (18.90)	40.95 (16.89)	44.00 (16.61)	53.17 (18.50)	55.92 (17.32)	57.70 (14.78)
Age group, n (%)							
<18 years	2184 (3.8)	1033 (4.5)	641 (5.5)	78 (2.7)	366 (2.2)	46 (2.5)	20 (1.0)
18–29 years	8509 (14.7)	3693 (16.2)	2543 (21.9)	538 (18.7)	1574 (9.5)	89 (4.8)	72 (3.7)
30–39 years	8972 (15.5)	3541 (15.5)	2496 (21.5)	579 (20.1)	2046 (12.3)	175 (9.5)	135 (6.9)
40–49 years	9362 (16.2)	3664 (16.1)	2205 (19.0)	555 (19.3)	2442 (14.7)	253 (13.8)	243 (12.4)
50–64 years	16 103 (27.9)	6231 (27.3)	2706 (23.3)	780 (27.1)	4937 (29.6)	626 (34.1)	823 (42.0)
65–74 years	7065 (12.2)	2658 (11.7)	689 (5.9)	230 (8.0)	2683 (16.1)	363 (19.8)	442 (22.6)
75–84 years	3620 (6.3)	1279 (5.6)	239 (2.1)	76 (2.6)	1647 (9.9)	195 (10.6)	184 (9.4)
≥85 years	891 (1.5)	303 (1.3)	54 (0.5)	22 (0.8)	440 (2.6)	44 (2.4)	28 (1.4)
Missing	1042 (1.8)	386 (1.7)	60 (0.5)	19 (0.7)	518 (3.1)	46 (2.5)	13 (0.7)
Sex, n (%)							
Female	30 782 (53.3)	12 856 (56.4)	6115 (52.6)	1721 (59.8)	8487 (51.0)	829 (45.1)	774 (39.5)
Male	26 939 (46.6)	9920 (43.5)	5515 (47.4)	1152 (40.0)	8160 (49.0)	1008 (54.9)	1184 (60.4)
Missing	27 (<0.1)	12 (<0.1)	3 (<0.1)	4 (<0.1)	6 (<0.1)	0 (0.0)	2 (<0.1)
Race, n (%)							
African American	13 183 (22.8)	3473 (15.2)	3178 (27.3)	790 (27.5)	4675 (28.1)	551 (30.0)	516 (26.3)
Asian	1848 (3.2)	639 (2.8)	438 (3.8)	100 (3.5)	555 (3.3)	41 (2.2)	75 (3.8)
Caucasian	29 074 (50.3)	13 746 (60.3)	4653 (40.0)	1337 (46.5)	7538 (45.3)	849 (46.2)	951 (48.5)
Missing	13 643 (23.6)	4930 (21.6)	3364 (28.9)	650 (22.6)	3885 (23.3)	396 (21.6)	418 (21.3)
Ethnicity, n (%)							
Hispanic	11 932 (20.7)	3942 (17.3)	3378 (29.0)	646 (22.5)	3298 (19.8)	332 (18.1)	336 (17.1)
Non- Hispanic	38 988 (67.5)	15 485 (68.0)	7121 (61.2)	1987 (69.1)	11 648 (69.9)	1294 (70.4)	1453 (74.1)
Missing	6828 (11.8)	3361 (14.7)	1134 (9.7)	244 (8.5)	1707 (10.3)	211 (11.5)	171 (8.7)
Smoking status, n (%)							
Current smoker	413 (0.7)	193 (0.8)	128 (1.1)	13 (0.5)	61 (0.4)	13 (0.7)	5 (0.3)
Previously smoked	740 (1.3)	417 (1.8)	111 (1.0)	27 (0.9)	145 (0.9)	25 (1.4)	15 (0.8)
Never smoked	2831 (4.9)	1468 (6.4)	750 (6.4)	121 (4.2)	404 (2.4)	50 (2.7)	38 (1.9)
Missing	53 764 (93.1)	20 710 (90.9)	10 644 (91.5)	2716 (94.4)	16 043 (96.3)	1749 (95.2)	1902 (97.0)
Obese, n (%)*	10 952 (19.0)	4905 (21.5)	1366 (11.7)	580 (20.2)	3246 (19.5)	406 (22.1)	449 (22.9)
Insurance, n (%)							
Commercial	29 145 (50.5)	13 134 (57.6)	5672 (48.8)	1482 (51.5)	7243 (43.5)	758 (41.3)	856 (43.7)
Medicaid	8652 (15.0)	2341 (10.3)	2223 (19.1)	542 (18.8)	2891 (17.4)	312 (17.0)	343 (17.5)
Medicare	8774 (15.2)	3173 (13.9)	788 (6.8)	245 (8.5)	3674 (22.1)	435 (23.7)	459 (23.4)
Other payor type	4004 (6.9)	1282 (5.6)	1071 (9.2)	211 (7.3)	1188 (7.1)	129 (7.0)	123 (6.3)
Uninsured	4833 (8.4)	1542 (6.8)	1731 (14.9)	281 (9.8)	1069 (6.4)	111 (6.0)	99 (5.1)

Continued

Table 1 Continued

	Subcohort						
	All patients (N=57 748)	Outpatient (n=22 788)	ER on diagnosis (n=11 633)	ER (n=2877)	Hospitalisation without ICU (n=16 653)	ICU without ventilation (n=1837)	ICU with ventilation (n=1960)
Missing	2340 (4.1)	1316 (5.8)	148 (1.3)	116 (4.0)	588 (3.5)	92 (5.0)	80 (4.1)
Month of COVID-19 diagnosis, n (%)							
February 2020	115 (0.2)	61 (0.3)	3 (<0.1)	4 (0.1)	34 (0.2)	5 (0.3)	8 (0.4)
March 2020	8197 (14.2)	1527 (6.7)	1893 (16.3)	527 (18.3)	3288 (19.7)	306 (16.7)	656 (33.5)
April 2020	18 591 (32.2)	6018 (26.4)	3480 (29.9)	926 (32.2)	6684 (40.1)	676 (36.8)	807 (41.2)
May 2020	14 188 (24.6)	6154 (27.0)	2703 (23.2)	729 (25.3)	3755 (22.5)	477 (26.0)	370 (18.9)
June 2020	14 832 (25.7)	7846 (34.4)	3073 (26.4)	597 (20.8)	2826 (17.0)	371 (20.2)	119 (6.1)
July 2020	1825 (3.2)	1182 (5.2)	481 (4.1)	94 (3.3)	66 (0.4)	2 (0.1)	0 (0.0)
Mean weighted CCI (SD)	1.20 (2.06)	1.09 (2.03)	0.56 (1.32)	0.88 (1.76)	1.64 (2.31)	2.15 (2.46)	2.13 (2.39)

*No patient data were missing for obesity.

CCI, Charlson Comorbidity Index; ER, emergency room; ICU, intensive care unit.

was higher than the proportions with cardiovascular or mental health LTOs. New respiratory LTOs were diagnosed more frequently during the earlier time window across subcohorts, except in the outpatient subcohort where the proportion of patients diagnosed was the same in both time windows (both 8.1%; [table 2](#)). No clear temporal trends were noted for diagnosis of cardiovascular or mental health LTOs, with similar proportions of patients with new cardiovascular and mental health LTOs observed in the >30–≤90-day and >90–≤180-day windows for each subcohort ([table 2](#)). The proportions of patients with LTOs in more than one category (ie, ‘respiratory and cardiovascular’, ‘respiratory and mental health’, ‘mental health and cardiovascular’ or ‘respiratory, cardiovascular and mental health’) were lower than the proportions of patients with LTOs in a single category, suggesting that a diagnosis in one category did not necessarily lead to a diagnosis in another.

Regarding individual conditions, the prevalence of newly diagnosed pneumonia, dyspnoea and respiratory failure in the >90–≤180-day window closely followed the pattern of initial COVID-19 severity (as proxied by treatment setting), with most cases being diagnosed in the ‘ICU with ventilation’ subcohort (online supplemental table 4). Similarly, although encephalopathy, confusion or disorientation, cardiac arrhythmia and myocardial infarction were less common, the prevalence of these conditions also increased with increasing COVID-19 severity. Full details of conditions that were diagnosed in the >30–≤90-day and >90–≤180-day windows following COVID-19 diagnosis or hospital discharge are provided in online supplemental table 4.

Modeling

The most striking potential covariate associated with increased risk of newly diagnosed respiratory conditions at >30–≤90 days and >90–≤180 days post COVID-19 diagnosis or hospital discharge was increasing severity of illness according to increasing hospitalisation severity, using the outpatient subcohort as the reference group ([figure 2](#) and online supplemental table 5). ICU with ventilation was associated with increased risk of a novel respiratory condition diagnosis compared with the outpatient subcohort at >30–≤90 days (RR 2.64, 95% CI 2.27 to 3.04) and >90–≤180 days post COVID-19 diagnosis or hospital discharge (RR 1.86, 95% CI 1.55 to 2.21); in addition, ICU without ventilation was associated with increased risk during the >30–≤90-day time window (RR 1.69, 95% CI 1.39 to 2.03), while ER was associated with increased risk at both >30–≤90 days (RR 1.39, 95% CI 1.17 to 1.65) and >90–≤180 days (RR 1.33, 95% CI 1.10 to 1.58) post COVID-19 diagnosis or hospital discharge. By contrast, patients with an ER visit on the COVID-19 diagnosis date were less likely than those in the outpatient subcohort to be diagnosed with a new respiratory condition at >30–≤90 days (RR 0.64, 95% CI 0.56 to 0.74) and 90–180 days post COVID-19 diagnosis or hospital discharge (RR 0.56, 95% CI 0.48 to 0.65). Additional covariates associated with increased risk of new respiratory conditions were older patient age and obesity. A COVID-19 diagnosis during or prior to April 2020 exhibited a non-significant trend towards increased risk of new respiratory condition occurrence compared with later diagnosis, which may reflect changes in treatment algorithms over time. Full results are presented in online supplemental table 5.

Table 2 Long-term outcomes that were diagnosed >30–≤90 days or >90–≤180 days post COVID-19 by subcohort

Condition, n (%)	Outpatient (N=22788)		ER on diagnosis (N=11 633)		ER (N=2877)		Hospitalisation without ICU (N=16 653)		ICU without ventilation (N=1837)		ICU with ventilation (N=1960)	
	>30–≤90 days	>90–≤180 days	>30–≤90 days	>90–≤180 days	>30–≤90 days	>90–≤180 days	>30–≤90 days	>90–≤180 days	>30–≤90 days	>90–≤180 days	>30–≤90 days	>90–≤180 days
Respiratory	1846 (8.1)	1846 (8.1)	520 (4.5)	484 (4.2)	322 (11.2)	311 (10.8)	2124 (12.8)	1599 (9.6)	296 (16.1)	222 (12.1)	547 (27.9)	381 (19.4)
CV	439 (1.9)	475 (2.1)	81 (0.7)	92 (0.8)	54 (1.9)	57 (2.0)	585 (3.5)	618 (3.7)	102 (5.6)	99 (5.4)	128 (6.5)	132 (6.7)
Mental health	940 (4.1)	1081 (4.7)	231 (2.0)	293 (2.5)	144 (5.0)	154 (5.4)	819 (4.9)	866 (5.2)	119 (6.5)	118 (6.4)	167 (8.5)	144 (7.3)
Cancer	137 (0.6)	151 (0.7)	24 (0.2)	30 (0.3)	18 (0.6)	14 (0.5)	95 (0.6)	106 (0.6)	9 (0.5)	12 (0.7)	22 (1.1)	14 (0.7)
Respiratory and CV	131 (0.6)	124 (0.5)	16 (0.1)	30 (0.3)	17 (0.6)	19 (0.7)	201 (1.2)	181 (1.1)	30 (1.6)	26 (1.4)	63 (3.2)	47 (2.4)
Respiratory and mental health	205 (0.9)	216 (0.9)	65 (0.6)	63 (0.5)	48 (1.7)	33 (1.1)	262 (1.6)	245 (1.5)	45 (2.4)	34 (1.9)	83 (4.2)	50 (2.6)
Mental health and CV	53 (0.2)	49 (0.2)	8 (0.1)	9 (0.1)	4 (0.1)	2 (0.1)	57 (0.3)	64 (0.4)	10 (0.5)	11 (0.6)	7 (0.4)	11 (0.6)
Respiratory, CV and mental health	57 (0.3)	52 (0.2)	9 (0.1)	10 (0.1)	6 (0.2)	9 (0.3)	98 (0.6)	84 (0.5)	17 (0.9)	15 (0.8)	30 (1.5)	31 (1.6)
No new conditions* (respiratory, CV or mental health)	20 066 (88.1)	19 879 (87.2)	10 908 (93.8)	10 886 (93.6)	2438 (84.7)	2427 (84.4)	13 841 (83.1)	14 228 (85.4)	1439 (78.3)	1499 (81.6)	1331 (67.9)	1473 (75.2)

*Only conditions that appeared >30–≤180 days after COVID-19 diagnosis or hospital discharge are included; pre-existing conditions are excluded. CV, cardiovascular; ER, emergency room; ICU, intensive care unit.

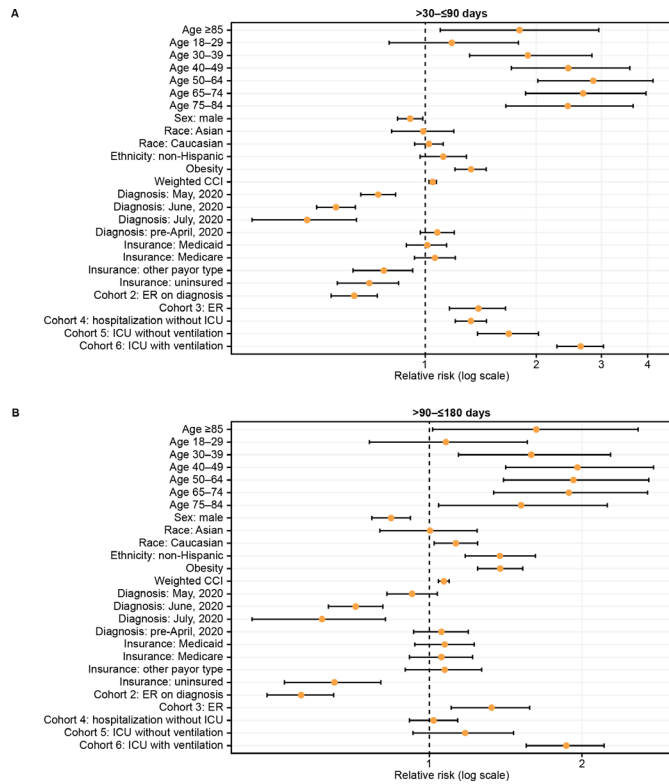


Figure 2 Relative risk of new respiratory conditions occurring from >30 days to ≤180 days after COVID-19 diagnosis or hospital discharge. Relative risk of new respiratory conditions occurring at (A) >30–≤90 days and (B) >90–≤180 days after COVID-19 diagnosis or hospital discharge. Graphs represent relative risk and 95% CIs. Reference groups are: <18 years (age), female (sex), African American (race), Hispanic (ethnicity), non-obese (obesity), diagnosis in February and March 2020 (diagnosis month), commercial (insurance) and cohort 1: outpatient (cohort). CCI, Charlson Comorbidity Index; ER, emergency room; ICU, intensive care unit.

Increasing hospitalisation severity was also found to be associated with increased risk of a new cardiovascular condition occurring post COVID-19 diagnosis or hospital discharge (figure 3 and online supplemental table 5). Notably, ICU with ventilation was associated with increased risk of the occurrence of novel cardiovascular conditions compared with the outpatient subcohort at >30–≤90 days (RR 3.16, 95% CI 1.83 to 5.18) and >90–≤180 days post COVID-19 diagnosis or hospital discharge (RR 2.65, 95% CI 1.49 to 4.43), while ICU without ventilation was associated with increased risk during the >90–≤180-day time window (RR 2.41, 95% CI 1.25 to 4.23). Similar to the findings regarding respiratory conditions, patients with an ER visit on the COVID-19 diagnosis date were less likely than outpatients to be diagnosed with novel cardiovascular conditions in both the >30–≤90-day (RR 0.45, 95% CI 0.27 to 0.71) and >90–≤180-day windows (RR 0.59, 95% CI 0.38 to 0.89). Additional covariates associated with an increased risk of new cardiovascular conditions occurring included older patient age and non-Hispanic

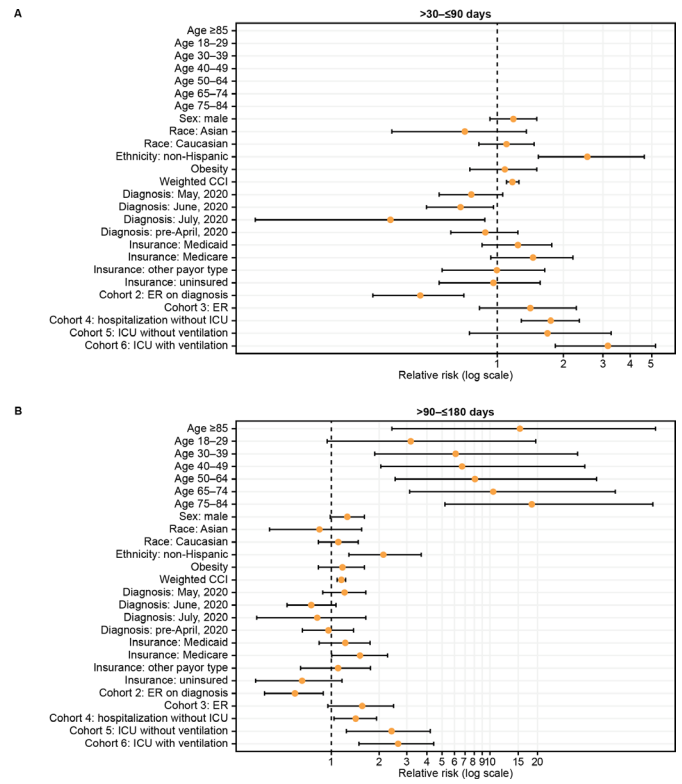


Figure 3 Relative risk of new cardiovascular conditions occurring from >30 days to ≤180 days after COVID-19 diagnosis or hospital discharge. Relative risk of new cardiovascular conditions occurring at (A) >30–≤90 days and (B) >90–≤180 days after COVID-19 diagnosis or hospital discharge. Graphs represent relative risk and 95% CIs. Reference groups are: <18 years (age), female (sex), African American (race), Hispanic (ethnicity), non-obese (obesity), diagnosis in February and March 2020 (diagnosis month), commercial (insurance) and cohort 1: outpatient (cohort). Relative risk in the >30–≤90-day time window was not calculated as no new diagnoses were made in the reference group (<18 years) during this time. CCI, Charlson Comorbidity Index; ER, emergency room; ICU, intensive care unit.

ethnicity. Full results are presented in online supplemental table 5.

The risk of a new mental health condition occurring post COVID-19 diagnosis or hospital discharge also increased according to increasing hospitalisation severity (figure 4 and online supplemental table 5). ICU with ventilation was associated with increased risk of a new mental health condition occurring compared with the outpatient subcohort at >30–≤90 days (RR 1.89, 95% CI 1.51 to 2.35) and >90–≤180 days post COVID-19 diagnosis or hospital discharge (RR 1.52, 95% CI 1.20 to 1.91) and ICU without ventilation was similarly associated with increased risk of a new mental health condition diagnosis during the >90–≤180-day window (RR 1.34, 95% CI 1.02 to 1.73). Of note, compared with those <18 years, all age groups examined appeared to be at higher risk of the occurrence of new mental health conditions at >30–≤90 days post COVID-19 diagnosis or hospital discharge. In the >90–≤180-day window, only the 65–74 and 75–84 years

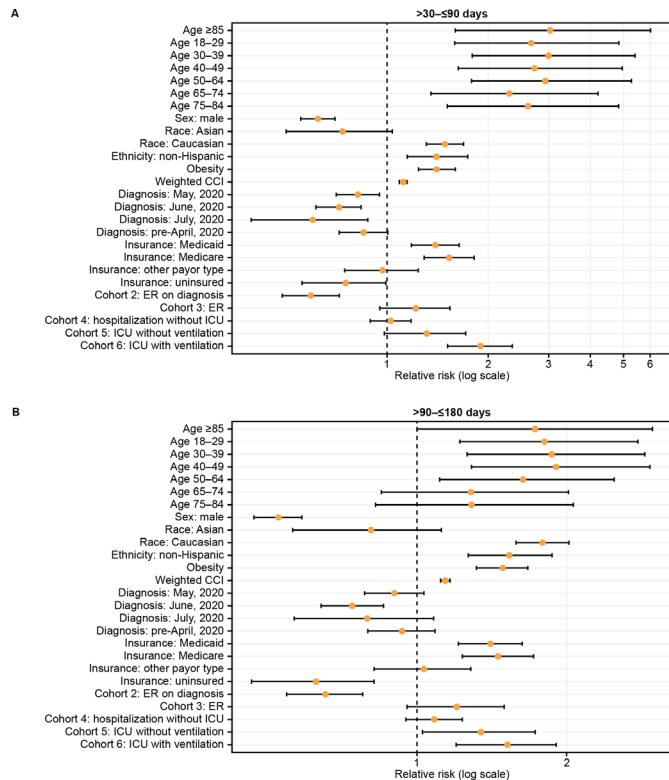


Figure 4 Relative risk of new mental health conditions occurring from >30 days to ≤180 days after COVID-19 diagnosis or hospital discharge. Relative risk of new mental health conditions occurring at (A) >30-≤90 days and (B) >90-≤180 days after COVID-19 diagnosis or hospital discharge. Graphs represent relative risk and 95% CIs. Reference groups are: <18 years (age), female (sex), African American (race), Hispanic (ethnicity), non-obese (obesity), diagnosis in February and March 2020 (diagnosis month), commercial (insurance) and cohort 1: outpatient (cohort). CCI, Charlson Comorbidity Index; ER, emergency room; ICU, intensive care unit.

age groups were not at higher risk. Additional covariates associated with increased risk of a new mental health condition occurring included obesity, Caucasian race and non-Hispanic ethnicity. See online supplemental table 5 for full results.

Sensitivity analysis

With the exception of older age, COVID-19 severity (proxied by treatment setting) did not predict a new cancer diagnosis up to 180 days after COVID-19 diagnosis or hospital discharge (online supplemental figure 1 and supplemental table 6), giving confidence in the results of the original analysis.

DISCUSSION

By using EHRs of over 55 000 patients from hospitals and clinics across the USA, this study set out to examine the types of new LTOs (ie, only those that were identified after COVID-19 diagnosis or hospital discharge) associated with long COVID-19 and to identify potential underlying factors that may contribute to their occurrence.

Severe disease was found to predict an increased likelihood of a new LTO diagnosis, whereby increasing hospitalisation severity was associated with increased risk of new respiratory (eg, pneumonia), cardiovascular (eg, myocardial infarction) and mental health conditions (eg, confusion or disorientation). In severely affected patients with COVID-19, some LTOs were diagnosed between three and 6 months after hospital discharge, suggesting that the overall COVID-19 burden extends far beyond the acute infection phase. In addition, although patients with severe disease were most at risk of presenting with new LTOs, non-hospitalised patients also experienced a relatively high incidence of LTOs, suggesting that even patients with mild disease are at risk of adverse long-term effects associated with COVID-19.

Although the data show a clear general trend of increased LTOs that correlated with COVID-19 severity (proxied by treatment setting), the specificity of this effect to COVID-19 is unclear, as ICU survivors commonly develop a range of new conditions on discharge collectively referred to as ‘postintensive care syndrome’, regardless of their underlying diagnosis.¹¹ Nonetheless, preventing the development of more severe disease, where possible, may decrease the likelihood of health problems post infection and would be expected to simultaneously increase the probability of survival. Together, these effects would have a cumulative positive impact on both patients and healthcare systems.

Interestingly, the ‘ER on diagnosis’ subcohort exhibited a reduced incidence of LTOs compared with the outpatient subcohort. The reasons for this are not clear but are likely due in part to the lower mean age and reduced incidence of comorbidities in this subcohort relative to the other subcohorts. In addition, it is possible that in the context of the pandemic, when primary care physicians had more limited personal protective equipment and other resources, these patients were directed to the ER to be tested for COVID-19, despite not having severe enough disease to warrant an ER visit. Finally, depending on the hospital setting and processes in place, asymptomatic patients who attended the ER for non-COVID-19 reasons may have tested positive while there, which may have led to the inclusion of milder COVID-19 cases in this subcohort.

Previous studies have examined the link between COVID-19 severity and LTOs. A study of 2469 hospitalised patients with COVID-19 in Wuhan, China, showed that more severe disease correlated with increased risk of LTOs up to 6 months after infection, including fatigue, sleep difficulties and anxiety or depression.²³ Anxiety or depression was observed in 23% of patients in that study compared with ~10% in our study; this difference is likely because our study was limited to newly diagnosed disorders in both inpatients and outpatients, while the previous study included new or worsening symptoms in hospitalised patients only. A separate, large study of patients with COVID-19 that used a US EHR database (N=2 363 79) to examine 6-month outcomes (inpatients and outpatients)

reported that ~7% of patients had a first anxiety disorder compared with ~17% that had any anxiety disorder and that increased incidence was correlated with increased disease severity.¹³ A further study compared 73 435 non-hospitalised patients with COVID-19 who were users of the Veterans Health Administration with 4 990 835 control patients and reported an increased risk of incident sequelae including, but not limited to, respiratory, cardiovascular and mental health disorders after a median follow-up duration of 126 and 130 days, respectively.¹⁹ Smaller, single-site hospital studies in the UK have reported similar trends between disease severity and shorter-term outcomes, with breathlessness commonly reported up to 12 weeks post COVID-19.^{24 25} In addition, self-reported data in patients with COVID-19 (N=4182) showed that upper respiratory complaints (eg, shortness of breath) and cardiac symptoms (eg, palpitations, tachycardia) were commonly reported in patients with long COVID-19 (symptoms lasting ≥ 28 days)²⁶ and data from a separate study using wearable devices provided further evidence of prolonged tachycardia in symptomatic patients with COVID-19.²⁷ The current study builds on these previous reports and provides additional evidence of a link between COVID-19 severity (proxied by treatment setting) and increased risk of developing LTOs, using a large dataset from both hospitalised and non-hospitalised patients. In addition, our study provides a detailed summary of the incidence of a wide range of specific health conditions that occurred up to 6 months after COVID-19 diagnosis or hospital discharge, providing a useful resource to better understand and characterise the range of conditions that constitute long COVID-19.

Our study categorised three major classes of LTOs that occur in patients with long COVID-19: respiratory, cardiovascular and mental health. This is broadly in keeping with a previous retrospective cohort study in England that followed 48 780 patients hospitalised with COVID-19, who had significantly higher rates of respiratory and cardiovascular disease after a mean follow-up of 140 days.²⁸ In addition, a retrospective study that used a large administrative all-payer database including 27 589 inpatients and 46 857 outpatients demonstrated that post COVID-19, patients were more likely to experience a range of conditions, including respiratory, nervous and circulatory system conditions, than outpatient control patients.²⁹ A greater understanding of the conditions that characterise long COVID-19 is needed to better anticipate the future healthcare burden of COVID-19 and to optimise strategies to minimise long COVID-19 development. In this regard, signals detected in the current study such as lung fibrosis, as well as other factors including paediatric long COVID-19, vaccination effects and healthcare utilisation, are topics that may warrant future analysis. In particular, a greater understanding of the long-term economic consequences of COVID-19 and the impact of long COVID-19 on patient quality of life is needed.

A major limitation of this analysis is that treatment setting is used as a proxy for COVID-19 severity; therefore,

it is difficult to tease out the effect of treatment setting procedures (eg, invasive ventilation) from the underlying COVID-19 severity. Furthermore, our analysis did not distinguish short-term outcomes from chronic health conditions. Additional limitations include missing information on smoking status, the restriction of follow-up to only 6 months, the lack of a COVID-19-negative control group, the possibility of missing data (eg, patients may have sought care for an LTO not captured in the Optum deidentified COVID-19 EHR dataset), the lack of information on COVID-19 treatments received and the lack of laboratory values or other biomarkers to better characterise disease. Finally, capture of health conditions relies on International Classification of Disease-10 (ICD-10) codes, whereas some conditions of interest (eg, anosmia, ageusia and brain fog) lack specific ICD-10 codes and other conditions are known to be undercaptured. The B97.29 diagnosis code includes other coronaviruses in addition to SARS-CoV-2 and may therefore be a potential limitation of our study; however, the majority of our COVID-19 cohort (>85%) was diagnosed from April to July using the official U07.1 diagnosis code that is specific to COVID-19, meaning it is unlikely that a substantial number of infections, if any, were from other coronaviruses.

CONCLUSIONS

Although LTOs were reported in patients across all subcohorts, increased risk of new respiratory, cardiovascular and mental health conditions was observed with increasing COVID-19 severity, using treatment setting as a proxy. Strikingly, the risk of new conditions being diagnosed remained high up to 6 months post COVID-19 diagnosis or hospital discharge, suggesting that the burden of COVID-19 extends far beyond the acute infection phase. Future research is warranted to understand specific factors that lead to the occurrence of new LTOs in patients with COVID-19 and to distinguish between the relative effect of COVID-19 severity versus any general effects that may occur after acute critical illness.

Acknowledgements The authors thank Shemra Rizzo for valuable contributions. Third-party medical writing assistance, under the direction of the authors, was provided by John Bett, PhD, from Ashfield MedComms, an Ashfield Health company.

Contributors All authors were involved in drafting and revising the manuscript, approved the final version and agree to being accountable for all aspects of the work. NJ contributed to the conception of the research question, study design, analysis and data interpretation. XC contributed to study design, analysis and data interpretation. UB contributed to the conception of the research question, study design, analysis and data interpretation. KZ contributed to the conception of the research question, design of the analysis, selection of outcomes and data interpretation. DC contributed to the conception of the research question, design of the analysis and selection of outcomes. LT contributed to study design and data interpretation. MB contributed to data interpretation. MN contributed to selection and categorisation of key complications for study design. VY contributed to the study design, acquisition, analysis and data interpretation.

Funding This work was supported by F. Hoffmann-La Roche Ltd.

Competing interests NJ and UB are employees of F. Hoffmann-La Roche Ltd. MB is an employee of Roche Nederland BV. UB and MB hold shares in F. Hoffmann-La Roche Ltd. XC, DC, LT, MN and VY are employees of Genentech, Inc. and hold shares

in F. Hoffmann-La Roche Ltd. KZ is a former employee of Genentech, Inc. and holds shares in F. Hoffmann-La Roche Ltd.

Patient consent for publication Not applicable.

Ethics approval The use of the Optum deidentified COVID-19 EHR dataset was reviewed by the New England Institutional Review Board (IRB) and was determined to be exempt from broad IRB approval, as this study did not involve human subject research.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Data were licensed from Optum and interested researchers may contact Optum for data access requests. All interested researchers can access the data in the same manner as the authors. The authors had no special access privileges.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access 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, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Vincent Yau <http://orcid.org/0000-0001-5301-5300>

REFERENCES

- Centers for Disease Control and Prevention. Estimated disease burden of COVID-19, 2021. Available: <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/burden.html> [Accessed May 2021].
- Carfi A, Bernabei R, Landi F, *et al.* Persistent symptoms in patients after acute COVID-19. *JAMA* 2020;324:603–5.
- Tenforde MW, Kim SS, Lindsell CJ, *et al.* 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.
- Nalbandian A, Sehgal K, Gupta A, *et al.* Post-acute COVID-19 syndrome. *Nat Med* 2021;27:601–15.
- Centers for Disease Control and Prevention. Post-COVID conditions, 2021. Available: <https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects.html> [Accessed May 2021].
- Augustin M, Schommers P, Stecher M, *et al.* Post-COVID syndrome in non-hospitalised patients with COVID-19: a longitudinal prospective cohort study. *Lancet Reg Health Eur* 2021;6:100122.
- Mullard A. Long COVID's long R&D agenda. *Nat Rev Drug Discov* 2021;20:329–31.
- National Institutes of Health. *NIH launches new initiative to study "Long COVID"*. Bethesda, Maryland: National Institutes of Health, 2021. <https://www.nih.gov/about-nih/who-we-are/nih-director/statements/nih-launches-new-initiative-study-long-covid> [Accessed July 2021]
- Mayor N, Tsang R, Joy M, *et al.* Long covid: coding is caring. *BMJ* 2021;373:n1262.
- Marx V. Scientists set out to connect the dots on long COVID. *Nat Methods* 2021;18:449–53.
- Brown SM, Bose S, Banner-Goodspeed V, *et al.* Approaches to addressing Post-Intensive care syndrome among intensive care unit survivors. A narrative review. *Ann Am Thorac Soc* 2019;16:947–56.
- Myall KJ, Mukherjee B, Castanheira AM, *et al.* Persistent Post-COVID-19 interstitial lung disease. An observational study of corticosteroid treatment. *Ann Am Thorac Soc* 2021;18:799–806.
- Taquet M, Geddes JR, Husain M, *et al.* 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *Lancet Psychiatry* 2021;8:416–27.
- Puntmann VO, Carerj ML, Wieters I, *et al.* Outcomes of cardiovascular magnetic resonance imaging in patients recently recovered from coronavirus disease 2019 (COVID-19). *JAMA Cardiol* 2020;5:1265–73.
- University of Birmingham. Why are some people more susceptible to the long-term health effects of COVID-19? 2021. Available: <https://www.birmingham.ac.uk/research/quest/21st-century-healthcare/long-covid.aspx> [Accessed May 2021].
- Mount Sinai Hospital. Center for Post-COVID care, 2021. Available: <https://www.mountsinai.org/about/covid19/center-post-covid-care> [Accessed May 2021].
- Chawla D, Rizzo S, Zalocusky K. Descriptive epidemiology of 16,780 hospitalized COVID-19 patients in the United States. *medRxiv* 2021.
- Incerti D, Rizzo S, Li X, *et al.* Prognostic model to identify and quantify risk factors for mortality among hospitalised patients with COVID-19 in the USA. *BMJ Open* 2021;11:e047121.
- Al-Aly Z, Xie Y, Bowe B. High-Dimensional characterization of post-acute sequelae of COVID-19. *Nature* 2021;594:259–64.
- Centers for Disease Control and Prevention. People with certain medical conditions, 2021. Available: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html> [Accessed July 2021].
- R Core Team. R: a language and environment for statistical computing Vienna, Austria, 2020. Available: <https://www.r-project.org/> [Accessed May 2021].
- Lüdtke D. Statistical functions for regression models (version 0.18.1), 2021. Available: <https://CRAN.R-project.org/package=sjstats> [Accessed May 2021].
- 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.
- Halpin SJ, McIvor C, Whyatt G, *et al.* Postdischarge symptoms and rehabilitation needs in survivors of COVID-19 infection: a cross-sectional evaluation. *J Med Virol* 2021;93:1013–22.
- Arnold DT, Hamilton FW, Milne A, *et al.* Patient outcomes after hospitalisation with COVID-19 and implications for follow-up: results from a prospective UK cohort. *Thorax* 2021;76:399–401.
- Sudre CH, Murray B, Varsavsky T, *et al.* Attributes and predictors of long COVID. *Nat Med* 2021;27:626–31.
- Radin JM, Quer G, Ramos E, *et al.* Assessment of prolonged physiological and behavioral changes associated with COVID-19 infection. *JAMA Netw Open* 2021;4:e2115959.
- Ayoubkhani D, Khunti K, Nafilyan V, *et al.* Post-covid syndrome in individuals admitted to hospital with covid-19: retrospective cohort study. *BMJ* 2021;372:n693.
- Chevinsky JR, Tao G, Lavery AM, *et al.* Late conditions diagnosed 1–4 months following an initial coronavirus disease 2019 (COVID-19) encounter: a matched-cohort study using inpatient and outpatient administrative Data-United states, 1 March–30 June 2020. *Clin Infect Dis* 2021;73:S5–S16.