Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Risks and burdens of incident diabetes in long COVID: a cohort study

Yan Xie, Ziyad Al-Aly

Summary

Background There is growing evidence suggesting that beyond the acute phase of SARS-CoV-2 infection, people with COVID-19 could experience a wide range of post-acute sequelae, including diabetes. However, the risks and burdens of diabetes in the post-acute phase of the disease have not yet been comprehensively characterised. To address this knowledge gap, we aimed to examine the post-acute risk and burden of incident diabetes in people who survived the first 30 days of SARS-CoV-2 infection.

Methods In this cohort study, we used the national databases of the US Department of Veterans Affairs to build a cohort of 181,280 participants who had a positive COVID-19 test between March 1, 2020, and Sept 30, 2021, and survived the first 30 days of COVID-19; a contemporary control (n=4,118,441) that enrolled participants between March 1, 2020, and Sept 30, 2021; and a historical control (n=4,286,911) that enrolled participants between March 1, 2018, and Sept 30, 2019. Both control groups had no evidence of SARS-CoV-2 infection. Participants in all three comparison groups were free of diabetes before cohort entry and were followed up for a median of 352 days (IQR 245–406). We used inverse probability weighted survival analyses, including predefined and algorithmically selected high dimensional variables, to estimate post-acute COVID-19 risks of incident diabetes, antihyperglycaemic use, and a composite of the two outcomes. We reported two measures of risk: hazard ratio (HR) and burden per 1000 people at 12 months.

Findings In the post-acute phase of the disease, compared with the contemporary control group, people with COVID-19 exhibited an increased risk (HR 1·40, 95% CI 1·36–1·44) and excess burden (13·46, 95% CI 12·11–14·84, per 1000 people at 12 months) of incident diabetes; and an increased risk (1·85, 1·78–1·92) and excess burden (12·35, 11·36–13·38) of incident antihyperglycaemic use. Additionally, analyses to estimate the risk of a composite endpoint of incident diabetes or antihyperglycaemic use yielded a HR of 1·46 (95% CI 1·43–1·50) and an excess burden (12·11–14·84, per 1000 people at 12 months) of incident diabetes; and an increased risk (1·85, 1·78–1·92) and excess burden (13·46, 95% CI 12·11–14·84, per 1000 people at 12 months) of incident diabetes; and an increased risk (1·85, 1·78–1·92) and excess burden (13·46, 95% CI 12·11–14·84, per 1000 people at 12 months) of incident diabetes, antihyperglycaemic use, and a composite of the two outcomes. We reported two measures of risk: hazard ratio (HR) and burden per 1000 people at 12 months.

Interpretation In the post-acute phase, we report increased risks and 12-month burdens of incident diabetes and antihyperglycaemic use in people with COVID-19 compared with a contemporary control group of people who were enrolled during the same period and had not contracted SARS-CoV-2, and a historical control group from a pre-pandemic era. Post-acute COVID-19 care should involve identification and management of diabetes.

Funding US Department of Veterans Affairs and the American Society of Nephrology.

Introduction

A growing body of evidence suggests that beyond the first 30 days, the acute phase of the disease, people with COVID-19 could experience post-acute sequelae—referred to as long COVID—which can involve pulmonary and extrapulmonary organ system manifestations, including diabetes outcomes. Although diabetes and other glycometabolic abnormalities have been widely reported during the acute phase of COVID-19, less is known about the risk and burden of diabetes and related outcomes in the post-acute phase of COVID-19. A detailed assessment of the risk and burden of diabetes in the post-acute phase of COVID-19 is needed to inform post-acute COVID-19 care strategies.
Research in context

Evidence before this study
We searched PubMed for human studies published between Dec 1, 2019, and Sept 6, 2021, using terms “COVID-19”, “SARS-CoV-2” or “long COVID”, and “diabetes”, with no language restrictions. Small studies (<1000 people) limited to short follow-up periods (up to 3 months) showed that people with COVID-19 might be at increased risk of incident diabetes. A large-scale in-depth assessment of the risks and burdens of incident diabetes over a longer time horizon has not been done. In this study, we aimed to examine the post-acute risk and burden of diabetes in people who survived the first 30 days of SARS-CoV-2 infection.

Added value of this study
In this study involving 181 280 people with COVID-19, 4118 441 contemporary controls, and 4286 911 historical controls, we provide estimates of risks and 12-month burdens of incident diabetes outcomes. Our results suggest that beyond the first 30 days of infection, COVID-19 survivors exhibited increased risks and burdens of incident diabetes and antihyperglycaemic use. The risks and burdens were significant among those who were non-hospitalised and increased in a graded fashion according to the care setting of the acute phase of the disease (that is whether people were non-hospitalised, hospitalised, or admitted to intensive care during the acute phase of COVID-19). The risks and associated burdens were evident in comparisons versus both the contemporary control group and the historical control group.

Implications of all the available evidence
Altogether, there is evidence to suggest that beyond the acute phase of COVID-19, survivors might be at an increased risk of developing incident diabetes, and increased risk of incident antihyperglycaemic use in the post-acute phase of the disease. Diabetes should be considered as a facet of the multifaceted long COVID syndrome. Post-acute care strategies of people with COVID-19 should integrate screening and management of diabetes.
Outcomes
Post-acute COVID-19 diabetes outcomes were examined in the period of follow-up from 30 days after T₀ up to the end of follow-up. Diabetes status was defined based on the ICD-10 codes (E08.X to E13.X) or a HbA₁c measurement of more than 6.4% (46 mmol/mol), identified based on the Logical Observation Identifiers Names and Codes (LOINC). Anti-hyperglycaemic use was defined based on prescription record of diabetes medications for more than 30 days. A composite endpoint was also defined as the first occurrence of diabetes or anti-hyperglycaemic use.

Covariates
We used both predefined covariates and algorithmically selected high-dimensional covariates to adjust for the difference in baseline characteristics between groups. Predefined covariates were selected based on previous knowledge.¹²⁻²¹ Covariates were assessed within 1 year before T₀. Predefined baseline variables included age, race (White, Black, or other race), sex, area deprivation index, BMI, smoking status (current smoker, former smoker, or never smoker), use of long-term care (including nursing homes and assisted-living centres), number of outpatient and inpatient encounters, and number of HbA₁c measurements. Comorbidities such as cancer, cardiovascular disease, cerebrovascular disease, chronic lung disease, dementia, HIV, hyperlipidaemia, and peripheral artery disease were also included as predefined covariates. Additionally, we also adjusted for laboratory test results including estimated glomerular filtration rate (eGFR) and HbA₁c; vital signs including systolic and diastolic blood pressure; and medications including the use of steroids. Missingness of BMI, blood pressure, eGFR, and HbA₁c were 1·02%, 1·28%, 6·20%, and 15·43%, respectively. Mean imputations conditional on eGFR, and HbA₁c were 1·02%, 1·28%, 6·20%, and 6·43%, respectively. Mean imputations conditional on age, race, sex, and group assignment were applied to missing values and continuous variables were transformed into restricted cubic spline functions to account for the potential non-linear relationships.

To further enhance the adjustment of potential confounding, and to complement our list of prespecified variables, we algorithmically selected and adjusted for potential confounders from data domains including diagnoses, medications, and laboratory test results.¹⁶ We obtained all patient encounter data, prescription data, and laboratory data for the cohort of participants within 1 year before T₀. We classified more than 70 000 ICD-10 diagnosis codes into 540 diagnostic categories based on the Clinical Classifications Software Refined (version 2021.1), which is developed as part of the Healthcare Cost and Utilization Project sponsored by the Agency for Healthcare Research and Quality.²⁰ We classified 3425 medications, on the basis of the VA drug classification system, into 543 medication classes.²¹ In total, 62 laboratory test abnormalities from 38 laboratory measurements were identified on the basis of LOINC. Because rare conditions occurring in less than 100 people in a group might not be sufficiently substantial to describe the characteristics of the group, only diagnoses, medications, or laboratory test abnormalities with an event of more than 100 within each group, which were not included as predefined variables, were used to further estimate the univariate relative risk for COVID-19 group assignment.¹ The top 100 variables with the strongest univariate relative risk were selected.²³ The selection process was done independently for COVID-19 versus contemporary control groups, and COVID-19 versus historical control groups.

Statistical analyses
Baseline characteristics of the COVID-19 and control groups, as well as standardised mean differences between the groups were reported. To estimate the association between COVID-19 and post-acute diabetes outcomes, high dimensional propensity scores were used to adjust for the difference between the COVID-19 and control groups at baseline. For each study group, a logistic regression, including predefined and 100 algorithmically selected high dimensional variables, was used to estimate the propensity score as the probability of assignment to the target population, which was defined as VHA users in 2019 (the year before the first COVID-19 infection occurred in the study population). The inverse probability weight for each participant was then constructed as the propensity score from the previous logistic regression divided by 1 minus the propensity score.²⁵ Inverse probability weighting was then applied to a Cox survival model to estimate the association between COVID-19 and diabetes outcomes. Two measures of risks were estimated, including the adjusted hazard ratios (HRs) and excess burdens. To generate the excess burdens, burdens of diabetes outcomes at 12 months in each group were estimated based on the survival probability at 12 months of follow-up. Excess burdens per 1000 people at 12 months from COVID-19 compared with controls was estimated based on the difference on survival probability between groups and transformed as event rate difference. Comparisons were done between COVID-19 and contemporary control groups, and independently between COVID-19 and historical control groups. The analyses were then repeated in subgroups based on age (≤65 years and >65 years), race (White and Black); subgroup analyses for other race category were not done because of the heterogeneity within this category), sex (male and female), BMI categories (>18·5 to ≤25 kg/m²; >25 to ≤30 kg/m²; and >30 kg/m²), area deprivation index quartiles, and diabetes risk score quartiles. A diabetes risk score was built using logistic regression to predict the probability of having a composite diabetes outcome within 1 year. The risk score was built within control groups based on diabetes risk factors including age, race, sex, BMI, HbA₁c, cardiovascular disease, hypertension, and hyperlipidaemia status. The risk score was then applied to the COVID-19 group to evaluate the risk of diabetes outcomes before exposure to COVID-19. 
To gain a better understanding of which subgroups with COVID-19 are more likely to have post-acute COVID-19 diabetes events, we estimated the effect of risk factors including diabetes risk scores, age, race, cardiovascular diseases, hypertension, hyperlipidaemia, prediabetes status (HbA1c >5.6% and <6.4%), and BMI categories on diabetes outcomes within 30-day survivors of COVID-19. We constructed logistic regressions within each COVID-19 subgroup to estimate the probability of assignment to the target population, conditional on covariates other than the subgrouping definition. Inverse probability weightings were then computed, and survival models were used to examine the HRs and burdens of these risk factors on diabetes outcomes.

We then separated the COVID-19 group into three mutually exclusive groups based on the care setting of the acute phase of the disease; that is whether people were non-hospitalised, hospitalised, or admitted to intensive care during the first 30 days after a COVID-19 positive test. Logistic regressions were applied to each care setting group to estimate the inverse probability weights. Cox survival models with inverse probability weighting were then applied and HRs, burdens, and excess burdens were reported.

To test the robustness of our findings, we applied an alternative analytic plan. Only cohort participants with complete data and at least 12 months of follow-up were selected and censored at 12 months (COVID-19 group n=62,110 and contemporary control group n=127,7659). Multinomial logistic regression adjusting for predefined covariates was used to estimate the propensity scores for cohort participants. Average treatment effect weights were then constructed from the propensity score with stabilisation based on proportions of each group in the overall cohort. Weighted logistic regressions were then applied to estimate the odds ratios and predicted probabilities of having the outcome. Variance was estimated through generalised estimating equation, which considers the within-participant correlation after weightings.

We also did multiple additional sensitivity analyses to test the robustness of results to changes in specification of our primary approach. First, we repeated the analyses while additionally adjusting for the month of cohort enrolment, in consideration of the putative presence of a temporal confounding effect. Second, we defined outcomes based on their second occurrence during the follow-up. Third, we used 300 algorithmically selected high dimensional variables (instead of the 100 used in the primary analyses) to adjust for potential additional confounders. Fourth, conversely, we estimated the association by using only predefined covariates (ie, without the use of high dimensional variables). Fifth, instead of inverse probability weighting, we used overlap weighting to estimate the association.33,34 Sixth, we applied the doubly robust adjustment method to further adjust for covariates after applying inverse probability weighting. Seventh, to further account for missing data, we applied multiple imputation to generate ten imputed datasets based on fully conditional specification regression method and estimated results.35 Eight, to remove the influence of steroid use during the acute phase of the infection, we additionally adjusted for steroid use during the acute phase of the infection. Finally, to reduce the bias associated with increased surveillance for COVID-19 patients during follow-up, we additionally adjusted for the number of outpatient visits, number of hospitalisations, and number of HbA1c measurements during the follow-up as time varying variables.

To evaluate the success of our approach, we first tested the association between COVID-19 and the risk of death as a positive outcome control—where established evidence suggests an association is expected. To detect the presence of spurious biases, we first examined the association between COVID-19 and risks of diagnostic codes based outcomes including hearing aid use and acne, and, separately, risk of laboratory-based outcomes including serum albumin of more than 5 g/dL, total protein of more than 8.5 g/dL, serum potassium of more than 5.1 mmol/L, serum calcium of more than 10.5 mg/dL, and high-density lipoprotein of less than 40 mg/dL as negative outcome controls—where there is no evidence to suggest that an association is expected. Successful reproduction of established knowledge (positive outcome control), and the successful application of negative controls, would reduce concerns about biases related to cohort building, study design, analytic approach, outcome ascertainment, residual confounding, and other latent biases.36,37

Robust sandwich estimators were used to estimate variances when weightings were applied. For all analyses, a 95% CI that excluded unity or a p value of less than 0.05 was considered evidence of statistical significance. Analyses were done using SAS Enterprise Guide (version 8.2) and results were visualised using SAS Enterprise Guide (version 8.2) and R (version 4.0.4).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

There were 4,299,721 US veterans in the cohort overall recruited from March 1, 2020, to Sept 30, 2021; 181,280 were in the COVID-19 group and 4,118,441 were in the contemporary control group. The median follow-up time was 352 (IQR 244–406) days in the COVID-19 group and 352 (245–406) days in the contemporary control group, corresponding to 181,280 person-years and 376,315 person-years of follow-up, respectively.

To test the consistency of the results, we also built a historical cohort of 4,286,911 participants followed up for a median of 352 (IQR 245–406) days, corresponding to 3,916,979 person-years of follow-up.
The demographic and health characteristics of the historical control group, contemporary control group, and COVID-19 group before weighting are provided in the appendix (pp 8–9); characteristics after weighting are provided in the table. The absolute numbers and incident rates for outcomes before and after weighting are also provided in the appendix (pp 10–11). Most incident diabetes outcomes were type 2 diabetes; 0.68% and 0.71% of the ICD-based outcomes in the COVID-19 group were diabetes type 1 in the unweighted and weighted cohort, respectively (appendix p 11).

For all analyses, we provide two measures of risk: first, we estimated the adjusted HRs of incident diabetes outcomes; and second, we estimated the excess burden from the difference between the incident event rates per 1000 people at 12 months in the COVID-19 and control groups. Assessment of covariate balance after application of inverse probability weighting suggested that standardised mean differences are less than 0.1 (indicating good balance) for predefined covariates, high dimensional covariates selected by our algorithm, and those not selected (table and appendix p 3).

### Table: Demographic and health characteristics of the COVID-19, contemporary control, and historical control groups after adjustment

| Baseline characteristics | COVID-19 (n=181 280) | Contemporary control (n=4 118 441) | Historical control (n=4 286 911) | Absolute standardised difference |
|--------------------------|----------------------|-----------------------------------|---------------------------------|---------------------------------|
|                          |                      | COVID-19 and contemporary control* | COVID-19 and historical control* |
| Age, years               | 60.92 (17.02)        | 61.5 (17.08)                      | 61.49 (17.13)                   | 0.01                            | 0.01                            |
| Race                     |                      |                                   |                                 |                                 |
| White                    | 138 949 (76.65%)     | 319 488 (77.58%)                  | 332 624 (77.59%)                | 0.02                            | 0.02                            |
| Black                    | 34 015 (18.76%)      | 73 605 (17.91%)                   | 76 654 (17.88%)                 | 0.02                            | 0.02                            |
| Other†                   | 8 314 (4.59%)        | 18 865 (4.51%)                    | 19 415 (4.53%)                  | 0.00                            | 0.00                            |
| Sex                      |                      |                                   |                                 |                                 |
| Male                     | 159 666 (88.08%)     | 365 034 (88.75%)                  | 380 476 (88.74%)                | 0.02                            | 0.02                            |
| Female                   | 21 614 (11.92%)      | 46 347 (11.25%)                   | 48 285 (11.26%)                 | 0.02                            | 0.02                            |
| Smoking status           |                      |                                   |                                 |                                 |
| Never                    | 77 577 (42.8%)       | 184 024 (44.68%)                  | 193 978 (44.72%)                | 0.04                            | 0.04                            |
| Former                   | 61 748 (34.06%)      | 126 746 (33.19%)                  | 140 554 (33.14%)                | 0.02                            | 0.02                            |
| Current                  | 41 858 (23.09%)      | 91 142 (22.13%)                   | 94 936 (22.15%)                 | 0.02                            | 0.02                            |
| BMI, kg/ m²              | 29.2 (6.06)          | 29.15 (5.98)                      | 29.15 (5.98)                    | 0.00                            | 0.00                            |
| Area deprivation index‡  | 54.17 (18.97)        | 53.89 (19.06)                     | 53.90 (19.06)                   | 0.02                            | 0.02                            |
| Clinical characteristics  |                      |                                   |                                 |                                 |
| Outpatient encounter§    |                      |                                   |                                 |                                 |
| Zero or one              | 92 214 (50.87%)      | 213 471 (51.9%)                   | 222 020 (51.79%)                | 0.02                            | 0.02                            |
| Two                      | 48 435 (26.75%)      | 111 845 (27.16%)                  | 114 700 (26.77%)                | 0.01                            | 0.00                            |
| Three or more            | 40 583 (22.39%)      | 86 482 (20.94%)                   | 91 285 (21.44%)                 | 0.04                            | 0.02                            |
| Number of HbA 1c measure‡| 0.43 (0.62)          | 0.42 (0.62)                       | 0.42 (0.62)                     | 0.03                            | 0.03                            |
| Long-term care¶          | 1017 (0.56%)         | 16 062 (0.39%)                    | 17 233 (0.40%)                  | 0.02                            | 0.02                            |
| Estimated glomerular filtration rate, mL/min per 1.73m² | 81.67 (19.77) | 81.12 (19.45) | 81.12 (19.47) | 0.02 | 0.02 |
| HbA1c                    | 5.53% (0.35)         | 5.54% (0.34)                      | 5.53% (0.35)                    | 0.02                            | 0.02                            |
| HbA1c, mmol/mol          | 36.94 (3.83)         | 37.05 (3.72)                      | 36.94 (3.83)                    | 0.02                            | 0.02                            |
| Systolic blood pressure, mm Hg | 131.63 (12.44) | 131.73 (12.31) | 131.71 (12.37) | 0.01 | 0.01 |
| Diastolic blood pressure, mm Hg | 78.32 (7.55) | 78.24 (7.51) | 78.25 (7.53) | 0.01 | 0.01 |
| Cancer                   | 9330 (5.15%)         | 207 940 (5.05%)                   | 217 818 (5.08%)                 | 0.00                            | 0.00                            |
| Cardiovascular disease   | 15 030 (8.29%)       | 339 277 (8.24%)                   | 356 457 (8.32%)                 | 0.00                            | 0.00                            |
| Cerebrovascular disease  | 57 30 (3.16%)        | 124 418 (3.02%)                   | 130 879 (3.05%)                 | 0.01                            | 0.01                            |
| Chronic lung disease     | 16 942 (0.35%)       | 369 671 (0.98%)                   | 386 937 (0.93%)                 | 0.01                            | 0.01                            |
| Dementia                 | 467 (2.58%)          | 98 678 (2.40%)                    | 103 958 (2.43%)                 | 0.01                            | 0.01                            |
| HIV                      | 75 (0.42%)           | 16 227 (0.39%)                    | 17 019 (0.40%)                  | 0.00                            | 0.00                            |
| Hyperlipidaemia          | 40 092 (2.08%)       | 1 069 312 (25.96%)                | 1 119 141 (26.11%)              | 0.03                            | 0.02                            |
| Peripheral artery disease| 1026 (0.57%)         | 22 075 (0.54%)                    | 22 707 (0.55%)                  | 0.00                            | 0.00                            |
| Steroid prescription     | 2779 (1.53%)         | 58 152 (1.41%)                    | 61 121 (1.43%)                  | 0.01                            | 0.01                            |

Data are mean (SD) or n (%). *Standardised difference of less than 0.10 is considered good balance. †Latinx, Asian, American Indian Native Hawaiian, and patients of other races. ‡Area deprivation index is a measure of socioeconomic disadvantage, with a range from low to high disadvantage of 0–100. §Data collected within 1 year of cohort enrolment. ¶Nursing homes and assisted-living centers.
Compared to the contemporary control group, 30-day survivors of COVID-19 exhibited an increased risk (HR 1·40, 95% CI 1·36–1·44) and excess burden (13·46, 95% CI 12·11–14·84, per 1000 people at 12 months) of incident diabetes; and an increased risk (1·85, 1·78–1·92) and excess burden (12·35, 11·36–13·38) of incident antihyperglycaemic use. Analyses to estimate the risk of a composite endpoint of incident diabetes or antihyperglycaemic use yielded a HR of 1·46 (95% CI 1·43–1·50) and an excess burden of 18·03 (16·59–19·51) per 1000 people at 12 months (figure 1 and appendix pp 4, 12).

Subgroup analyses suggested that COVID-19 was associated with an increased risk of diabetes outcomes across age (≤65 years and >65 years), race (White and Black), BMI categories (>18·5 to ≤25 kg/m², >25 to ≤30 kg/m², and >30 kg/m²), and area deprivation index quartiles. We then examined the associations according to diabetes risk score quartiles; the results suggested that COVID-19 was associated with an increased risk of diabetes across all risk score quartiles, including the lowest risk score quartile (appendix pp 13–14).

We then further examined the risks and burdens of post-acute incident diabetes, antihyperglycaemic use, and the composite outcome by the severity of disease during the acute phase of the infection (non-hospitalised, hospitalised, and admitted to intensive care); demographic and health characteristics of these groups before and after weighting are provided in the appendix (pp 15–18). Assessment of covariate balance after application of weights suggested covariates were well balanced. Compared with the contemporary control group, the risks and burdens of post-acute of incident diabetes, antihyperglycaemic use, and the composite outcome increased according to the severity of the acute infection (figure 2 and appendix p 19).

We examined the associations between COVID-19 and diabetes in analyses considering a historical control.
The results suggested that COVID-19 was associated with an increased risk of diabetes outcomes in comparisons of COVID-19 versus the overall historical control group, and across all the subgroups examined (appendix pp 5–6, 20–22), and were consistent with those evaluating the COVID-19 versus contemporary control groups.

We then did analyses by care setting of the acute phase of the COVID-19 infection compared with the historical control group. The results suggested that the risks of diabetes outcomes exhibited a graded increase according to the intensity of care during the acute phase of the infection and were consistent with analyses considering the COVID-19 group versus the contemporary control group (appendix pp 7, 23–27).

To gain a deeper understanding of who is at most risk of post-acute diabetes outcomes, we did analyses among people who survived the first 30 days of COVID-19 to identify characteristics of individuals who were at highest risk of incident diabetes, antihyperglycaemic use, and the composite outcome. We found there was a graded increase in risks and burdens with increasing quartile of diabetes risk score (figure 3A). People older than 65 years had higher risks and burdens than those younger than 65 years. People with cardiovascular disease, hypertension, and hyperlipidaemia also had higher risks than those without these conditions.

The outcomes were ascertained from day 30 after COVID-19 infection until the end of follow-up. Adjusted hazard ratios and 95% CIs are presented in a base 10 logarithmic scale. Excess burden per 1000 people at 12 months and 95% CIs are also presented.
65 years. Black participants exhibited higher risks and burdens than White participants. Those with cardiovascular disease, hypertension, hyperlipidaemia, or prediabetes also exhibited higher risks and burdens than people without these conditions. Compared to those with a BMI of >18·5 kg/m² to ≤25 kg/m², there was a graded increase in risks and burdens in those with BMIs of >25 kg/m² and ≤30 kg/m² and or in those with a BMI of >30 kg/m² (figure 3B and appendix p 28).

To test the robustness of our results, we applied an alternative analytic approach where we used predefined covariates based inverse probability weighted logistic regression within participants with at least 1 year of follow-up. The risk and burden of diabetes outcomes were consistent with the main findings (appendix p 29).

All sensitivity analyses produced results consistent with the primary analyses (appendix p 30).

To test whether our approach would reproduce established associations, we examined death as a positive outcome control; the results suggested that COVID-19 was associated with higher risk of death (HR 1·49, 95% CI 1·44—1·55; appendix p 31).

We then tested the association between COVID-19 and the risks of hearing aid use and—independently—risk of acne as two ICD-10 based negative outcome controls where no previous knowledge suggests an association is expected. The results suggested no association between COVID-19 and the risk of hearing aid use or acne. We additionally tested the association between COVID-19 and laboratory-based negative outcome controls including serum albumin of more than 5 g/dL, total protein of more than 8·5 g/dL, serum potassium of more than 5·1 mmol/L, serum calcium of more than 10·5 mg/dL, and high-density lipoprotein of less than 40 mg/dL. The results suggested no association with any of the laboratory-based negative outcome controls (appendix p 31).

**Discussion**

In this study involving participants with COVID-19, contemporary controls, and historical controls, we provide evidence that suggests that beyond the first 30 days of infection, COVID-19 survivors exhibited increased risks and burdens of incident diabetes, and antihyperglycaemic use. The risks and burdens of all outcomes were significant among those non-hospitalised and increased in a graded fashion according to the care setting of the acute phase of the infection. The risks and burdens were also consistent in comparisons versus a historical control group. Altogether, our results indicate that beyond the acute phase of COVID-19, survivors are at an increased risk of developing incident diabetes and antihyperglycaemic use; therefore diabetes should be considered as a component of the multifaceted long COVID. Post-acute care strategies of people with COVID-19 should also integrate screening and management of diabetes.

The implications of our findings are clear. In the post-acute phase of the disease, COVID-19 was significantly associated with increased risk of incident diabetes. Although the risks and burdens increased according to the severity of the acute infection (as proxied by the care setting), they were evident and not trivial among people who were not hospitalised for COVID-19—this group represents most people with COVID-19. For example, the excess burden of diabetes among non-hospitalised individuals was 8·28 per 1000 people at 12 months. Given the large and growing number of people infected with COVID-19 (>450 million people globally as of March 15, 2022), these absolute numbers might translate into substantial overall population level burdens and could further strain already overwhelmed health systems. Governments and health systems around the world should be prepared to screen and manage the glycometabolic sequelae of COVID-19. Although the optimal composition of post-acute COVID clinics is still not clear, evidence from this report indicated that those should include attention and care for diabetes.

Our approach examines the risks and burdens of diabetes in comparisons versus a contemporary control group exposed to the same contextual forces of the pandemic (eg, economic, social, and environmental stressors) and a historical control group from a pre-pandemic era that represents a baseline unaffected by the pandemic. COVID-19 consistently exhibited an increased risk of diabetes in comparisons versus both the contemporary and historical control groups, suggesting enhanced vulnerability to diabetes among people with COVID-19.

Our subgroup analyses suggest that even people with a low risk of diabetes before exposure to COVID-19 exhibited increased risk compared to both contemporary and historical controls. In addition, our analyses of who is at risk of diabetes among people with COVID-19 suggest that the relationship between COVID-19 and diabetes exhibited a graded association according to baseline risk of diabetes suggesting that diabetes could manifest in people at low risk (compared with controls), and COVID-19 could likely amplify baseline risks and further accelerate manifestation of disease among individuals already at high risk.

Studies on the link between COVID-19 and diabetes are generally limited by short follow-up and most investigate outcomes in hospitalised individuals. Evidence in children and young adults is mixed. A study of two large databases of more than 2·5 million children (aged <18 years) suggested that those with COVID-19 exhibited a higher risk of new diabetes than those without COVID-19. Additionally, the risk of new diabetes was higher in COVID-19 than in those with pre-pandemic acute respiratory infections. This study did not report the proportion of type 1 or type 2 diabetes. An analysis, which has not yet been peer reviewed, of 1·8 million people aged younger than 35 years suggested increased risk of type 1 diabetes within, but not beyond, the first 30 days after SARS-CoV-2 infection. Studies in...
adults are generally more concordant and show evidence of increased risk of diabetes in people with COVID-19.\textsuperscript{41–43} Our study sheds light on this and provides evidence of increased risk in adults among both non-hospitalised and hospitalised individuals at 1 year after COVID-19 diagnosis; and that most (>99%) of diagnoses of diabetes in our cohort relate to type 2 diabetes.

The mechanism(s) underpinning the association between COVID-19 and risk of diabetes are not entirely clear. Several pancreatic cell types express three proteins (angiotensin converting enzyme 2 receptor protein, TMPRSS2 enzyme protein, and neuropilin 1) on which SARS-CoV-2 depends for its entry into human cells.\textsuperscript{39} Evidence suggests that SARS-CoV-2 can infect and replicate in insulin-producing pancreatic beta cells subsequently resulting in impaired production and secretion of insulin.\textsuperscript{40–43} However, in-vitro SARS-CoV-2-infected human pancreatic islets exhibit largely non-cytopathic modest cellular perturbations and inflammatory responses – suggesting that direct infection of pancreatic cells is – on its own – unlikely to fully explain new onset diabetes in people with COVID-19.\textsuperscript{42} Other potential explanations include autonomic dysfunction, hyperactivated immune response or autoimmunity, and persistent low-grade inflammation leading to insulin resistance.\textsuperscript{44–46} It is also possible that people with COVID-19 might have differentially experienced some of the broader contextual changes (social, economic, environmental, and other) that characterised the pandemic and that might have indirectly contributed to shaping the outcomes evaluated in this study.\textsuperscript{47,48}

There are several strengths of this study. We leveraged the breadth and depth of the US Department of Veterans Affairs electronic health-care databases to build a large national cohort of veterans, without a history of diabetes, to investigate the association between COVID-19 and risks of diabetes outcomes. We tested the association using two large controls (contemporary and historical controls), an approach that allowed us to deduce that the associations between COVID-19 and risks of diabetes are not related to the broader temporal changes between the pre-pandemic and the pandemic eras, but rather related (possibly through both a direct and indirect pathway) to exposure to COVID-19 itself. Our covariates specification approach included 22 predefined variables selected based on previous evidence and 100 algorithmically selected variables from high dimensional data domains including diagnostic codes, prescription records, and laboratory test results. We evaluated several incident diabetes outcomes across the continuum of the severity scale, including diabetes diagnoses and initiation of antihyperglycaemic therapy. We tested robustness of our approach in multiple sensitivity analyses, and successfully applied positive and negative outcome controls. We provided estimates of risks on both the ratio scale (HRs) and the absolute scale (burden per 1000 people at 12 months). The absolute scale also reflects the contribution of baseline risk and provides an estimate of potential harm that is more easily explainable to the general public than risk reported on the ratio scale (eg, HR).

This study has several limitations. The demographic composition of our cohort (comprised mostly of White males) could limit the generalisability of the findings. Although we leveraged the breadth and depth of the vast electronic health-care databases to build our cohorts, required well defined criteria for cohort entry, and defined health characteristics based on validated definitions, we cannot rule out misclassification bias; in particular, misclassification of diabetes type. Although we adjusted (through inverse probability weighting) for a large set of predefined covariates and 100 algorithmically selected high dimensional covariates, we cannot completely rule out residual confounding. We required a positive COVID-19 test for enrolment in the COVID-19 group. For the contemporary control group, it is possible that some of those enrolled might have contracted SARS-CoV-2 and were not tested for it, and if these people were present in large numbers within the contemporary control group, this might have biased the results towards the null hypothesis. Although we took care to balance the exposure groups at baseline, and did analyses adjusting for health resources use during follow-up, we cannot rule out the possibility that some of the cases were undiagnosed diabetes cases that were formally diagnosed after COVID-19. Lastly, as the pandemic continues (in the USA and in several areas around the globe), as new variants emerge, and as treatment strategies for acute COVID-19 continue to evolve, it is likely that the epidemiology of post-acute COVID-19 sequelae, including diabetes, will likely also change over time.\textsuperscript{49}

In conclusion, we suggest that in the post-acute phase of the disease, people with COVID-19 exhibit increased risk and burden of diabetes, and antihyperglycaemic use. The risks and burdens were evident among those who were non-hospitalised during the acute phase of the infection and increased according to the severity of the acute infection as proxied by the care setting (non-hospitalised, hospitalised, and admitted to intensive care). Taken together, current evidence suggests that diabetes is a facet of the multifaceted long COVID syndrome and that post-acute care strategies of people with COVID-19 should include identification and management of diabetes.

**Contributors**

ZA-A was responsible for the research and study design; administrative, technical, or material support; and supervision and mentorship. YX acquired the data and did the statistical analysis. YX and ZA-A were responsible for the data analysis and interpretation, drafting the manuscript, and critical revision of the manuscript. Each author contributed important intellectual content during manuscript drafting or revision, and accept accountability for the overall work by ensuring that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved. ZA-A takes responsibility that this study has been reported honestly, accurately, and transparently; that no important aspects of the study have been omitted, and that any discrepancies from the study as planned have been explained. Both authors had full access to all the data, and both have
verified the accuracy of all underlying data. Both authors had final responsibility for the decision to submit for publication.

Declaration of interests
YZ and ZA-A declare support from the US Department of Veterans Affairs for the submitted work. YZ declares support for the American Society of Nephrology for the submitted work. ZA-A reports receiving consultation fees from GlaxoSmithKline and receipt of funding (unrelated to this work) from Tovix Pharmaceuticals. ZA-A is a Member Board of Directors for Veterans Research and Education Foundation of Saint Louis, associate editor for the Journal of the American Society of Nephrology, and is a member of multiple editorial boards.

Data sharing
The data that support the findings of this study are available from the US Department of Veterans Affairs (VA), Office of Research and Development, VA Information Resource Center by emailing VIREC@va.gov.

Acknowledgments
This study used data from the VA COVID-19 shared data resource. This research was funded by the US Department of Veterans Affairs to ZA-A. ZA-A is a Member Board of Directors for Veterans Research and Education Foundation of Saint Louis, associate editor for the Journal of the American Society of Nephrology, and is a member of multiple editorial boards.

This research was funded by the US Department of Veterans Affairs to ZA-A. Veterans Affairs or the US Government. The contents of this manuscript do not represent the views of the US Department of Veterans Affairs or the US Government.

References
1 Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of COVID-19. Nature 2021; 594: 559–64.
2 Montefusco L, Ben Nasser M, D’Addio F, et al. Acute and long-term disruption of glycometabolic control after SARS-CoV-2 infection. Nat Metab 2021; 3: 774–85.
3 Xie Y, Bowe B, Maddukuri G, Al-Aly Z. Comparative evaluation of clinical manifestations and risk of death in patients admitted to hospital with COVID-19 and seasonal influenza: cohort study. BMJ 2020; 367: n4677.
4 Accili D. Can COVID-19 cause diabetes? Nat Metab 2021; 3: 123–25.
5 Saltish T, Kapoor N, Cao Y, Tapp RJ, Zimmet P. Proportion of newly diagnosed diabetes in COVID-19 patients: a systematic review and meta-analysis. Diabetes Obes Metab 2021; 23: 870–74.
6 Rubino F, Amiel SA, Zimmet P, et al. New-onset diabetes in COVID-19. N Engl J Med 2020; 383: 789–90.
7 Ayoubkhani D, Khunti K, Nafliany V, et al. Post-COVID syndrome in individuals admitted to hospital with COVID-19: retrospective cohort study. BMJ 2021; 372: n693.
8 Steenblock C, Schwarz PEH, Ludwig B, et al. COVID-19 and metabolic disease: mechanisms and clinical management. Lancet Diabetes Endocrinol 2021; 9: 786–98.
9 Khunti K, Del Prato S, Mathieu C, Kahn SE, Gabbay RA, Buse JB. COVID-19, hyperglycemia, and new-onset diabetes. Diabetes Care 2021; 44: 2645–55.
10 Bowe B, Xie Y, Xian H, Liu M, Al-Aly Z. Geographic variation and US County characteristics associated with rapid kidney function decline. Kidney Int Rep 2016; 2: 5–17.
11 Xie Y, Bowe B, Li T, Xian H, Yan Y, Al-Aly Z. Higher blood urea nitrogen is associated with increased risk of incident diabetes mellitus. Kidney Int 2018; 93: 741–52.
12 Bowe B, Xie Y, Yan Y, Al-Aly Z. Burden of cause-specific mortality associated with PM2.5 air pollution in the United States. JAMA Netw Open 2019; 2: e1915834.
13 Xie Y, Bowe B, Yan Y, Xian H, Li T, Al-Aly Z. Estimates of all cause mortality and cause-specific mortality associated with proton pump inhibitors among US veterans: cohort study. BMJ 2019; 365: h15580.
14 Bowe B, Xie Y, Xian H, Li T, Al-Aly Z. Association between monocyt count and risk of incident CKD and progression to ESRD. Clin J Am Soc Nephrol 2017; 12: 603–11.
15 Xie Y, Bowe B, Gibson A, et al. Comparative effectiveness of the sodium-glucose co-transporter-2 inhibitor empagliflozin vs other antihyperglycemics on risk of major adverse kidney events. Diabetologia 2020; 43: 2785–95.
16 Xie Y, Bowe B, Gibson AK, McGill JB, Maddukuri G, Al-Aly Z. Comparative effectiveness of sodium-glucose cotransporter 2 inhibitors vs sulfonylureas in patients with type 2 diabetes. JAMA Intern Med 2021; 181: 1043–51.
17 Bowe B, Cai M, Xie Y, Gibson AK, Maddukuri G, Al-Aly Z. Acute kidney injury in a national cohort of hospitalized US veterans with COVID-19. Clin J Am Soc Nephrol 2020; 16: 14–25.
18 Xie Y, Bowe B, Gibson AK, et al. Comparative effectiveness of SGLT2 Inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and sulfonylureas on risk of kidney outcomes: emulation of a target trial using health care databases. Diabetes Care 2020; 43: 2589–69.
19 Xie Y, Bowe B, Gibson AK, McGill JB, Maddukuri G, Al-Aly Z. Clinical implications of estimated glomerular filtration rate dip following sodium-glucose cotransporter-2 inhibitor initiation on cardiovascular and kidney outcomes. J Am Heart Assoc 2021; 10: e020237.
20 Department of Veterans Affairs. COVID-19: shared data resource 2020. https://vhacdwdweb1001.ocha.med.va.gov/phenotype/index. php?COVID-19-Shared_Data_Resource#Acknowledgements__COVID-19_Shared_Data_Resource (accessed Dec 30, 2021).
21 Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible—the Neighborhood Atlas. N Engl J Med 2018; 378: 2456–58.
22 Xie Y, Bowe B, Al-Aly Z. Burdens of post-acute sequelae of COVID-19 by severity of acute infection, demographics and health status. Nat Commun 2021; 12: 6571.
23 Bowe B, Xie Y, Xie Y, Al-Aly Z. Kidney outcomes in long COVID. J Am Soc Nephrol 2021; 32: 2851–62.
24 Xie Y, Xie Y, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. Nat Med 2022; published online Feb 7. https://doi.org/10.1038/s41591-022-01689-3.
25 Xie Y, Xie Y, Al-Aly Z. Risks of mental health outcomes in people with COVID-19: cohort study. BMJ 2022; 377: o068991.
26 Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 2009; 20: 512–22.
27 Wei Y, Wang Y, Di Q, et al. Short term exposure to fine particulate matter and hospital admission risks and costs in the Medicare population: time stratified, case crossover study. BMJ 2019; 367: l2528.
28 Aubert CE, Schnipper JL, Roumelt M, et al. Best definitions of multimorbidity to identify patients with high health care resource utilization. Mayo Clin Proc Innov Qual Outcomes 2020; 4: 40–49.
29 Agency for Healthcare Research and Quality. Clinical Classifications Software Refined (CCSR). https://www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp (accessed Dec 30, 2021).
30 Olvey EL, Clausercche S, Malone DC. Comparison of critical drug-drug interaction listings: the Department of Veterans Affairs medical system and standard reference compendia. Clin Pharmacol Ther 2010; 87: 48–51.
31 Greene M, Steinman MA, McNicholl IR, Valcour V. Polypharmacy, drug-drug interactions, and potentially inappropriate medications in older adults with human immunodeficiency virus infection. J Am Geriatr Soc 2014; 62: 447–53.
32 Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011; 46: 399–424.
33 Thomas LE, Li F, Pencina MJ. Overlap weighting: a propensity score method that mimics attributes of a randomized clinical trial. JAMA 2020; 323: 2417–18.
34 Li F, Thomas LE, Li F. Addressing extreme propensity scores via the overlap weights. Am J Epidemiol 2019; 188: 250–57.
35 van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res 2007; 16: 219–42.
36 Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology 2018; 29: 383–88.
37 Shi X, Miao W, Tchetgen ET. A selective review of negative control methods in epidemiology. Curr Epidemiol Rep 2020; 7: 190–202.
38 WHO. WHO coronavirus (COVID-19) dashboard. https://covid19.who.int/ (accessed March 15, 2022).
39 Barrett CE, K. A., Alvarez P, et al. Risk for newly diagnosed diabetes >30 days after SARS-CoV-2 infection among persons aged <18 years—United States, March 1, 2020–June 28, 2021. Morb Mortal Wkly Rep 2021; 70: 59–65.
40 McKeigue PM, et al. Relation of incident type 1 diabetes to recent COVID-19 infection: cohort study using e-health record linkage in Scotland. medRxiv 2022; published online Feb 11. https://doi.org/10.1101/2022.02.11.22270785 (preprint).

41 Sathish T, Kapoor N, Cao Y, Tapp RJ, Zimmet P. Proportion of newly diagnosed diabetes in COVID-19 patients: a systematic review and meta-analysis. Diabetes Obes Metab 2021; 23: 870–74.

42 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.

43 Collaborative O, et al. Rates of serious clinical outcomes in survivors of hospitalisation with COVID-19: a descriptive cohort study within the OpenSAFELY platform. medRxiv 2021; published online Jan 25. https://doi.org/10.1101/2021.01.22.21250304 (preprint).

44 Ziegler CGK, Allon SJ, Nyquist SK, et al. SARS-CoV-2 receptor ACE2 is an interferon – stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. Cell 2020; 181: 1016–35.

45 Müller JA, Groß R, Conzelmann C, et al. SARS-CoV-2 infects and replicates in cells of the human endocrine and exocrine pancreas. Nat Metab 2021; 3: 149–65.

46 Yang L, Han Y, Nilsson-Payant BE, et al. A human pluripotent stem cell-based platform to study SARS-CoV-2 tropism and model virus infection in human cells and organoids. Cell Stem Cell 2020; 27: 125–136.

47 Wu CT, Lidsky PV, Xiao Y, et al. SARS-CoV-2 infects human pancreatic β cells and elicits β cell impairment. Cell Metab 2021; 33: 1563–76.

48 Tang X, Uhl S, Zhang T, et al. SARS-CoV-2 infection induces beta cell transdifferentiation. Cell Metab 2021; 33: 1577–91.

49 van der Heide V, Jangra S, Cohen P, et al. Limited extent and consequences of pancreatic SARS-CoV-2 infection. Cell Rep 2022; 38: 110598.

50 Bowe B, Xie Y, Gibson AK, et al. Ambient fine particulate matter air pollution and the risk of hospitalization among COVID-19 positive individuals: cohort study. Environ Int 2021; 154: 106564.

51 Xie Y, Bowe B, Yan Y, Cai M, Al-Aly Z. County-level contextual characteristics and disparities in life expectancy. Mayo Clin Proc 2021; 96: 92–104.

52 Cai M, Bowe B, Xie Y, Al-Aly Z. Temporal trends of COVID-19 mortality and hospitalisation rates: an observational cohort study from the US Department of Veterans Affairs. BMJ Open 2021; 11: e047369.