**PROTOCOL v0.7:**

**Effect of metformin on the risk of COVID-19 among people living with type 2 diabetes mellitus in England: study protocol for an observational study in OpenSAFELY-TPP**

**Short title:** *Is metformin protective against severe COVID and Long COVID?*

*Alain Amstutz1,2,3, Venexia Walker1, Paul Madley-Dowd1, Rachel Denholm1, Mengxuan Zou1, Will Hulme4,...*

*+ Others from OpenSAFELY*

*+ Others from (PROACT) EU-RESPONSE consortium*

*…, Jonathan Sterne1*

1 Electronic Health Records Group, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

2 Oslo Centre for Biostatistics & Epidemiology, Oslo University Hospital, Oslo, Norway

3 Division of Clinical Epidemiology, Department of Clinical Research, University Hospital Basel and University of Basel, Basel, Switzerland

4 Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

…

**Current Protocol Version Date: July 20th, 2025**

| *Version* | *Notes* |
| --- | --- |
| *0.1* | *Background, objectives & table 1* |
| *0.2* | *Variable definition tables* |
| *0.3* | *Updated protocol regarding design specifications* |
| *0.4* | *Changed causal question after initial causal question seemed unfeasible* |
| *0.5* | *More details re modified landmark approach, structured protocol according to TTE framework, and clarified treatment strategies* |
| *0.6* | *Clarified modified landmark approach; minor modifications to covariates* |
| *0.7* | *Structured protocol according to RECORD-PE, added sensitivity analyses, finalized eligibility criteria* |

# 

# **Background**

Metformin is the most commonly prescribed first-line treatment for Type 2 diabetes mellitus (T2DM) and is an inexpensive, readily accessible medication with a well-known safety profile. Metformin was explored as a promising drug repurposing candidate for the prevention and treatment of severe COVID-19 and Long COVID-19 due to its hypothesized anti-inflammatory[1–5](https://www.zotero.org/google-docs/?tM1Bgo) and even antiviral properties[5,6](https://www.zotero.org/google-docs/?VUq0rt).

Several randomised trials have evaluated the use of metformin during the acute phase of COVID-19 among a broad at-risk population, summarized in a recent meta-analysis[7](https://www.zotero.org/google-docs/?tvRNHX). The analysis concludes that metformin has little to no effect on mortality, some (but uncertain) effect on severe COVID-19, i.e. COVID-19-related admission to hospital, and a potential effect in reducing the risk of Long COVID.

This is in contrast to results from observational studies and meta-analyses thereof[8–11](https://www.zotero.org/google-docs/?un93b2). There are several possible explanations. First, people with T2DM were largely excluded from the randomized trials since randomising such patients to metformin or placebo is likely unfeasible and unethical. T2DM is an important risk factor for severe COVID-19 and appropriate blood glucose control with metformin (or other antidiabetic treatments) could itself impact the course of the disease[12,13](https://www.zotero.org/google-docs/?SiOo2t). Second, most patients in observational studies were admitted to hospital - while patients from the randomized trials were recruited in the primary care setting, presenting with non-severe COVID-19. Third, most observational studies used a prevalent-user design comparing people receiving (long-term) metformin. This design is addressing a causal question of “being on metformin vs not being on metformin”, which cannot be translated to clinical decision-making nor how a randomized trial investigating the effect of metformin on COVID outcomes would be designed. It also may introduce bias under certain circumstances[14](https://www.zotero.org/google-docs/?wefN70). Fourth, other design-related and confounding biases in the observational studies could be an explanation.

Long COVID, or post-acute sequelae of SARS-CoV-2 infection (PASC), is a health condition representing signs and symptoms that continue or develop after acute COVID-19. It may include both ongoing symptomatic COVID-19 (from 4 to 12 weeks) and post-COVID-19 syndrome (12 weeks or more). Common symptoms include fatigue, shortness of breath, and cognitive dysfunction, but are highly variable and wide ranging.[15–19](https://www.zotero.org/google-docs/?4cNXYa) Long COVID poses a significant public health burden with lasting physical health, mental health, and societal impacts - and at present without any established treatments.[20–22](https://www.zotero.org/google-docs/?uJQys5)

We aim to estimate the effect of starting metformin versus not starting metformin in the primary care setting, among adults living with T2DM in England, on COVID-19 related hospitalisations or death, COVID-19 diagnosis, and Long COVID diagnosis.

# **Objectives**

***Primary Objective***

1. To estimate the effect of starting metformin compared to not starting metformin (nor any other antidiabetic) on the risk of COVID-related hospitalisation or death, among adults with T2DM

***Secondary Objectives***

1. To estimate the effect of starting metformin compared to not starting metformin (nor any other antidiabetic) on the risk of COVID-19 diagnosis, among adults with T2DM
2. To estimate the effect of starting metformin compared to not starting metformin (nor any other antidiabetic) on the risk of Long COVID, among adults with T2DM
3. To estimate the treatment effect in relevant subgroups of age, sex, ethnicity, deprivation status, obesity and HbA1c

# 

# **Methods**

## Data Source

We will use primary care records managed by the General Practitioner (GP) software provider The Phoenix Partnership (TPP), linked to Office of National Statistics (ONS) death registration data (containing information on all deaths including those due to con-COVID-19 and those occurring in the community), the national coronavirus testing records from the Second Generation Surveillance System (SGSS), the national vaccine register (National Immunisation Management System [NIMS]), and the NHS Secondary Use Service (SUS) data, through OpenSAFELY, a data analytics platform created by our team on behalf of NHS England to address urgent COVID-19 research questions (<https://opensafely.org>). OpenSAFELY-TPP includes pseudonymised data for approximately 40% of the English population, including coded diagnoses, medications and physiological parameters.

## Study design and Population

### Study design choice rationale

We will use a population-based cohort design with a landmark approach to address our research question. We aim to estimate the effect of starting metformin versus not starting metformin on COVID-19 outcomes**.** When studying the effect of initiation of a treatment versus no treatment, we face the difficulty of comparing treatment strategies that are not distinguishable at time zero (baseline)[23](https://www.zotero.org/google-docs/?OCqpdp). To classify eligible participants into treatment groups, i.e. if they initiated or not, we need to look into their future up to a clinically sensible time point when most should have initiated. This may introduce immortal time bias. We use two elements to mitigate immortal time bias for our scenario. First, we apply a landmark approach[24](https://www.zotero.org/google-docs/?Gf3Rsv). In a landmark approach, a landmark time point is selected and anyone who was lost to follow-up or died prior to this time is excluded from further analysis. In our case this is 6 months after T2DM diagnosis. For the remaining, those who initiate treatment by the landmark time are classified as the treated, those who do not initiate by the landmark time are classified as the untreated. This avoids immortal time bias by aligning time zero with treatment assignment. Second, we move the treatment assignment window to a period where no outcomes of interest can occur (01.08.2018 - 01.08.2019), i.e. before the start of the COVID-19 pandemic. The rationale for setting the end of eligibility period to 01.08.2019 is to ensure the last possible 6-month landmark window finishes before the start of the pandemic in the UK (01.02.2020).

Our design accounts for the special circumstances of the pandemic - that the incidence of COVID-related outcomes was zero (or close to zero) before February 2020. Therefore, there were no COVID-related outcomes within 6 months of diagnoses of T2DM that occurred up to mid-2019. It follows that a ‘landmark’ analysis starting 6 months after diagnoses of T2DM between mid-2018 and mid-2019 will not exclude any COVID-related outcomes.

[Figure 1](#_90ntf0y1ofs) depicts a graphical illustration of the study design.

### Study population

From the raw data, we will select individuals using the following inclusion and exclusion criteria. All criteria are assessed at each individuals’ baseline date, which is their T2DM diagnosis.

**Inclusion Criteria:**

1. Aged ≥ 18 years & ≤ 85 years
2. Diagnosis of T2DM between 01.08.2018 to 01.08.2019
3. Registered for at least a year with a GP in England who uses TPP software

**Exclusion Criteria:**

1. Hospitalized, defined as having an admission to hospital record but no discharge record at baseline
2. In a care home, defined as having a record of long-stay nursing/care home at baseline
3. In palliative care, defined as having a record of palliative care within 6 months prior to baseline
4. Any use of metformin or another major glucose-lowering medication (SGLT2, DPP4, sulfonylurea, thiazolidinedione, GLP-1, meglitinides, alpha-glucosidase-inhibitors, insulin) prior to baseline
5. HbA1c above 75 mmol/mol, most recent, within 2 years prior to baseline
6. Known hypersensitivity / intolerance to metformin
7. Any history of moderate to severe renal impairment (eGFR of <30ml/min/1.73 m2; chronic kidney diseases stage 4/5)
8. Any history of advanced decompensated liver cirrhosis
9. Use of the following medications in the past 14 days: Cimetidine, hydroxychloroquine, tafenoquine, patiromer, ranolazine, Monoamine Oxide Inhibitors (Phenelzine, Tranylcypromine, Selegiline, Isocarboxazide, moclobemide), Alpha-1 antagonists, Sotalol, Clonidine, Phosphodiesterase 5 inhibitors, Methyldopa, Prazosin, Terasozin, Doxazosin)

Details about each eligibility variable, their data sources, lookback period, and codelists are available in [table 1](#_8xndjqjwl0r1). In addition, we apply the following data quality assurance and data completeness criteria:

**Data quality assurance criteria:**

1. Remove individuals whose year of birth is missing
2. Remove individuals whose year of birth is after their year of death
3. Remove individuals whose year of birth is after the study end date or who were older than 110 years when the UK National Health Services (NHS) was established (1838, since NHS was established in 1948)
4. Remove individuals whose date of death is on or before 01/01/1900 or after current date (Sys.Date())
5. Remove men whose records contain pregnancy codes
6. Remove men whose records contain hormone replacement therapy or combined contraceptive pill medication codes
7. Remove women whose records contain prostate cancer codes

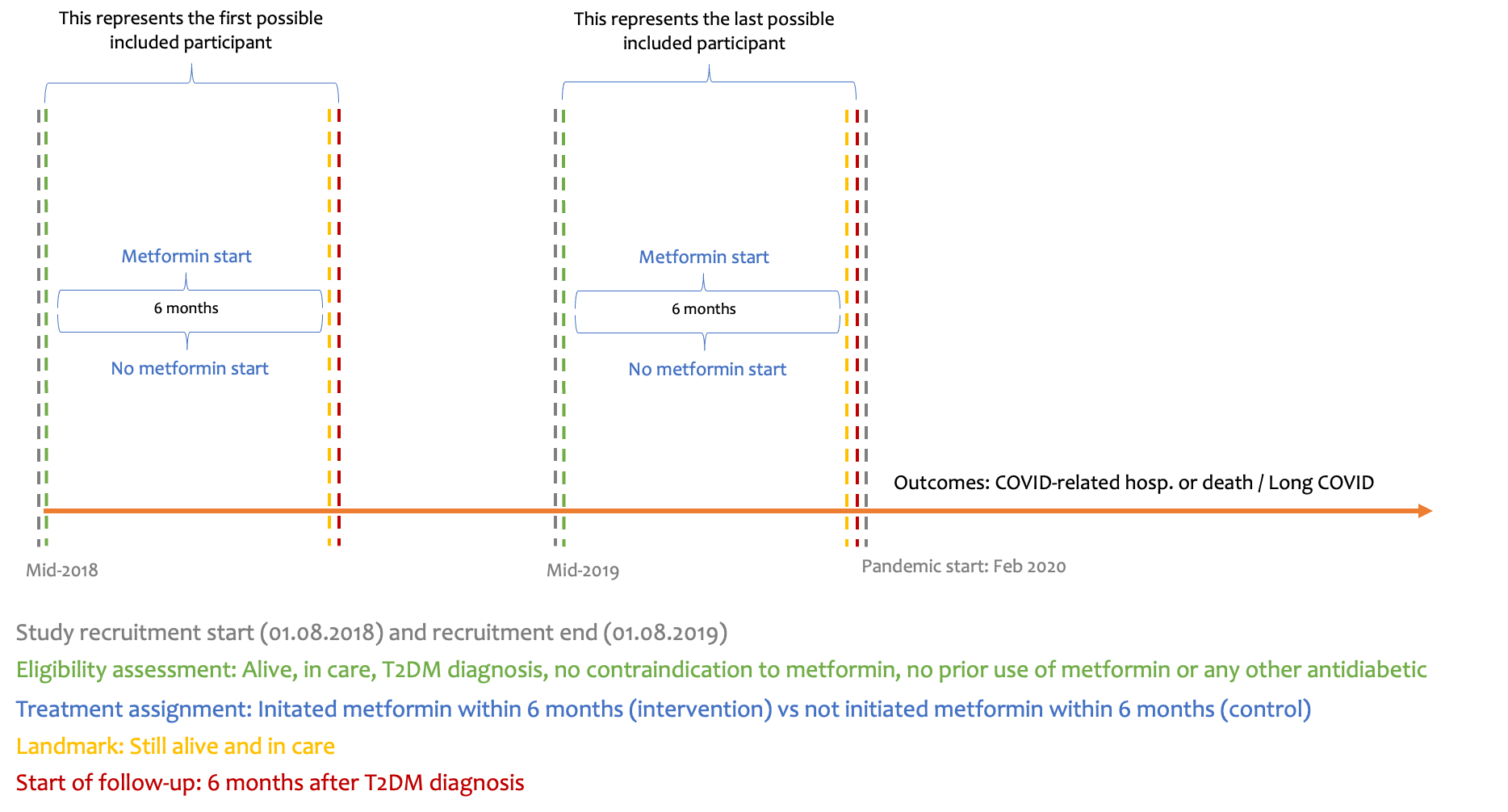
**Data completeness criteria:**

Patients will be included only if they meet all completeness criteria:

1. Known sex (female/male)
2. Known deprivation (Index of Multiple Deprivation [IMD]) at baseline date
3. Known sustainability and transformation partnership (STP) region, an NHS administrative geographical area, at baseline

Details regarding data quality and data completeness criteria variables are presented in [table 2](#_p1f3vb2jjigu).

### Figure 1: Illustration of the study design



## 

## Study measures

### Exposure

We will compare the treatment strategies of initiating metformin monotherapy within 6 months of T2DM diagnosis versus not initiating metformin within 6 months of T2DM diagnosis. Metformin monotherapy is the most common first-line T2DM treatment option across all cardiovascular risk groups[25](https://www.zotero.org/google-docs/?6mPrz0). We allow for any metformin titration regimen and maintenance dosage, and exclude participants initiating any other antidiabetic within 6 months of T2DM diagnosis.

Details regarding the exposure variable, including all antidiabetic codelists used, are presented in [table 3](#_j5gdramnobzx).

### Outcomes

The primary outcome is COVID-19-related hospitalisation or COVID-19-related death. COVID-19 related hospitalisations, obtained from secondary care SUS data (including emergency care attendance), are defined as any (emergency) hospitalisation listing a COVID-19 diagnosis in any position. Deaths are identified using linked ONS death registration data. COVID-19-related death is defined as a death where the underlying or contributory cause on the death certificate is COVID-19. If a person is hospitalised before death, the date of hospitalisation will be used.

The secondary outcome ‘COVID-19 diagnosis’ was defined using an established algorithm combining info from SGSS, HES-APC, SUS and ONS(ref): Any event indicating a COVID diagnosis, i.e., a positive SARS-CoV-2 test, a COVID code in primary care, a COVID diagnostic code in secondary care, or a COVID-related death.

The secondary outcome ‘Long COVID’ was defined using the Long COVID diagnosis codes in primary care,introduced by NHS, supplemented by viral fatigue codes.

For sensitivity analyses, we will use diabetes-related deaths as a positive control outcome and hospitalization due to bone fracture as a negative control outcome.

Details regarding all outcome variables, including their codelists used, are presented in [table 4](#_73wvl4vaqzej).

### Follow-up

Individuals will be followed from baseline (T2DM diagnosis) plus 6 months (landmark) until the earliest of: outcome, death, loss to follow-up, maximum follow-up period (2 years) or administrative end of study (01.03.2022; end of systematic SARS-CoV-2 mass testing in the UK).

### Covariates

The covariates, which were pre-specified as potentially important confounding variables are listed below, with further details in [table 5](#_pi9b0i6f4gmj), and their relationship depicted in [figure 2b](#_l7nwqplujsn). Unless otherwise specified, variables are created using diagnostic codes present ever in a patients’ medical record and are defined at baseline (T2DM diagnosis):

* Age
* Sex
* Ethnicity (White, Black or Black British, Asian or Asian British, Mixed, Other)
* Calendar time of T2DM diagnosis
* Index of Multiple Deprivation ([IMD], derived as quintiles, from the patient’s postcode at lower super output area level)
* Rural/urban area
* Health care worker occupation
* Primary care consultation rate in previous year
* Smoking status
* Obesity (BMI 30+ or obesity diagnosis)
* HbA1c in mmol/mol (<42; 42-58; 59-75; missing), most recent value in past 2 years
* Diabetes-related complications (diabetic foot, nephropathy, neuropathy, retinopathy)
* Total cholesterol/high-density lipoprotein cholesterol ratio (below 3.5:1, 3.5:1 to 5:1, above 5:1; missing), most recent value in past 2 years
* Prediabetes (clinical code or HbA1c in prediabetes range [42-47.9 mmol/mol])
* Comorbidities (acute myocardial infarction, stroke, other arterial embolism, venous thromboembolism, heart failure, angina, dementia, cancer, hypertension, depression, chronic obstructive pulmonary disease, liver disease, chronic kidney disease, polycystic ovary syndrome)

All comorbidities are identified through SNOMED CT codes in primary care records and ICD-10 in secondary care records. In case of convergence or disclosure risk, we may collapse categories, e.g. for ethnicity or smoking status.

### Missing data

Individuals with missing ethnicity, HbA1c, Total Chol/HDL ratio, rural/urban area, and smoking status are included with a missing indicator. Individuals without recorded codes for comorbidities, and health care worker status are assumed as not having such a diagnosis/status.

### Reporting

The study will be reported according to the REporting of studies Conducted using Observational Routinely collected health Data for PharmacoEpidemiological research (RECORD-PE) reporting guideline[26](https://www.zotero.org/google-docs/?5Ynb9y). The dataset and analysis code are developed and shared in [this](https://github.com/opensafely/metformin_covid) public GitHub repository.

## Statistical analysis

We will provide flowcharts showing the number of participants meeting each inclusion and exclusion criteria, a comprehensive baseline characteristics table for each treatment strategy group, and histograms of the propensity score for receiving metformin for each treatment strategy group for descriptive purposes.

We will estimate the observational analogue of the intention-to-treat effect, i.e., we will ignore the fact that some people may deviate from their strategy (stop metformin in intervention or start metformin in control) and include their outcomes. We will estimate 2-year hazard ratios for each outcome, using region-stratified Cox models, and standard Wald 95% confidence intervals as well as cumulative incidence (risk) curves, risk ratio and risk differences, derived from pooled logistic regression. We will present models adjusted for age and sex only, as well as fully adjusted models, including all of the confounders listed in the covariates list above. Non-COVID-related deaths will be treated as a censoring event for the outcome ‘COVID-related hospitalization or death’ and all-cause deaths will be treated as a censoring event for the outcome ‘Long COVID’ to account for the competing risk. We will also estimate the per-protocol effect of starting metformin monotherapy vs never starting any metformin on the primary outcome. We will artificially censor individuals in the control group deviating from their assigned treatment strategy (i.e. starting metformin) and use inverse probability of censoring weights to account for this informative censoring. We will use the same covariates as for the primary outcome model to estimate the weights.

### Subgroup analyses

We will estimate the treatment effect of the primary outcome in relevant subgroups:

* Age (below 60 vs 60 years and above)
* Sex (female vs male)
* Ethnicity (white versus non-white)
* IMD (most deprived vs other)
* Obesity (yes/no)
* HbA1c (below 59 mmol/mol vs 59 mmol/mol and above)

### Sensitivity analyses

To investigate residual confounding of our primary analysis model, we use diabetes-related deaths as a positive control outcome and hospitalization due to bone fracture as a negative control outcome (see section Outcomes above).

We will also quantify and characterize the individuals excluded from analyses due to the landmark design.

## Definition of all variables

Below we outline the details of all variables used in this study, including their codelists.

[Table 1](#_8xndjqjwl0r1):Eligibility variables

[Table 2](#_p1f3vb2jjigu): Data completeness and quality assurance variables

[Table 3](#_j5gdramnobzx): Exposure/treatment variables

[Table 4](#_73wvl4vaqzej): Outcome variables

[Table 5](#_pi9b0i6f4gmj): Covariates/Confounders

### Table 1. Eligibility variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Name** | **Type** | **Definition** | **Data sources *(***[***tpp schema***](https://docs.opensafely.org/ehrql/reference/schemas/tpp/#tpp-schema) ***in brackets)*** | **Notes** |
| New type 2 diabetes mellitus (T2DM) diagnosis | elig\_date\_t2dm | Date | Date | 1) Primary care: Diagnoses *(clinical\_events):*  - [diabetes\_type1\_ctv3\_clinical](https://www.opencodelists.org/codelist/user/hjforbes/type-1-diabetes/674fbd7a/)  - [diabetes\_type2\_ctv3\_clinical](https://www.opencodelists.org/codelist/user/hjforbes/type-2-diabetes/3530d710/)  - [diabetes\_other\_ctv3\_clinical](https://www.opencodelists.org/codelist/user/hjforbes/other-or-nonspecific-diabetes/0311f0a6/)  - [diabetes\_gestational\_ctv3\_clinical](https://www.opencodelists.org/codelist/user/hjforbes/gestational-diabetes/1ed423d1/)  - [diabetes\_diagnostic\_ctv3\_clinical](https://www.opencodelists.org/codelist/user/hjforbes/nondiagnostic-diabetes-codes/50f30a3b/)  - [hba1c\_new\_codes](https://www.opencodelists.org/codelist/user/alainamstutz/hba1c-bristol/7482c0c7/)  2) Primary care: Drugs *(medications):*  - [insulin\_dmd](https://www.opencodelists.org/codelist/opensafely/insulin-medication/2020-04-26/)  - [antidiabetic\_drugs\_snomed\_clinical](https://www.opencodelists.org/codelist/opensafely/antidiabetic-drugs/2020-07-16/)  - [non\_metformin\_dmd](https://www.opencodelists.org/codelist/user/r_denholm/non-metformin-antidiabetic-drugs_bristol/7207eb58/)  3) Secondary care: Diagnoses *(apcs):*  - [diabetes\_type1\_icd10](https://www.opencodelists.org/codelist/opensafely/type-1-diabetes-secondary-care/2020-09-27/)  - [diabetes\_type2\_icd10](https://www.opencodelists.org/codelist/user/r_denholm/type-2-diabetes-secondary-care-bristol/0b7f6cd4/)  - [diabetes\_gestational\_icd10](https://www.opencodelists.org/codelist/user/alainamstutz/gestational-diabetes-icd10-bristol/474e7a09/) | In ehrQL: search for **first code** before studyend\_date, apply diagnosis window later in R  See [Appendix 1](?tab=t.0#heading=h.2yudo3gkfjlx) regarding the algorithm to define Diabetes types.  **The final diabetes variable (elig\_date\_t2dm) is extracted through the adjudication algorithm later in R.** |
| Any metformin use, before baseline | elig\_date\_metfin\_first (= exp\_date\_metfin\_first) | Date | Date | Primary care (medications):  - [metformin\_dmd](https://www.opencodelists.org/codelist/user/john-tazare/metformin-dmd/48e43356/) | In ehrQL: search for **first code** before elig\_date\_t2dm |
| Any other antidiabetic use, before baseline | elig\_date\_\*\*\*\_first | Date | Date | Primary care (medications)\*\*\*:  - [sulfonylurea\_dmd](https://www.opencodelists.org/codelist/user/alex-walker/sulfonylureas-dmd/6e0ab9fd/)  - [dpp4\_dmd](https://www.opencodelists.org/codelist/user/alex-walker/dpp-4-inhibitors-dmd/57ae06be/)  - [tzd\_dmd](https://www.opencodelists.org/codelist/user/alainamstutz/thiazolidinedione-bristol-dmd/7de688dc/)  - [sglt2\_dmd](https://www.opencodelists.org/codelist/user/alex-walker/sglt-2-inhibitors-dmd/2029a068/)  - [glp1\_dmd](https://www.opencodelists.org/codelist/user/alainamstutz/glp1-bristol-dmd/7c628abb/)  - [meglitinides\_dmd](https://www.opencodelists.org/codelist/user/alainamstutz/meglitinides-bristol-dmd/79afcac5/)  - [agi\_dmd](https://www.opencodelists.org/codelist/user/alainamstutz/alpha-glucosidase-inhibitors-bristol-dmd/4790e45b/)  - [insulin\_dmd](https://www.opencodelists.org/codelist/opensafely/insulin-medication/2020-04-26/) | In ehrQL: search for **first code** before elig\_date\_t2dm |
| Known hypersensitivity and / or intolerance to metformin, before baseline | elig\_date\_metfin\_allergy\_first | Date | Date | Primary care (clinical\_events):  - [metformin\_allergy\_snomed\_clinical](https://www.opencodelists.org/codelist/user/alainamstutz/metformin-intolerance-bristol/5af86d52/) | In ehrQL: search for **first code** before elig\_date\_t2dm |
| Clinical history of moderate to severe renal impairment (eGFR of <30ml/min/1.73 m2; chronic stage 4/5), before baseline | elig\_date\_ckd\_45\_first | Date | Date | 1) Primary care (clinical\_events):  - [ckd\_snomed\_clinical\_45](https://www.opencodelists.org/codelist/nhsd-primary-care-domain-refsets/ckdatrisk1_cod/08a67d83/)  2) HES APC (apcs):   * ckd\_stage4\_icd10 -> ["N184"] * ckd\_stage5\_icd10 -> ["N185"] | In ehrQL: search for **first code** before elig\_date\_t2dm |
| Clinical history of advance decompensated liver cirrhosis, before baseline | elig\_date\_liver\_cirrhosis\_first | Date | Date | 1) Primary care (clinical\_events):   * [advanced\_decompensated\_cirrhosis\_snomed\_clinical](https://www.opencodelists.org/codelist/opensafely/condition-advanced-decompensated-cirrhosis-of-the-liver/071038cf/) * [ascitic\_drainage\_snomed\_c](https://www.opencodelists.org/codelist/opensafely/procedure-ascitic-drainage/39388836/)linical   2) HES APC (apcs):   * [advanced\_decompensated\_cirrhosis\_icd10](https://www.opencodelists.org/codelist/opensafely/condition-advanced-decompensated-cirrhosis-of-the-liver-and-associated-conditions-icd-10/00e40554/) | In ehrQL: search for **first code** before elig\_date\_t2dm |
| Use of the following medications in the last 14 days: Cimetidine, hydroxychloroquine, dolutegravir, patiromer, ranolazine, Monoamine Oxide Inhibitors (Phenelzine, Tranylcypromine, Selegiline, moclobemide), Sotalol, Clonidine, Methyldopa, Prazosin, Doxazosin, before baseline | elig\_date\_metfin\_interaction\_last | Date | Date | Primary care (medications):   * [metformin\_interaction\_](https://www.opencodelists.org/codelist/user/alainamstutz/metformin-drug-drug-interaction-bristol-dmd/76baa07d/)dmd | In ehrQL: search for **first code** before elig\_date\_t2dm |
| Hospitalized, within 2 days prior to baseline | elig\_bin\_hosp | Binary | T, F | HES APC (apcs):  - any admission date within 2 days prior to baseline | Baseline: elig\_date\_t2dm |
| Care home resident, at baseline | elig\_bin\_carehome\_status | Binary | T, F | 1) Primary care (clinical\_events):  - [carehome](https://www.opencodelists.org/codelist/primis-covid19-vacc-uptake/longres/v1/)  2) Primary care (addresses):  - care\_home\_is\_potential\_match  - care\_home\_requires\_nursing  - care\_home\_does\_not\_require\_nursing | Baseline: elig\_date\_t2dm |
| Palliative care, within 6 months prior to baseline | elig\_date\_palliative | Date | Date | 1) Primary care (clinical\_events):  - [palliative\_snomed](https://www.opencodelists.org/codelist/nhsd-primary-care-domain-refsets/palcare_cod/20241205/) | Baseline: elig\_date\_t2dm |

### Table 2. Data completeness and quality assurance variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Name** | **Type** | **Definition** | **Data sources (**[**tpp schema**](https://docs.opensafely.org/ehrql/reference/schemas/tpp/#tpp-schema) **in brackets)** | **Notes** |
| Known sex | qa\_bin\_is\_female\_or\_male | Binary | T, F | Primary care (patients) | Based on entire EHR |
| Known age between 18 and 110 inclusive | qa\_bin\_was\_adult | Binary | T, F | Primary care (patients) | Based elig\_date\_t2dm |
| Alive | qa\_bin\_was\_alive | Binary | T, F | Primary care (patients) | Based elig\_date\_t2dm |
| Known deprivation (IMD) | qa\_bin\_known\_imd | Binary | T, F | Primary care (practice\_registration) | Based elig\_date\_t2dm |
| Known region | *based on same variable as below* |  |  |  |  |
| Registered in an English GP with TPP for at least past 12 months | qa\_bin\_was\_registered | Binary | T, F | Primary care (practice\_registration) | 1 year = 365.25 days, taking into account leap years. Based elig\_date\_t2dm |
| Year of Birth | qa\_num\_birth\_year | Numeric | continuous (year only) | Primary care (patients) | Based on entire EHR |
| Date of Death | qa\_date\_of\_death | Date | Date | Death certificates (ons\_deaths) | Based on entire EHR |
| Pregnancy | qa\_bin\_pregnancy | Binary | T, F | Primary care (clinical\_events):  [pregnancy\_snomed\_clinical](https://www.opencodelists.org/codelist/user/RochelleKnight/pregnancy_and_birth_snomed/1d46bcb3/) | Based on entire EHR, |
| Combined oral contraceptive pill | qa\_bin\_cocp | Binary | T, F | Primary care (medications):  [cocp\_dmd](https://www.opencodelists.org/codelist/user/elsie_horne/cocp_dmd/1666a7a3/) | Based on entire EHR |
| Hormone replacement therapy | qa\_bin\_hrt | Binary | T, F | Primary care (medications):  [hrt\_dmd](https://www.opencodelists.org/codelist/user/elsie_horne/hrt_dmd/19196799/) | Based on entire EHR |
| Prostata cancer diagnosis | qa\_bin\_prostate\_cancer | Binary | T, F | 1) Primary care (clinical\_events):  [prostate\_cancer\_snomed\_clinical](https://www.opencodelists.org/codelist/user/RochelleKnight/prostate_cancer_snomed/0437497e/)  3) HES APC (apcs):  [prostate\_cancer\_icd10](https://www.opencodelists.org/codelist/user/RochelleKnight/prostate_cancer_icd10/6b27d648/)  3) Death certificates (ons\_deaths):  [prostate\_cancer\_icd10](https://www.opencodelists.org/codelist/user/RochelleKnight/prostate_cancer_icd10/6b27d648/) | Based on entire EHR  Re [ons\_deaths](https://docs.opensafely.org/ehrql/reference/schemas/tpp/#ons_deaths): Search in underlying\_cause\_of\_death as well as in all other 15 cause\_of\_death fields |

### 

### Table 3. Treatment/Exposure variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Name** | **Type** | **Definition** | **Data sources (**[**tpp schema**](https://docs.opensafely.org/ehrql/reference/schemas/tpp/#tpp-schema) **in brackets)** | **Notes** |
| Metformin prescription | exp\_bin\_metfin | Binary | T, F | Primary care (medications):  - [metformin\_mono\_dmd](https://www.opencodelists.org/codelist/user/alainamstutz/metformin-without-other-antidiabetic-bristol-dmd/40fca29a/#full-list) | First Metformin monotherapy prescription before studyend\_date. Define/assign treatment strategy in R. |
| No antidiabetic prescription | exp\_bin\_treat\_nothing | Binary | T, F | Primary care (medications), none of these:  - [metformin\_dmd](https://www.opencodelists.org/codelist/user/alex-walker/metformin-dmd/258f2054/)  - [sulfonylurea\_dmd](https://www.opencodelists.org/codelist/user/alex-walker/sulfonylureas-dmd/6e0ab9fd/)  - [dpp4\_dmd](https://www.opencodelists.org/codelist/user/alex-walker/dpp-4-inhibitors-dmd/57ae06be/)  - [tzd\_dmd](https://www.opencodelists.org/codelist/user/alainamstutz/thiazolidinedione-bristol-dmd/7de688dc/)  - [sglt2\_dmd](https://www.opencodelists.org/codelist/user/alex-walker/sglt-2-inhibitors-dmd/2029a068/)  - [glp1\_dmd](https://www.opencodelists.org/codelist/user/alainamstutz/glp1-bristol-dmd/7c628abb/)  - [meglitinides\_dmd](https://www.opencodelists.org/codelist/user/alainamstutz/meglitinides-bristol-dmd/79afcac5/)  - [agi\_dmd](https://www.opencodelists.org/codelist/user/alainamstutz/alpha-glucosidase-inhibitors-bristol-dmd/4790e45b/)  - [insulin\_dmd](https://www.opencodelists.org/codelist/opensafely/insulin-medication/2020-04-26/) | First antidiabetic prescription before studyend\_date. Define/assign treatment strategy in R. |

### 

### Table 4. Outcome and censoring variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Name** | **Type** | **Definition** | **Data sources *(***[***tpp schema***](https://docs.opensafely.org/ehrql/reference/schemas/tpp/#tpp-schema) ***in brackets*)** | **Notes** |
| First COVID-19 related hospitalization or death, after baseline | out\_date\_severecovid | Date | First/any covid-19 (emergency) hospitalization or covid-19 death | 1) HES APC (apcs): ["U071", "U072", "U109"]  2) Emergency care (emergency\_care\_attendances):  ["1240751000000100", "1325171000000109", "1325181000000106"]  3) Death registry from ONS (ons\_death):  ["U071", "U072", "U109"] | Including clinically diagnosed COVID-19.  Search in all 25 emergency diagnosis fields.  Search in underlying\_cause\_of\_death as well as in all other 15 cause\_of\_death fields |
| First COVID-19 event diagnosis, after baseline | out\_date\_covid | Date | First covid-19 event | 1) Primary care *(clinical\_events):*  -  [covid\_primary\_care\_code](https://www.opencodelists.org/codelist/opensafely/covid-identification-in-primary-care-probable-covid-clinical-code/2020-07-16/)  - [covid\_primary\_care\_positive\_test](https://www.opencodelists.org/codelist/opensafely/covid-identification-in-primary-care-probable-covid-positive-test/2020-07-16/)  - [covid\_primary\_care\_sequelae](https://www.opencodelists.org/codelist/opensafely/covid-identification-in-primary-care-probable-covid-sequelae/2020-07-16/)  2) SGSS *(sgss\_covid\_all\_tests)*  3) HES APC *(apcs):*  [Covid\_codes\_incl\_clin\_diag](https://codelists.opensafely.org/codelist/opensafely/covid-identification/2020-06-03/), but adapted to ["U071", "U072", "U109"] to add additionally: "Multisystem inflammatory syndrome associated with COVID-19, unspecified"  4) Emergency care (emergency\_care\_attendances):  covid\_emergency -> ["1240751000000100", "1325171000000109", "1325181000000106"] | SGSS: COVID-19 tests results from SGSS (the Second Generation Surveillance System)  HES APC: including clinically diagnosed covid-19.  Adaptation based on: <https://github.com/opensafely/comparative-booster-spring2023/blob/main/analysis/codelists.py> |
| First Long COVID diagnosis, after baseline | out\_date\_longvocid\_virfat | Binary; Date | First/any recorded Long COVID diagnosis | Primary care (clinical\_events):  [long\_covid\_diagnostic\_codes](https://www.opencodelists.org/codelist/opensafely/nice-managing-the-long-term-effects-of-covid-19/64f1ae69/) [long\_covid\_referral\_codes](https://www.opencodelists.org/codelist/opensafely/referral-and-signposting-for-long-covid/12d06dc0/) [long\_covid\_assessment\_codes](https://www.opencodelists.org/codelist/opensafely/assessment-instruments-and-outcome-measures-for-long-covid/79c0fa8a/)  [post\_viral\_fatigue\_codes](https://www.opencodelists.org/codelist/user/alex-walker/post-viral-syndrome/70708fd6/) | Including post viral fatigue codes, see current long covid definition/repo [here](https://github.com/opensafely/long-covid) |
| De-registration date, after baseline | cens\_date\_dereg | Date | Deregistration date from TPP | Primary care (practice\_registrations) |  |
| *Sensitivity analyses, negative & positive outcome control:* In ehrQL: search for **first code** after baseline, then apply outcome window in R | | | | | |
| Hospitalization due to fracture, after landmark/pandemic start | out\_date\_fracture | Date | First/any diagnosis, after pandemic start | 1) HES APC (apcs):  [fracture\_icd10](https://www.opencodelists.org/codelist/bristol/fractures/565037f8/) |  |
| Diabetes-related death, after landmark/pandemic start | out\_date\_dm\_death | Date | First/any diagnosis, after pandemic start | Death registry from ONS (ons\_death):  [diabetes\_type2\_icd10](https://www.opencodelists.org/codelist/user/r_denholm/type-2-diabetes-secondary-care-bristol/0b7f6cd4/) |  |

### 

### Table 5. Demographic variables, covariates and potential confounders

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Name** | **Type** | **Definition** | **Data sources** | **Notes** |
| ***‘Same-day/most recent’ baseline covariates*** | | | | | |
| Sex | cov\_cat\_sex | Categorical | Male, Female | Primary care (patients) | Any other code than M and F is excluded by design (see qa\_bin\_is\_female\_or\_male) |
| Age at baseline | cov\_num\_age  derived:  cov\_cat\_age | Continuous | "18-39", "40-59", "60-79", "80+" | Primary care (patients) | Will be modelled using splines  Baseline: elig\_date\_t2dm |
| Ethnicity | cov\_cat\_ethnicity | Categorical | White  Black  Asian  Mixed  Other  Unknown | Primary care (clinical\_events) and SUS:  [ethnicity\_snomed](https://www.opencodelists.org/codelist/opensafely/ethnicity-snomed-0removed/2e641f61/) | Additionally included ethnicity\_from\_sus; codelist for primary care (snomed), based on recent publication: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11234682/>  Based on entire EHR |
| Index of Multiple Deprivation, at baseline | cov\_cat\_deprivation\_5 | Categorical | 5 categories: 1 (most deprived) - 5 (least deprived) | Primary care (addresses) | Based on the Index of Multiple Deprivation score: <https://docs.opensafely.org/legacy/study-def-variables/#cohortextractor.patients.address_as_of>  Baseline: pandemicstart\_date for Target Trial 1 (SeqTrial) and elig\_date\_t2dm for Target Trial 2 (modified Landmark)  <https://data.cdrc.ac.uk/dataset/index-multiple-deprivation-imd>  The IMD combines multiple aspects of deprivation into an overall score. These aspects include:   * Household overcrowding * Homelessness * Housing affordability * Distance to key amenities like schools, supermarkets, and GP surgeries |
| Region, at baseline | strat\_cat\_region | Categorical | East  London  Midlands  North East and Yorkshire  North West  South East  South West | Primary care (practice\_registrations) | Based on the Sustainable and Transformation Partnerships (STP) names and codes: <https://geoportal.statistics.gov.uk/documents/bec635f6c83e4582bcf76ce02c2be840/about>  <https://docs.opensafely.org/legacy/study-def-variables/#cohortextractor.patients.registered_practice_as_of>  Baseline: elig\_date\_t2dm |
| Rural/urban classification, at baseline | cov\_cat\_rural\_urban | Categorical | Urban conurbation  Urban city or town  Rural town or village | Primary care (addresses) | Baseline: elig\_date\_t2dm |
| Smoking status, on or before (most recent) baseline | cov\_cat\_smoking\_status | Categorial | Smoker  Ever  Never  Unknown/Missing | Primary care (clinical\_events): A combination of [Smoking\_clear](https://www.opencodelists.org/codelist/opensafely/smoking-clear/2020-04-29/) & [Ever\_smoking](https://www.opencodelists.org/codelist/user/alainamstutz/ever-smoking-bristol/00731978/) | Baseline: elig\_date\_t2dm |
|  |  |  |  |  |  |
| Participant is a healthcare worker | cov\_bin\_healthcare\_worker | Binary | T, F | NHS England COVID-19 data store (see: <https://docs.opensafely.org/study-def-variables/#cohortextractor.patients.with_healthcare_worker_flag_on_covid_vaccine_record>) | Based on vaccination data, connected to the day of receiving a vaccination => at date in the future, but still a good proxy to use, since this status is long-term |
| Consultation rate, in year prior to baseline | cov\_num\_consrate | Numeric | Number of primary care contacts | Primary care | # QC: restrict to max 365 (average of one per day)  Baseline: elig\_date\_t2dm |
| BMI value, on or before (most recent) baseline (max. 2 years back) | cov\_cat\_bmi\_groups;  cov\_num\_bmi | Categorical;  Numeric | <18; 18-24; 25-29; 30+; missing  continuous | Primary care (clinical\_events) | Cov\_cat\_bmi\_groups: NA coded as missing indicator (“missing”)  Cov\_num\_bmi: biologically implausible BMI < 12 and > 70 replace with NA  Baseline: elig\_date\_t2dm |
| HbA1c value, on or before (most recent) baseline (max. 2 years back) | cov\_cat\_hba1c\_mmol\_mol  cov\_num\_hba1c\_mmol\_mol | Categorical  Numeric | <42; 42-58; 59-75; 75+; missing  continuous | Primary care (clinical\_events):  [hba1c\_snomed](https://www.opencodelists.org/codelist/opensafely/glycated-haemoglobin-hba1c-tests-numerical-value/5134e926/) | Implausible values set to missing:  < 0 or > 120  Categories based on  <https://www.southtees.nhs.uk/resources/the-hba1c-test/>  Baseline: elig\_date\_t2dm |
| Total cholesterol/high-density lipoprotein [HDL] cholesterol ratio [TC/HDL], on or before (most recent) baseline (max. 2 years back) | cov\_cat\_tc\_hdl\_ratio  cov\_num\_tc\_hdl\_ratio | Categorical  Numeric | "below 3.5:1"; "3.5:1 to 5:1"; "above 5:1"; missing  continuous | Primary care (clinical\_events)  [cholesterol\_snomed](https://www.opencodelists.org/codelist/opensafely/cholesterol-tests-numerical-value/7e3a22f3/)  [hdl\_cholesterol\_snomed](https://www.opencodelists.org/codelist/bristol/hdl-cholesterol/64775990/) | TC/HDL values are derived from the recorded total and HDL cholesterol values.  Implausible values set to missing:   1. total Cholesterol values < 1.75 or > 20 2. HDL values < 0.4 or > 5 3. ratio values < 1 or > 50   Reference values/limits: <https://doi.org/10.1093/ije/dyz099>  <https://www.urmc.rochester.edu/encyclopedia/content?ContentTypeID=167&ContentID=lipid_panel_hdl_ratio#:~:text=Most%20healthcare%20providers%20want%20the,1%20is%20considered%20very%20good.>  Baseline: elig\_date\_t2dm |
| ***‘Ever’ baseline covariates (any history of)*** | | | | | |
| Obesity, on or before baseline | cov\_bin\_obesity | Binary | T if BMI>=30 (most recent in the past 2 years) or any obesity code/diagnosis; F otherwise | 1) Primary care (clinical\_events):  - [bmi\_obesity\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/bmi_obesity_snomed/0764e9b4/)  2) HES APC (apcs):  - [bmi\_obesity\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/bmi_obesity_icd10/6e55767e/) | Baseline: elig\_date\_t2dm |
| Acute myocardial infarction, on or before baseline | cov\_bin\_ami | Binary | T, F | 1) Primary care (clinical\_events):  - [ami\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/ami_snomed/36d11028/)  2) HES APC (apcs):  - [ami\_prior\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/ami_prior_icd10/360a5c99/)  - [ami\_icd10](https://www.opencodelists.org/codelist/user/RochelleKnight/ami_icd10/3dea268d/) | Baseline: elig\_date\_t2dm |
| All stroke, on or before baseline | cov\_bin\_all\_stroke | Binary | T, F | 1) Primary care (clinical\_events):   * [stroke\_isch\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/stroke_isch_snomed/1cfae964/) * [stroke\_sah\_hs\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/stroke_sah_hs_snomed/6adc02f9/)   2) HES APC (apcs):   * [stroke\_isch\_icd10](https://www.opencodelists.org/codelist/user/RochelleKnight/stroke_isch_icd10/278d734e/) * [stroke\_sah\_hs\_icd10](https://www.opencodelists.org/codelist/user/RochelleKnight/stroke_sah_hs_icd10/0e7e0019/) | Baseline: elig\_date\_t2dm |
| Other arterial embolism, on or before baseline | cov\_bin\_other\_arterial\_embolism | Binary | T, F | 1) Primary care (clinical\_events):  - [other\_arterial\_embolism\_snomed\_clinical](https://www.opencodelists.org/codelist/user/tomsrenin/other_art_embol/3838d352/)  2) HES APC (apcs):  - [other\_arterial\_embolism\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/other_arterial_embolism_icd10/463adc5d/) | Baseline: elig\_date\_t2dm |
| Venous thromboembolism events, on or before baseline | cov\_bin\_vte | Binary | T, F | 1) Primary care *(clinical\_events):*  - [portal\_vein\_thrombosis\_snomed\_clinical](https://www.opencodelists.org/codelist/user/tomsrenin/pvt/51484687/),  [dvt\_dvt\_snomed\_clinical](https://www.opencodelists.org/codelist/user/tomsrenin/dvt_main/1c76a027/),  [dvt\_icvt\_snomed\_clinical](https://www.opencodelists.org/codelist/user/tomsrenin/dvt_icvt/03672cf2/),  [dvt\_pregnancy\_snomed\_clinical](https://www.opencodelists.org/codelist/user/tomsrenin/dvt-preg/6a57b9bc/),  [other\_dvt\_snomed\_clinical](https://www.opencodelists.org/codelist/user/tomsrenin/dvt-other/1f29601d/),  [pe\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/pe_snomed/6d8ec2ef/)  2) HES APC *(apcs):*  - [portal\_vein\_thrombosis\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/portal_vein_thrombosis_icd10/22606950/),  [dvt\_dvt\_icd10](https://www.opencodelists.org/codelist/user/RochelleKnight/dvt_dvt_icd10/24dab358/),  [dvt\_icvt\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/dvt_icvt_icd10/30a4dcad/),  [dvt\_pregnancy\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/dvt_pregnancy_icd10/6576830d/),  [other\_dvt\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/other_dvt_icd10/547f4fba/),  [icvt\_pregnancy\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/icvt_pregnancy_icd10/3b6fdc85/),  [pe\_icd10](https://www.opencodelists.org/codelist/user/RochelleKnight/pe_icd10/59ac59b8/) | Baseline: elig\_date\_t2dm |
| Heart failure, on or before baseline | cov\_bin\_hf | Binary | T, F | 1) Primary care (clinical\_events):  [hf\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/hf_snomed/33579ca3/)  2) HES APC (apcs):  [hf\_icd10](https://www.opencodelists.org/codelist/user/RochelleKnight/hf_icd10/72bbcced/) | Baseline: elig\_date\_t2dm |
| Angina, on or before baseline | cov\_bin\_angina | Binary | T, F | 1) Primary care (clinical\_events):  [angina\_snomed\_clinical](https://www.opencodelists.org/codelist/user/hjforbes/angina_snomed/52df16a2/)  2) HES APC (apcs):  [angina\_icd10](https://www.opencodelists.org/codelist/user/RochelleKnight/angina_icd10/756e8ce3/) | Baseline: elig\_date\_t2dm |
| Dementia, on or before baseline | cov\_bin\_dementia | Binary | T, F | 1) Primary care (clinical\_events):  [dementia\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/dementia_snomed/7bd3364c/)  [dementia\_vascular\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/dementia_vascular_snomed/0eb67607/)  2) HES APC (apcs):  [dementia\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/dementia_icd10/2df21cb7/)  [dementia\_vascular\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/dementia_vascular_icd10/27c5e93c/) | Baseline: elig\_date\_t2dm |
| Cancer, on or before baseline | cov\_bin\_cancer | Binary | T, F | 1) Primary care (clinical\_events):  [cancer\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/cancer_snomed/23271cdf/)  2) HES APC (apcs):  [cancer\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/cancer_icd10/55460349/) | Baseline: elig\_date\_t2dm |
| Hypertension, on or before baseline | cov\_bin\_hypertension | Binary | T, F | 1) Primary care (clinical\_events):  [hypertension\_snomed\_clinical](https://www.opencodelists.org/codelist/nhsd-primary-care-domain-refsets/hyp_cod/20210127/)  2) HES APC (apcs):  [hypertension\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/hypertension_icd10/1a48296e/) | Baseline: elig\_date\_t2dm |
| Depression, on or before baseline | cov\_bin\_depression | Binary | T, F | 1) Primary care (clinical\_events):  [depression\_snomed\_clinical](https://www.opencodelists.org/codelist/user/hjforbes/depression-symptoms-and-diagnoses/499814eb/)  2) HES APC (apcs):  [depression\_icd10](https://www.opencodelists.org/codelist/user/kurttaylor/depression_icd10/4dc56a05/) | Baseline: elig\_date\_t2dm. |
| Chronic obstructive pulmonary disease , on or before baseline | cov\_bin\_copd | Binary | T, F | 1) Primary care (clinical\_events):  [copd\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/copd_snomed/419c1000/)  2) HES APC (apcs):  [copd\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/copd_icd10/5aab8335/) | Baseline: elig\_date\_t2dm |
| Liver disease, on or before baseline | cov\_bin\_liver\_disease | Binary | T, F | 1) Primary care (clinical\_events):  [liver\_disease\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/liver_disease_snomed/5c978f9c/)  2) HES APC (apcs):  [liver\_disease\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/liver_disease_icd10/75a702d1/) | Baseline: elig\_date\_t2dm |
| Chronic kidney disease, on or before baseline | cov\_bin\_chronic\_kidney\_disease | Binary | T, F | 1) Primary care (clinical\_events):  [ckd\_snomed\_clinical](https://www.opencodelists.org/codelist/user/elsie_horne/ckd_snomed/25d9dcd5/)  2) HES APC (apcs):  [ckd\_icd10](https://www.opencodelists.org/codelist/user/elsie_horne/ckd_icd10/0cca69a0/) | Baseline: elig\_date\_t2dm |
| Polycystic Ovary Syndrome, on or before baseline | cov\_bin\_pcos | Binary | T, F | 1) Primary care (clinical\_events):  [pcos\_snomed\_clinical](https://www.opencodelists.org/codelist/user/alainamstutz/pcos-bristol/796d6073/)  2) HES APC (apcs):  [pcos\_icd10](https://www.opencodelists.org/codelist/user/alainamstutz/pcos-icd10-bristol/605ded3e/) | Baseline: elig\_date\_t2dm |
| Prediabetes, on or before baseline | cov\_bin\_prediabetes | Binary | T, F | Primary care (clinical\_events):  - [prediabetes\_snomed](https://www.opencodelists.org/codelist/opensafely/prediabetes-snomed/6bdbb7dd/) or  - [hba1c\_new\_codes](https://www.opencodelists.org/codelist/user/alainamstutz/hba1c-bristol/7482c0c7/), where((clinical\_events.numeric\_value>=42) & (clinical\_events.numeric\_value<=47.9)) | preDM HbA1c measure in period before index\_date in preDM range (mmol/mol): 42-47.9.  Baseline: elig\_date\_t2dm |
| Diabetes-related complications (Diab. Foot/Neuro/Nephro/Ret), on or before baseline | cov\_bin\_diabetescomp | Binary | T, F | 1) Primary care (clinical\_events):  [diabetescomp\_snomed\_clinical](https://www.opencodelists.org/codelist/user/alainamstutz/diabetes-complications-bristol/33a486c0/)  2) HES APC (apcs):  [diabetescomp\_icd10](https://www.opencodelists.org/codelist/user/alainamstutz/diabetes-complications-icd10-bristol/1a95138b/) | Baseline: elig\_date\_t2dm |

### 

### Figure 2a. Directed Acyclic Graph, all relationships

This DAG focuses on the primary outcome (severe COVID). We conduct region-stratified analyses, thus, region is not part of the DAG. It would have the same representation as “Rural/urban”.

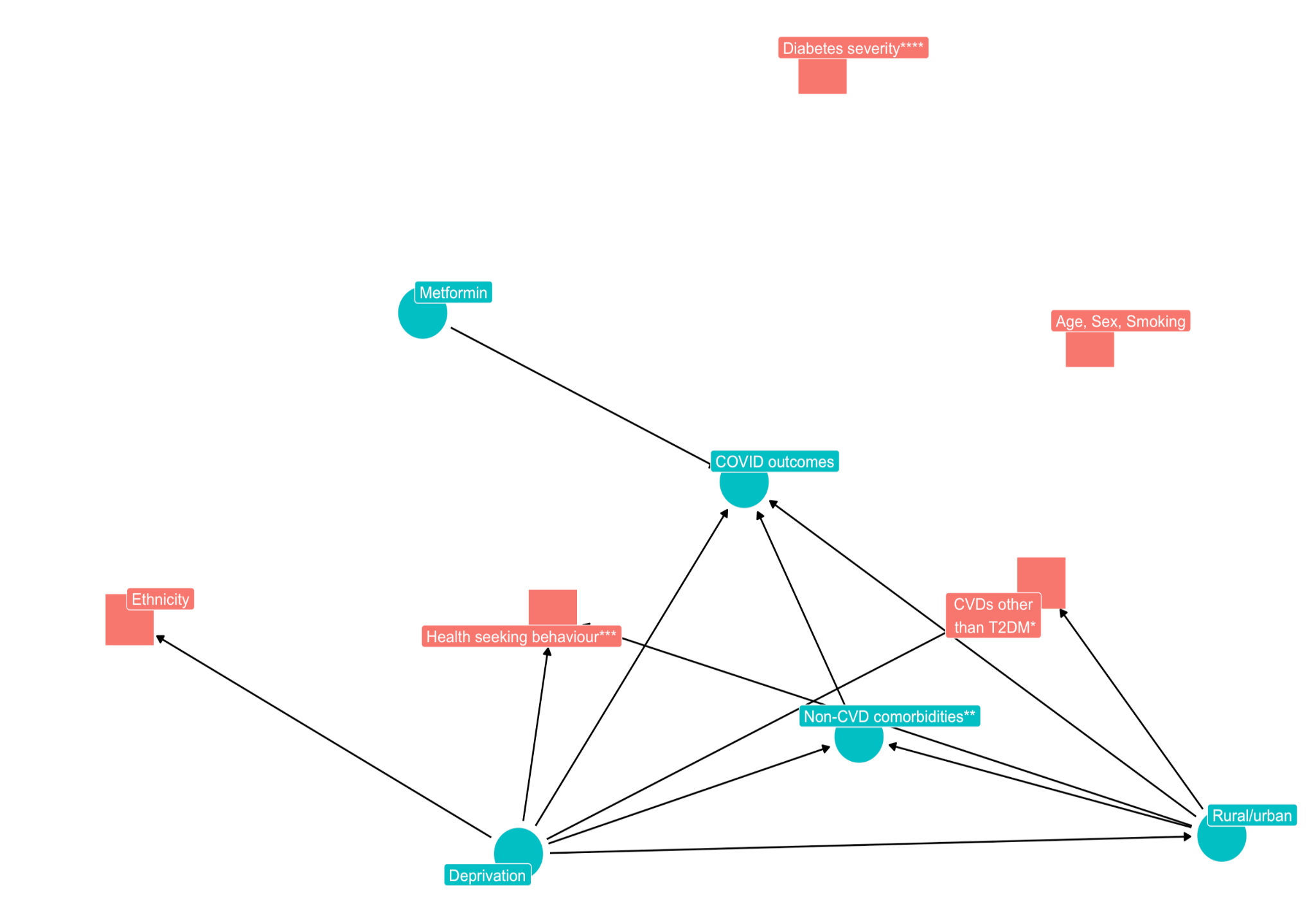
CVD = Cardiovascular disease; DAG created using [dagify()](https://github.com/opensafely/metformin_covid/blob/main/analysis/dag.R)

\* CVDs other than T2DM: Total Chol/HDL ratio, obesity, preDM, PCOS, stroke, angina, acute myocardial infarction, other arterial embolism, venous thromboembolism events, heart failure, hypertension

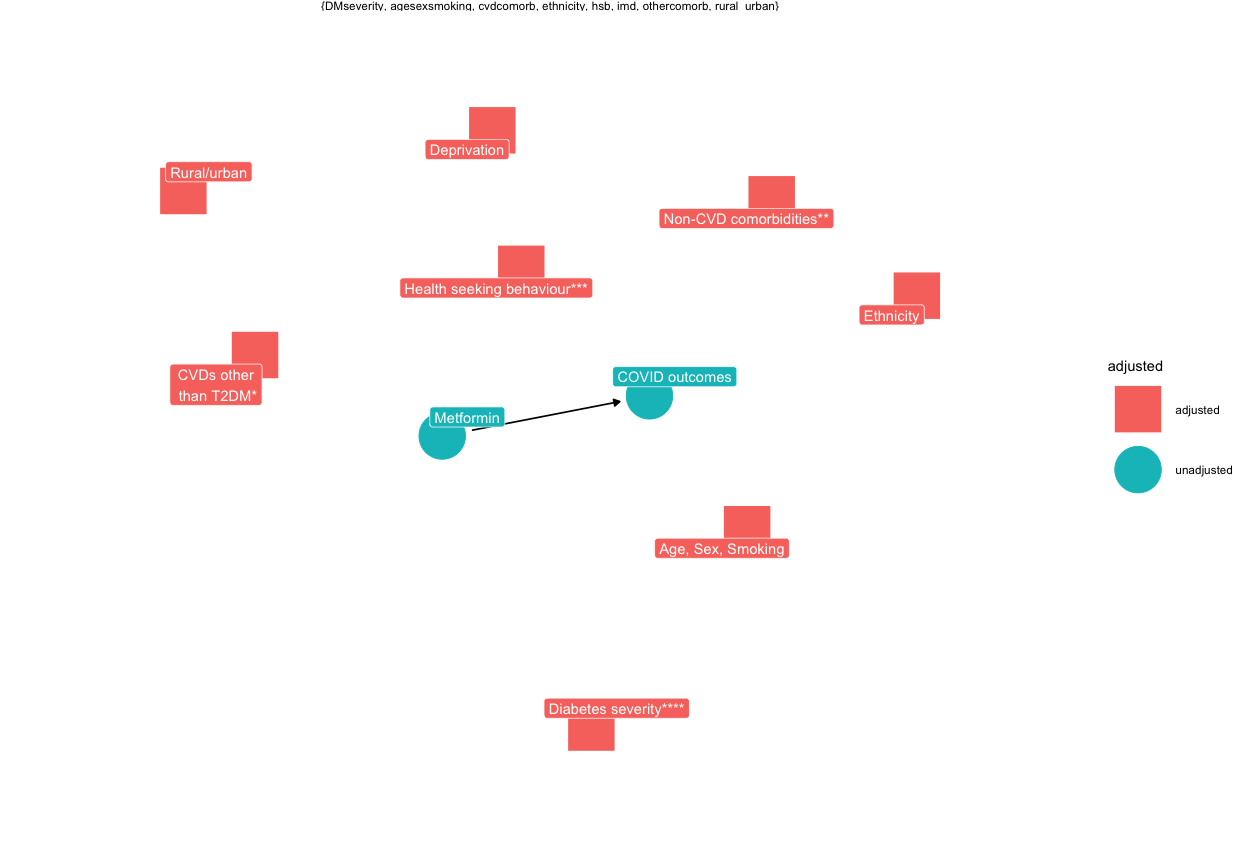
\*\* non-CVD comorbidities: dementia, cancer, chronic obstructive pulmonary disease, liver disease,

chronic kidney disease, depression

\*\*\* Health seeking behaviour; proxies used: Is a healthcare worker; number of primary care consultation rate in previous year

\*\*\*\* HbA1c levels and T2DM complications

### Figure 2b. Directed Acyclic Graph, minimal set of confounders (in red)



# **Limitations**

Our study has limitations:

1. **Landmark design:** The landmark design has two limitations. First, date of eligibility (baseline) and start of follow-up are misaligned. Depending on their clinical vulnerability/susceptibility, participants may have a different risk of reaching certain outcomes across the two groups, and therefore introduce selection bias at the landmark date. In our case, selection bias is not possible, because, in addition, we moved the time period of treatment allocation before any outcome of interest (i.e. COVID-related hospitalization or COVID-related death) can happen. However, potential selection bias due to depletion of susceptibles (e.g. regarding all-cause death) is still possible. We will quantify and characterize the individuals excluded from analyses within 6 months of T2DM. Second, estimates from this analysis have to be interpreted as conditional on still being alive at the time of reaching the 6-month landmark.
2. **Misclassification:** Possible misclassifications of cause of death on death certificates. Possible misclassification of exposure since we only have metformin prescription but not dispensing nor intake.
3. **Confounding by indication:** Despite carefully defining a suitable study population with an indication for metformin (new T2DM diagnosis with no contraindication to metformin, no extremely high HbA1c that violate positivity assumption; checking overlap of propensity scores), adjusting for important confounders, and sensitivity analysis with control outcomes, we will not be able to completely rule out residual confounding by indication.
4. **Changes over time:** Interaction with lockdown/COVID measures that affect risk of SARS-CoV-2 infection, vaccinations, and changing risk behaviours among people living with T2DM, affecting the two groups differently.
5. **Underreporting** of Long COVID codes in primary care
6. **Multiple testing** across subgroup analyses with low power

# 

# **Administrative**

## Funding

The OpenSAFELY platform is principally funded by grants from: NHS England [2023-2025]; The Wellcome Trust (222097/Z/20/Z) [2020-2024]; MRC (MR/V015737/1) [2020-2021]. Additional contributions to OpenSAFELY have been funded by grants from: MRC via the National Core Study programme, Longitudinal Health and Wellbeing strand (MC\_PC\_20030, MC\_PC\_20059) [2020-2022] and the Data and Connectivity strand (MC\_PC\_20058) [2021-2022]; NIHR and MRC via the CONVALESCENCE programme (COV-LT-0009, MC\_PC\_20051) [2021-2024]; NHS England via the Primary Care Medicines Analytics Unit [2021-2024]. This study is also supported by the European Union’s Horizon 2020 research and innovation programme.

## Patient and public involvement and engagement

OpenSAFELY: OpenSAFELY has involved patients and the public in various ways: we developed a public website that provides a detailed description of the platform in language suitable for a lay audience (https://opensafely.org); we have participated in two citizen juries exploring public trust in OpenSAFELY; we have co-developed an explainer video (https://www.opensafely.org/about/); we have patient representation who are experts by experience on our OpenSAFELY Oversight Board; we have partnered with Understanding Patient Data to produce lay explainers on the importance of large datasets for research; we have presented at various online public engagement events to key communities (e.g., Healthcare Excellence Through Technology; Faculty of Clinical Informatics annual conference; NHS Assembly; HDRUK symposium); and more. To ensure the patient voice is represented, we are working closely to decide on language choices with appropriate medical research charities (e.g., Association of Medical Research Charities). We will share information and interpretation of our findings through press releases, social media channels, and plain language summaries.

# 

# **References**

[1. Scheen AJ. Metformin and COVID-19: From cellular mechanisms to reduced mortality. Diabetes & Metabolism. 2020 Nov 1;46(6):423–6.](https://www.zotero.org/google-docs/?i6Wl3c)

[2. Sharma S, Ray A, Sadasivam B. Metformin in COVID-19: A possible role beyond diabetes. Diabetes Research and Clinical Practice [Internet]. 2020 Jun 1 [cited 2025 Jun 2];164. Available from: https://www.diabetesresearchclinicalpractice.com/article/S0168-8227(20)30433-2/fulltext](https://www.zotero.org/google-docs/?i6Wl3c)

[3. Samuel SM, Varghese E, Büsselberg D. Therapeutic Potential of Metformin in COVID-19: Reasoning for Its Protective Role. Trends in Microbiology. 2021 Oct 1;29(10):894–907.](https://www.zotero.org/google-docs/?i6Wl3c)

[4. Zhu Z, Zeng Q, Liu Q, Wen J, Chen G. Association of Glucose-Lowering Drugs With Outcomes in Patients With Diabetes Before Hospitalization for COVID-19: A Systematic Review and Network Meta-analysis. JAMA Network Open. 2022 Dec 6;5(12):e2244652.](https://www.zotero.org/google-docs/?i6Wl3c)

[5. Siedner MJ, Sax PE. Repurposing Revisited: Exploring the Role of Metformin for Treatment of COVID-19. Clinical Infectious Diseases. 2024 Aug 15;79(2):292–4.](https://www.zotero.org/google-docs/?i6Wl3c)

[6. Bramante CT, Beckman KB, Mehta T, Karger AB, Odde DJ, Tignanelli CJ, et al. Favorable Antiviral Effect of Metformin on SARS-CoV-2 Viral Load in a Randomized, Placebo-Controlled Clinical Trial of COVID-19. Clinical Infectious Diseases. 2024 Aug 15;79(2):354–63.](https://www.zotero.org/google-docs/?i6Wl3c)

[7. Chowdhury SR, Islam N, Zhou Q, Hasan MK, Chowdhury MR, Siemieniuk RA, et al. Metformin for covid-19: systematic review and meta-analysis of randomised controlled trials. bmjmed [Internet]. 2025 May 21 [cited 2025 Jun 2];4(1). Available from: https://bmjmedicine.bmj.com/content/4/1/e001126](https://www.zotero.org/google-docs/?i6Wl3c)

[8. Zhan K, Weng L, Qi L, Wang L, Lin H, Fang X, et al. Effect of Antidiabetic Therapy on Clinical Outcomes of COVID-19 Patients With Type 2 Diabetes: A Systematic Review and Meta-Analysis. Ann Pharmacother. 2023 Jul;57(7):776–86.](https://www.zotero.org/google-docs/?i6Wl3c)

[9. Petrelli F, Grappasonni I, Nguyen CTT, Tesauro M, Pantanetti P, Xhafa S, et al. Metformin and Covid-19: a systematic review of systematic reviews with meta-analysis. Acta Biomed. 2023 Aug 30;94(S3):e2023138.](https://www.zotero.org/google-docs/?i6Wl3c)

[10. Song ZH, Huang QM, Xu SS, Zhou JB, Zhang C. The Effect of Antihyperglycemic Medications on COVID-19: A Meta-analysis and Systematic Review from Observational Studies. Ther Innov Regul Sci. 2024 Jul;58(4):773–87.](https://www.zotero.org/google-docs/?i6Wl3c)

[11. Keels JN, McDonald IR, Lee CS, Dwyer AA. Antidiabetic agent use and clinical outcomes in patients with diabetes hospitalized for COVID-19: a systematic review and meta-analysis. Front Endocrinol (Lausanne). 2024;15:1482853.](https://www.zotero.org/google-docs/?i6Wl3c)

[12. Harding JL, Oviedo SA, Ali MK, Ofotokun I, Gander JC, Patel SA, et al. The bidirectional association between diabetes and long-COVID-19 - A systematic review. Diabetes Res Clin Pract. 2023 Jan;195:110202.](https://www.zotero.org/google-docs/?i6Wl3c)

[13. Su Y, Yuan D, Chen DG, Ng RH, Wang K, Choi J, et al. Multiple early factors anticipate post-acute COVID-19 sequelae. Cell. 2022 Mar 3;185(5):881-895.e20.](https://www.zotero.org/google-docs/?i6Wl3c)

[14. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. Nat Rev Rheumatol. 2015 Jul;11(7):437–41.](https://www.zotero.org/google-docs/?i6Wl3c)

[15. Xie J, López-Güell K, Dedman D, Duarte-Salles T, Kolde R, López-Blasco R, et al. Incidence of post-acute COVID-19 symptoms across healthcare settings in seven countries: an international retrospective cohort study using routinely-collected data. eClinicalMedicine [Internet]. 2024 Nov 1 [cited 2025 Mar 21];77. Available from: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(24)00482-6/fulltext](https://www.zotero.org/google-docs/?i6Wl3c)

[16. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. Sci Rep. 2021 Aug 9;11:16144.](https://www.zotero.org/google-docs/?i6Wl3c)

[17. Xie Y, Choi T, Al-Aly Z. Postacute Sequelae of SARS-CoV-2 Infection in the Pre-Delta, Delta, and Omicron Eras. New England Journal of Medicine. 2024 Aug 7;391(6):515–25.](https://www.zotero.org/google-docs/?i6Wl3c)

[18. Thaweethai T, Jolley SE, Karlson EW, Levitan EB, Levy B, McComsey GA, et al. Development of a Definition of Postacute Sequelae of SARS-CoV-2 Infection. JAMA. 2023 Jun 13;329(22):1934–46.](https://www.zotero.org/google-docs/?i6Wl3c)

[19. Gentilotti E, Górska A, Tami A, Gusinow R, Mirandola M, Rodríguez Baño J, et al. Clinical phenotypes and quality of life to define post-COVID-19 syndrome: a cluster analysis of the multinational, prospective ORCHESTRA cohort. EClinicalMedicine. 2023 Aug;62:102107.](https://www.zotero.org/google-docs/?i6Wl3c)

[20. Al-Aly Z, Agarwal A, Alwan N, Luyckx VA. Long COVID: long-term health outcomes and implications for policy and research. Nat Rev Nephrol. 2023;19(1):1–2.](https://www.zotero.org/google-docs/?i6Wl3c)

[21. Perlis RH, Trujillo KL, Safarpour A, Santillana M, Ognyanova K, Druckman J, et al. Research Letter: Association between long COVID symptoms and employment status. medRxiv. 2022 Nov 18;2022.11.17.22282452.](https://www.zotero.org/google-docs/?i6Wl3c)

[22. Cutler DM. The Costs of Long COVID. JAMA Health Forum. 2022 May 12;3(5):e221809.](https://www.zotero.org/google-docs/?i6Wl3c)

[23. Hernán MA, Sterne JAC, Higgins JPT, Shrier I, Hernández-Díaz S. A Structural Description of Biases That Generate Immortal Time. Epidemiology. 2025 Jan 1;36(1):107–14.](https://www.zotero.org/google-docs/?i6Wl3c)

[24. Morgan CJ. Landmark analysis: A primer. J Nucl Cardiol. 2019 Apr 1;26(2):391–3.](https://www.zotero.org/google-docs/?i6Wl3c)

[25. NICE. Type 2 diabetes in adults: choosing medicines. visual-summary-short-version-choosing-medicines-for-firstline-treatment-pdf-10956472094.pdf [Internet]. [cited 2025 Jun 25]. Available from: https://www.nice.org.uk/guidance/ng28/resources/visual-summary-short-version-choosing-medicines-for-firstline-treatment-pdf-10956472094](https://www.zotero.org/google-docs/?i6Wl3c)

[26. Langan SM, Schmidt SA, Wing K, Ehrenstein V, Nicholls SG, Filion KB, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ. 2018 Nov 14;363:k3532.](https://www.zotero.org/google-docs/?i6Wl3c)

[27. Hansford HJ, Cashin AG, Jones MD, Swanson SA, Islam N, Dahabreh IJ, et al. Development of the TrAnsparent ReportinG of observational studies Emulating a Target trial (TARGET) guideline. BMJ Open. 2023 Sep 12;13(9):e074626.](https://www.zotero.org/google-docs/?i6Wl3c)

[28. CKS is only available in the UK [Internet]. NICE. NICE; [cited 2025 Mar 21]. Available from: https://www.nice.org.uk/cks-uk-only](https://www.zotero.org/google-docs/?i6Wl3c)

[29. Thygesen JH, Tomlinson C, Hollings S, Mizani MA, Handy A, Akbari A, et al. COVID-19 trajectories among 57 million adults in England: a cohort study using electronic health records. The Lancet Digital Health. 2022 Jul 1;4(7):e542–57.](https://www.zotero.org/google-docs/?i6Wl3c)

[30. England NHS. NHS England » Commissioning guidance for post-COVID services for adults, children and young people [Internet]. [cited 2025 Mar 21]. Available from: https://www.england.nhs.uk/long-read/commissioning-guidance-for-post-covid-services-for-adults-children-and-young-people/](https://www.zotero.org/google-docs/?i6Wl3c)

[31. Henderson AD, Butler-Cole BF, Tazare J, Tomlinson LA, Marks M, Jit M, et al. Clinical coding of long COVID in primary care 2020-2023 in a cohort of 19 million adults: an OpenSAFELY analysis. EClinicalMedicine. 2024 Jun;72:102638.](https://www.zotero.org/google-docs/?i6Wl3c)

[32. COVID-19 rapid guideline: managing the long-term effects of COVID-19.](https://www.zotero.org/google-docs/?i6Wl3c)

[33. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV. A clinical case definition of post-COVID-19 condition by a Delphi consensus. The Lancet Infectious Diseases. 2022 Apr;22(4):e102–7.](https://www.zotero.org/google-docs/?i6Wl3c)

[34. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK - Office for National Statistics [Internet]. [cited 2025 Mar 21]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/30march2023](https://www.zotero.org/google-docs/?i6Wl3c)

[35. Harris V, Holmes J, Gbinigie-Thompson O, Rahman NM, Richards DB, Hayward G, et al. Health outcomes 3 months and 6 months after molnupiravir treatment for COVID-19 for people at higher risk in the community (PANORAMIC): a randomised controlled trial. The Lancet Infectious Diseases. 2025 Jan 1;25(1):68–79.](https://www.zotero.org/google-docs/?i6Wl3c)

[36. Taylor K, Eastwood S, Walker V, Cezard G, Knight R, Arab MA, et al. Diabetes following SARS-CoV-2 infection: Incidence, persistence, and implications of COVID-19 vaccination. A cohort study of fifteen million people [Internet]. medRxiv; 2023 [cited 2024 Jul 8]. p. 2023.08.07.23293778. Available from: https://www.medrxiv.org/content/10.1101/2023.08.07.23293778v1](https://www.zotero.org/google-docs/?i6Wl3c)

[37. Amstutz A. Introducing ‘diabetes-algo’ - A reusable action for OpenSAFELY users | Bennett Institute for Applied Data Science [Internet]. 2025 [cited 2025 Jun 1]. Available from: https://4407af29.bennettoxford.pages.dev/blog/2025/05/introducing-diabetes-algo-a-reusable-action-for-opensafely-users/](https://www.zotero.org/google-docs/?i6Wl3c)

[38. Eastwood SV, Mathur R, Atkinson M, Brophy S, Sudlow C, Flaig R, et al. Algorithms for the Capture and Adjudication of Prevalent and Incident Diabetes in UK Biobank. PLOS ONE. 2016 Sep 15;11(9):e0162388.](https://www.zotero.org/google-docs/?i6Wl3c)

# 

# 

# **Appendix Figure S1:** Diabetes type adjudication algorithm

We use clinician-verified Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) and International Classification of Disease Codes Version 10 (ICD-10) code lists for diabetes phenotypes and medications and extract these from primary care and hospital admission data. We subsequently apply a diabetes diagnostic adjudication algorithm developed for OpenSAFELY and previously published[36–38](https://www.zotero.org/google-docs/?VtHT5Z).

