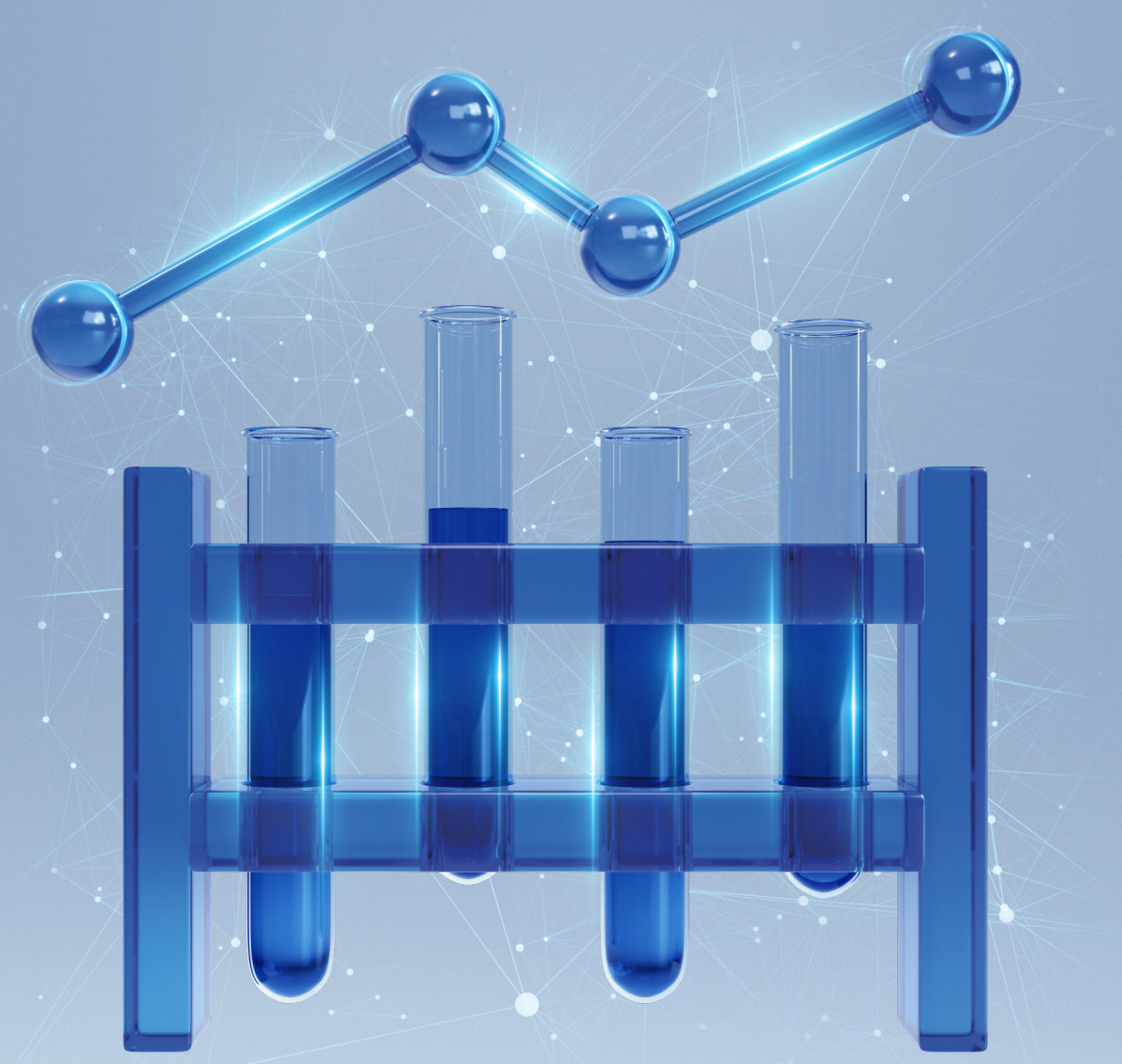
CATAPULT
Medicines Discovery



Data Access
Quick Guide to
CPRD

February 2019

CPRD (Clinical Practice Research Datalink) provides real-world retrospective and prospective research services based on UK population electronic medical records.

For official guidance visit the CPRD website.

About CPRD Data

Data held by CPRD encompasses more than 3 billion consultations on over 35 million patients lives, of which over 10 million are currently registered patients, and with at least 20 years of follow-up for 25% of patients. The data contains de-identified primary care data collected from a network of over 1200 GP practices across the UK with linkage to a range of other health related data. The primary care data are near real-time, longitudinal and representative of the UK population and can be used for observational research and to recruit patients to prospective clinical studies.

CPRD data are used by global regulators including the FDA, academic researchers and the life science industry for observational and interventional public health studies, with over 2,000 peer reviewed publications being published to date.

CPRD is jointly sponsored by the Medicines and Healthcare products Regulatory Agency (MHRA) and the National Institute for Health Research (NIHR).

Types of Data

Primary Care Data

The CPRD database contains coded and anonymised EHR data from a UK-wide network of over 1,200 primary care practices capturing information on:

- Demographic data
- Diagnoses and symptoms
- GP prescribed drugs and appliances
- Vaccination history
- Laboratory tests
- Referrals to hospital and specialist care

The data doesn't include any free text entered in primary care.

CPRD collects fully coded patient electronic health records from GP Practices using the Vision® or EMIS® software systems. These data are provided as two separate databases (CPRD GOLD and CPRD Aurum respectively) with metadata. CPRD can advise on which database may be best to select for a particular research question.

Specification documents for the two databases can be found at the CPRD website.

CPRD Linked Data

Linkage of CPRD primary care data with other patient level datasets is available for English practices who have consented to participate in the linkage scheme. Anonymised primary care patient data can be individually linked to a number of secondary care and other health and area-based datasets. Data linkage is undertaken by NHS Digital in England. CPRD is expanding its healthcare data and research services to increase both the cover of primary care data and the number of data sets that are linked and made available on a routine basis to the research community.

As of June 2018, CPRD GOLD linkage data include patients from 411 practices (56% of contributing CPRD GOLD practices in the UK, 10,553,586 patients). CPRD Aurum linkage data include patients from 232 practices (43% of contributing CPRD Aurum practices in England, 6,566,869 patients). Additional Aurum practices will be added to the scheme with each linkage dataset release. Please see CPRD linked data, for up-to-date linkage information.

Linked datasets include:

- Hospital Episode Statistics (HES) Admitted Patient Care (HES APC) data
- HES Outpatient (HES OP) data
- HES Accident and Emergency (HES A&E) data
- HES Diagnostic Imaging Dataset (HES DID)
- Death registration data from the Office for National Statistics (ONS)
- Cancer data from Public Health England (PHE):
    - Cancer Registration
    - Systemic Anti-Cancer Therapy (SACT) Dataset
    - National Radiotherapy Dataset (RTDS)
    - Cancer Patient Experience Survey (CPES)
- Mental Health Dataset (MHDS)
- Small area level data:
    - Index of Multiple Deprivation (IMD)
    - Townsend Deprivation Index
    - Carstairs Index
    - Rural-Urban Classification

Observational Research Services

Access to CPRD data

Access to CPRD data for observational research studies is subject to protocol approval by an Independent Scientific Advisory Committee (ISAC). ISAC protocol review is carried out to ensure that investigators using the databases for research have viable plans which do not raise governance concerns and do reach an acceptable scientific standard. Approval is required if access to anonymised patient-level data (primary care data, linked datasets such as HES) are being requested for research purposes. Some other uses of data are exempt from ISAC review.

Potential CPRD clients are requested to provide information (using a standard New client request for access to CPRD data form) to CPRD's Senior Management Team to determine the organisation's suitability to access CPRD data. Research organisations must demonstrate that they meet data governance and security requirements and have the relevant expertise to analyse the data. Organisations can only use the data for health research and CPRD is unable to license data to

organisations engaging in activities that may undermine patient and public trust in CPRD's data stewardship role.

All protocols must be submitted to the ISAC Secretariat using the Protocol Application Form, along with a brief curriculum vitae and conflict of interest statement for each researcher named on the application. Template files are available at the CPRD website.

Exemptions

Some uses of data, such as conducting analyses for submissions to regulatory bodies, are exempt from review, such as:

- Counts of patients receiving prescriptions for a certain substance
- Counts of prescriptions for a certain substance(s)
- Distribution of number of prescriptions per patient (mean, median, minimum, maximum, interquartile range, etc.)
- Distribution of prescription duration
- Distribution of patient time at risk
- Counts of patients with a particular clinical event

Pricing

Multi-Study annual Licence (MSL) to CPRD primary care data – data from CPRD GOLD and CPRD Aurum can be downloaded by the organisation for any number of individual ISAC-approved studies and for some internal studies that are exempt from ISAC review. Price includes cost for online access to CPRD GOLD and CPRD Aurum, and mandatory training for two designated users from the MSL holder organisation:

- Commercial Multi-study Licence Fee: from £150,000 - £330,000
- Non-Commercial Multi-study Licence Fee: £75,000
- Prices exclude VAT
- Additional costs will be incurred for data linkages

Patient level primary care dataset for individual studies - to access study-specific CPRD primary care datasets. Customers will be provided with the most suitable primary care data for their study which may be sourced from CPRD GOLD, CPRD Aurum or both databases:

- Commercial Dataset Licence Fee: from £30,000 - £60,000
- Non-Commercial Dataset Licence Fee: £15,000
- Prices exclude VAT

For more information on CPRD's criteria that determines the exact pricing level for your organisation you should contact CPRD directly.

Feasibility Counts and Feasibility Studies

CPRD can provide simple feasibility counts free-of-charge to clients so they can assess if CPRD data are suitable for their research needs. These are limited to counts of patients or events recorded in a specified period and including no more than three medical and/or prescribing definitions combined in a single request.

A feasibility study is a study where the intended purpose is to assess the feasibility of conducting a future study. This could be a future observational study using CPRD data, a prospective observational study involving enhanced data collection, or an interventional study. Sample sizes for feasibility studies are limited to 50,000 patients. A feasibility study could include percentages (rather than just counts) to assess the distribution/prevalence of key events, exposures, and outcomes in CPRD data. However, a feasibility study approved under this route should not include any hypothesis testing or tests of significance.

Feasibility studies using CPRD data are not subject to a full scientific protocol review by the ISAC, provided they meet the definition and scope of a feasibility study, and subject to the restrictions detailed in the Feasibility Study Application Guidance Notes. Approval for feasibility studies may be requested by submitting the Feasibility Study Application Form and any other required documentation to the ISAC Secretariat. Applicants wishing to submit a Feasibility Study must first contact enquiries@cprd.com to discuss the study prior to submitting an application.

Commissioned Research Services

CRPD have epidemiological and statistical experts who can offer commissioned research services including feasibility studies, aggregated data to support regulatory submissions, validation questionnaires and research studies.

Non-Standard Linkage Services

If a dataset is not already linked to CPRD primary care data as standard, CPRD can be commissioned to undertake a non-standard linkage. Researchers wishing to use the CPRD non-standard linkage service should submit a completed Non-Standard Linkage Service Application Form after reading the associated guidance. Timings and costs will vary depending on the details of the specific linkage. Overall costs will

depend on CPRD time for the linkage process, standard CPRD charges for data extraction and provision, and pass through costs from third-parties, for example the NHS Digital DARS application fee.

Interventional Research Services

CPRD offers research services to support EHR-enabled real world studies including study planning, patient recruitment, study execution and supplementary data collection to augment research, based on CPRD's primary care datasets and a recruitment pool of over 10 million patients registered at CPRD's network of GP practices across the UK.

CPRD can provide the following services:

1. Study Planning

CPRD can support a sponsor's trial feasibility and protocol optimisation by extracting data from patient EHRs using near real-time estimates to obtain real-world insights on patient characteristics and geographical location. Searches on the data based on required criteria can help define the patient population and result in optimisations that have a significant impact on patient recruitment, reducing the need for protocol amendments later in the study and help define the parameters required for the study endpoints.

2. Patient Recruitment

CPRD's study recruitment service uses data derived from primary care EHRs to locate eligible patients. The GP is subsequently provided with a pre-screened list of pseudonymised patients to identify and refer into secondary recruitment sites or directly consent patients into a clinical study in primary care.

3. Study Execution

This service uses an integrated platform to undertake site initiation, database configuration for study data capture, provide tailor-made dashboards and interfaces for use by GPs, patients, CPRD study managers and sponsors.

The output of this service may be an analysis ready dataset including any additional data collected in the study such as patient questionnaires or extend to study data analysis. Long term (5 years+) follow-up can be provided.

4. Supplementary Data Collection

This is a service that provides data augmenting research, based on CPRD's primary care datasets. Researchers may seek to supplement or validate findings from an observational research study using CPRD data (GP validation study) or patient questionnaires that may or may not become a component of a clinical trial.