7-9-2023

Diabetes status and other factors as correlates of risk for thrombotic and thromboembolic events during SARS-CoV-2 infection: A nationwide retrospective case-control study usingCerner Real-World Data™

Erin M. Tallon
Children's Mercy Kansas City

Mary Pat Gallagher

Vincent S. Staggs
Children's Mercy Hospital

Diana Ferro
Children's Mercy Kansas City

Deepa Badrinath Murthy

See next page for additional authors

Recommended Citation

Tallon EM, Gallagher MP, Staggs VS, et al. Diabetes status and other factors as correlates of risk for thrombotic and thromboembolic events during SARS-CoV-2 infection: A nationwide retrospective case-control study using Cerner Real-World Data™. BMJ Open. 2023;13(7):e071475. Published 2023 Jul 9. doi:10.1136/bmjopen-2022-071475

BMJ Open

Diabetes status and other factors as correlates of risk for thrombotic and thromboembolic events during SARS-CoV-2 infection: A nationwide retrospective case–control study using *Cerner Real-World Data™*

Erin M Tallon ,[1,2] Mary Pat Gallagher,[3] Vincent S Staggs,[4,5] Diana Ferro,[2,6] Deepa Badrinath Murthy,[3] Osagie Ebekozien,[7,8] Mikhail N Kosiborod,[9,10] Marcus Lind,[11,12,13] Camila Manrique-Acevedo,[14] Chi-Ren Shyu,[1,15] Mark A Clements[2,5]

To cite: Tallon EM, Gallagher MP, Staggs VS, *et al*. Diabetes status and other factors as correlates of risk for thrombotic and thromboembolic events during SARS-CoV-2 infection: A nationwide retrospective case–control study using *Cerner Real-World Data™*. *BMJ Open* 2023;**13**:e071475. doi:10.1136/bmjopen-2022-071475

► 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-2022-071475).

EMT and MPG are joint first authors.

Received 29 December 2022
Accepted 19 June 2023

For numbered affiliations see end of article.

Correspondence to
Dr Erin M Tallon;
etallon@cmh.edu

ABSTRACT

Objectives We sought to examine in individuals with SARS-CoV-2 infection whether risk for thrombotic and thromboembolic events (TTE) is modified by presence of a diabetes diagnosis. Furthermore, we analysed whether differential risk for TTEs exists in type 1 diabetes mellitus (T1DM) versus type 2 diabetes mellitus (T2DM).

Design Retrospective case–control study.

Setting The December 2020 version of the *Cerner Real-World Data* COVID-19 database is a deidentified, nationwide database containing electronic medical record (EMR) data from 87 US-based health systems.

Participants We analysed EMR data for 322 482 patients >17 years old with suspected or confirmed SARS-CoV-2 infection who received care between December 2019 and mid-September 2020. Of these, 2750 had T1DM; 57 811 had T2DM; and 261 921 did not have diabetes.

Outcome TTE, defined as presence of a diagnosis code for myocardial infarction, thrombotic stroke, pulmonary embolism, deep vein thrombosis or other TTE.

Results Odds of TTE were substantially higher in patients with T1DM (adjusted OR (AOR) 2.23 (1.93–2.59)) and T2DM (AOR 1.52 (1.46–1.58)) versus no diabetes. Among patients with diabetes, odds of TTE were lower in T2DM versus T1DM (AOR 0.84 (0.72–0.98)).

Conclusions Risk of TTE during COVID-19 illness is substantially higher in patients with diabetes. Further, risk for TTEs is higher in those with T1DM versus T2DM. Confirmation of increased diabetes-associated clotting risk in future studies may warrant incorporation of diabetes status into SARS-CoV-2 infection treatment algorithms.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Strengths of this study include its large, nationwide sample and the rigorous processes used to identify individuals with diabetes versus those without diabetes.
⇒ While the use of diagnosis codes to identify individuals who experienced thrombotic and thromboembolic events (TTE) is likely to have captured the vast majority of TTE cases, this case identification method may have misclassified or missed some TTE cases.
⇒ Data pertaining to medication use were not included because the validity of medication data in electronic medical records varies substantially across medication classes.
⇒ Additional studies using pharmacy claims data are needed to evaluate the impact of patients' medication exposures on TTE outcomes in individuals with SARS-CoV-2 infection.

INTRODUCTION

As of December 2022, over 653 million people have been diagnosed with COVID-19 worldwide, with more than 6 million deaths reported.[1] Higher rates of COVID-19-associated mortality are reported in adults with type 1 (T1DM) and type 2 diabetes mellitus (T2DM).[2 3] The risk of COVID-19-associated mortality in patients with diabetes increases with age, with rare exceptions reported in young adults with T2DM.[3] Impaired immunity, a proinflammatory state, and increased prevalence of cardiovascular disease are postulated as reasons for increased mortality risk in patients with diabetes.[4] Endothelial dysfunction and increased risk for thrombotic and thromboembolic events (TTE) are associated with both diabetes and COVID-19 and are reported in patients with COVID-19 independent of diabetes diagnosis[5–7] as well as in adults with T1DM or T2DM independent of COVID-19 illness.[8] This hypercoagulable

state can be triggered either by infection generally or by SARS-CoV-2 infection directly.

To date, little is known about whether individuals with acute COVID-19 illness with pre-existing T1DM or T2DM have higher risk for TTE than those without diabetes. Higher risk for TTE could partially explain differences in outcomes, warranting additional study into targeted interventions or therapeutics for patients with COVID-19 and diabetes.

SUBJECTS, MATERIALS AND METHODS

Study design and data source

We conducted a case–control study using Oracle Cerner's nationwide deidentified COVID-19 database, curated from the *Cerner Real-World Data* (CRWD) clinical data warehouse.[9] We stratified patients by diabetes status (T1DM, T2DM or no diabetes) to summarise demographic and clinical characteristics of patients with and without diabetes. We then examined these characteristics as correlates of TTE in patients with suspected or confirmed SARS-CoV-2 infection. We also examined the impact of diabetes type and other characteristics on the likelihood of TTE in the subset of patients with diabetes.

The December 2020 version (2020Q3 CRWD COVID database release) of Oracle Cerner's COVID-19 database contains longitudinal electronic medical record (EMR) data for 490 373 patients seen or admitted at 87 US-based health systems between 1 December 2019 and mid-September 2020. (Ambiguity in the latter date is due to date shifting in the deidentified database.) The database includes patients from CRWD with a healthcare encounter associated with a positive result for a COVID-related laboratory test or a diagnosis code associated with suspected or confirmed SARS-CoV-2 infection.[10 11]

Oracle Cerner curated its COVID-19 database to capture as many SARS-CoV-2 infection cases as possible.[10 11] The diagnosis and laboratory test codes that Oracle Cerner designated as 'qualifying' codes for the purposes of inclusion in the data set were intentionally kept broad (online supplemental tables 1 and 2). This allowed for full capture despite the lack of COVID-19-specific codes at the onset of the pandemic. Oracle Cerner flagged a subset of these codes as more apt to indicate definitive SARS-CoV-2 infection (online supplemental tables 1 and 2).[10] We considered individuals assigned one of the codes in this subset at the time of the index qualifying encounter (IQE) that qualified them for inclusion in the database to be 'confirmed' SARS-CoV-2 infection cases. All other individuals were considered 'suspected' infection cases. In this study, we therefore analysed two separate cohorts: (1) all patients with either suspected or confirmed SARS-CoV-2 infection, and (2) only patients with confirmed SARS-CoV-2 infection. Following a primary analysis with the first cohort, we conducted a sensitivity analysis with the second cohort. The database also contains patients' historical data dating back to 1 January 2015.

Diabetes status definitions

Using diagnosis, medication and laboratory data specified in the validated SUrveillance, PREvention, and ManagEment of Diabetes Mellitus (SUPREME-DM) algorithm, we identified a well-defined subset of sample patients with diabetes of any type.[12] We used an online conversion tool (*ICD10Data.com*) to map the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10) codes in Oracle Cerner's database to ICD-9 codes in the SUPREME-DM algorithm (online supplemental table 3).

Patients from the SUPREME-DM cohort who met criteria from a validated algorithm for identifying T1DM cases (viz the Klompas Optimized Algorithm) were identified as having T1DM.[12] Individuals meeting criteria for the US Food and Drug Administration Sentinel Initiative Workgroup's T2DM Algorithm (Option 1) were identified as having T2DM.[12] Patients who did not have diabetes were those with no diabetes-related ICD codes or medications,[12] no HbA1c result ≥6.5% (≥48 mmol/mol), no fasting plasma glucose ≥126 mg/dL and no random plasma glucose ≥200 mg/dL (online supplemental figure 1).

Outcomes

The primary outcome was TTE (of any type), documented using ICD-9 and ICD-10 codes, in association with the IQE. In Oracle Cerner's COVID-19 database, every diagnosis code assigned to a patient is linked to the encounter at which the condition was recorded. A patient was defined as a TTE case if a diagnosis code for one or more of the following conditions was assigned at the patient's IQE: (a) myocardial infarction (MI), (b) thrombotic stroke (TS), (c) pulmonary embolism (PE), (d) deep vein thrombosis (DVT), (e) superficial or unspecified TTE, (f) arterial TTE or (g) other TTE (online supplemental table 4). We also examined factors associated with each of four subtypes of TTE (MI, TS, PE and DVT) in the overall sample, as well as factors associated with TTE (of any type) in the subset of patients with diabetes.

Patient characteristics

For each patient, we extracted diabetes status, age, sex, race and ethnicity, body mass index (BMI), encounter type, length of stay (inpatient encounters only), HbA1c, history of TTE and health system census division (table 1).

Only two out of 53 332 patients <18 years old experienced a TTE; therefore, patients <18 years old were excluded. Sex was categorised as male or female. History of TTE was identified using ICD-9 and ICD-10 codes for MI, TS, PE, DVT and other TTEs documented prior to the IQE service date (online supplemental appendix 1). Patients with no TTE documented prior to the IQE were presumed to have no history of TTE. Census division was classified according to US Census Bureau designations.

BMI was recorded, or calculated from height and weight as needed, using data collected at the IQE if available, or at the closest visit within 12 months of the IQE. HbA1c was defined as the HbA1c measurement

Table 1 Patient and health system characteristics for 322 482 individuals with suspected or confirmed COVID-19

Characteristic	Overall	No diabetes	Type 1 diabetes	Type 2 diabetes
Total, n	322 482	261 921	2750	57 811
Age, years (median (IQR))*	48 (32, 64)	42 (29, 59)	41 (29, 57)	65 (55, 75)
Age categories, years, n (%)				
18–29	67 808 (21.0)	66 358 (25.3)	726 (26.4)	724 (1.3)
30–39	55 634 (17.3)	52 711 (20.1)	587 (21.3)	2336 (4.0)
40–49	47 348 (14.7)	41 072 (15.7)	450 (16.4)	5826 (10.1)
50–59	49 510 (15.4)	37 624 (14.4)	395 (14.4)	11 491 (19.9)
60–69	42 333 (13.1)	27 443 (10.5)	306 (11.1)	14 584 (25.2)
70–79	32 797 (10.2)	19 038 (7.3)	187 (6.8)	13 572 (23.5)
80–89	26 823 (8.3)	17 515 (6.7)	97 (3.5)	9211 (15.9)
≥90	229 (0.1)	160 (0.1)	2 (0.1)	67 (0.1)
Sex, n (%)				
Male	144 264 (44.9)	114 330 (43.8)	1392 (50.7)	28 542 (49.5)
Female	177 351 (55.0)	146 935 (56.1)	1351 (49.1)	29 065 (50.3)
Unknown	867 (0.3)	656 (0.3)	7 (0.3)	204 (0.4)
Race/ethnicity, n (%)				
Non-Hispanic White	126 200 (39.1)	100 301 (38.3)	1404 (51.1)	24 495 (42.4)
Hispanic	112 729 (35.0)	94 295 (36.0)	665 (24.2)	17 769 (30.7)
Non-Hispanic Black	45 479 (14.1)	35 328 (13.5)	439 (16.0)	9712 (16.8)
AAPI | AIAN | NHO	17 799 (5.5)	14 121 (5.4)	112 (4.1)	3566 (6.2)
Unknown ethnicity, White	5403 (1.7)	4678 (1.8)	50 (1.8)	675 (1.2)
Non-Hispanic, Unknown race	3997 (1.2)	3139 (1.2)	44 (1.6)	814 (1.4)
Unknown ethnicity, Other race	2038 (0.6)	1767 (0.7)	13 (0.5)	258 (0.4)
Unknown ethnicity, Black	1309 (0.4)	1112 (0.4)	8 (0.3)	189 (0.3)
Unknown ethnicity and unknown race	7528 (2.3)	7180 (2.7)	15 (0.5)	333 (0.6)
BMI, kg/m^2 (median (IQR))	28.5 (24.6, 33.7)	28.1 (24.2, 32.9)	25.7 (22.1, 30.5)	30.9 (26.5, 36.6)
BMI categories, kg/m^2, n (%)				
<18.5	6238 (1.9)	5339 (2.0)	173 (6.3)	726 (1.3)
18.5–24.9	68 954 (21.4)	59 038 (22.5)	1019 (37.1)	8897 (15.4)
25.0–29.9	84 662 (26.3)	68 550 (26.2)	737 (26.8)	15 375 (26.6)
30.0–34.9	58 506 (18.1)	44 614 (17.0)	383 (13.9)	13 509 (23.4)
35.0–39.9	30 164 (9.4)	21 674 (8.3)	193 (7.0)	8297 (14.4)
≥40.0	26 377 (8.2)	17 323 (6.6)	140 (5.1)	8914 (15.4)
Unknown	47 581 (14.8)	45 383 (17.3)	105 (3.8)	2093 (3.6)
Encounter type, n (%)				
Emergency	155 954 (48.4)	140 613 (53.7)	638 (23.2)	14 703 (25.4)
Inpatient (includes admission for observation)	128 079 (39.7)	84 814 (32.4)	2053 (74.7)	41 212 (71.3)
Urgent care encounter	38 449 (11.9)	36 494 (13.9)	59 (2.1)	1896 (3.3)
Length of stay (inpatient encounters), days (median (IQR))	3.3 (1.9, 6.3)	2.9 (1.7, 5.3)	3.9 (2.0, 7.2)	4.6 (2.3, 8.7)
In-hospital mortality, n (%)†	4314 (1.3)	1607 (0.6)	69 (2.5)	2638 (4.6)

Continued

Table 1 Continued

Characteristic	Overall	No diabetes	Type 1 diabetes	Type 2 diabetes
HbA1c, % (median (IQR))	6.2 (5.5, 7.8)	5.4 (5.2, 5.7)	9.3 (7.6, 11.5)	7.2 (6.3, 8.8)
HbA1c, mmol/mol (median (IQR))	44 (37, 62)	36 (33, 39)	78 (60, 102)	55 (45, 73)
History of TTE, n (%)	39 162 (12.1)	18 477 (7.1)	669 (24.3)	20 016 (34.6)
TTE (any type) at the COVID encounter, n (%)	14 720 (4.6)	8144 (3.1)	225 (8.2)	6351 (11.0)
Myocardial infarction at the COVID encounter, n (%)	5520 (1.7)	2650 (1.0)	103 (3.7)	2767 (4.8)
Thrombotic stroke at the COVID encounter, n (%)	4388 (1.4)	2373 (0.9)	55 (2.0)	1960 (3.4)
Pulmonary embolism at the COVID encounter, n (%)	2855 (0.9)	2029 (0.8)	35 (1.3)	791 (1.4)
Deep vein thrombosis at the COVID encounter, n (%)	2376 (0.7)	1378 (0.5)	37 (1.3)	961 (1.7)
Superficial/unspecified TTE at the COVID encounter, n (%)	461 (0.1)	275 (0.1)	13 (0.5)	173 (0.3)
Arterial TTE at the COVID encounter, n (%)	191 (0.1)	109 (0.0)	4 (0.1)	78 (0.1)
Other TTE at the COVID encounter, n (%)	905 (0.3)	506 (0.2)	13 (0.5)	386 (0.7)
Census division, n (%)				
East North Central	15 622 (4.8)	12 003 (4.6)	216 (7.9)	3403 (5.9)
East South Central	3034 (0.9)	2591 (1.0)	27 (1.0)	416 (0.7)
Middle Atlantic	59 295 (18.4)	48 042 (18.3)	483 (17.6)	10 770 (18.6)
Mountain	54 588 (16.9)	43 407 (16.6)	647 (23.5)	10 534 (18.2)
New England	7497 (2.3)	6109 (2.3)	82 (3.0)	1306 (2.3)
Pacific	39 976 (12.4)	31 382 (12.0)	347 (12.6)	8247 (14.3)
South Atlantic	103 048 (32.0)	87 815 (33.5)	536 (19.5)	14 697 (25.4)
West North Central	13 497 (4.2)	10 917 (4.2)	164 (6.0)	2416 (4.2)
West South Central	25 925 (8.0)	19 655 (7.5)	248 (9.0)	6022 (10.4)

*Individuals <18 years old were excluded from analysis.
†In-hospital mortality was defined as a discharge disposition of 'Expired' at the COVID-related encounter.
AAPI | AIAN | NHO, Asian American/Pacific Islander, American Indian/Alaska Native, non-Hispanic Other; BMI, body mass index; HbA1c, haemoglobin A1c; TTE, thrombotic or thromboembolic event.

documented closest to and within 12 months of the IQE. The 12-month timeframe for BMI and HbA1c improved data capture and reduced the need for imputation of missing values. Codes for extracting BMI and HbA1c data are found in online supplemental table 5. Outlier detection and removal are described in online supplemental appendix 2. For each patient, we also extracted historical weight, BMI and HbA1c values (ie, values documented closest, and at least 1 year prior, to the IQE).

Statistical analysis

Oracle Cerner's data science ecosystem, called *HealtheDataLab*, is hosted by Amazon Web Services and powered by Apache Spark V.2.4.4 (Apache Software Foundation, Wilmington, Delaware). We used Python V.3.7.6 (Python Software Foundation) and R V.4.0.2 (R Foundation for Statistical Computing, Vienna, Austria) for data extraction and analysis.

We used the *mice* package in R to create 15 complete data sets using multivariate imputation by chained equations under a missing at random assumption.[13] Variables in the imputation model included diabetes status, age, health system ID, encounter type, presence of each type of TTE (ie, MI, TS, PE, DVT, superficial or unspecified TTE, arterial TTE and other TTE), history of TTE, sex, race, ethnicity, height, historical weight, current weight, historical BMI, current BMI, historical HbA1c and current HbA1c.

Historical values (defined as values documented >1 year prior to the IQE) for weight, BMI and HbA1c

were included in the model to inform the imputation of missing current values (defined as values documented within 1 year of the IQE). When historical values were missing but current values were present, historical values were set equal to the current values. When historical BMI or current BMI were missing, they were calculated as derived values after imputing missing height and corresponding weight values.

There were no missing data for COVID-19 status, diabetes status, age, encounter type, health system ID or census division. Frequency of missing data for other covariates is found in online supplemental table 6. Because most patients (~86%) without diabetes did not have an HbA1c result, we did not impute HbA1c for those patients nor include HbA1c in statistical models that included patients without diabetes. All other missing values in the data set were imputed.

We used the *GLMMadaptive* package in R to fit for each imputed data set a logistic mixed model to examine diabetes status, age, race and ethnicity, sex, current BMI and history of TTE as correlates of TTE at the IQE in patients with suspected or confirmed COVID-19. The model included a random health system intercept to adjust for clustering of patients within systems. GLMMadaptive uses adaptive Gauss-Hermite quadrature for model fitting.[14] We then conducted a sensitivity analysis by fitting a similar model to data from the subset of patients with confirmed COVID-19. In a further analysis of the full sample of patients with suspected or confirmed COVID-19, we added interaction terms for (a) diabetes type X age, (b) diabetes type X BMI and (c) diabetes type X history of TTE.

We fitted similar models to examine correlates of MI, TS, PE and DVT (ie, modelled as four separate outcomes) at the IQE in patients with suspected or confirmed COVID-19. Each model was limited to patients who experienced (1) the specific outcome being modelled or (2) no TTE of any kind.

Lastly, we fitted a logistic mixed model to examine correlates of TTE in patients with diabetes. This model included HbA1c and the explanatory variables common to the models described above.

Age was analysed as a categorical variable in increments generally spanning 10 years (eg, ages 18–29, 30–39, 40–49, …, 80+). BMI (in kg/m^2) was grouped into six categories: <18.5, 18.5–24.9, 25.0–29.9, 30.0–34.9, 35.0–39.9 and ≥40.0. Race and ethnicity were categorised as (1) non-Hispanic White (NHW); (2) non-Hispanic Black (NHB); (3) Hispanic; and (4) Asian American/Pacific Islander, American Indian/Alaska Native or non-Hispanic Other (AAPI | AIAN | NHO). HbA1c was analysed as a continuous variable, as a percentage.

We used the *parameters* package in R to pool regression analyses according to Rubin's rules.[15] To quantify hypothesis testing error and control for false discoveries resulting from multiple hypothesis testing, we used the *q-value* package in R to calculate a q value (the Bayesian analogue of a p value[16]) for each test. We used a q value threshold of 0.05 to identify findings unlikely to have occurred by chance under the hypothesised null model.

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

RESULTS

Characteristics of patients with suspected or confirmed COVID-19

Of the 322 482 patients with suspected or confirmed COVID-19 infection, 261 921 (81.2%) did not have diabetes; 2750 (0.9%) had T1DM; and 57 811 (17.9%) had T2DM. Median age was 48 years (IQR 32–64), median BMI was 28.5 kg/m^2 (IQR 24.6–33.7) and 44.9% were male. Approximately 12% of individuals (n=39 162) had experienced a previous TTE. Median HbA1c for individuals with T1DM was 9.3% (IQR 7.6%–11.5%) (78 mmol/mol (IQR 60–102 mmol/mol)); for those with T2DM, median HbA1c was 7.2% (IQR 6.3%–8.8%) (55 mmol/mol (IQR 45–73 mmol/mol)).

Baseline probability of experiencing a TTE at the IQE was 4.6%. Baseline probability of experiencing MI at the IQE was 1.8%; for stroke, 1.4%; for PE, 0.9%; and for DVT, 0.8%. Table 1 contains demographic and clinical characteristics for these individuals, stratified by diabetes type. Baseline characteristics stratified by diabetes status and by outpatient versus inpatient status are found in online supplemental table 7. Inpatient mortality at encounters associated with each type of TTE is stratified by diabetes status in online supplemental table 8.

Characteristics of patients with confirmed COVID-19

Of 116 370 patients with confirmed COVID-19, 93 098 (80.0%) did not have diabetes; 802 (0.7%) had T1DM; and 22 470 (19.3%) had T2DM. Median age was 47 years (IQR 32–62), median BMI was 29.4 kg/m^2 (IQR 25.4–34.5) and 46.5% were male. Approximately 9.5% of individuals (n=11 094) had experienced a previous TTE. Median HbA1c for individuals with T1DM was 9.5% (IQR 7.7%–12.0%) (80 mmol/mol (IQR 61–108 mmol/mol)); for those with T2DM, median HbA1c was 7.5% (IQR 6.5%–9.3%) (58 mmol/mol (IQR 48–78 mmol/mol)). Baseline probability of experiencing a TTE at the IQE was 3.3%.

Baseline probability of experiencing MI at the IQE was 1.3%; for TS, 0.8%; for PE, 0.8%; and for DVT, 0.6%. Online supplemental table 9 contains demographic and clinical characteristics for these patients, stratified by diabetes type. Inpatient mortality at encounters associated with each type of TTE in individuals with confirmed COVID-19 is stratified by diabetes status in online supplemental table 10.

Table 2 Adjusted ORs: Correlates of thrombotic and thromboembolic events of any type

Characteristic	Suspected or confirmed COVID-19		Confirmed COVID-19	
	Adjusted OR (95% CI)*	Q value	Adjusted OR (95% CI)*	Q value
Diabetes status				
No diabetes	1 [Reference)		1 [Reference)	
Type 1 diabetes	2.23 (1.93 to 2.59)	<0.001	3.32 (2.53 to 4.36)	<0.001
Type 2 diabetes	1.52 (1.46 to 1.58)	<0.001	1.96 (1.82 to 2.12)	<0.001
Age (years)				
18–29	1 (Reference)		1 (Reference)	
30–39	2.02 (1.78 to 2.30)	<0.001	1.67 (1.32 to 2.11)	<0.001
40–49	3.85 (3.42 to 4.34)	<0.001	2.91 (2.35 to 3.59)	<0.001
50–59	6.06 (5.42 to 6.78)	<0.001	4.07 (3.33 to 4.98)	<0.001
60–69	9.00 (8.05 to 10.06)	<0.001	5.93 (4.85 to 7.24)	<0.001
70–79	11.22 (10.03 to 12.55)	<0.001	8.41 (6.87 to 10.29)	<0.001
80+	13.40 (11.97 to 14.99)	<0.001	10.59 (8.64 to 12.99)	<0.001
Sex				
Female	1 (Reference)		1 (Reference)	
Male	1.32 (1.28 to 1.37)	<0.001	1.47 (1.37 to 1.57)	<0.001
Race and ethnicity				
NHW	1 (Reference)		1 (Reference)	
NHB	1.05 (0.99 to 1.11)	0.015	1.17 (1.06 to 1.30)	<0.001
Hispanic	0.78 (0.74 to 0.82)	<0.001	0.87 (0.79 to 0.95)	0.001
AAPI | AIAN | NHO	0.98 (0.91 to 1.05)	0.082	1.24 (1.09 to 1.42)	<0.001
BMI (kg/m^2)				
18.5–24.9	1 (Reference)		1 (Reference)	
<18.5	1.16 (1.04 to 1.28)	0.001	1.10 (0.87 to 1.38)	0.061
25.0–29.9	0.94 (0.90 to 0.99)	0.002	0.87 (0.79 to 0.95)	0.001
30.0–34.9	0.95 (0.90 to 1.00)	0.012	0.88 (0.79 to 0.97)	0.002
35.0–39.9	0.94 (0.87 to 1.00)	0.009	0.90 (0.80 to 1.02)	0.018
≥40.0	0.99 (0.93 to 1.07)	0.112	1.04 (0.91 to 1.19)	0.072
History of TTE				
Not present	1 (Reference)		1 (Reference)	
Present	2.69 (2.58 to 2.79)	<0.001	2.81 (2.61 to 3.03)	<0.001

*Unadjusted ORs are found in online supplemental table 11.
AAPI | AIAN | NHO, Asian American/Pacific Islander, American Indian/Alaska Native, non-Hispanic Other; BMI, body mass index; NHB, non-Hispanic Black; NHW, non-Hispanic White; TTE, thrombotic or thromboembolic event.

Risk of TTE of any type in patients with suspected or confirmed COVID-19

Results from the fully adjusted model in the overall study cohort showed that the odds of TTE were substantially higher for both T1DM (vs no diabetes; adjusted OR (AOR) 2.23; 95% CI 1.93 to 2.59; Q<0.001) and T2DM (vs no diabetes; AOR 1.52; 95% CI 1.46 to 1.58; Q<0.001) (table 2).

We observed increased odds of TTE in males (vs females; AOR 1.32; 95% CI 1.28 to 1.37; Q<0.001) and in patients with a history of TTE (AOR 2.69; 95% CI 2.58 to 2.79; Q<0.001). Meaningful associations were identified between TTE and older age (table 2). Risk of TTE was lower for Hispanic individuals (vs NHW; AOR 0.78; 95% CI 0.74 to 0.82; Q<0.001). There was some tendency for patients with decreased BMI (<18.5 kg/m^2) to have slightly higher odds of TTE (vs BMI 18.5–24.9 kg/m^2; AOR 1.16; 95% CI 1.04 to 1.28; Q=0.001) and for patients in higher BMI categories to have slightly lower odds of TTE (table 2).

The sensitivity analysis carried out by fitting the fully adjusted model to data from the cohort of patients with confirmed COVID-19 showed that, relative to patients who were NHW, individuals who were NHB (AOR 1.17; 95% CI 1.06 to 1.30; Q<0.001) or AAPI | AIAN | NHO (AOR 1.24; 95% CI 1.09 to 1.42;

Table 3 Adjusted ORs: Interaction effects for diabetes status by age group, body mass index category and history of thrombosis

Effect of diabetes by age group*							
	Ages 18–29	Ages 30–39	Ages 40–49	Ages 50–59	Ages 60–69	Ages 70–79	Age 80+
No diabetes diagnosis	1 (No diabetes, main effect AOR)	1.93	3.41	5.34	8.80	11.67	14.76
T1DM	3.69 (T1DM, main effect AOR)	7.76	16.34	10.29	18.28	23.96	19.91
T2DM	3.20 (T2DM, main effect AOR)	5.78	9.68	13.77	17.13	19.78	20.53
Effect of diabetes status by BMI category†							
	BMI 18.5–24.9 kg/m^2	BMI<18.5 kg/m^2	BMI 25.0–29.9 kg/m^2	BMI 30.0–34.9 kg/m^2	BMI 35.0–39.9 kg/m^2	BMI≥40.0 kg/m^2	
No diabetes diagnosis	1 (No diabetes, main effect AOR)	1.12	0.93	0.96	0.98	1.16	
T1DM	3.69 (T1DM, main effect AOR)	3.21	3.21	3.41	2.55	4.41	
T2DM	3.20 (T2DM, main effect AOR)	3.79	3.13	3.04	2.83	2.68	
Effect of diabetes status by history of thrombosis‡							
	History of thrombosis, none	History of thrombosis, present					
No diabetes diagnosis	1 (No diabetes, main effect AOR)	3.30					
T1DM	3.69 (T1DM, main effect AOR)	9.64					
T2DM	3.20 (T2DM, main effect AOR)	6.81					

Beta coefficients and q values for the fully adjusted model are found in online supplemental table 12.
*The referent group was the group of individuals who did not have diabetes and who were aged 18–29.
†The referent group was the group of individuals who did not have diabetes and whose BMI was 18.5–24.9 kg/m^2.
‡The referent group was the group of individuals who did not have diabetes and who had no history of thrombosis.
AOR, adjusted OR; BMI, body mass index; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

Q<0.001) were at increased risk of TTE (table 2). The odds of TTE remained substantially higher for both T1DM (vs no diabetes; AOR 3.32; 95% CI 2.53 to 4.36; Q<0.001) and T2DM (vs no diabetes; AOR 1.96; 95% CI 1.82 to 2.12; Q<0.001) (table 2).

Interaction effects in patients with suspected or confirmed COVID-19

Results from the model that estimated interaction effects (ie, between diabetes type and age, diabetes type and BMI and diabetes type and history of TTE) showed that increased TTE risk associated with diabetes was most pronounced in individuals with T1DM who were <50 years old and in individuals with T2DM who were <60 years old (table 3). The increased risk for TTE associated with BMI≥40.0 kg/m^2 was attenuated in individuals with T2DM. Risk associated with history of TTE was highest in individuals who did not have diabetes (table 3).

Risk of specific TTEs in patients with suspected or confirmed COVID-19

Risk for MI, risk for TS and risk for DVT were substantially higher in patients with diabetes versus those without diabetes (table 4). Odds of PE were higher in individuals with T1DM (vs no diabetes; AOR 1.30; 95% CI 0.92 to 1.83; Q=0.021) and lower in individuals with T2DM (vs no diabetes; AOR 0.79; 95% CI 0.72 to 0.87; Q<0.001). Clear associations existed between older age and all four TTE subtypes (table 4). Protective factors across all TTE types included female sex and no history of thrombosis (table 4).

Risk of TTE of any type in patients with diabetes with suspected or confirmed COVID-19

In the adjusted model fit to data for only patients with diabetes, odds of TTE were lower in patients with T2DM (vs T1DM; AOR 0.84; 95% CI 0.72 to 0.98; Q=0.005) (table 5). Each one percentage point higher HbA1c was associated with an estimated 5% higher average odds

Table 4 Adjusted ORs: Correlates of myocardial infarction, thrombotic stroke, pulmonary embolism and deep vein thrombosis in individuals with suspected or confirmed COVID-19

	Myocardial infarction		Thrombotic stroke		Pulmonary embolism		Deep vein thrombosis	
Characteristic	**Adjusted OR* (95% CI)**	**Q value**	**Adjusted OR* (95% CI)**	**Q value**	**Adjusted OR* (95% CI)**	**Q value**	**Adjusted OR* (95% CI)**	**Q value**
Diabetes status								
No diabetes	1 (Reference)		1 (Reference)		1 (Reference)		1 (Reference)	
Type 1 diabetes	3.36 (2.72 to 4.15)	<0.001	1.89 (1.43 to 2.50)	<0.001	1.30 (0.92 to 1.83)	0.021	1.97 (1.41 to 2.75)	<0.001
Type 2 diabetes	1.92 (1.80 to 2.04)	<0.001	1.49 (1.39 to 1.59)	<0.001	0.79 (0.72 to 0.87)	<0.001	1.29 (1.17 to 1.42)	<0.001
Age (years)								
18–29	1 (Reference)		1 (Reference)		1 (Reference)		1 (Reference)	
30–39	2.36 (1.72 to 3.24)	<0.001	3.30 (2.34 to 4.65)	<0.001	1.66 (1.36 to 2.01)	<0.001	1.86 (1.40 to 2.48)	<0.001
40–49	7.14 (5.38 to 9.47)	<0.001	8.15 (5.93 to 11.21)	<0.001	2.43 (2.02 to 2.93)	<0.001	3.45 (2.65 to 4.50)	<0.001
50–59	13.87 (10.58 to 18.18)	<0.001	13.29 (9.76 to 18.09)	<0.001	3.28 (2.75 to 3.92)	<0.001	5.20 (4.05 to 6.68)	<0.001
60–69	21.31 (16.29 to 27.88)	<0.001	23.30 (17.18 to 31.61)	<0.001	3.95 (3.31 to 4.72)	<0.001	7.58 (5.92 to 9.70)	<0.001
70–79	28.71 (21.93 to 37.58)	<0.001	29.38 (21.64 to 39.88)	<0.001	5.08 (4.24 to 6.08)	<0.001	8.31 (6.46 to 10.68)	<0.001
80+	36.77 (28.08 to 48.15)	<0.001	37.89 (27.92 to 51.43)	<0.001	4.73 (3.92 to 5.70)	<0.001	8.53 (6.61 to 11.02)	<0.001
Sex								
Female	1 (Reference)		1 (Reference)		1 (Reference)		1 (Reference)	
Male	1.53 (1.44 to 1.61)	<0.001	1.11 (1.04 to 1.18)	<0.001	1.26 (1.16 to 1.35)	<0.001	1.37 (1.26 to 1.49)	<0.001
Race and ethnicity								
NHW	1 (Reference)		1 (Reference)		1 (Reference)		1 (Reference)	
NHB	0.93 (0.85 to 1.02)	0.021	1.25 (1.13 to 1.37)	<0.001	0.95 (0.85 to 1.07)	0.057	1.19 (1.05 to 1.34)	0.001
Hispanic	0.85 (0.79 to 0.93)	<0.001	0.85 (0.78 to 0.93)	<0.001	0.57 (0.51 to 0.64)	<0.001	0.68 (0.60 to 0.77)	<0.001
AAPI | AIAN | NHO	1.16 (1.05 to 1.29)	0.001	1.00 (0.88 to 1.14)	0.119	0.74 (0.62 to 0.88)	<0.001	0.78 (0.64 to 0.94)	0.002
BMI (kg/m^2)								
18.5–24.9	1 (Reference)		1 (Reference)		1 (Reference)		1 (Reference)	
<18.5	1.27 (1.09 to 1.48)	0.001	0.92 (0.77 to 1.11)	0.055	0.92 (0.70 to 1.20)	0.072	1.32 (1.02 to 1.70)	0.006
25.0–29.9	0.92 (0.85 to 0.99)	0.005	0.89 (0.82 to 0.96)	0.001	1.05 (0.94 to 1.17)	0.057	1.08 (0.96 to 1.22)	0.029
30.0–34.9	0.92 (0.84 to 1.00)	0.009	0.82 (0.75 to 0.90)	<0.0001	1.29 (1.15 to 1.45)	<0.001	1.34 (1.19 to 1.52)	<0.001
35.0–39.9	0.91 (0.82 to 1.01)	0.014	0.71 (0.63 to 0.80)	<0.0001	1.44 (1.26 to 1.66)	<0.001	1.26 (1.08 to 1.47)	0.001
≥40.0	0.96 (0.85 to 1.08)	0.065	0.62 (0.54 to 0.71)	<0.0001	1.81 (1.58 to 2.08)	<0.001	1.49 (1.27 to 1.75)	<0.001
History of TTE								
Not present	1 (Reference)		1 (Reference)		1 (Reference)		1 (Reference)	
Present	2.15 (2.02 to 2.28)	<0.001	3.09 (2.89 to 3.30)	<0.001	2.68 (2.45 to 2.92)	<0.001	3.09 (2.82 to 3.39)	<0.001

*Unadjusted ORs are found in online supplemental table 13.
AAPI | AIAN | NHO, Asian American/Pacific Islander, American Indian/Alaska Native, non-Hispanic Other; BMI, body mass index; NHB, non-Hispanic Black; NHW, non-Hispanic White; TTE, thrombotic or thromboembolic event.

Table 5 Adjusted ORs: Correlates of thrombotic and thromboembolic events in individuals with diabetes with suspected or confirmed COVID-19

Characteristic	Adjusted OR (95% CI)*	Q value
Diabetes status		
Type 1 diabetes	1 (Reference)	
Type 2 diabetes	0.84 (0.72 to 0.98)	0.005
Haemoglobin A1c (%)†	1.05 (1.04 to 1.07)	<0.001
Age (years)		
18–29	1 (Reference)	
30–39	2.02 (1.32 to 3.10)	<0.001
40–49	3.47 (2.33 to 5.17)	<0.001
50–59	4.76 (3.22 to 7.04)	<0.001
60–69	6.23 (4.21 to 9.21)	<0.001
70–79	7.39 (5.00 to 10.94)	<0.001
80+	7.71 (5.20 to 11.43)	<0.001
Sex		
Female	1 (Reference)	
Male	1.22 (1.15 to 1.28)	<0.001
Race and ethnicity		
NHW	1 (Reference)	
NHB	1.02 (0.94 to 1.11)	0.077
Hispanic	0.91 (0.84 to 0.98)	0.002
AAPI | AIAN | NHO	1.11 (1.00 to 1.24)	0.008
BMI (kg/m^2)		
18.5–24.9	1 (Reference)	
<18.5	1.16 (0.95 to 1.41)	0.024
25.0–29.9	0.97 (0.90 to 1.05)	0.063
30.0–34.9	0.94 (0.87 to 1.02)	0.025
35.0–39.9	0.88 (0.80 to 0.97)	0.002
≥40.0	0.85 (0.77 to 0.94)	<0.001
History of TTE		
Not present	1 (Reference)	
Present	2.23 (2.12 to 2.36)	<0.001

*Unadjusted ORs are found in online supplemental table 14.
†Haemoglobin A1c was analysed as a continuous variable.
AAPI | AIAN | NHO, Asian American/Pacific Islander, American Indian/Alaska Native, non-Hispanic Other; BMI, body mass index; NHB, non-Hispanic Black; NHW, non-Hispanic White; TTE, thrombotic or thromboembolic event.

of TTE (AOR 1.05; 95% CI 1.04 to 1.07; Q<0.001). A negative relationship existed between BMI and TTE; odds of TTE were lowest in patients with BMI≥40.0 kg/m^2 (vs 18.5–24.9 kg/m^2; AOR 0.85; 95% CI 0.77 to 0.94; Q<0.001). Protective factors included female sex and no history of TTE (table 5).

DISCUSSION

This analysis is the first to rigorously stratify patients based on diabetes status (T1DM, T2DM and no diabetes) and investigate diabetes status as a risk factor for arterial and venous TTEs in a large cohort of adult patients with COVID-19. Compared with patients with no diabetes, risk for TTE was substantially higher for patients with diabetes of either type. When odds of experiencing MI, TS, PE or DVT were analysed separately, both the T1DM and T2DM cohorts had demonstrably higher risk for MI, stroke and DVT. Relative to patients without diabetes, risk of PE was slightly elevated in the T1DM cohort and lower in patients with T2DM. Predictors for TTE included increased age, history of TTE and male sex. BMI≥30 kg/m^2 was associated with increased risk for venous thrombosis (PE and DVT), but not arterial thrombosis (MI and TS). Overall, although increased age, decreased BMI (<18.5 kg/m^2), increased BMI (≥40.0 kg/m^2) and history of thrombosis were associated with higher risk of TTE, none of these factors exerted an interaction effect that impacted the increased TTE risk associated with diabetes. In the analysis restricted to patients with diabetes, we found that the odds of TTE were highest in patients with T1DM. HbA1c had a positive relationship with TTE.

Early in the pandemic, high rates of venous TTE were reported in intensive care unit (ICU) patients with COVID-19.[17–19] A meta-analysis of prospective studies showed greater frequency of venous TTE in ICU patients with COVID-19 versus those without COVID-19.[20] Subsequent meta-analyses reported high prevalence of venous and arterial TTEs (and higher subsequent mortality) in adults with COVID-19.[21 22] Diabetes, cardiovascular disease and hypertension are frequent comorbidities in hospitalised patients with severe COVID-19; and T1DM and T2DM are associated with increased risk of mortality in hospitalised adults with COVID-19.[4]

Numerous retrospective studies report on arterial and venous thrombosis in COVID-19 and note the prevalence of diabetes, hypertension and elevated BMI in affected populations; but whether TTEs were more frequent in patients with COVID-19 with diabetes was unclear. One meta-analysis reported increased age and hypertension, but not diabetes, as risk factors for stroke in COVID-19 disease[23] while another identified diabetes as a risk factor for stroke in COVID-19.[24] Differences in case definitions may account for the variability across studies.

Our findings agree with those from a small prospective study of 169 patients hospitalised with COVID-19 pneumonia that showed increased risk of thromboembolism in adults with T2DM or stress hyperglycaemia.[25] That single-centre study, however, included only 32 patients with T2DM. A study of 4451 hospitalised patients with COVID-19 noted higher frequency of TTEs in those with diabetes but did not evaluate whether diabetes was an independent risk factor for TTEs.[26] A third analysis reported results from a cohort of 1105 adults hospitalised with COVID-19, including 138 adults with pre-existing T2DM. While this study did not report on TTEs, it did report that coagulopathy was a major extrapulmonary risk factor for mortality in the T2DM population but not in the cohort without diabetes.[27] Unlike our study, none

of the aforementioned studies evaluated patients with T1DM.

As reported in the general population, a BMI≥30 kg/m^2 increased risk for venous TTEs (PE and DVT).[28] The same relationship was not identified for arterial TTEs (MI and stroke). In fact, increasing BMI associated with substantially decreased risk of stroke. Although we did not analyse history of anticoagulant medication, this finding may reflect higher baseline use of anticoagulant therapies in patients with increased BMI. Further examination revealed that median BMI was similar across age categories and in those with and without a history of TTE when stratified by TTE status (present/absent) at the IQE (online supplemental figures 2 and 3). A slightly negative correlation existed between BMI and HbA1c in T1DM; correlation between BMI and HbA1c was slightly positive in T2DM (online supplemental figure 4).

The present results are of clinical relevance because they suggest that diabetes status may be useful to include in models that predict risk for TTEs in individuals with SARS-CoV-2 infection. Accurate risk prediction tools are needed to design and test tailored care pathways that provide heightened monitoring and treatment to individuals at highest risk. Appropriate risk stratification allows clinicians to anticipate individuals' needs and optimise outcomes through proactive engagement of at-risk patients.

Strengths and limitations

Strengths of this study include the very large sample size and utilisation of a nationwide EMR database. We used rigorous algorithms to identify patients with and without diabetes and to distinguish T1DM from T2DM. As well, we adjusted for clustering of patients within health systems and for demographic and clinical characteristics that impact risk for TTEs. This study is the first to examine factors associated with various types of arterial and venous TTEs in patients with and without diabetes. It is the only study to assess risk for TTEs in patients with T1DM with concurrent COVID-19.

Certain limitations also warrant consideration. While Oracle Cerner's COVID-19 database contains (shifted) encounter dates and start times, it does not contain timestamps that indicate the day and time at which TTEs occurred. For this reason, we were unable to evaluate the extent to which glucose management during hospitalisation may have impacted risk for TTEs. Some factors known to broadly impact TTE risk, such as smoking history and malignancy, were not evaluated. Data pertaining to medication use were not included because validity of EMR-based medication data varies substantially across medication classes (eg, non-steroidal anti-inflammatory drugs vs platelet inhibitors),[29] healthcare institutions and clinical settings.[30] Given the highly fragmented nature of medication records in EMR data, future studies should evaluate data from pharmacy claims databases, which contain the totality of patients' medication exposures, to investigate the impact of anticoagulant, antiplatelet, statin, high-dose steroid and oral contraceptive use on TTE outcomes in individuals with SARS-CoV-2 infection.

EMR data are often used, but not primarily documented, for research purposes. Rigorous case definitions are needed to identify medical conditions and clinical outcomes, as well as distinguish between diabetes types, in EMR data. The high specificity of the algorithms we used to identify patients' diabetes status resulted in an inability to definitively determine diabetes status for 112 248 (ie, 25.8% of) adult patients in the database. Further research is needed to identify methods that optimally balance sensitivity and specificity of EMR-based diabetes case definitions, as well as validate methods for identifying patients with other types of diabetes.

Similar to other published studies that use EMR data, we used ICD-10 codes to create case definitions for MI, TS, PE and DVT.[31–33] An important limitation of this approach is that few such case definitions have been validated via manual review of patient records. While we believe that our approach captured the vast majority of TTE cases, it is possible that some cases were not identified or were misclassified. We acknowledge that our findings may have been different if more rigorous methods were used to identify TTE cases (eg, presence of diagnosis codes combined with confirmatory laboratory results); however, requiring presence of multiple criteria also results in the loss of many true positive cases due to differences in documentation practices across health systems.

Finally, risk estimates could be biased if coding practices differed substantially across sites. Our analysis was limited to healthcare systems using Oracle Cerner's EMR; therefore, these findings may not generalise to healthcare settings that do not use Oracle Cerner's EMR.

CONCLUSION

This case–control study using a nationwide database identified diabetes as a risk factor for TTE in a large cohort of adult patients (n=322 482) with COVID-19. Compared with patients without diabetes, risk for TTE was substantially higher for patients with T1DM and T2DM. Further research is needed to determine whether this increased risk plays a role in the increased COVID-19-associated mortality reported in patients with diabetes.

Author affiliations
[1]Institute for Data Science and Informatics, University of Missouri, Columbia, Missouri, USA
[2]Department of Pediatrics - Division of Pediatric Endocrinology and Diabetes, Children's Mercy Kansas City, Kansas City, Missouri, USA
[3]Department of Pediatrics, Hassenfeld Children's Hospital at NYU Langone Health, New York, New York, USA
[4]Department of Pediatrics - Division of Health Services and Outcomes Research, Children's Mercy Kansas City, Kansas City, Missouri, USA
[5]Department of Pediatrics, University of Missouri–Kansas City School of Medicine, Kansas City, Missouri, USA

[6]Bambino Gesù Children's Hospital, IRCCS, Rome, Italy
[7]T1D Exchange, Boston, Massachusetts, USA
[8]School of Population Health, University of Mississippi, Jackson, Mississippi, USA
[9]Department of Cardiovascular Disease, Saint Luke's Mid America Heart Institute, Kansas City, Missouri, USA
[10]University of Missouri-Kansas City School of Medicine, Kansas City, Missouri, USA
[11]Department of Molecular and Clinical Medicine, University of Gothenburg, Gothenburg, Sweden
[12]Department of Medicine, NU-Hospital Group, Trollhättan and Uddevalla, Sweden
[13]Department of Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden
[14]Department of Medicine - Division of Endocrinology, Diabetes and Metabolism, University of Missouri, Columbia, Missouri, USA
[15]Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, Missouri, USA

Acknowledgements We thank the members of the T1D Exchange Publications Committee—especially Drs Carla Demeterco-Berggren (Rady Children's Hospital, University of California, San Diego, California, USA), Alissa J Roberts (University of Washington and Seattle Children's Research Institute, Seattle, Washington, USA), Sarah K Lyons (Baylor College of Medicine and Texas Children's Hospital, Houston, Texas, USA), Ruth S Weinstock (SUNY Upstate Medical University, Syracuse, New York, USA) and Shideh Majidi (Children's National Hospital, Washington, DC, USA)—for their valuable feedback pertaining to this manuscript.

Contributors MPG first conceptualised the idea for this manuscript. EMT, MPG and MAC formulated the study analysis plan and the aims of the study. EMT, MPG, VSS and MAC made significant contributions to the conceptualisation and design of the study, as well as to the analysis and interpretation of data. C-RS obtained the funding and provided the computational resources needed to conduct the study. VSS, C-RS and MAC provided research supervision and oversight. VSS formulated the statistical methodology. EMT and VSS completed the formal analysis and data visualisation. EMT was responsible for computer code development and implementation. EMT and MAC curated and verified the data and evaluated the reproducibility of the data and results. MPG and VSS assisted with data verification. EMT, MPG and DBM wrote the original manuscript draft. EMT, MPG, VSS, DF, OE, MNK, ML, CM-A, C-RS and MAC reviewed and edited the manuscript. All authors read and approved the final manuscript. EMT is the guarantor of this work and takes responsibility for the integrity and accuracy of the data curation and analysis. EMT and MAC accessed and verified the underlying data reported in this manuscript.

Funding ET was supported by the National Institutes of Health (grant number 5T32LM012410).

Disclaimer This manuscript's content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Competing interests MAC is the Chief Medical Officer at Glooko. He receives research support from Dexcom and Abbott Diabetes Care. MNK receives, or has received, research grant support from AstraZeneca and Boehringer Ingelheim, as well as other research support form AstraZeneca. He receives, or has received, research honoraria from AstraZeneca, Boehringer Ingelheim and Novo Nordisk. He serves, or has served, as a consultant and/or advisor for Alnylam, Amgen, AstraZeneca, Bayer, Boehringer Ingelheim, Cytokinetics, Eli Lilly, Esperion Therapeutics, Janssen, Lexicon, Merck (Diabetes and Cardiovascular), Novo Nordisk, Pharmacosmos, Sanofi and Vifor Pharma. He has received support for attending meetings and/or travel from AstraZeneca, Bayer, Boehringer Ingelheim and Novo Nordisk, and he participated on an advisory board for Applied Therapeutics.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

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. The deidentified data that support the findings of this study are proprietary and are available from Oracle Cerner. Restrictions apply to the availability of these data, which were used under licence for this study. Upon publication, additional information pertaining to statistical methods, methods used to generate results summaries and ancillary results will be provided upon request. This information will be shared with endocrinologists, healthcare professionals, data scientists and informaticians who are interested in our analytical methods or who are conducting studies pertaining to thrombotic and thromboembolic events in individuals with SARS-CoV-2 infection.

ORCID iD
Erin M Tallon http://orcid.org/0000-0003-1353-6632

REFERENCES

1 Johns Hopkins University & Medicine. Coronavirus resource center. 2022 Available: https://coronavirus.jhu.edu
2 Altonen BL, Arreglado TM, Leroux O, *et al*. Characteristics, comorbidities and survival analysis of young adults hospitalized with COVID-19 in New York City. *PLoS One* 2020;15:e0243343.
3 Dennis JM, Mateen BA, Sonabend R, *et al*. Type 2 diabetes and COVID-19-related mortality in the critical care setting: a national cohort study in England, March-July 2020. *Diabetes Care* 2021;44:50–7.
4 Longmore DK, Miller JE, Bekkering S, *et al*. Diabetes and overweight/obesity are independent, nonadditive risk factors for in-hospital severity of COVID-19: an international, multicenter retrospective meta-analysis. *Diabetes Care* 2021;44:1281–90.
5 Divani AA, Andalib S, Di Napoli M, *et al*. Coronavirus disease 2019 and stroke: clinical manifestations and pathophysiological insights. *J Stroke Cerebrovasc Dis* 2020;29:104941.
6 Li X, Weber NC, Cohn DM, *et al*. Effects of hyperglycemia and diabetes mellitus on coagulation and hemostasis. *J Clin Med* 2021;10:2419.
7 Wang W, Sun Q, Bao Y, *et al*. Analysis of risk factors for thromboembolic events in 88 patients with COVID-19 pneumonia in Wuhan, China: a retrospective descriptive report. *Med Sci Monit* 2021;27:e929708.
8 Petrauskiene V, Falk M, Waernbaum I, *et al*. The risk of venous thromboembolism is markedly elevated in patients with diabetes. *Diabetologia* 2005;48:1017–21.
9 Ehwerhemuepha L, Carlson K, Moog R, *et al*. Cerner Real-World Data (CRWD) - a de-identified multicenter electronic health records database. *Data Brief* 2022;42:108120.
10 Oracle Cerner Corporation. Cerner Real-World Data (CRWD) 2020Q3 COVID database data dictionary; 2020.
11 Qeadan F, Mensah NA, Tingey B, *et al*. The risk of clinical complications and death among pregnant women with COVID-19 in the Cerner COVID-19 cohort: a retrospective analysis. *BMC Pregnancy Childbirth* 2021;21:305.
12 Raebel MA, Schroeder EB, Goodrich G, *et al*. Mini-Sentinel statistical methods: validating type 1 and type 2 diabetes mellitus in the Mini-Sentinel Distributed Database using the SUrveillance, PREvention, and ManagEment of Diabetes Mellitus (SUPREME-DM) DataLink. 2016. Available: https://www.sentinelinitiative.org/sites/default/files/Methods/Mini-Sentinel_Methods_Validating-Diabetes-Mellitus_MSDD_Using-SUPREME-DM-DataLink.pdf
13 van BuurenS, Groothuis–Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011;45:1–67.
14 Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J Comput Grap Stat* 1995;4:12–35.
15 Rubin DB. Multiple imputation for nonresponse in surveys. Hoboken, NJ, USA: John Wiley & Sons, Inc, 1987.
16 Storey JD. The positive false discovery rate: a bayesian interpretation and the Q-value. *Ann Statist* 2003;31:2013–35.
17 Giannis D, Barish MA, Goldin M, *et al*. Incidence of venous thromboembolism and mortality in patients with initial presentation of COVID-19. *J Thromb Thrombolysis* 2021;51:897–901.

18 Lodigiani C, Iapichino G, Carenzo L, *et al*. Venous and arterial thromboembolic complications in COVID-19 patients admitted to an academic hospital in Milan, Italy. *Thromb Res* 2020;191:9–14.
19 Zhang L, Feng X, Zhang D. Deep vein thrombosis in hospitalized patients with COVID-19 in Wuhan, China. *Circulation* 2020;142:114–28.
20 Mai V, Tan BK, Mainbourg S, *et al*. Venous thromboembolism in COVID-19 compared to non-COVID-19 cohorts: a systematic review with meta-analysis. *Vascul Pharmacol* 2021;139:106882.
21 Kollias A, Kyriakoulis KG, Lagou S, *et al*. Venous thromboembolism in COVID-19: a systematic review and meta-analysis. *Vasc Med* 2021;26:415–25.
22 Malas MB, Naazie IN, Elsayed N, *et al*. 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.
23 Shahjouei S, Naderi S, Li J, *et al*. Risk of stroke in hospitalized SARS-CoV-2 infected patients: a multinational study. *EBioMedicine* 2020;59:102939.
24 Nannoni S, de Groot R, Bell S, *et al*. Stroke in COVID-19: a systematic review and meta-analysis. *Int J Stroke* 2021;16:137–49.
25 Calvisi SL, Ramirez GA, Scavini M, *et al*. Thromboembolism risk among patients with diabetes/stress hyperglycemia and COVID-19. *Metabolism* 2021;123:154845.
26 Nemetski SM, Ip A, Josephs J, *et al*. Clotting events among hospitalized patients infected with COVID-19 in a large multisite cohort in the United States. *PLoS One* 2022;17:e0262352.
27 Chen X, Chen Y, Wu C, *et al*. Coagulopathy is a major extrapulmonary risk factor for mortality in hospitalized patients with COVID-19 with type 2 diabetes. *BMJ Open Diabetes Res Care* 2020;8:e001851.
28 Mi Y, Yan S, Lu Y, *et al*. Venous thromboembolism has the same risk factors as atherosclerosis: a PRISMA-compliant systemic review and meta-analysis. *Medicine (Baltimore)* 2016;95:e4495.
29 Rowan CG, Flory J, Gerhard T, *et al*. Agreement and validity of electronic health record prescribing data relative to pharmacy claims data: a validation study from a US electronic health record database. *Pharmacoepidemiol Drug Saf* 2017;26:963–72.
30 Pevnick JM, Shane R, Schnipper JL. The problem with medication reconciliation. *BMJ Qual Saf* 2016;25:726–30.
31 Kim Y-E, Huh K, Park Y-J, *et al*. Association between vaccination and acute myocardial infarction and ischemic stroke after COVID-19 infection. *JAMA* 2022;328:887–9.
32 Katsoularis I, Fonseca-Rodríguez O, Farrington P, *et al*. Risk of acute myocardial infarction and ischaemic stroke following COVID-19 in sweden: a self-controlled case series and matched cohort study. *Lancet* 2021;398:599–607.
33 Pellathy T, Saul M, Clermont G, *et al*. Accuracy of identifying hospital acquired venous thromboembolism by administrative coding: implications for big data and machine learning research. *J Clin Monit Comput* 2022;36:397–405.