

UK Biobank

COVID-19 vaccination record data

Version 1.1

<http://www.ukbiobank.ac.uk/>

November 2023



This document details the COVID-19 vaccination data received from national data sources for UK Biobank participants.

Contents

1. Understanding the dataset.....	3
1.1. Data providers	3
1.2. Data coverage	3
2. Available data	3
2.1. Overview.....	3
2.2. Data on the Data Portal	3
2.3. Data Portal table structure	4
3. Summary fields.....	4
4. Clinical classifications	4
4.1. SNOMED codes	4
5. Data quality.....	5
6. Data cleaning.....	5
7. Sample SQL	5

1. Understanding the dataset

This document provides information about the COVID-19 vaccination data available in UK Biobank. These data are restricted and can be used for COVID-19 research only; principal investigators must contact the Access Management Team (AMT) to request authorisation to access the data.

1.1. Data providers

Data on COVID-19 vaccinations administered in England are provided to UK Biobank by the [Data Access Request Service \(DARS\)](#), managed by [NHS England](#). The dataset is called [COVID-19 Vaccination Status](#).

COVID-19 vaccination data from Scotland and Wales are not currently available but are planned to be added in due course. They will be merged into the existing data structure described below.

1.2. Data coverage

Information about the date ranges covered in this dataset can be found on the [Data providers and dates](#) page in the Showcase. The sporadic nature of vaccination appointments leads to date clustering at specific times of year; therefore, UK Biobank's standard definition for censoring dates is not applicable to this dataset.

Vaccination data from each provider will be made available in a static release; no ongoing updates are planned at this time.

2. Available data

2.1. Overview

Only records for administered vaccinations are published; UK Biobank does not have data on vaccine refusal, appointment cancellations or similar categories. Only records related to COVID-19 vaccinations are present in the dataset.

2.2. Data on the Data Portal

After receiving authorisation to access the vaccination data (see Section 1), the data on the Data Portal can be obtained by adding [Data-Field 32040](#) to an application basket via the Access Management System (AMS). This gives access to the covid19_vaccination table.

Please see the Data Access Guide on the [Accessing UK Biobank data](#) page on Showcase for details of how to access the Data Portal.

2.3. Data Portal table structure

The COVID-19 vaccination data on the Data Portal are contained in a single table. Participants can have multiple records in the covid19_vaccination table – each record represents one administered vaccination. The table has the following columns:

column	description	encoding
eid	Encrypted UK Biobank participant ID	
source	Data source for this record	Data-Coding 1970
vacc_date	Date the vaccine was administered on	Data-Coding 819
product	Vaccine product administered to participant	
procedure	Vaccination procedure that took place	
procedure_uni	Procedure translated to a unified format across different providers, identifying whether the record indicates a primary course or booster procedure	Data-Coding 1960

3. Summary fields

[Data-Field 32040](#) gives a count of vaccination records per participant in the Data Portal table.

[Data-Field 32041](#) groups vaccination date by instance. For each participant, instance 0 represents their first administered vaccination, instance 1 represents their second, and so on. Note that, in the covid19_vaccination Data Portal table from which these data are derived, there are a small number of participants with multiple vaccination records on the same date (see [Section 5](#) below for more information). For this summary field, these records have been condensed into single instances per date and participant.

4. Clinical classifications

4.1. SNOMED codes

The English data use SNOMED codes for product and procedure fields.

NHS England hosts a [SNOMED CT browser](#), which researchers may find useful in understanding these fields.

5. Data quality

Data are published as closely as possible to how they are received and, in some cases, anomalies may be observed.

In particular, while exact duplicates are excluded, a small number of records appear to be duplicates with a minor variation, for example in the product or procedure code. UK Biobank is investigating whether certain records among these duplicates are updates or corrections but is not currently able to determine this. Researchers are advised to be aware of such anomalies and adapt their analyses accordingly.

6. Data cleaning

Other than checking validity and linkage, there has been no detailed data cleaning, and no validation against other sources such as self-reported data has been attempted.

NHSE provides UK Biobank with records that represent only administered vaccinations. However, a small number of records contain data that should be present for non-administered vaccinations only. These records are invalid, so they have been removed prior to publication.

A small number of vaccination dates have been recoded according to [Data-Coding 819](#). This replaces dates that are erroneous and/or sensitive (e.g. participant's date of birth) with consistent coded values.

7. Sample SQL

Tables on the Data Portal can either be downloaded in full, or queries can be run through the Portal to access specific subsets of data.

Some examples of extracting particular data using SQL queries on the Portal are given below:

(1) Search for participant eids with primary course vaccinations

```
select eid from covid19_vaccination
where procedure_uni = 'primary'
```

(2) As in (1) but searching specifically for first vaccinations in English data by using the appropriate SNOMED code

```
select eid from covid19_vaccination
where procedure_uni = 'primary'
and procedure = '1324681000000101'
and source = 'NHSE'
```