Cerebral venous thrombosis: a retrospective cohort study of 513,284 confirmed COVID-19 cases and a comparison with 489,871 people receiving a COVID-19 mRNA vaccine

Maxime Taquet,[1,2] Masud Husain,[3,4] John R Geddes,[1,2] Sierra Luciano,[5] Paul J Harrison[1,2,*]

[1]Department of Psychiatry, University of Oxford, Oxford, UK

[2]Oxford Health NHS Foundation Trust, Oxford, UK

[3]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

[4]Oxford University Hospitals NHS Foundation Trust, Oxford, UK

[5]TriNetX, Cambridge MA, USA

*Corresponding author: University Department of Psychiatry, Warneford Hospital, Oxford

OX3 7JX, UK. paul.harrison@psych.ox.ac.uk

Abstract

Using an electronic health records network we estimated the absolute incidence of cerebral venous thrombosis (CVT) in the two weeks following COVID-19 diagnosis (N=513,284), or influenza (N=172,742), or receipt of the BNT162b2 or mRNA-1273 COVID-19 vaccines (N=489,871). The incidence of portal vein thrombosis (PVT) was also assessed in these groups, as well as the baseline CVT incidence over a two-week period. The incidence of CVT after COVID-19 diagnosis was 39.0 per million people (95% CI, 25.2–60.2). This was higher than the CVT incidence after influenza (0.0 per million people, 95% CI 0.0–22.2, adjusted RR=6.73, P=.003) or after receiving BNT162b2 or mRNA-1273 vaccine (4.1 per million people, 95% CI 1.1–14.9, adjusted RR=6.36, P<.001). The relative risks were similar if a broader definition of CVT was used. For PVT, the incidence was 436.4 per million people (382.9-497.4) after COVID-19, 98.4 (61.4-157.6) after influenza, and 44.9 (29.7-68.0) after BNT162b2 or mRNA-1273. The incidence of CVT following COVID-19 was higher than the incidence observed across the entire health records network (0.41 per million people over any 2-week period). Laboratory test results, available in a subset of the COVID-19 patients, provide preliminary evidence suggestive of raised D-dimer, lowered fibrinogen, and an increased rate of thrombocytopenia in the CVT and PVT groups. Mortality was 20% and 18.8% respectively. These data show that the incidence of CVT is significantly increased after COVID-19, and greater than that observed with BNT162b2 and mRNA-1273 COVID-19 vaccines. The risk of CVT following COVID-19 is also higher than the latest estimate from the European Medicines Agency for the incidence associated with ChAdOx1 nCoV-19 vaccine (5.0 per million people, 95% CI 4.3–5.8). Although requiring replication and corroboration, the present data highlight the risk of serious thrombotic events in COVID-19, and can help contextualize the risks and benefits of vaccination in this regard.

There are concerns about a possible association between vaccines against SARS-CoV-2 and cerebral venous thrombosis (CVT, also called cerebral venous sinus thrombosis; Silvis et al, 2017). The concern has focused primarily on ChAdOx1 nCoV-19 (Astra Zeneca), and more recently the Ad26.COV2-S vaccine (Janssen). The current risk with ChAdOx1 nCoV-19 is estimated at approximately 5 per million vaccinated individuals. Emerging data suggest that the association reflects a 'vaccine-induced thrombotic thrombocytopaenia' (VITT) (Greinacher et al, 2021; Schultz et al, 2021). Governments and medical regulators have reacted by restricting the use of the two vaccines in different subgroups of the population, based on a risk-benefit analysis. Yet one key component of the risk-benefit calculation that is crucial to understand the context of the risk is currently unknown: the absolute risk of CVT following a diagnosis of COVID-19. To date there are only a few case reports of CVT post-COVID-19 (Dakay et al 2021).

Here, using an electronic health records network, we estimated the incidence of CVT occurring in confirmed COVID-19 cases and compared this incidence to two other groups: people who received a COVID-19 mRNA vaccine, and a cohort of patients with influenza. We also compared the COVID-19 incidence to that observed in the whole network population, and with the latest estimate for the risk following the ChAdOx1 nCoV-19 vaccine, using the European Medicines Agency (EMA) data. We also examined the rate of portal vein thrombosis (PVT), another diagnosis associated with thrombosis in the venous system and thought to occur in VITT (Schultz et al, 2021).

Methods

We used TriNetX Analytics, a federated electronic health records network recording anonymized data from 59 healthcare organizations, primarily in the USA, totaling 81 million patients. For details, see Taquet et al. (2021), and the supplement.

A cohort of all patients who had a confirmed diagnosis of COVID-19 (ICD-10 code U07.1) between January 20, 2020 and March 25, 2021 was defined for study. The absolute risk of a diagnosis of CVT

(ICD-10 code I67.6) was calculated by identifying those patients in the cohort who had the diagnosis in the two weeks following their diagnosis of COVID-19. The absolute risk of patients with PVT (ICD-10 code I81) was also calculated. For the whole COVID-19 cohort, and for cases with CVT or PVT following COVID-19, baseline characteristics are reported. We identified patients who had a reported high D-dimer (> 5mg/L), low fibrinogen (< 200 mg/dL), or thrombocytopenia (any of the ICD-10 codes D69.49, D69.59, D69.6) within the 2 weeks after their COVID-19 diagnosis. We also assessed how many of them had died by the time of the analysis (April 14, 2021).

Two control cohorts based on other index events were used for comparison: a diagnosis of influenza (ICD-10 codes J09-J11) between January 20, 2020 and March 25, 2021, and the injection of a first dose of the two vaccines administered to this predominantly US population: the BNT162b2 ('Pfizer-BioNTech') vaccine or the mRNA-1273 ('Moderna') vaccine before March 25, 2021. We excluded from these cohorts any patients who had a diagnosis of COVID-19 on or after January 20, 2020. The absolute risk of CVT and PVT in each cohort was assessed in the same way as for the COVID-19 cohort. We calculated the relative risk (RR) of a CVT diagnosis and of a PVT diagnosis in the two weeks after COVID-19 diagnosis compared to the other index events (i.e. influenza, or vaccination). In addition, the overall incidence of CVT in the population was calculated by dividing the number of patients diagnosed with a CVT in the study period (January 20, 2020 to March 25, 2021) by the total population size and reporting it over 2 weeks by assuming a uniform distribution of CVT events throughout the study period. For comparison, we also report the rates of CVT and splanchnic thrombosis (PVT and thrombosis of other splanchnic veins) after ChAdOx1 nCoV-19 based on the European Medicines Agency data (https://www.ema.europa.eu/en/news/astrazenecas-covid-19-vaccine-ema-finds-possible-link-very-rare-cases-unusual-blood-clots-low-blood; accessed 14 April 2021).

We carried out two secondary analyses. First, to see whether the findings were specific to CVT, we calculated the incidence and relative risks for PVT; since this is not coded as such in the EMA data, we used 'splanchnic thrombosis' for that comparison. Second, the analyses were repeated after broadening the diagnostic criteria for CVT to include I63.6 (cerebral infarction due to central thrombosis, non-

pyogenic), G08 (intracranial and intraspinal phlebitis and thrombophlebitis), O22.5 (CVT in pregnancy) and O87.3 (CVT in the puerperium), in line with recent studies that have taken this approach to CVT in other settings (Handley and Emsley, 2020; Otite et al, 2020).

Statistical analyses

Fisher's exact tests were used to compare characteristics (baseline and laboratory) and death rates between patients with COVID-19 who had a CVT (or PVT) compared to patients with COVID-19 who did not. Fisher's exact tests were also used to test the null hypothesis that the relative risks of CVT and PVT in the two weeks after COVID-19 vs. influenza and vs. mRNA vaccine were equal to 1. Confidence intervals for absolute risks were based on Wilson score intervals. Confidence intervals for relative risks were based on Wald confidence limits, with small sample adjustment if the number of events in any cohort was lower than 5. Statistical significance was set at a 2-sided P value < 0.05. Analyses were performed using R version 3.6.3.

Further details about TriNetx, cohort definitions, and statistical analyses can be found in the supplement.

Results

513,284 patients with a confirmed diagnosis of COVID-19 were included in this study (54.8% females, mean [SD] age 46.6 [21.4]; Table 1 and Table S1 in the supplement). Of these, 20 were diagnosed with a CVT in the two weeks following their diagnosis (absolute risk: 39.0 per million people, 95% CI 25.2–60.2). The risk was significantly higher among patients with a history of cardiovascular diseases (Table 1), specifically cerebral/precerebral artery stenosis/occlusion, and intracranial hemorrhage. Among the 20 events, 6 were observed in patients under the age of 30, 4 between 30 and 39, 2 between 40 and 49, 2 between 50 and 59, 3 between 60 and 69, and 3 between 70 and 79. Three patients also had a CVT prior to their COVID-19 diagnosis, one between 4 and 8 weeks beforehand, and the other 2 more than 8 weeks prior.

The risk of being diagnosed with a CVT was significantly higher in the two weeks after COVID-19 compared to influenza (N=172,742; 0.0 per million people, 95% CI 0.0–22.2, adjusted RR=6.73, P=.003) or after receiving an mRNA vaccine (N=489,871; 4.1 per million people, 95% CI 1.1–14.9, adjusted RR=6.36, P<.001; Figure 1A). In the latter group, 2 cases were observed. 1 was a patient after the BNT162b2 vaccine (out of 331,503 people), and 1 was a patient where it was undetermined whether they had received BNT162b2 or mRNA-1273. The risk associated with COVID-19 was also higher than: (a) that currently reported by the EMA following vaccination with the ChAdOx1 nCoV-19 vaccine (currently 169 cases out of 34 million people, or 5.0 per million people, 95% CI 4.3–5.8); (b) the overall incidence observed in the TriNetX network (0.41 per million people over any 2-week period), or (c) the historical incidence of CVT in the USA (range: 13.9 to 20.2 per million per year, or 0.53 to 0.77 per million in any 2-week period; Otite et al, 2020).

We repeated the above analyses for PVT. The absolute incidence in the two weeks after COVID-19 diagnosis was 436.4 per million people (95% CI 382.9-497.4). This was significantly higher than after influenza (98.4 per million people, 95% CI 61.4–157.6, RR=4.43, 95% CI 2.71–7.26, P<.001) or after receiving an mRNA vaccine (44.9 per million people, 95% CI 29.7–68.0, RR=9.72, 95% CI 6.27–15.0, P<0.001; Figure 1B). In the latter group, 22 cases were observed. 11 occurred after BNT162b2 (out of 331,503 people), 2 following mRNA-1273 (out of 70,939); it is unknown which vaccine the other 9 had received. The incidence of PVT in COVID-19 was also higher than that reported for splanchnic thrombosis by the EMA following vaccination with ChAdOx1 nCoV-19 (53 cases out of 34 million people, or 1.6 per million people, 95% CI 1.2–2.0) or the overall incidence observed in our dataset (4.1 per million people over any 2-week period).

Laboratory data were available for a subset of the COVID-19 patients. Although the data do not cover most patients with a diagnosis of CVT, they suggest that patients with CVT after COVID-19 were significantly more likely to have elevated D-dimer level than patients with COVID-19 who did not have CVT, whereas patients with PVT after COVID-19 were significantly more likely to have low fibrinogen

level and thrombocytopenia (Table 2). The death rate among patients with CVT in the two weeks after COVID-19 was 20.0% (4 out of 20 patients, 95% CI 8.0–41.6%; Figure S1 in the appendix) and that among patients with PVT after COVID-19 was 18.3% (41 out of 224 patients, 95% CI 13.8–23.9%; Figure S1 in the appendix) and were significantly higher than among patients with COVID-19 who did not have those events (P<.001).

When the definition of CVT in terms of ICD-10 codes was broadened, the incidence of CVT in the two weeks after COVID-19 was 171.4 per million people (95% CI 139.2–211.2), which was significantly higher than after a diagnosis of influenza (52.1 per million people, 95% CI 27.4–99.0; RR=3.3, 95% CI 1.7–6.5; P<.001) or after receiving an mRNA vaccine (22.7 per million people, 95% CI 13.0–39.8; RR=7.5, 95% CI 4.1–13.8; P<.001); Figure S2 in the appendix). The majority of the extra cases came from the G08 diagnostic category.

Discussion

In a large electronic health records network, we report the absolute incidence of CVT in the 14 days after COVID-19 diagnosis and show that this is substantially greater than for the comparison groups. Although the magnitude of the risk cannot be quantified with confidence (see below), the risk after COVID-19 is approximately 8-10 times that reported for the vaccines, and about 100-fold increased compared to the population rate. The increased rate of CVT in COVID-19 is notable, being much more marked than the increased risks for other forms of stroke and cerebral haemorrhage (Taquet et al, 2021). The PVT data highlight that COVID-19 is associated with thrombotic events that are not limited to the cerebral vasculature.

All the relative risks should be interpreted with caution. First, the magnitude of the COVID-19 risk versus the population baseline, or versus influenza, is not based on cohorts which were matched for age or other demographic factors. For the same reason, we cannot conclude that the mRNA vaccines studied here are associated with an increased risk of CVT; far larger samples are needed to address this question.

Second, we have no information about diagnostic accuracy or completeness, though this is likely to be less of an issue for CVT or PVT compared to many diagnoses since radiological confirmation is typically needed. Third, the absence of key haematological laboratory data from many patients limits our ability to comment on whether the mechanism of CVT after COVID-19 is likely to be similar or different from that observed after ChAdOx1 nCoV-19, especially regarding anti-platelet factor 4 (PF4) antibodies (Greinacher et al 2021; Schultz et al 2021). Finally, we cannot directly compare the risks of CVT associated with ChAdOx1 nCoV-19 with any of the other vaccines, or with COVID-19, since we are using data collected by the EMA monitoring system, not from the electronic health records network. (No patients in the network had received ChAdOx1 nCoV-19, reflecting the fact that the network is almost entirely US-based).

In summary, COVID-19 is associated with a markedly increased incidence of CVT compared to the general population, patients with influenza, and people who have received BNT162b2 or mRNA-1273 vaccines. The risk with COVID-19 also appears greater than with ChAdOx1 nCoV-19, although as noted this conclusion is tentative. The rarity of CVT in all populations means that larger sample sizes are required to confirm the results, and complementary study designs are needed to aid interpretation. Nevertheless, the current data highlight the risk of serious thrombotic events in COVID-19, and can help contextualize and inform debate about the risk-benefit ratio for current COVID-19 vaccines.

Declaration of interests: SL is an employee of TriNetX.

Data sharing: The TriNetX system returned the results of the analyses as .csv files, which were downloaded and archived. Data presented will be freely accessible after peer review at: https://osf.io/h2mt7. In addition, TriNetX will grant access to researchers if they have a specific concern (through a third-party agreement option).

References

Dakay K, Cooper J, Bloomfield J, Overby P, Mayer SA, Nuoman R, Sahni R, Gulko E, Kaur G, Santarelli J, Gandhi CD, Al-Mufti F. Cerebral venous thrombosis in COVID-19 infection: A case series and review of the literature. *J Stroke Cerebrovasc Dis* 2021; 30: 105434.

Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. Thrombotic thrombocytopenia after ChAdOx1 nCov-19 vaccination. *N Engl J Med*, AOL 9 April 2021.

Handley JD, Emsley HCA. Validation of ICD-10 codes shows intracranial venous thrombosis incidence to be higher than previously reported. *Health Information Management J* 2020; 49 (1): 58-61.

Otite FO, Patel S, Sharma R, Khandwala P, Desai D, Latorre JG, Akano EO, Anikpezie N, Izzy S, Malik AM, Yavagai D, Khandewal P, Chaturvedi S. Trends in incidence and epidemiologic characteristics of cerebral venous thrombosis in the United States. *Neurology* 2020;95(16):e2200–13.

Schultz NH, Sorvoli IH, Michelsen AE, Munthe LA, Lund-Johansen F, Ahlen MT, Wiedmann M, Aarnoldt A-H, Skatter TH, Tjonnfjord GE, Holme PA. Thrombosis and thrombocytopenia after ChAdOx1 nCoV-19 vaccination. *N Engl J Med*, AOL 9 April 2021.

Silvis SM, Aguiar de Sousa D, Ferro JM, Coutinho JM. Cerebral venous thrombosis. *Nat Rev Neurol* 2017; 13: 555-565.

Taquet M, Geddes JR, Husain M, Luciano S, Harrison PJ. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *Lancet Psychiatry*, AOL 6 April 2021;Available from: http://dx.doi.org/10.1016/S2215-0366(21)00084-5

Table 1 – Baseline characteristics of the whole COVID-19 cohort and the groups who received a diagnosis of CVT or PVT in the two weeks after COVID-19 diagnosis. The Fisher exact P-value for CVT and PVT groups compared to the whole COVI-19 cohort is shown.

	All patients with COVID-19	Patients with COVID-19 and CVT		Patients with COVID-19 and PVT	
	n (%) | mean (SD)	n (%) | mean (SD)	P	n (%) | mean (SD)	P
Sample size, n	513284	20	-	224	-
Age, mean (SD), y	46.6 (21.4)	43.9 (20.6)	0.57	58.1 (14.6)	<.001
Sex					
Female	281209 (54.8)	12 (60.0)	0.82	94 (42.0)	<.001
Male	229706 (44.8)	8 (40.0)	0.82	130 (58.0)	<.001
Race					
White	312248 (60.8)	12 (60.0)	1	150 (67.0)	0.064
Black	92427 (18.0)	4 (20.0)	0.77	33 (14.7)	0.22
Asian	13816 (2.7)	2 (10.0)	0.1	3 (1.3)	0.3
Other	3224 (0.6)	0 (0.0)	1	1 (0.4)	1
Unknown	91569 (17.8)	2 (10.0)	0.56	37 (16.5)	0.66
Comorbidities at baseline					
Obesity	89436 (17.4)	4 (20.0)	0.77	60 (26.8)	<.001
Hypertension	149056 (29.0)	5 (25.0)	0.81	129 (57.6)	<.001
CKD	34598 (6.7)	2 (10.0)	0.39	54 (24.1)	<.001
Ischemic heart diseases	47603 (9.3)	5 (25.0)	0.033	47 (21.0)	<.001
Cardiac failure	28863 (5.6)	2 (10.0)	0.31	27 (12.1)	<.001
Arterial diseases	37153 (7.2)	4 (20.0)	0.052	50 (22.3)	<.001
Venous diseases	32030 (6.2)	3 (15.0)	0.13	156 (69.6)	<.001
(Pre-)Cerebral art. stenosis/occlusion	20264 (3.9)	4 (20.0)	0.0071	19 (8.5)	0.0017
Intracranial haemorrhage	3925 (0.8)	3 (15.0)	<.001	7 (3.1)	0.0018
Dementia	11728 (2.3)	0 (0.0)	1	3 (1.3)	0.5
Chronic lower resp. disease	87590 (17.1)	5 (25.0)	0.37	62 (27.7)	<.001
Connective tissue disorders	9291 (1.8)	0 (0.0)	1	12 (5.4)	<.001
Liver disease	31862 (6.2)	1 (5.0)	1	163 (72.8)	<.001
Diabetes mellitus	74689 (14.6)	5 (25.0)	0.2	80 (35.7)	<.001
Malignancy	39278 (7.7)	3 (15.0)	0.19	92 (41.1)	<.001
Past CVT	83 (0.02)	3 (15.0)	<.001	0 (0.0)	1
Past PVT	635 (0.1)	0 (0.0)	1	123 (54.9)	<.001

Table 2 – Laboratory characteristics of the patients in each group. P values are from Fisher's exact test, comparing the CVT and PVT groups to the whole COVID-19 cohort.

	All patients with COVID-19	Patients with COVID-19 and CVT		Patients with COVID-19 and PVT	
	n (%)	n (%)	P	n (%)	P
D-dimer > 5 mg/L n/n with measurement (%)	2035/67212 (3.0)	2/6 (33.3)	0.013	5/74 (6.8)	0.074
Fibrinogen < 200 mg/dL n/n with measurement (%)	1138/19414 (5.9)	1/6 (16.7)	0.3	23/51 (45.1)	<.001
Thrombocytopenia (ICD-10 codes D69.49, D69.59, D69.6)	9323 (1.8)	1 (5.0)	0.31	69 (30.8)	<.001
Death	16091 (3.1)	4 (20.0)	0.0031	41 (18.3)	<.001

Figure 1 – Incidence of CVT (A) and PVT (B) per million people in the two weeks after different health events. The numbers in parentheses on the right of each bar represent the 95% confidence intervals. Data for the ChAdOx1 nCoV-19 vaccine are presented for reference and inferred from the European Medicines Agency data (posted 7 April 2021).
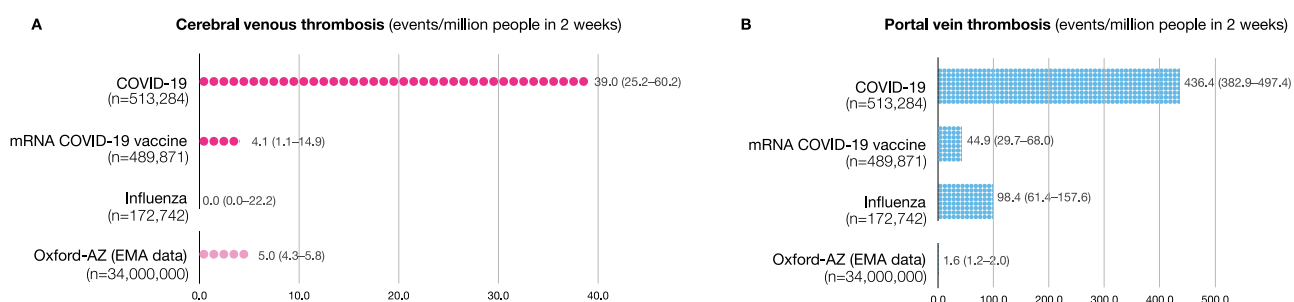
Appendix

Cerebral venous thrombosis: a retrospective cohort study of 513,284 confirmed COVID-19 cases and a comparison with 489,871 people receiving a COVID-19 mRNA vaccine

Maxime Taquet,[1,2] Masud Husain,[3,4] John R Geddes,[1,2] Sierra Luciano,[5] Paul J Harrison[1,2,*]

[1]Department of Psychiatry, University of Oxford

[2]Oxford Health NHS Foundation Trust, Oxford, UK

[3]Nuffield Department of Clinical Neurosciences, University of Oxford, UK

[4]Oxford University Hospitals NHS Foundation Trust, Oxford, UK

[5]TriNetX, Cambridge MA, USA

*Corresponding author: University Department of Psychiatry, Warneford Hospital, Oxford OX3 7JX, UK.

paul.harrison@psych.ox.ac.uk

Supplementary methods

TriNetX network

This section provides a version of our previous description of the network.[1]

Legal and ethical status

TriNetX's Analytics network is compliant with the Health Insurance Portability and Accountability Act (HIPAA), the US federal law which protects the privacy and security of healthcare data. TriNetX is certified to the ISO 27001:2013 standard and maintains an Information Security Management System (ISMS) to ensure the protection of the healthcare data it has access to and to meet the requirements of the HIPAA Security Rule. Any data displayed on the TriNetX Platform in aggregate form, or any patient level data provided in a data set generated by the TriNetX Platform, only contains de-identified data as per the de-identification standard defined in Section §164.514(a) of the HIPAA Privacy Rule. The process by which the data is de-identified is attested to through a formal determination by a qualified expert as defined in Section §164.514(b)(1) of the HIPAA Privacy Rule. This formal determination by a qualified expert, refreshed in December 2020, supersedes the need for TriNetX's previous waiver from the Western Institutional Review Board (IRB). The network contains data that are provided by participating Health Care Organizations (HCOs), each of which represents and warrants that it has all necessary rights, consents, approvals and authority to provide the data to TriNetX under a Business Associate Agreement (BAA), so long as their name remains anonymous as a data source and their data are utilized for research purposes. The data shared through the TriNetX Platform are attenuated to ensure that they do not include sufficient information to facilitate the determination of which HCO contributed which specific information about a patient.

Acquisition of data, quality control, and other procedures

The data are stored onboard a TriNetX appliance – a physical server residing at the institution's data centre or a virtual hosted appliance. The TriNetX platform is a fleet of these appliances connected into a federated network able to broadcast queries to each appliance. Results are subsequently collected and aggregated.

Once the data are sent to the network, they are mapped to a standard and controlled set of clinical terminologies and undergo a data quality assessment including 'data cleaning' that rejects records which do not meet the TriNetX quality standards. HIPAA compliance of the clinical patient data is achieved using de-identification. Different data modalities are available in the network. They include demographics (coded to HL7 version 3 administrative standards), diagnoses (represented by ICD-10-CM codes), procedures (coded in ICD-10-PCS or CPT), and measurements (coded to LOINC). While extensive information is provided about patients' diagnoses and procedures, other variables (such as socioeconomic and lifetime factors) are not comprehensively represented.

The data from a typical HCO generally go back around 7 years, with some going back 13 years. The data are continuously updated. HCOs update their data at various times, with most refreshing every 1, 2, or 4 weeks.

The data come primarily (>93%) from HCOs in the USA, with the remainder coming from India, Australia, Malaysia, Taiwan, Spain, UK, and Bulgaria. Only 1.8% of patients with COVID-19 are contributed from HCOs outside the USA. As noted above, to comply with legal frameworks and ethical guidelines guarding against data re-identification, the identity of participating HCOs and their individual contribution to each dataset are not disclosed to researchers.

Data quality assessment followed a standardised strategy wherein the data are reviewed for conformance (adherence to specified standards and formats), completeness (quantifying data presence or absence) and plausibility (believability of the data from a clinical perspective). There are pre-defined metrics for each of the above assessment categories. Results for these metrics are visualised and reviewed for each new site that joins the network as well as on an ongoing basis. Any identified issue is communicated to the data provider and resolved before continuing data collection.

The basic formatting of contributed data is also checked (e.g. to ensure that dates are properly represented). Records are checked against a list of required fields (e.g., patient identifier) and rejects those records for which the required information is missing. Referential integrity checking is done to ensure that data spanning multiple database tables can be successfully joined together. As the data are refreshed, changes in volume of data over time is monitored to

ensure data validity. At least one non-demographic fact for each patient is required for them to be counted in the dataset. Patient records with only demographics information are discarded.

The software also undergoes quality control. The engineers testing the software are independent from the engineers developing it. Each test code is checked by two independent testing engineers. Each piece of software is tested extensively against a range of synthetic data (i.e. generated for the purpose of testing) for which the expected output is established independently. If the software fails to return this output, then the software is deemed to have failed the test and is examined and modified accordingly. For statistical software (including that used for propensity score matching, for Kaplan-Meier analysis, etc), an additional quality control step is implemented. Two independent codes are written in two different programming languages (typically R and python) and the statistical results are compared. If discrepancies are identified, then the codes are deemed to have failed the test and are examined and modified accordingly. All the code is reviewed independently by another engineer.

The test strategy follows three levels of granularity:

1. Unit tests: These test specific blocks, or units, of code that perform specific actions (e.g. querying the database).
2. Integration tests: These ensure that different components are working together correctly.
3. End-to-end tests: These tests run the entire system and check the final output.

Some comments on advantages and disadvantages of EHR data

The advantage of EHR data, like those in TriNetX, over insurance claim data is that both insured and uninsured patients are included. An advantage of EHR data over survey data is that they represent the diagnostic rates in the population presenting to healthcare facilities. This provides an accurate account of the burden of specific diagnoses on healthcare systems. The downside of relying on diagnoses is that they obviously do not account for undiagnosed patients who might be suffering from the illness but did not seek medical attention (or in whom the diagnosis was missed). A general limitation of EHR data is that a patient may be seen in different HCOs for different parts of their care and if one HCO is not part of the federated network then part of their medical records may not be available.

Using a network of HCOs (rather than a single HCO) limits this possibility but does not fully remove it. Finally, historical data before the start of EHRs (or the addition of an HCO to the network) may be incomplete.

Cohorts definition and index events

The two control cohorts used consisted of patients who received an mRNA vaccine and patients with a diagnosis of influenza. Specifically, patients who received the vaccine were those who had any of the following procedure codes in their electronic health records:

- 91300: "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Coronavirus disease [COVID-19]) vaccine, mRNA-LNP, spike protein, preservative free, 30 mcg/0.3mL dosage, diluent reconstituted, for intramuscular use"

- 0001A: "Immunization administration by intramuscular injection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Coronavirus disease [COVID-19]) vaccine, mRNA-LNP, spike protein, preservative free, 30 mcg/0.3mL dosage, diluent reconstituted; first dose"

- 0002A: "Immunization administration by intramuscular injection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Coronavirus disease [COVID-19]) vaccine, mRNA-LNP, spike protein, preservative free, 30 mcg/0.3mL dosage, diluent reconstituted; second dose"

- 91301: "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Coronavirus disease [COVID-19]) vaccine, mRNA-LNP, spike protein, preservative free, 100 mcg/0.5mL dosage, for intramuscular use"

- 0011A: "Immunization administration by intramuscular injection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Coronavirus disease [COVID-19]) vaccine, mRNA-LNP, spike protein, preservative free, 100 mcg/0.5mL dosage; first dose"

- 0012A: "Immunization administration by intramuscular injection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Coronavirus disease [COVID-19]) vaccine, mRNA-LNP, spike protein, preservative free, 100 mcg/0.5mL

- 2468231: "SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein"

Patients with influenza were those who had any of the following diagnoses:

- J09: Influenza due to certain identified influenza viruses

- J10: Influenza due to other identified influenza virus

- J11: Influenza due to unidentified influenza virus.

Because some patients with the control index event might have had COVID-19 at a different point in time, we excluded from the control cohorts all those who had COVID-19 at any point in time. To avoid any contamination between cohorts, COVID-19 as an exclusion criterion was defined in the broader sense to be all patients with a confirmed diagnosis of COVID-19 (ICD-10 code U07.1) but also patients with an unconfirmed COVID-19 diagnosis (U07.2), a recorded positive PCR test for COVID-19, or any of the following recorded on or after January 20, 2020: Pneumonia due to SARS-associated coronavirus (J12.81), Other coronavirus as the cause of disease classified elsewhere (B97.29), or Coronavirus infection unspecified (B34.2). Inclusion of the latter three diagnostic codes captures patients who receive a COVID-19 diagnosis in the early stage of the pandemic when the ICD code for COVID-19 (U07) was not yet defined. Specifically, the following codes were excluded from the control cohorts if they occurred on or after January 20, 2020:

- U07.1: COVID-19, virus identified

- U07.2: COVID-19, virus not identified

- J12.81: Pneumonia due to SARS-associated coronavirus

- B97.29: Other coronavirus as the cause of disease classified elsewhere

- B34.2: Coronavirus infection, unspecified

- Positive SARS-CoV-2 RNA in Respiratory specimen

- Positive SARS-CoV-2 RNA in Unspecified specimen

- Positive SARS-CoV-2 N gene in Respiratory specimen

- Positive SARS-CoV-2 N gene in Unspecified specimen

- Positive SARS-CoV-2 RdRp gene in Respiratory specimen

- Positive SARS-CoV-2 E gene in Respiratory specimen

- Positive SARS-CoV-2 E gene in Unspecified specimen

- Positive SARS-CoV-2 RNA panel in Respiratory specimen

- Positive SARS-CoV-2 RNA panel in Unspecified specimen

- Positive SARS-CoV-2 RNA in Nasopharynx

- Positive SARS coronavirus 2 and related RNA

- Positive SARS-related coronavirus RNA in Respiratory specimen

- Positive SARS coronavirus 2 ORF1ab in Respiratory specimen

Baseline characteristics code

When reporting baseline characteristics, the following ICD-10 codes are used:

- Obesity: E66

- Hypertension: I10-I16

- Chronic kidney disease: N18

- Ischemic heart disease: I20-I25

- Heart failure: I50

- Disease of the arteries, arterioles, or capillaries: I70-I79

- Disease of (non-cerebral) veins: I80-I87

- Cerebral/Pre-cerebral artery stenosis/occlusion: I63 (cerebral infarction), I65 (Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction), I66 (Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction)

- Intracranial hemorrhage: I60 (Nontraumatic subarachnoid hemorrhage), I61 (Nontraumatic intracerebral hemorrhage), I62 (Other and unspecified nontraumatic intracranial hemorrhage)

- Dementia: F01 (Vascular dementia), F02 (Dementia in other diseases classified elsewhere), F03 (Unspecified dementia), G30 (Alzheimer's disease), G31.0 (Frontotemporal dementia), and G31.83 (Dementia with Lewy bodies)

- Chronic lower respiratory diseases: J40-J47

- Connective tissue disorders: M30-M36

- Liver diseases: K70-K77

- Diabetes mellitus: E08-E13

- Malignancy: C00-C14 (Malignant neoplasms of lip, oral cavity and pharynx), C15-C26 (Malignant neoplasms of digestive organs), C30-C39 (Malignant neoplasms of respiratory and intrathoracic organs),

C40-C41 (Malignant neoplasms of bone and articular cartilage), C43-C44 (Melanoma and other malignant neoplasms of skin), C45-C49 (Malignant neoplasms of mesothelial and soft tissue), C50 (Malignant neoplasms of breast), C51-C58 (Malignant neoplasms of female genital organs), C60-C63 (Malignant neoplasms of male genital organs), C64-C68 (Malignant neoplasms of urinary tract), C69-C72 (Malignant neoplasms of eye, brain and other parts of central nervous system), C73-C75 (Malignant neoplasms of thyroid and other endocrine glands), C76-C80 (Malignant neoplasms of ill-defined, other secondary and unspecified sites), C7A (Malignant neuroendocrine tumors), C7B (Secondary neuroendocrine tumors), C81-C96 (Malignant neoplasms of lymphoid, hematopoietic and related tissue)

Supplementary figures

Cerebral venous thrombosis

Portal vein thrombosis
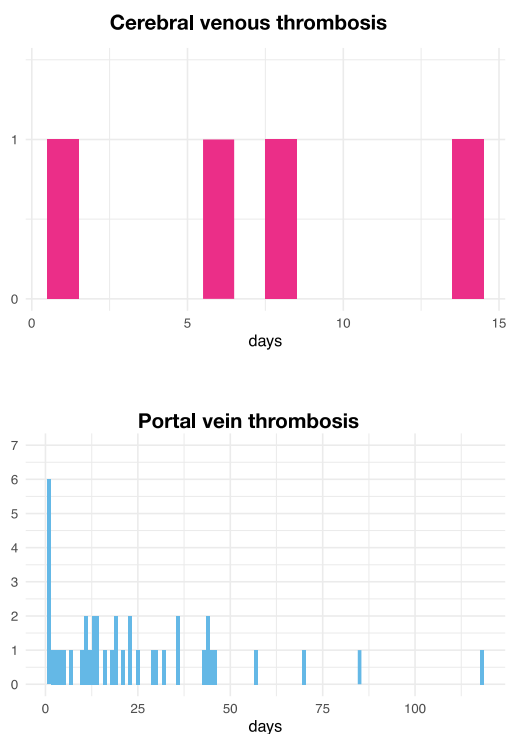
Fig. S1 – Distribution of the day of recorded death relative to the index event for patients who died after having had a CVT post COVID-19 (top) or a PVT post COVID-19 (bottom).

CVT, extended definition (events/million people in 2 weeks)
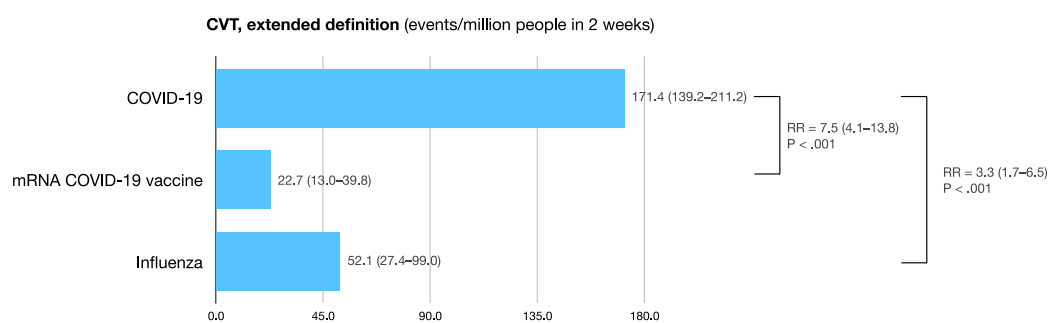
Fig. S2 - Incidence of CVT (defined using an extended list of diagnostic codes) per million people in the two weeks after different health events. The numbers in parentheses on the right of each bar represent the 95% confidence intervals. The relative risks and corresponding P-values from Fisher's exact test are presented on the right.

Supplementary table

Table S1 – Demographics of the patients who were diagnosed with COVID-19, who received an mRNA vaccine, and who were diagnosed with influenza.

	COVID-19	**mRNA vaccine**	**Influenza**
Age, mean (SD), y	46.6 (21.4)	62.2 (17.8)	26.4 (22.7)
Sex, % Female	54.8%	59.1%	54.1%

References

1. Taquet M, Geddes JR, Husain M, Luciano S, Harrison PJ. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. Lancet Psychiatry [Internet] 2021;Available from: http://dx.doi.org/10.1016/S2215-0366(21)00084-5