



# Summary of the CPRD-COVID-19 linked data to be made available for research

Version 1.3

Date: 28 May 2021

## Overview

CPRD-linked COVID-19 datasets will comprise:

1. Public Health England (PHE) Second Generation Surveillance System (SGSS) COVID-19 positive virology test data
2. PHE COVID-19 Hospitalisation in England Surveillance System (CHESS)
3. Intensive Care National Audit and Research Centre (ICNARC) data on COVID-19 intensive care admissions

We plan to release quarterly updates of these data, along with quarterly updates of the linked Hospital Episodes Statistics Admitted Patient Care, Office for National Statistics mortality data, and small area deprivation data to support COVID-19 research.

## 1. Second Generation Surveillance System (SGSS)

SGSS is the national laboratory reporting system used in England to capture routine laboratory data on infectious diseases and antimicrobial resistance. The SARS-CoV-2 testing started in UK laboratories on 24/02/2020, with the SGSS data reflecting testing (swab samples, PCR test method) offered to those in hospital and NHS key workers (i.e. Pillar 1). The CPRD-SGSS linked data contain positive tests results only.

The linked data comprise:

- Patid & pracid
- Specimen identifier
- Organism Species Name (all SARS-CoV-2)
- Specimen date
- Lab Report Date
- Reporting Lab ID

The first release of CPRD-SGSS linked data covers the period 01/03/2020 – 31/07/2020, and includes:

- 121,843 unique CPRD Aurum patients<sup>1</sup>
- 26,166 unique CPRD GOLD patients<sup>1</sup>

## 2. COVID-19 Hospitalisation in England Surveillance System (CHESS)

PHE established CHESS across all NHS Trusts in England on 15/03/2020 to collect epidemiological data on COVID-19 infection in persons requiring hospitalisation and ICU/HDU admission. Trends in hospital and critical care admission rates need to be interpreted in the context of testing recommendations, which changed over time.

The CPRD-CHESS linked data comprise >90 variables, including:

- Patid & pracid
- Swab/lab test date & result
- Reason & date for admission
- Complications & comorbidities
- # days patient mechanically ventilated
- Specific anti-COVID-19 treatment
- Obesity & BMI

- Work as a healthcare worker
- Outcome & date of outcome
- If died, cause of death

The first release of CPRD-CHESS linked data contains:

- Admissions to end July 2020
- 10,484 unique CPRD Aurum patients<sup>2</sup>
- 2,406 unique CPRD GOLD patients<sup>1</sup>

---

<sup>1</sup> Sept 2020 build

<sup>2</sup> Aug 2020 build



### 3. Intensive Care National Audit and Research Centre (ICNARC)

ICNARC is a national clinical audit covering all NHS adult, general intensive care and combined intensive care/high dependency units, and some additional specialist and non-NHS critical care units. Data on patients critically ill with confirmed COVID-19 admitted to critical care units will be linked to the CPRD data. Please note that these data have not yet been linked to CPRD data and will be available in coming months.

The CPRD-ICNARC linked data will comprise >70 variables, including:

- Age, sex
- Date of admission & discharge to hospital/unit
- BP & hypertension
- Height, weight & BMI
- Cardiovascular & respiratory support days
- Blood gas measurements, haemoglobin, platelet count, lactate
- Heart rate, respiratory rate
- Glasgow Coma Score
- APACHE II mortality prediction score & ICNARC physiology score
- Status & reason at discharge from hospital/unit

#### Data quality

The CPRD-SGSS and CPRD-CHESS linked data are being made available 'as is'; the CPRD has not been able to explore issues of data coverage, quality or completeness. We therefore urge our user community to share with us any issues or insights gained from using these data, which can be disseminated to other users via the documentation supporting the next quarterly update.

#### Availability of and access to linked data

As with our other standard linked patient-level datasets, linkage is available for patients at English practices who have consented to participate in the linkage scheme. Patients who have registered an opt-out will not contribute to the linked data. Access is dependent on approval of a study protocol, and all required linked data sources must be requested on the application form. These linked data are only provided by CPRD when part of a data extract is linked to CPRD primary care data.

Access to SGSS and CHESS will be priced at £4000 per study (or £7000 for studies requiring both datasets); ICNARC will be priced at £7000 per study. These costs are in addition to any other CPRD and linked data access/extraction fees. These datasets are not covered by existing licences, and data can only be released to organisations within the UK/EU/EEA.

The initial CPRD-SGSS and CPRD CHESS linked data form part of the set 20 linkages, and the source file describing patient eligibility (when finalised), will be available to Fob Holders on request from [enquiries@cprd.com](mailto:enquiries@cprd.com).

#### Further information

More information on the sources of data used in the coronavirus (COVID-19) surveillance systems to understand and monitor the pandemic in England can be found:

<https://www.gov.uk/government/publications/national-covid-19-surveillance-reports/sources-of-covid-19-systems>

Information about the UK COVID-19 data is available: <https://coronavirus.data.gov.uk/about-data>

COVID-19 testing data methodology note: <https://www.gov.uk/government/publications/coronavirus-covid-19-testing-data-methodology/covid-19-testing-data-methodology-note>