

Diabetes following SARS-CoV-2 infection: Incidence, persistence, and implications of COVID-19 vaccination. A cohort study of fifteen million people.

Authors

Kurt Taylor^{1*}, Sophie Eastwood^{2*}, Venexia Walker^{1,3,4*}, Genevieve Cezard^{5,6}, Rochelle Knight^{1,3,7,8}, Marwa Al Arab¹, Yinghui Wei⁹, Elsie M F Horne^{1,7}, Lucy Teece¹⁰, Harriet Forbes¹¹, Alex Walker¹², Louis Fisher¹², Jon Massey¹², Lisa E M Hopcroft¹², Tom Palmer^{1,2}, Jose Cuitun Coronado¹, Samantha Ip^{5,6}, Simon Davy¹², Iain Dillingham¹², Caroline Morton¹², Felix Greaves^{13,14}, John Macleod^{1,7,8}, Ben Goldacre¹², Angela Wood^{5,6,15,16,17,18}, Nishi Chaturvedi^{2^}, Jonathan A C Sterne^{1,7,19^*}, Rachel Denholm^{1,7,19^}

on behalf of the Longitudinal Health and Wellbeing and Data and Connectivity UK COVID-19 National Core Studies, CONVALESCENCE study, and the OpenSAFELY collaborative.

* authors contributed equally

^ authors contributed equally

Affiliations

¹Population Health Sciences, University of Bristol, Bristol, UK

²MRC Unit for Lifelong Health and Ageing, University College London, London, UK

³MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

⁴Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

⁵British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

⁶Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge UK

⁷NIHR Bristol Biomedical Research Centre, Bristol, UK

⁸The National Institute for Health and Care Research Applied Research Collaboration West (NIHR ARC West) at University Hospitals Bristol and Weston, UK

⁹Centre for Mathematical Sciences, School of Engineering, Computing and Mathematics, University of Plymouth, Plymouth, UK

¹⁰ Department of Population Health Sciences, University of Leicester, Leicester, UK

¹¹Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK

¹²The Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

¹³ National Institute for Health and Care Excellence, UK

¹⁴ Department of Primary Care and Public Health, Imperial College London, London, UK

¹⁵ British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK

¹⁶ National Institute for Health and Care Research Blood and Transplant Research Unit in Donor Health and Behaviour, University of Cambridge, Cambridge, UK

¹⁷ Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK

¹⁸ Cambridge Centre of Artificial Intelligence in Medicine

¹⁹ Health Data Research UK South-West, Bristol, UK

* **Corresponding author:** Jonathan A C Sterne, Jonathan.Sterne@bristol.ac.uk

Abstract

Background

Type 2 diabetes (T2DM) incidence is increased after diagnosis of COVID-19. The impact of vaccination on this increase, for how long it persists, and the effect of COVID-19 on other types of diabetes remain unclear.

Methods

With NHS England approval, we studied diabetes incidence following COVID-19 diagnosis in pre-vaccination (N=15,211,471, January 2020-December 2021), vaccinated (N =11,822,640), and unvaccinated (N=2,851,183) cohorts (June-December 2021), using linked electronic health records. We estimated adjusted hazard ratios (aHRs) comparing diabetes incidence post-COVID-19 diagnosis with incidence before or without diagnosis up to 102 weeks post-diagnosis. Results were stratified by COVID-19 severity (hospitalised/non-hospitalised) and diabetes type.

Findings

In the pre-vaccination cohort, aHRs for T2DM incidence after COVID-19 (compared to before or without diagnosis) declined from 3.01 (95% CI: 2.76,3.28) in weeks 1-4 to 1.24 (1.12,1.38) in weeks 53-102. aHRs were higher in unvaccinated than vaccinated people (4.86 (3.69,6.41)) versus 1.42 (1.24,1.62) in weeks 1-4) and for hospitalised COVID-19 (pre-vaccination cohort 21.1 (18.8,23.7) in weeks 1-4 declining to 2.04 (1.65,2.51) in weeks 52-102), than non-hospitalised COVID-19 (1.45 (1.27,1.64) in weeks 1-4, 1.10 (0.98,1.23) in weeks 52-102). T2DM persisted for 4 months after COVID-19 for ~73% of those diagnosed. Patterns were similar for Type 1 diabetes, though excess incidence did not persist beyond a year post-COVID-19.

Interpretation

Elevated T2DM incidence after COVID-19 is greater, and persists longer, in hospitalised than non-hospitalised people. It is markedly less apparent post-vaccination. Testing for T2DM after severe COVID-19 and promotion of vaccination are important tools in addressing this public health problem.

1 **Research in context**

2 *Evidence before this study*

3 We searched PubMed for population-based observational studies published between December 1st
4 2019 and July 12th 2023 examining associations between SARS-CoV-2 infection or COVID-19 diagnosis
5 (search string: SARS-CoV-2 or COVID* or coronavirus*) and subsequent incident diabetes (search term:
6 diabetes). Of nineteen relevant studies; eight had a composite outcome of diabetes types, six stratified
7 by diabetes type and five pertained to type-1-diabetes (T1DM) only. We did not identify any studies
8 relating to gestational or other types of diabetes. Eleven studies were from the US, three from the UK,
9 two from Germany, one from Canada, one from Denmark and one from South Korea.

10 Most studies described cumulative relative risks (for infection versus no infection) one to two years
11 post-SARS-CoV-2 infection of 1.2 to 2.6, though four studies found no associations with T1DM after the
12 post-acute period. All studies lacked the power to compare diabetes relative risk by type, severity, and
13 vaccination status in population subgroups. One study examined relative risks by vaccination status, but
14 this used a composite outcome of diabetes and hyperlipidaemia and was conducted in a predominantly
15 white male population.

16 Two studies of T1DM found no evidence of elevated risk beyond 30 days after COVID-19 diagnosis,
17 whilst two reported elevated risks at six months. Two studies of type 2 diabetes (T2DM) examined
18 relative risks by time period post-infection: one study of US insurance claims reported a persistent
19 association six months post-infection, whereas a large UK population-based study reported no
20 associations after 12 weeks. However, the latter study used only primary care data, therefore COVID-19
21 cases were likely to have been under-ascertained.

1 No large studies have investigated the persistence of diabetes diagnosed following COVID-19; key to
2 elucidating the role of stress/steroid-induced hyperglycaemia.

3 *Added value of this study*

4 This study, which is the largest to address the question to date, analysed linked primary and secondary
5 care health records with SARS-CoV-2 testing and COVID-19 vaccination data for 15 million people living
6 in England. This enabled us to compare the elevation in diabetes incidence after COVID-19 diagnosis by
7 diabetes type, COVID-19 severity and vaccination status, overall and in population subgroups.
8 Importantly, excess diabetes incidence by time period since infection could also be quantified. Since
9 healthcare in the UK is universal and free-at-the-point-of-delivery, almost the entire population is
10 registered with primary care. Therefore the findings are likely to be generalisable.

11 We found that, before availability of COVID-19 vaccination, a COVID-19 diagnosis (vs. no diagnosis) was
12 associated with increased T2DM incidence which remained elevated by approximately 30% beyond one
13 year after diagnosis. Though still present (with around 30% excess incidence at eight weeks), these
14 associations were substantially attenuated in unvaccinated compared with vaccinated people. Excess
15 incidence was greater in people hospitalised with COVID-19 than those who were not hospitalised after
16 diagnosis. T1DM incidence was elevated up to, but not beyond, a year post COVID-19. Around 73% of
17 people diagnosed with incident T2DM after COVID-19 still had evidence of diabetes four months after
18 infection.

19 *Implications of all the available evidence*

20 There is a 30-50% elevated T2DM incidence post-COVID-19, but we report the novel finding that there is
21 elevated incidence beyond one-year post-diagnosis. Elevated T1DM incidence did not appear to persist
22 beyond a year, which may explain why previous studies disagree. For the first time in a general-
23 population dataset, we demonstrate that COVID-19 vaccination reduces, but does not entirely

- 1 ameliorate, excess diabetes incidence after COVID-19. This supports a policy of universal vaccination and
- 2 suggests that other public health activities, such as enhanced diabetes screening after severe COVID-19,
- 3 may be warranted, particularly in unvaccinated people.

1 **Introduction**

2 At least 660 million people worldwide have been infected with severe acute respiratory syndrome
3 coronavirus 2 (SARS-CoV-2).^{1,2} Therefore reports of excess diabetes risk after COVID-19³⁻¹³ have alarming
4 public health implications. A 30% to 50% excess of type-2-diabetes (T2DM) incidence following SARS-
5 CoV-2 infection has been reported^{3,4,6,8,13,14}. In contrast, the only three studies sufficiently powered to
6 examine type-1-diabetes (T1DM) reported no association with SARS-CoV-2¹⁵⁻¹⁷. A key question is
7 whether excess diabetes after COVID-19 is driven by short-term causes of hyperglycaemia
8 (stress/steroid-induced) or is a durable consequence of infection. Yet most studies only examined
9 diabetes incidence at a fixed timepoint after COVID-19^{3,4,9-12} and those that stratified by time period
10 post-infection do not concur^{3,4,6,9,11,13}. COVID-19 severity is likely to be a key determinant of
11 downstream sequelae^{4-11,13,15,16}, only one study, in a predominantly white male cohort, examined the
12 impact of vaccination on diabetes after COVID¹³.

13 In the largest study to date, we quantified associations of COVID-19 with incident diabetes using a UK
14 database of linked COVID-19 testing data and primary and secondary care records. We investigated how
15 associations varied: at different timepoints up to two years post-diagnosis; by diabetes type; according
16 to vaccination availability and vaccination status; by COVID-19 severity; and within population
17 subgroups.

18 **Methods**

19 Information on data access can be found under “Data Sharing” at the end of this paper. Analyses were
20 performed according to a pre-specified protocol which can be found along with publicly available code
21 lists and analysis code: <https://github.com/opensafely/post-covid-diabetes>. This study was approved by
22 the Health Research Authority [REC reference 22/PR/0095] and by the University of Bristol's Faculty of

1 Health Sciences Ethics Committee [reference 117269] (see “Information governance and ethical
2 approval”).

3 **Study design and data source**

4 This observational cohort study included all adults registered with a primary care general practice using
5 TPP software in England. Data were linked, stored and analysed securely within the OpenSAFELY
6 platform (<https://opensafely.org/>), which includes individual-level, primary care electronic health
7 records from 24 million people, covering around 40% of the English population, securely linked to the
8 Second Generation Surveillance System (SGSS) for Pillar 1 and Pillar 2 SARS-COV-2 infection laboratory
9 testing data, NHS hospital admissions (Secondary Uses Services data) and the Office of National
10 Statistics death registry, including causes of death¹⁸. COVID-19 vaccination records (National
11 Immunisation Management System) are available within TPP primary care data.

12 The UK’s COVID-19 vaccine roll-out started on December 8th, 2020 with timing of eligibility in order of
13 priority groups determined by the Joint Committee on Vaccination and Immunisation, based on age,
14 clinical vulnerability and health and social care occupation¹⁹. All adults in England were eligible to
15 receive a first vaccination by June 18th 2021, and a second vaccination by August 2021²⁰.

16 **Outcomes**

17 Diabetes phenotypes were defined using primary care and hospital admission data (**Supplementary**
18 **Table 1**). An updated, clinician-verified SNOMED-CT and ICD-10 diabetes diagnostic adjudication
19 algorithm^{21,22} was used to define incident T1DM and T2DM, gestational diabetes, other or non-specific
20 diabetes, and diabetes unlikely (**Supplementary Figure 1**). In the pre-vaccination cohort, to differentiate
21 T2DM from temporary steroid- or stress-induced hyperglycaemia, a secondary outcome of ‘persistent
22 T2DM’ was defined as T2DM with continued treatment (≥ 2 prescriptions of glucose-lowering
23 medication) or elevated HbA1c levels (≥ 47.5 mmol)) 4 months after diagnosis.

1 **Exposures**

2 The date of COVID-19 diagnosis was defined as the earliest of the date of a positive SARS-CoV-2
3 polymerase chain reaction or antigen test in the SGSS system, a confirmed COVID-19 diagnosis in
4 primary or secondary care, or death with SARS-CoV-2 infection listed as primary or underlying cause.
5 People hospitalised with a COVID-19 diagnosis in the primary position within 28 days of first COVID-19
6 diagnosis were defined as 'hospitalised COVID-19', otherwise as 'non-hospitalised COVID-19'.

7 **Potential confounding variables**

8 Primary and secondary care records up to the cohort start date (see below) were used to define age,
9 sex, ethnicity, socioeconomic deprivation (area-based index of multiple deprivation quintiles), smoking
10 status and region. We defined additional potential confounding variables based on previous disease
11 diagnoses, comorbidities and medications (**Supplementary Table 2**).

12 **Study population**

13 We defined three cohorts, summarised in **Supplementary Table 3**. In the 'pre-vaccination' cohort,
14 follow-up started on January 1st 2020 (baseline) and ended on the earliest of December 14th 2021
15 (when the Omicron variant became dominant in England)²³, date that the outcome event was recorded,
16 and date of death. Exposure was defined as a COVID-19 diagnosis between baseline and the earliest of
17 eligibility for COVID-19 vaccination (based on age and clinical vulnerability) and date of first vaccination:
18 this exposure period was before the Delta variant became dominant in England. The other two cohorts
19 were followed during the period when the Delta variant was dominant in England: between June 1st
20 2021 (baseline) and December 14th 2021 (study end date). Follow-up in the 'vaccinated' cohort started
21 at the later of baseline and two weeks after a second COVID-19 vaccination and ended at the earliest of
22 the study end date, outcome event date, and date of death. The 'unvaccinated' cohort had not received
23 a COVID-19 vaccine by 12 weeks after they became eligible for vaccination. Follow-up started at the

1 later of baseline and 12 weeks after eligibility for vaccination and ended at the earliest of the study end
2 date, outcome event date, date of death, and date of first vaccination.

3 People eligible for each cohort had been registered with an English GP using TPP software for at least six
4 months before the cohort baseline, were alive and aged between 18 and 110 years at baseline, and had
5 known sex, region and area deprivation. People with a history of SARS-CoV-2 infection or COVID-19
6 diagnosis before the cohort baseline were excluded. In the vaccinated cohort, people who received a
7 COVID-19 vaccination before December 8th 2020, or a second dose before or less than three weeks after
8 their first dose, or received more than one type of vaccine before 7th May 2021, were excluded. In the
9 unvaccinated cohort, people who could not be assigned to a vaccination priority group as defined by the
10 Joint Committee on Vaccination and Immunisation (JCVI) were excluded.

11 People were excluded if they had previously been diagnosed with diabetes prior to the study start date.
12 For gestational diabetes, the study population was restricted to women.

13 **Statistical analyses**

14 We described baseline demographic and clinical characteristics of each cohort and calculated the
15 number of events per outcome, person-years of follow-up and incidence rates (per 100,000 person-
16 years) of events before and after all, hospitalised and non-hospitalised COVID-19.

17 For each outcome, we analysed time to first event. Cox models were fitted with calendar time scale
18 using the cohort-specific baseline as the origin (time zero). We estimated hazard ratios (HRs) comparing
19 follow-up after COVID-19 with follow-up before or without COVID-19, splitting follow-up time into
20 periods 1-4 weeks and 5-28 weeks after COVID-19 for all cohorts and additionally 29-52 and 53-102
21 weeks after COVID-19 for the pre-vaccination cohort. All models were stratified by region, to account for
22 between-region variation. For each outcome and cohort, we estimated: (i) age and sex; and (ii)
23 maximally (including all potential confounders) adjusted HRs. We conducted subgroup analyses

1 according to whether people had been hospitalised for COVID-19 within 28 days of COVID-19 diagnosis.
2 For T2DM only, we conducted additional subgroup analyses by age group (18-39 / 40-59 / 60-79 / 80-
3 110 years), sex, ethnicity, prior history of pre-diabetes and obesity. Absolute excess risks (AER) of T2DM
4 after COVID-19, weighted by the proportions of individuals in age and sex strata in the pre-vaccination
5 cohort, were derived. Further details of the statistical analyses and AER calculations are provided in the
6 supplementary material.

7 **Sensitivity analyses**

8 We performed sensitivity analyses for T2DM outcomes only. We repeated the main analysis in people
9 who had COVID-19 before the cohort start date (vaccinated and unvaccinated cohorts only), and
10 separating events on day 0 (day of COVID-19 diagnosis) from the rest of weeks 1-4, to explore the
11 influence of simultaneous reporting of the exposure and outcome.

12 **Results**

13 **Supplementary Figures 2-4** show participant selection into each cohort. Among 15,211,471 people in
14 the pre-vaccination cohort, 827,074 were diagnosed with COVID-19 during the study period (**Table 1**).
15 Corresponding numbers were 750,370 diagnosed with COVID-19 among 11,822,640 people in the
16 vaccinated cohort and 147,044 among 2,851,183 in the unvaccinated cohort. Compared to the
17 vaccinated cohort, people who remained unvaccinated were younger (61.5% versus 28.3% aged ≤ 40
18 years in the unvaccinated and vaccinated cohorts, respectively), more likely to be men (59% versus
19 47.3% respectively), and more likely to be from South Asian (9.6% versus 5.1%, respectively), Black (5.1%
20 versus 1.5%, respectively) and other (5.6% and 1.7%) ethnic backgrounds, and the most deprived
21 background (29.9% versus 15.7%, respectively). People in the unvaccinated cohort had fewer
22 comorbidities compared to those in the vaccinated cohort (**Supplementary Table 4**).

1 There were 133,323, 32,466, and 2,140 incident T2DM diagnoses in the pre-vaccination, vaccination and
2 unvaccinated cohorts, respectively (**Table 2**). Corresponding numbers of incident T1DM events were
3 14,917, 2410 and 45 respectively. In all cohorts, both T2DM and T1DM incidence was substantially
4 greater after hospitalised COVID-19 than before or without COVID-19 or after non-hospitalised COVID-
5 19. Among people in in the pre-vaccination cohort with follow-up data four months after diagnosis,
6 there were 110,580 incident T2DM diagnoses of which 81,729 (74%) were defined as persistent T2DM.
7 The proportion of T2DM diagnoses after COVID-19 defined as persistent was 3,186/4,385 (73%) and was
8 similar after hospitalised COVID-19 (754/1,097 (69%)) and non-hospitalised COVID-19 (2,432/3,288
9 (74%)).

10 **Incident T2DM after COVID-19 versus before or without COVID-19**

11 In the pre-vaccination and unvaccinated cohorts, maximally adjusted HRs (aHRs) comparing T2DM
12 incidence after COVID-19 with the incidence before or without COVID-19 were attenuated compared
13 with age-, sex- and region-adjusted HRs (**Table 3**). However, maximally adjusted and age-sex adjusted
14 aHRs were similar in the vaccinated cohort. In each cohort, T2DM incidence was elevated during the first
15 four weeks after COVID-19 diagnosis: this elevation was greater in the unvaccinated cohort (aHR 4.82
16 [95% CI 3.65,6.36]) than the vaccinated cohort (1.42 [1.25,1.62]) (**Figure 1**). In the pre-vaccination
17 cohort, T2DM incidence remained elevated 53-102 weeks post diagnosis (aHR 1.24 [1.12,1.37]). The
18 elevation in T2DM incidence 5-28 weeks after COVID-19 was less marked in the vaccinated and
19 unvaccinated cohorts (aHRs 1.17 [1.06,1.31] and 1.12 [0.74,1.68] respectively) than in the pre-
20 vaccination cohort (1.30 [1.24-1.38]). In the pre-vaccination cohort, the aHRs for persistent T2DM were
21 lower than for all T2DM during weeks 1-4, similar during weeks 5-28 and higher during weeks 29-52
22 (**Supplementary Figure 5**). In the unvaccinated cohort, aHRs for T2DM did not change markedly when
23 censoring at first vaccination was removed (**Supplementary Table 5 and Supplementary Figure 6**).

1 The aHRs comparing T2DM incidence after COVID-19 with the incidence before or without COVID-19
2 were higher for hospitalised than non-hospitalised COVID-19, in each cohort (for example, 21.0
3 [18.7,23.6] versus 1.45 [1.28,1.65] in weeks 1-4 in the pre-vaccination cohort) (**Table 3, Figure 1**). T2DM
4 incidence after hospitalised COVID-19 remained elevated beyond four weeks after COVID-19 diagnosis
5 in each cohort. In the pre-vaccination cohort, the aHR for incident T2DM after hospitalised COVID-19
6 was 2.04 (1.65,2.51) during weeks 53-102. In the vaccinated cohort, T2DM incidence was not markedly
7 elevated after non-hospitalised COVID-19. Patterns of aHRs estimated within shorter time intervals
8 during the first 16 weeks after COVID-19 were similar to the main analyses (**Supplementary Figure 7**).
9 The aHRs for persistent T2DM after hospitalised and non-hospitalised COVID-19 were consistent with
10 those for all incident T2DM, except that the aHR for persistent T2DM after non-hospitalised COVID-19
11 was close to 1 (**Supplementary Figure 5**). In each cohort, and after both hospitalised and non-
12 hospitalised COVID-19, a substantial proportion of T2DM diagnoses during weeks 1-4 were on the day of
13 COVID-19 diagnosis (**Supplementary Table 6**) and the aHRs were markedly higher on the day of COVID-
14 19 diagnosis than during the rest of weeks 1-4 (**Supplementary Figure 8**).

15 The aHRs comparing T2DM incidence after COVID-19 with the incidence before or without COVID-19
16 were higher in older than younger age groups during weeks 1-4. Between age-group differences in aHRs
17 were small in subsequent time periods (**Supplementary Table 7 and Supplementary Figure 9**). There
18 were no marked differences between aHRs for T2DM by sex or between ethnic groups. In the pre-
19 vaccination cohort, the aHR for T2DM during weeks 1-4 after COVID-19 diagnosis was higher for people
20 without than with obesity, and for people without than with pre-diabetes. In both the vaccinated and
21 unvaccinated cohorts, aHRs for T2DM in people with a COVID-19 diagnosis before the cohort start date
22 could not be estimated because there were too few (<50) incident T2DM events after a further COVID-
23 19 diagnosis during the cohort follow-up.

1 Absolute excess risks of T2DM by age and sex

2 Estimated excess risks of T2DM 6 months post-COVID-19, standardised to the age and sex distribution of
3 the pre-vaccination cohort, were 130, 56 and 94 per 100,000 people in the pre-vaccination, vaccinated
4 and unvaccinated cohorts respectively (**Supplementary Figure 10**). Absolute differences in risk were
5 higher in those aged over 60 years across, compared to younger age groups, in each cohorts. There was
6 little difference in excess risk by sex, ethnicity, and presence or absence of pre-diabetes and obesity.

7 Incident T1DM, gestational and other diabetes after COVID-19 versus before or without COVID-19

8 The aHRs comparing T1DM incidence after COVID-19 with the incidence before or without COVID-19
9 were higher during weeks 1-4 in the pre-vaccination and unvaccinated cohorts (1.73 [1.13,2.65] and
10 2.59 [1.21,5.55], respectively) than in the vaccinated cohort (1.34 [0.78-2.33]) (**Table 3 & Figure 2**). In
11 the pre-vaccination cohort, T1DM incidence remained elevated during weeks 29-52 (aHR 1.34 (1.09-
12 1.64)) but not during weeks 53-102 (0.91 [0.55-1.50]). In the pre-vaccination cohort, aHRs for T1DM
13 incidence were markedly higher after hospitalised COVID-19 than after non-hospitalised COVID-19.
14 There was no consistent evidence of excess gestational diabetes incidence after COVID-19 in any group.
15 The incidence of other types of diabetes was also elevated after COVID-19, most markedly during the
16 first year after hospitalised COVID-19 in the pre-vaccination cohort (**Figure 2**).

17 Discussion

18 Key findings

19 In the cohort exposed before availability of COVID-19 vaccination, T2DM incidence was three times
20 higher during the first four weeks after a diagnosis of COVID-19, compared with before or without
21 COVID-19. T2DM incidence remained elevated by 24% overall during the second year after diagnosis, 2-
22 fold higher in those hospitalised, and 10% higher in those not hospitalised with COVID-19. The majority

1 of incident T2DM after COVID-19 was persistent. Elevation in T2DM incidence after COVID-19 was
2 markedly attenuated in vaccinated compared with unvaccinated people (1.4-fold versus 4.8-fold during
3 weeks 1-4 after COVID-19 diagnosis). T1DM incidence was only elevated during the first year after
4 COVID-19 diagnosis. Gestational diabetes incidence did not appear elevated after COVID-19 diagnosis.

5 **Comparison with literature/ interpretation of findings**

6 Elevated diabetes incidence (mostly T2DM) up to a year after COVID-19 has been reported in a recent
7 systematic review²⁴. However, the largest previous UK study did not confirm elevation in diabetes
8 incidence beyond 12 weeks after COVID-19⁶. In that study, 20% of COVID-19 diagnoses were not
9 laboratory confirmed, potentially attenuating the association between COVID-19 and incident diabetes.
10 Additionally, people with pre-existing cardiovascular disease (CVD) were excluded, removing a group
11 susceptible to both COVID-19 and T2DM.

12 There was little variation in post COVID-19 hazard ratios for incident T2DM between subgroups defined
13 by ethnicity, sex, and pre-diabetes or obesity. This implies that absolute differences in excess risk after
14 COVID-19 are higher in ethnic groups at higher risk of T2DM and in people with pre-diabetes and
15 obesity.

16 Persistent T2DM, defined as being on glucose lowering medication or an HbA1c level consistent with
17 T2DM at 4 months post diagnosis, was observed in 73% of incident T2DM after COVID-19, similar to the
18 74% persistence of incident T2DM before or without COVID-19. Using a similar definition of persistence,
19 56% of all newly diagnosed T2DM persisted up to a year post COVID-19 in a previous study²⁵.
20 Additionally, 35% of newly diagnosed pre-diabetes post COVID-19 persisted at 6 months²⁵. These
21 previous studies were in hospitalised patients: we found similar levels of persistence in those
22 hospitalised and those not hospitalised with COVID-19.

1 No previous study has examined incident T2DM post COVID-19 diagnosis stratified by multiple periods
2 of follow up; some excluded events in the first 30 days³. Yet the very high T2DM incidence, including at
3 the time of COVID-19 diagnosis in those hospitalised, should not be overlooked, given that the majority
4 of these cases persist. Explanations for elevated T2DM incidence around the time of COVID-19 diagnosis
5 include routine tests unmasking undiagnosed diabetes, or acute infection stress precipitating
6 hyperglycaemia in those pre-disposed to diabetes. Incident diabetes later after COVID-19 diagnosis may
7 reflect direct pancreatic beta-cell damage, or systemic inflammatory responses in those not previously
8 at risk²⁶. The hazard ratio for incident T2DM between 5-28 weeks in the vaccinated cohort (1.17) was
9 lower than that for the pre-vaccination cohort (1.30) and almost identical to the hazard ratio of 1.16
10 observed for metabolic outcomes in a US cohort at 6 months¹³.

11 Pre-vaccination, T1DM incidence was elevated 1-4 weeks post COVID-19, and this elevation persisted
12 during weeks 5-52, after both hospitalised and (to a lesser extent) non-hospitalised COVID-19, but not
13 during the second year. This is consistent with previous studies in both adults (>18 years)⁴ and a meta-
14 analysis of children and adolescents,²⁷ but inconsistent with a Scottish study of children and young
15 adults which only observed an excess risk in the month after COVID-19¹¹. Thus, the balance of evidence
16 does not support the hypothesis that elevated T1DM incidence soon after COVID-19 diagnosis is simply
17 due to opportunistic testing for SARS-CoV-2 at the time of T1DM diagnosis or the short time lag
18 between onset of T1DM and diagnosis¹⁵. Previous studies largely censored follow up at around a year,
19 and spanned the pre and post vaccination periods. We found that the elevation in T1DM incidence after
20 COVID-19 was markedly attenuated in the vaccinated cohort.

21 We found no clear elevation in gestational diabetes incidence after COVID-19, though there was some
22 indication of small increases in the incidence of other forms of diabetes which persisted beyond 9
23 months. We are not aware of any previous studies examining these sub-types.

1 **Strengths and limitations**

2 A key strength is our sample size of around 15 million, enabling comparisons by diabetes type,
3 vaccination status and time period post-infection in many population sub-groups, which no other study
4 has done. Another strength is generalisability; the sample population was representative of the UK's
5 age, sex and ethnicity distribution. Linkage of primary and secondary care data with national SARS-CoV-2
6 testing data maximised and made more precise COVID-19 and other covariate ascertainment, and
7 enabled study of hospitalised and non-hospitalised groups. All analyses accounted precisely for calendar
8 time, so fluctuations in the incidence of COVID-19 and in the availability of health services will not have
9 affected our findings.

10 Several limitations are present. Mild or asymptomatic COVID-19 cases will have been under-ascertained
11 before widespread testing was available, biasing associations towards the null. Ascertainment of
12 prevalent, rather than incident, diabetes may have been more likely in those with COVID-19, especially
13 as diabetes pre-disposes to severe COVID-19. We addressed this possibility with sensitivity analyses
14 excluding diabetes diagnoses on the same day as COVID-19 diagnosis. Further, elevated diabetes
15 incidence after COVID-19 was evident in the non-hospitalised population, who were less likely to have
16 diabetes tests in the acute period. Conversely, there may have been under ascertainment of diabetes
17 due to reduced use of primary care services throughout the pandemic and cessation of normal screening
18 activities, which would bias associations towards the null. Finally, misclassification of diabetes diagnoses
19 is common in all settings ²¹ and may have particularly applied to acute hyperglycaemia in secondary care
20 settings in this study. We used a previously published algorithm to overcome this ²¹.

21 **Clinical implications**

22 Our findings have implications for the COVID-19, and potentially future, pandemics. England had around
23 2 million COVID-19 cases pre-vaccination, and 18 million post-vaccination ²⁸. Taking account of non-

1 vaccination, and using our estimates of absolute numbers of people with incident T2DM by vaccination
2 status, we estimate around 14,000 additional new cases of T2DM 6 months post COVID-19 over the
3 period of the pandemic. This contrasts with an estimated 56,000 new cases of all diabetes each year in
4 England between 2015 and 2020 ²⁹. Acknowledging the assumptions and uncertainties associated with
5 such calculations, and that this is markedly lower than a previous estimate ³⁰, this remains an alarmingly
6 high number of new cases of T2DM, with significant costs to the individual and society. Encouraging
7 vaccination, which markedly reduces risk of incident T2DM after COVID-19, is essential. Routine testing
8 for diabetes after severe COVID-19, particularly in people at prior elevated risk of diabetes, and ensuring
9 treatment and continued monitoring to identify those whose diabetes persists or resolves, should be
10 considered. That COVID-19 appears to increase incidence of T1DM ¹¹ and possibly T2DM ^{31,32} to a greater
11 extent than other respiratory infections adds weight to this recommendation. Our finding that T2DM
12 incidence remains elevated up to 2 years after COVID-19 in the unvaccinated cohort emphasises the
13 need to extend these analyses with longer follow up.

Table 1. Patient characteristics in the pre-vaccination, vaccinated and unvaccinated cohorts.

Characteristics		Pre-vaccination cohort (1 Jan 2020 to 14 Dec 2021)		Vaccinated cohort (1 June to 14 Dec 2021)		Unvaccinated cohort (1 June to 14 Dec 2021)	
		N (%)	COVID-19 diagnoses	N (%)	COVID-19 diagnoses	N (%)	COVID-19 diagnoses
All		15,211,471	827,074	11,822,640	750,370	2,851,183	147,044
Sex	Female	7,679,998 (50.5)	451,321	6,233,098 (52.7)	422,888	1,169,870 (41)	78,520
	Male	7,531,473 (49.5)	375,753	5,589,542 (47.3)	327,482	1,681,313 (59)	68,524
Age, years	18-29	2,767,824 (18.2)	216,367	1,570,150 (13.3)	83,851	877,616 (30.8)	39,918
	30-39	2,819,706 (18.5)	184,089	1,777,903 (15)	135,148	875,004 (30.7)	51,000
	40-49	2,691,915 (17.7)	168,773	1,988,754 (16.8)	215,123	549,468 (19.3)	33,780
	50-59	2,766,992 (18.2)	154,699	2,355,572 (19.9)	179,657	321,157 (11.3)	16,045
	60-69	2,004,640 (13.2)	66,193	1,876,023 (15.9)	85,212	144,857 (5.1)	4,629
	70-79	1,496,715 (9.8)	25,916	1,518,258 (12.8)	39,372	58,785 (2.1)	1,203
	80-89	582,058 (3.8)	8,685	625,214 (5.3)	10,195	19,557 (0.7)	396
	90+	81,621 (0.5)	2,352	110,766 (0.9)	1,812	4,739 (0.2)	73
Ethnicity	White	11,952,029 (78.6)	621,492	9,749,555 (82.5)	640,774	1,778,150 (62.4)	112,297
	Mixed	173,025 (1.1)	11,646	106,632 (0.9)	6,433	69,810 (2.4)	3,657
	South Asian	897,376 (5.9)	91,523	601,551 (5.1)	31,916	274,976 (9.6)	9,069
	Black	310,204 (2)	21,313	172,108 (1.5)	7,748	146,815 (5.1)	7,064
	Other	323,475 (2.1)	16,378	196,904 (1.7)	9,015	158,953 (5.6)	3,854
	Missing	1,555,362 (10.2)	64,722	995,890 (8.4)	54,484	422,479 (14.8)	11,103
Index of multiple deprivation quintile	1: Most deprived	2,881,855 (18.9)	199,514	1,854,399 (15.7)	112,517	851,173 (29.9)	44,116
	2	2,979,241 (19.6)	177,895	2,179,460 (18.4)	136,092	683,330 (24)	34,762
	3	3,278,622 (21.6)	165,299	2,616,291 (22.1)	161,666	570,594 (20)	29,246
	4	3,136,033 (20.6)	151,449	2,617,451 (22.1)	168,694	434,995 (15.3)	22,492
	5: Least deprived	2,935,720 (19.3)	132,917	2,555,039 (21.6)	171,401	311,091 (10.9)	16,428
Region	East	3,574,984 (23.5)	190,225	2,803,629 (23.7)	161,600	649,800 (22.8)	33,968
	East Midlands	2,684,304 (17.6)	157,426	2,084,802 (17.6)	138,836	470,137 (16.5)	28,240
	London	946,362 (6.2)	52,269	596,361 (5)	31,737	398,742 (14)	10,583
	North East	742,622 (4.9)	50,348	568,359 (4.8)	42,985	123,134 (4.3)	7,159
	North West	1,350,901 (8.9)	89,301	1,060,641 (9)	75,137	193,798 (6.8)	11,782
	South East	1,007,031 (6.6)	40,924	821,312 (6.9)	49,929	171,562 (6)	8,780
	South West	2,141,441 (14.1)	63,860	1,841,113 (15.6)	114,048	285,718 (10)	17,460

Characteristics	Pre-vaccination cohort (1 Jan 2020 to 14 Dec 2021)		Vaccinated cohort (1 June to 14 Dec 2021)		Unvaccinated cohort (1 June to 14 Dec 2021)	
	N (%)	COVID-19 diagnoses	N (%)	COVID-19 diagnoses	N (%)	COVID-19 diagnoses
West Midlands	576,420 (3.8)	42,691	399,516 (3.4)	24,722	148,724 (5.2)	7,417
Yorkshire/Humber	2,187,406 (14.4)	140,030	1,646,907 (13.9)	111,376	409,568 (14.4)	21,655
Smoking status						
Never smoker	7,105,315 (46.7)	423,039	5,635,383 (47.7)	372,057	1,174,015 (41.2)	59,053
Ever smoker	4,798,071 (31.5)	251,088	4,123,628 (34.9)	280,924	560,038 (19.6)	42,401
Current smoker	2,653,966 (17.4)	115,929	1,671,805 (14.1)	82,343	783,734 (27.5)	37,074
Missing	654,119 (4.3)	37,018	391,824 (3.3)	15,046	333,396 (11.7)	8,516
Care home resident	24,814 (0.2)	3,854	28,229 (0.2)	1,338	1,965 (0.1)	69

Table 2. Number of diabetes events in the pre-vaccination, vaccinated and unvaccinated cohorts, with person-years of follow-up, by COVID-19 severity. *Incidence rates are per 100,000 person-years.

Event	COVID-19 severity	Pre-vaccination cohort (N=15,211,471)		Vaccinated cohort (N=11,822,640)		Unvaccinated cohort (N=2,851,183)	
		Event/person-years	Incidence rate*	Event/person-years	Incidence rate*	Event/person-years	Incidence rate*
Type 2 diabetes	No COVID-19	127,607/28,752,729	444	31,601/5,330,475	593	1,960/990,420	198
	Hospitalised COVID-19	1,247/29,582	4,215	134/1,474	9,092	116/1,264	9,174
	Non-hospitalised COVID-19	4,469/828,838	539	731/138,303	529	64/21,827	293
Type 1 diabetes	No COVID-19	14,011/28,857,621	49	2,410/5,338,262	45	565/990,790	57
	Hospitalised COVID-19	54/30,993	174	†	NA	17/1,286	1,322
	Non-hospitalised COVID-19	852/833,322	102	†	NA	33/21,839	151
Gestational diabetes	No COVID-19	18,742/14,362,707	130	1,414/2,831,550	50	1,451/390,451	372
	Hospitalised COVID-19	29/13,632	213	†	NA	†	NA
	Non-hospitalised COVID-19	1,013/454,467	223	†	NA		
Other or non-specified diabetes	No COVID-19	12,630/28,860,799	44	4,862/5,337,688	91	402/990,792	41
	Hospitalised COVID-19	70/31,000	226	7/1,507	464	6/1,287	466
	Non-hospitalised COVID-19	510/833,882	61	108/138,543	78	17/21,844	78
Restricted to participants with at least 4 months follow-up							
Type 2 diabetes	No COVID-19	106,195/28,752,729	369				
	Hospitalised COVID-19	1,097/29,582	3,708				
	Non-hospitalised COVID-19	3,288/828,838	397				
Persistent type 2 diabetes	No COVID-19	78,543/28,752,729	273				
	Hospitalised COVID-19	754/29,582	2,549				
	Non-hospitalised COVID-19	2,432/828,838	293				

† insufficient events for estimation

Table 3. Adjusted hazard ratios (95% CI) comparing the incidence of diabetes events after COVID-19 with the incidence before or without COVID-19, in the pre-vaccination, vaccinated and unvaccinated cohorts, overall and according to COVID-19 severity. Hazard ratios are maximally adjusted unless otherwise stated.

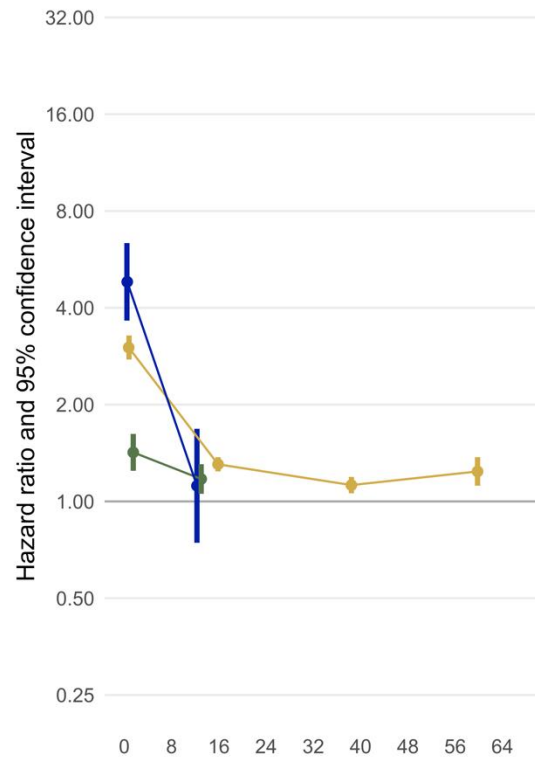
Outcome	Weeks since COVID-19	Pre-vaccination cohort	Vaccinated cohort	Unvaccinated cohort	
Type 2 diabetes		All	Persistent (treated for ≥4 months)		
All COVID-19, age/sex/region adjusted	1-4	4.25 (3.98-4.54)	3.47 (3.20-3.77)	1.44 (1.30-1.60)	8.99 (7.46-10.8)
	5-28	1.64 (1.57-1.71)	1.60 (1.52-1.67)	1.19 (1.09-1.29)	2.28 (1.74-2.98)
	29-52	1.45 (1.38-1.52)	1.52 (1.40-1.64)	-	-
	53-102	1.61 (1.48-1.76)	1.94 (1.63-2.31)	-	-
All COVID-19, maximally adjusted	1-4	3.01 (2.76-3.28)	2.10 (1.88-2.36)	1.42 (1.25-1.62)	4.82 (3.65-6.36)
	5-28	1.30 (1.24-1.37)	1.25 (1.17-1.33)	1.17 (1.06-1.31)	1.12 (0.74-1.68)
	29-52	1.12 (1.06-1.19)	1.10 (1.00-1.22)	-	-
	53-102	1.24 (1.12-1.37)	1.45 (1.18-1.78)	-	-
Hospitalised COVID-19, maximally adjusted	1-4	21.0 (18.7-23.6)	15.4 (13.2-18.0)	12.8 (9.89-16.6)	21.2 (15.2-29.5)
	5-28	2.38 (2.09-2.70)	2.24 (1.92-2.62)	2.64 (1.78-3.92)	3.72 (2.14-6.46)
	29-52	1.57 (1.33-1.85)	1.97 (1.56-2.49)	-	-
	53-102	2.04 (1.65-2.51)	2.12 (1.48-3.06)	-	-
Non-hospitalised COVID-19, maximally adjusted	1-4	1.45 (1.28-1.65)	1.02 (0.86-1.21)	1.07 (0.91-1.24)	1.69 (1.03-2.78)
	5-28	1.20 (1.14-1.27)	1.17 (1.09-1.25)	1.12 (1.00-1.25)	0.61 (0.34-1.12)
	29-52	1.08 (1.01-1.15)	1.01 (0.91-1.13)	-	-
	53-102	1.10 (0.98-1.23)	1.26 (0.98-1.62)	-	-
Type 1 diabetes					
All COVID-19, age/sex/region adjusted	1-4	2.20 (1.80-2.71)		1.49 (1.09-2.04)	4.48 (3.06-6.56)
	5-28	1.38 (1.25-1.53)		1.04 (0.79-1.37)	2.70 (1.75-4.15)
	29-52	1.33 (1.20-1.49)		-	-
	53-102	1.32 (1.05-1.67)		-	-
All COVID-19, maximally adjusted	1-4	1.73 (1.13-2.65)		1.34 (0.78-2.33)	2.59 (1.21-5.55)
	5-28	1.21 (0.99-1.48)		1.13 (0.73-1.73)	1.08 (0.44-2.67)
	29-52	1.34 (1.09-1.64)		-	-
	53-102	0.91 (0.55-1.50)		-	-
Hospitalised COVID-19, maximally adjusted	1-4	24.4 (12.3-48.5)		†	†
	5-28	2.84 (1.28-6.31)		†	†
	29-52	2.10 (0.79-5.62)		-	-
	53-102	†		-	-
Non-hospitalised COVID-19, maximally adjusted	1-4	1.18 (0.71-1.98)		1.24 (0.70-2.20)	†
	5-28	1.14 (0.93-1.41)		1.13 (0.74-1.74)	†
	29-52	1.29 (1.05-1.58)		-	-
	53-102	0.81 (0.48-1.38)		-	-
Gestational diabetes					
All COVID-19,	1-4	1.05 (0.82-1.34)		0.96 (0.64-1.45)	1.41 (0.97-2.05)
	5-28	1.12 (1.02-1.23)		0.82 (0.60-1.13)	2.05 (1.54-2.74)

Outcome	Weeks since COVID-19	Pre-vaccination cohort	Vaccinated cohort	Unvaccinated cohort
age/sex/region adjusted	29-52	1.26 (1.14-1.39)	-	-
	53-102	1.18 (0.95-1.46)	-	-
All COVID-19, maximally adjusted	1-4	1.07 (0.66-1.73)	0.56 (0.18-1.76)	0.67 (0.25-1.80)
	5-28	0.86 (0.69-1.07)	0.98 (0.54-1.80)	1.13 (0.58-2.21)
	29-52	0.99 (0.79-1.23)	-	-
	53-102	1.29 (0.87-1.93)	-	-
Hospitalised COVID-19, maximally adjusted	1-4	†	†	†
	5-28	†	†	†
	29-52	†	-	-
	53-102	†	-	-
Non-hospitalised COVID-19, maximally adjusted	1-4	1.02 (0.62-1.67)	0.53 (0.17-1.66)	0.71 (0.26-1.90)
	5-28	0.83 (0.67-1.04)	0.84 (0.44-1.57)	1.20 (0.61-2.35)
	29-52	0.98 (0.78-1.22)	-	-
	53-102	1.28 (0.85-1.93)	-	-
Other or non-specified diabetes				
All COVID-19, age/sex/region adjusted	1-4	1.77 (1.22-2.58)	0.97 (0.71-1.34)	†
	5-28	1.53 (1.33-1.76)	0.96 (0.77-1.21)	†
	29-52	1.44 (1.27-1.63)	-	-
	53-102	1.70 (1.39-2.09)	-	-
All COVID-19, maximally adjusted	1-4	1.42 (0.90-2.23)	0.89 (0.60-1.33)	†
	5-28	1.26 (1.06-1.49)	0.94 (0.72-1.22)	†
	29-52	1.15 (0.99-1.33)	-	-
	53-102	1.13 (0.88-1.46)	-	-
Hospitalised COVID-19, maximally adjusted	1-4	8.31 (5.26-13.10)	†	†
	5-28	1.80 (1.10-2.95)	†	†
	29-52	1.73 (1.13-2.62)	-	-
	53-102	0.81 (0.36-1.80)	-	-
Non-hospitalised COVID-19, maximally adjusted	1-4	1.20 (0.72-2.00)	0.88 (0.59-1.31)	†
	5-28	1.20 (1.00-1.44)	0.91 (0.70-1.20)	†
	29-52	1.08 (0.92-1.27)	-	-
	53-102	1.19 (0.91-1.55)	-	-

† Insufficient events for estimation

Figure 1. Maximally adjusted hazard ratios and 95% CIs comparing the incidence of type-2 diabetes events after COVID-19 (overall, hospitalised and non-hospitalised) with the incidence before or in the absence of COVID-19, in the pre-vaccination, vaccinated and unvaccinated cohorts. Points are plotted at the median time of the outcome event within each follow up period in each cohort.

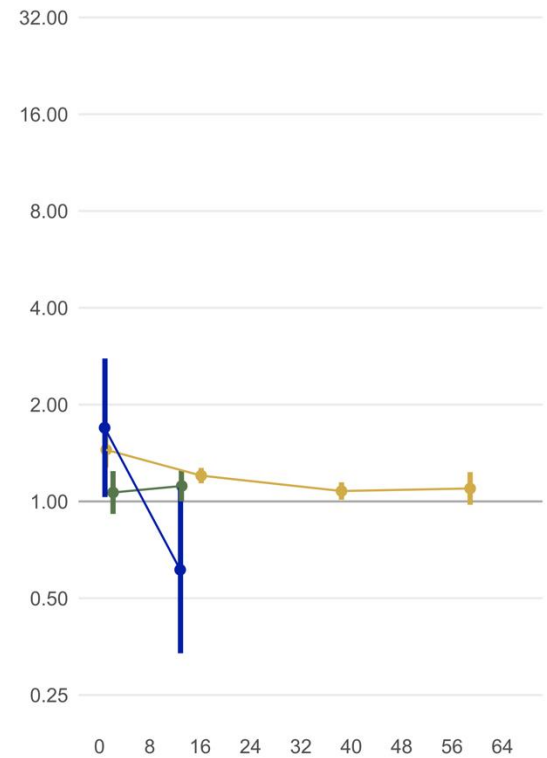
A: All COVID-19



B: Hospitalised COVID-19



C: Non-Hospitalised COVID-19

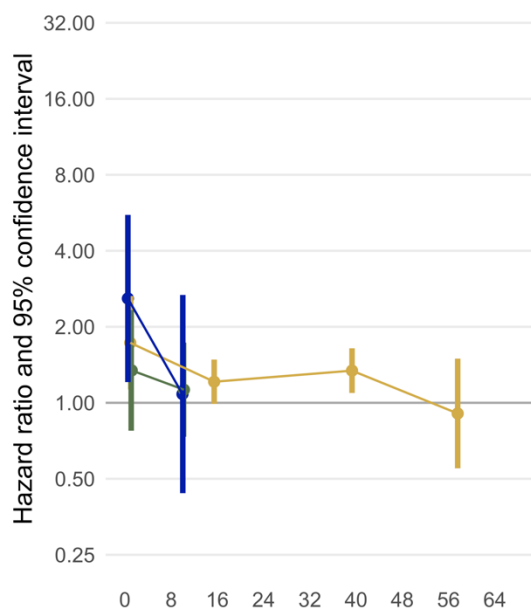


Weeks since COVID-19 diagnosis

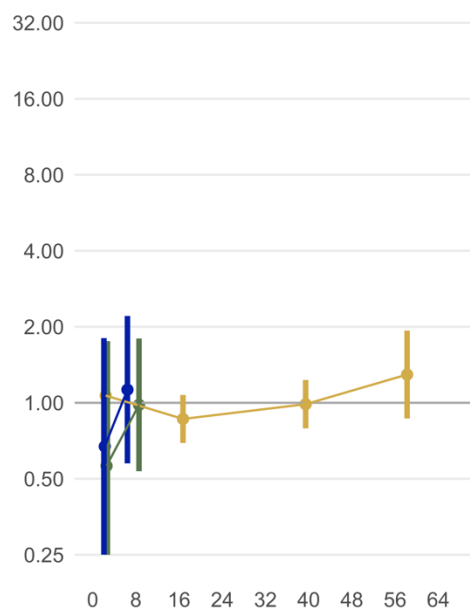
Cohort	Number of people	Total type 2 diabetes events	Events after COVID-19	After hospitalised COVID-19	After non-hospitalised COVID-19
Pre-vaccination (1 Jan 2020 to 14 Dec 2021)	15,211,471	133,332	5,722	1,247	4475
Vaccinated (1 Jun 2021 to 14 Dec 2021)	11,822,640	34,814	888	136	752
Unvaccinated (1 Jun 2021 to 14 Dec 2021)	2,851,183	2,286	188	122	66

Figure 2. Maximally adjusted hazard ratios and 95% CIs comparing the incidence of diabetes events after COVID-19 (overall, hospitalised and non-hospitalised) with the incidence before or in the absence of COVID-19, in the pre-vaccination, vaccinated and unvaccinated cohorts. Points are plotted at the median time of the outcome event within each follow up period in each cohort.

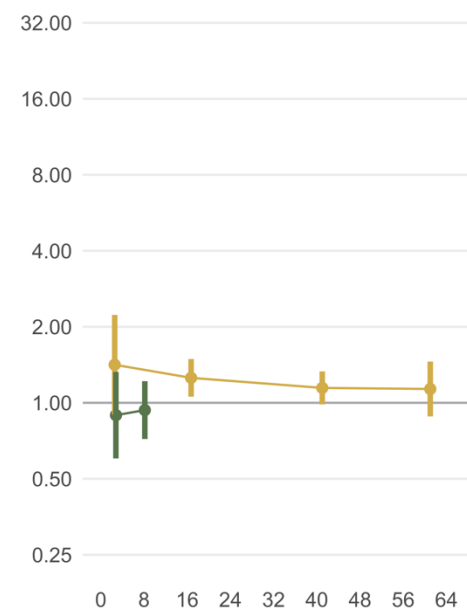
A: Type 1 Diabetes



B: Gestational Diabetes



C: Other or Non-Specified Diabetes



Weeks since COVID-19 diagnosis

Cohort	Number of people	Total Type 1 Diabetes events	Events after COVID-19	Total Gestational Diabetes events	Events after COVID-19	Total Other Diabetes events	Events after COVID-19
Pre-vaccination (1 Jan 2020 to 14 Dec 2021)	15,211,471	14,917	906	133,332	5,722	13,213	581
Vaccinated (1 Jun 2021 to 14 Dec 2021)	11,822,640	2,829	96	34,814	888	5,342	116
Unvaccinated (1 Jun 2021 to 14 Dec 2021)	2,851,183	713	53	2,286	188	449	23

1 **Contributors**

2 **Funding**

3 This study was supported by the COVID-19 Longitudinal Health and Wellbeing National Core Study,
4 funded by the UKRI Medical Research Council (MC_PC_20059); the COVID-19 Data and Connectivity
5 National Core Study, funded by the UKRI Medical Research Council; and by the CONVALESCENCE long
6 COVID study, funded by the UK National Institute for Health and Care Research (COVID-LT-009). This
7 work was also supported by Health Data Research UK, which is funded by the UK Medical Research
8 Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council,
9 Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government
10 Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh
11 Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. SVE is
12 funded by a Diabetes UK Sir George Alberti Research Training Fellowship (grant number 17/0005588).
13 GC and AWo are supported by the British Heart Foundation (RG/13/13/30194; RG/18/13/33946), BHF
14 Cambridge Centre of Research Excellence (RE/18/1/34212) and NIHR Cambridge Biomedical Research
15 Centre (BRC-1215-20014; NIHR203312). RK and VW are supported by the Medical Research Council
16 Integrative Epidemiology Unit at the University of Bristol [MC_UU_00011/1; MC_UU_00011/4]. RK, RD
17 and JACS are supported by the NIHR Bristol Biomedical Research Centre and by Health Data Research UK
18 South-West. YW was supported by an UKRI MRC Fellowship (MC/W021358/1) and received funding
19 from UKRI EPSRC Impact Acceleration Account (EP/X525789/1). SI and AWo were funded by a British
20 Heart Foundation–Turing Cardiovascular Data Science Award (BCDSA/100005). SI is funded by the
21 International Alliance for Cancer Early Detection, a partnership among Cancer Research UK
22 C18081/A31373, Canary Center at Stanford University, the University of Cambridge, OHSU Knight Cancer
23 Institute, University College London, and the University of Manchester. RK and JM were supported by

1 NIHR ARC West. RD and JACS were supported by Health Data Research UK. AWO is supported by the
2 Stroke Association (SA CV 20/100018). AWO is part of the BigData@Heart Consortium, funded by the
3 Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No 116074. The views
4 expressed are those of the authors and not necessarily those of the NIHR or the Department of Health
5 and Social Care.

6 **Authors contributions**

7 Author contributions are reported below in line with the Contributor Roles Taxonomy (CRediT).

8 Conceptualization: Nishi Chaturvedi, Rachel Denholm, Sophie Eastwood, Jonathan A C Sterne, Angela M
9 Wood, Venexia Walker

10 Methodology: Nishi Chaturvedi, Rachel Denholm, Sophie Eastwood, Jonathan A C Sterne, Angela M
11 Wood, Venexia Walker

12 Software: Louis Fisher, Jon Massey, Lisa Hopcroft, Tom Palmer, Simon Davy, Iain Dillingham, Catherine
13 Murton, Felix Greaves, Ben Goldacre

14 Validation: Kurt Taylor, Venexia Walker, Genevieve Cezard, Rochelle Knight, Marwa Al Arab, Yinghui
15 Wei, Elsie Horne, Lucy Teece, Jose Cuitun Coronado, Rachel Denholm

16 Investigation: Kurt Taylor, Venexia Walker, Genevieve Cezard, Rochelle Knight, Marwa Al Arab, Yinghui
17 Wei, Elsie Horne, Lucy Teece, Harriet Forbes, Jose Cuitun Coronado, Sam Ip, Rachel Denholm

18 Data curation: Kurt Taylor, Venexia Walker, Genevieve Cezard, Rochelle Knight, Marwa Al Arab, Yinghui
19 Wei, Elsie Horne, Lucy Teece, Harriet Forbes, Jose Cuitun Coronado, Rachel Denholm

20 Writing - Original Draft: Kurt Taylor, Sophie Eastwood, Nishi Chaturvedi, Jonathan A C Sterne, Rachel
21 Denholm

1 Writing - Review & Editing: Kurt Taylor, Sophie Eastwood, Venexia Walker, Genevieve Cezard, Rochelle
2 Knight, Marwa Al Arab, Yinghui Wei, Elsie Horne, Lucy Teece, Harriet Forbes, Alex Walker, Louis Fisher,
3 Jon Massey, Lisa Hopcroft, Tom Palmer, Jose Cuitun Coronado, Sam Ip, Simon Davy, Iain Dillingham,
4 Catherine Murton, Felix Greaves, John Macleod, Ben Goldacre, Angela Wood, Nishi Chaturvedi, Jonathan
5 A C Sterne, Rachel Denholm

6 Visualization: Kurt Taylor, Venexia Walker, Genevieve Cezard, Rochelle Knight, Marwa Al Arab, Yinghui
7 Wei, Elsie Horne, Lucy Teece, Jose Cuitun Coronado, Rachel Denholm

8 Project administration: Kurt Taylor, Jonathan A C Sterne, Venexia Walker, Rachel Denholm

9 Funding acquisition: Nishi Chaturvedi, Angela M Wood, Jonathan A C Sterne

10 **Declarations of interest**

11 NC is compensated by AstraZeneca for membership of Data Monitoring and Safety Committees for
12 clinical trials. The other authors report no conflicts.

13 **Data sharing**

14 Access to the underlying identifiable and potentially re-identifiable pseudonymised electronic health
15 record data is tightly governed by various legislative and regulatory frameworks, and restricted by best
16 practice. The data in OpenSAFELY is drawn from General Practice data across England where TPP is the
17 data processor.

18 TPP developers initiate an automated process to create pseudonymised records in the core OpenSAFELY
19 database, which are copies of key structured data tables in the identifiable records. These
20 pseudonymised records are linked onto key external data resources that have also been pseudonymised
21 via SHA-512 one-way hashing of NHS numbers using a shared salt. Bennett Institute for Applied Data

1 Science developers and PIs holding contracts with NHS England have access to the OpenSAFELY
2 pseudonymised data tables as needed to develop the OpenSAFELY tools.
3 These tools in turn enable researchers with OpenSAFELY data access agreements to write and execute
4 code for data management and data analysis without direct access to the underlying raw
5 pseudonymised patient data, and to review the outputs of this code. All code for the full data
6 management pipeline—from raw data to completed results for this analysis—and for the OpenSAFELY
7 platform as a whole is available for review at <https://github.com/OpenSAFELY>.

8 **Acknowledgements**

9 We are very grateful for all the support received from the TPP Technical Operations team throughout
10 this work, and for generous assistance from the information governance and database teams at NHS
11 England and the NHS England Transformation Directorate. We thank the CONVALESCENCE Study Long
12 COVID PPIE group for their input and for sharing their experiences and expertise throughout the
13 duration of the project.

14 **Information governance and ethical approval**

15 NHS England is the data controller for OpenSAFELY-TPP; TPP is the data processor; all study authors
16 using OpenSAFELY have the approval of NHS England. This implementation of OpenSAFELY is hosted
17 within the TPP environment which is accredited to the ISO 27001 information security standard and is
18 NHS IG Toolkit compliant ([https://digital.nhs.uk/data-and-information/looking-after-information/data-
19 security-and-information-governance/data-security-and-protection-toolkit](https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/data-security-and-protection-toolkit)).

20 Patient data are pseudonymised for analysis and linkage using industry standard cryptographic hashing
21 techniques; all pseudonymised datasets transmitted for linkage onto OpenSAFELY are encrypted; access
22 to the platform is via a virtual private network (VPN) connection, restricted to a small group of

1 researchers; the researchers hold contracts with NHS England and only access the platform to initiate
2 database queries and statistical models; all database activity is logged; only aggregate statistical outputs
3 leave the platform environment following best practice for anonymisation of results such as statistical
4 disclosure control for low cell counts (<https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/isb1523-anonymisation-standard-for-publishing-health-and-social-care-data>).

8 The OpenSAFELY research platform adheres to the obligations of the UK General Data Protection
9 Regulation (GDPR) and the Data Protection Act 2018. In March 2020, the Secretary of State for Health
10 and Social Care used powers under the UK Health Service (Control of Patient Information) Regulations
11 2002 (COPI) to require organisations to process confidential patient information for the purposes of
12 protecting public health, providing healthcare services to the public and monitoring and managing the
13 COVID-19 outbreak and incidents of exposure; this sets aside the requirement for patient consent
14 (<https://web.archive.org/web/20200421171727/https://www.gov.uk/government/publications/covid-19-notification-to-gps-and-nhs-england-to-share-information>). This was extended in November 2022 for
15 the NHS England OpenSAFELY COVID-19 research platform
16 (<https://www.gov.uk/government/publications/coronavirus-covid-19-notification-to-organisations-to-share-information/coronavirus-covid-19-notice-under-regulation-34-of-the-health-service-control-of-patient-information-regulations-2002>). In some cases of data sharing, the common law duty of
17 confidence is met using, for example, patient consent or support from the Health Research Authority
18 Confidentiality Advisory Group (<https://www.hra.nhs.uk/about-us/committees-and-services/confidentiality-advisory-group/>).

23 Taken together, these provide the legal bases to link patient datasets on the OpenSAFELY platform. GP
24 practices, from which the primary care data are obtained, are required to share relevant health

- 1 information to support the public health response to the pandemic, and have been informed of the
- 2 OpenSAFELY analytics platform.
- 3

1 References

- 2
- 3 1. Thompson EJ, Williams DM, Walker AJ, Mitchell RE, Niedzwiedz CL, Yang TC, et al. Long COVID
4 burden and risk factors in 10 UK longitudinal studies and electronic health records. *Nat Commun.*
5 2022 Jun 28;13(1):3528.
- 6 2. World Health Organization. WHO Coronavirus (COVID-19) Dashboard [Internet]. Available from:
7 <https://covid19.who.int>
- 8 3. Xie Y, Al-Aly Z. Risks and burdens of incident diabetes in long COVID: a cohort study. *The Lancet*
9 *Diabetes & Endocrinology.* 2022 May;10(5):311–21.
- 10 4. Bull-Otterson L, Baca S, Saydah S, Boehmer TK, Adjei S, Gray S, et al. Post-COVID Conditions Among
11 Adult COVID-19 Survivors Aged 18–64 and ≥65 Years — United States, March 2020–November 2021.
12 *MMWR Morb Mortal Wkly Rep.* 2022 May 27;71(21):713–7.
- 13 5. Rathmann W, Kuss O, Kostev K. Incidence of newly diagnosed diabetes after Covid-19. *Diabetologia.*
14 2022 Jun;65(6):949–54.
- 15 6. Rezel-Potts E, Douiri A, Sun X, Chowienczyk PJ, Shah AM, Gulliford MC. Cardiometabolic outcomes
16 up to 12 months after COVID-19 infection. A matched cohort study in the UK. Jia W, editor. *PLoS*
17 *Med.* 2022 Jul 19;19(7):e1004052.
- 18 7. Kompaniyets L, Bull-Otterson L, Boehmer TK, Baca S, Alvarez P, Hong K, et al. Post-COVID-19
19 Symptoms and Conditions Among Children and Adolescents - United States, March 1, 2020-January
20 31, 2022. *MMWR Morb Mortal Wkly Rep.* 2022 Aug 5;71(31):993–9.
- 21 8. Barrett CE, Koyama AK, Alvarez P, Chow W, Lundeen EA, Perrine CG, et al. Risk for Newly Diagnosed
22 Diabetes >30 Days After SARS-CoV-2 Infection Among Persons Aged <18 Years - United States,
23 March 1, 2020-June 28, 2021. *MMWR Morb Mortal Wkly Rep.* 2022 Jan 14;71(2):59–65.
- 24 9. Daugherty SE, Guo Y, Heath K, Dasmariñas MC, Jubilo KG, Samranvedhya J, et al. Risk of clinical
25 sequelae after the acute phase of SARS-CoV-2 infection: retrospective cohort study. *BMJ.* 2021 May
26 19;373:n1098.
- 27 10. Ayoubkhani D, Khunti K, Nafilyan V, Maddox T, Humberstone B, Diamond I, et al. Post-covid
28 syndrome in individuals admitted to hospital with covid-19: retrospective cohort study. *BMJ.* 2021
29 Mar 31;372:n693.
- 30 11. Kendall EK, Olaker VR, Kaelber DC, Xu R, Davis PB. Association of SARS-CoV-2 Infection With New-
31 Onset Type 1 Diabetes Among Pediatric Patients From 2020 to 2021. *JAMA Netw Open.* 2022 Sep
32 1;5(9):e2233014.
- 33 12. Birabaharan M, Kaelber DC, Pettus JH, Smith DM. Risk of new-onset type 2 diabetes in 600 055
34 people after COVID-19: A cohort study. *Diabetes Obes Metab.* 2022 Jun;24(6):1176–9.
- 35 13. Al-Aly Z, Bowe B, Xie Y. Long COVID after breakthrough SARS-CoV-2 infection. *Nat Med.* 2022
36 Jul;28(7):1461–7.

- 1 14. Choi JH, Kim KM, Song K, Seo GH. Risk for Newly Diagnosed Type 2 Diabetes Mellitus after COVID-19
2 among Korean Adults: A Nationwide Matched Cohort Study. *Endocrinol Metab.* 2023 Apr
3 30;38(2):245–52.
- 4 15. McKeigue PM, McGurnaghan S, Blackbourn L, Bath LE, McAllister DA, Caparrotta TM, et al. Relation
5 of Incident Type 1 Diabetes to Recent COVID-19 Infection: Cohort Study Using e-Health Record
6 Linkage in Scotland. *Diabetes Care.* 2022 Jul 26;dc220385.
- 7 16. Pietropaolo M, Hotez P, Giannoukakis N. Incidence of an Insulin-Requiring Hyperglycemic Syndrome
8 in SARS-CoV-2-Infected Young Individuals: Is It Type 1 Diabetes? *Diabetes.* 2022 Dec 1;71(12):2656–
9 63.
- 10 17. Noorzae R, Junker TG, Hviid AP, Wohlfahrt J, Olsen SF. Risk of Type 1 Diabetes in Children Is Not
11 Increased After SARS-CoV-2 Infection: A Nationwide Prospective Study in Denmark. *Diabetes Care.*
12 2023 Jun 1;46(6):1261–4.
- 13 18. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with
14 COVID-19-related death using OpenSAFELY. *Nature.* 2020 Aug;584(7821):430–6.
- 15 19. UKHSA. COVID-19: the green book, chapter 14a [Internet]. Available from:
16 <https://www.gov.uk/government/publications/covid-19-the-green-book-chapter-14a>
- 17 20. Horne EMF, Hulme WJ, Keogh RH, Palmer TM, Williamson EJ, Parker EPK, et al. Waning effectiveness
18 of BNT162b2 and ChAdOx1 covid-19 vaccines over six months since second dose: OpenSAFELY
19 cohort study using linked electronic health records. *BMJ.* 2022 Jul 20;e071249.
- 20 21. Eastwood SV, Mathur R, Atkinson M, Brophy S, Sudlow C, Flaig R, et al. Algorithms for the Capture
21 and Adjudication of Prevalent and Incident Diabetes in UK Biobank. Herder C, editor. *PLoS ONE*
22 [Internet]. 2016 Sep 15 [cited 2022 Jul 18];11(9):e0162388. Available from:
23 <https://dx.plos.org/10.1371/journal.pone.0162388>
- 24 22. Eastwood SV, Mathur R, Sattar N, Smeeth L, Bhaskaran K, Chaturvedi N. Ethnic differences in
25 guideline-indicated statin initiation for people with type 2 diabetes in UK primary care, 2006–2019:
26 A cohort study. Mauricio D, editor. *PLoS Med* [Internet]. 2021 Jun 29 [cited 2022 Jul
27 18];18(6):e1003672. Available from: <https://dx.plos.org/10.1371/journal.pmed.1003672>
- 28 23. UKHSA. Omicron daily overview: 17 December 2021 [Internet]. Available from:
29 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1042100/20211217_OS_Daily_Omicron_Overview.pdf
30
- 31 24. Wong R, Lam E, Bramante CT, Johnson SG, Reusch J, Wilkins KJ, et al. Does COVID-19 Infection
32 Increase the Risk of Diabetes? *Current Evidence.* *Curr Diab Rep.* 2023 Aug;23(8):207–16.
- 33 25. Cromer SJ, Colling C, Schatoff D, Leary M, Stamou MI, Selen DJ, et al. Newly diagnosed diabetes vs.
34 pre-existing diabetes upon admission for COVID-19: Associated factors, short-term outcomes, and
35 long-term glycaemic phenotypes. *J Diabetes Complications.* 2022 Apr;36(4):108145.
- 36 26. Al-Aly Z. Diabetes after SARS-CoV-2 infection. *Lancet Diabetes Endocrinol.* 2023 Jan;11(1):11–3.

- 1 27. D'Souza D, Empringham J, Pechlivanoglou P, Uleryk EM, Cohen E, Shulman R. Incidence of Diabetes
2 in Children and Adolescents During the COVID-19 Pandemic: A Systematic Review and Meta-
3 Analysis. *JAMA Netw Open*. 2023 Jun 30;6(6):e2321281.
- 4 28. UK Health Security Agency. Coronavirus (COVID-19) in the UK [Internet]. [cited 2023 Jul 1]. Available
5 from: <https://coronavirus.data.gov.uk/details/cases?areaType=nation&areaName=England>
- 6 29. Public Health England. Diabetes prevalence estimates for local populations [Internet]. 2023.
7 Available from: [https://www.gov.uk/government/publications/diabetes-prevalence-estimates-for-](https://www.gov.uk/government/publications/diabetes-prevalence-estimates-for-local-populations)
8 [local-populations](https://www.gov.uk/government/publications/diabetes-prevalence-estimates-for-local-populations)
- 9 30. Zaccardi F, Khunti K. Risk of Diabetes Following COVID-19: Translating Evidence Into Clinical and
10 Public Health Actions. *The Journal of Clinical Endocrinology & Metabolism*. 2022 Sep
11 28;107(10):e4248–9.
- 12 31. Lu JY, Wilson J, Hou W, Fleysher R, Herold BC, Herold KC, et al. Incidence of new-onset in-hospital
13 and persistent diabetes in COVID-19 patients: comparison with influenza. *eBioMedicine*. 2023
14 Apr;90:104487.
- 15 32. Holman N, Barron E, Young B, Gregg EW, Khunti K, Valabhji J, et al. Comparative Incidence of
16 Diabetes Following Hospital Admission for COVID-19 and Pneumonia: A Cohort Study. *Diabetes*
17 *Care*. 2023 May 1;46(5):938–43.

18