CPRD Research Data Governance Feasibility Study Guidance
Completing the Feasibility Study Application Form

This guidance document has been produced to help applicants to complete a Feasibility Study application to access CPRD data. Applications not completed in accordance with this guidance will be returned as invalid.

General Information:

- CPRD feasibility study applications must be completed and submitted via CPRD's electronic research application portal (eRAP) (https://www.erap.cprd.com/)
- All feasibility study applications must be supported by a research team comprising of at least two members.
- During CPRD's scientific review process, all research seeking access to CPRD primary care and/or linked data for public health research are reviewed for study feasibility and research team expertise/experience, the public health benefits/risks of the research and potential information governance risks (risks to patient confidentiality and privacy).
- For the benefit of patients and the public, the following information on approved RDG applications are published on CPRD's website 3 months after approval - Study title, Date of approval, Study protocol number, Lay Summary, Technical Summary, Health outcomes to be measured, Collaborators and Linked data requested. For further information on the publication of Approved studies using CPRD Data, please go to https://www.cprd.com/approved-studies-using-cprd-data.
- Where research is undertaken under a CPRD multi-study licence (MSL), CPRD may make the following information on approved RDG applications available to the administrators of the annual licence, if requested – study protocol number, study title, data sources requested, and all collaborators listed on the application. This information will be used by the licence administrators for data monitoring purposes.

Key Information:

Feasibility studies using CPRD data are not subject to a full scientific review under the CPRD Research Data Governance (RDG) process, provided they meet the definition and scope of a feasibility study and are subject to the restrictions detailed below.

- A feasibility study is a study where the intended purpose is to assess the potential or likelihood of conducting a future study in a particular area of interest. This may include assessing the feasibility of a future observational study using CPRD data, or a prospective observational study involving enhanced data collection such as questionnaires or bio-samples, or an interventional study (e.g., a pragmatic trial). Applicants may also use CPRD data to assess the feasibility of a future study that does not involve using CPRD data or services.
- Unlike simple feasibility counts, a feasibility study may include percentages (rather than just counts) to assess the distribution/prevalence of key events, exposures, and outcomes in CPRD data.

- By submitting a feasibility study application, applicants are:
    1. Declaring that they have read and understood the guidance on completing a Feasibility Study Application Form.
    2. Confirming that the submitted feasibility study application, and any supporting documents, are accurate.
    3. Agreeing to abide by all contractual obligations in relation to access to CPRD data and any other linked data where applicable.
    4. Agreeing to publication of summary information, should the application be approved, in accordance with CPRD's Transparency Policy.
- The use of superscripts, subscripts, footnotes, special characters and symbols should be avoided unless necessary. The use of graphs, charts, graphics, and images is not permitted.
- Each section of the application form must be completed in full. It is not possible to submit applications with incomplete required sections.
- Direct communication between applicants and reviewers is not permitted. All communication regarding applications must take place via the RDG Secretariat (rdg@cprd.com), and only individuals named on the application may make enquiries.

Restrictions:

- Sample sizes for feasibility studies are limited to 50,000 patients. Where more than 50,000 patients meet the cohort definition criteria, a random sample of 50,000 will be provided to applicants. This restriction on sample size does not apply to multi-study licence holders or clients who are commissioning CPRD to undertake the feasibility study on their behalf.
- Work proposed under the feasibility study route should not include any hypothesis testing or tests of significance.
- Applicants can request data based on up to 3 of the following eligibility criteria:
    1. Cohort definition based on a single code list representing a diagnosis/prescription (code list to be provided by the applicant as an exported text file)
    2. Age-group (specified as an age range or threshold e.g., 40 years or younger at date of diagnosis/prescription or start of the study period etc.)
    3. Sex
- Feasibility studies are permitted using the following data sources listed below. Further Information on the range of linked data available via CPRD can be found here https://cprd.com/cprd-linked-data.
    - CPRD primary care data (CPRD GOLD and CPRD Aurum)
    - ONS Death Registration Data
    - HES Admitted Patient Care, Accident and Emergency, Outpatient, and/or Diagnostic Imaging Dataset
    - NCRAS Cancer Registration data (excluding SACT and Radiotherapy Data)
    - CPRD Mother Baby Links
    - CPRD Pregnancy Registers
    - Practice Level Index of Multiple Deprivation
    - Patient Level Index of Multiple Deprivation
    - CPRD Ethnicity Record
- For clients without a multi-study licence for a specific dataset, data will be released under a Standard Dataset Agreement, with non-negotiable Terms and Conditions for Feasibility Studies.
- Findings from approved feasibility studies can be shared with third parties and published in internal reports, peer-reviewed academic journals or included in conference presentations without the need for a full RDG protocol submission. However, findings should only be presented in the context of a feasibility study – for example, it would not be appropriate to present the findings in a publication as an incidence or prevalence study.

ALL FEASIBILITY STUDY APPLICATIONS MUST BE COMPLETED AND SUBMITTED VIA THE CPRD ELECTRONIC RESEARCH APPLICATION PORTAL (eRAP) www.erap.cprd.com

Part 1: Application Form

SECTION A: GENERAL INFORMATION ABOUT THE PROPOSED FEASIBILITY STUDY

Question 1: Study Title *[255 characters/max. 50 words]*

Please note that the study title of approved feasibility studies is published on the CPRD website for the benefit of patients and the public, to inform them of how anonymised patient data collected by CPRD are used for research.

Reviewer Assessment Criteria

In this section, reviewers will assess whether the study title clearly describes the focus and purpose of the research as outlined in the application.

Application Requirements

Applicants **must** include the words **"feasibility study"** in the study title. e.g. Feasibility study to evaluate the use of CPRD data for cost effectiveness analysis of condition X.

The study title must provide a concise overview of the aim/s of the feasibility study.

Avoid including acronyms in the title, defining these at first use.

The study title must be no longer than 255 characters, including spaces. Only letters a to z, numbers 0 to 9, spaces and the following characters are allowed: "%&()-_+=:;',.?

Question 2: Chief Investigator

Please note that the name and affiliation of the Chief Investigator on approved feasibility studies is published on the CPRD website for the benefit of patients and the public, to inform them of how CPRD's anonymised patient data are used for research.

The Chief Investigator will take responsibility for ensuring that the research is undertaken with full adherence to CPRD RDG guidelines, and any contractual terms and conditions. As such, students and junior researchers are not eligible to act as the Chief Investigator on CPRD feasibility studies.

The full name, job title, organisation name, and e-mail address for corresponding with the Chief investigator must be included. The organisational affiliation of the Chief Investigator will be the sponsor of the proposed study.

All applicants must indicate whether they have statistical experience, experience of handling large datasets and/or experience practicing in UK primary or secondary care, when registering for an account on the CPRD electronic research application portal (eRAP). Research teams without statistical experience, experience of handling large datasets and/or experience practicing in UK primary or secondary care are encouraged to collaborate with other researchers with experience in the relevant areas, prior to submitting their research for RDG review. Please note that at least one member of the research team must be listed as accessing the data for the study to be conducted.

Question 3: The Corresponding Applicant

Please note that the name and affiliation of the Corresponding Applicant on approved feasibility studies is published on the CPRD website for the benefit of patients and the public, to inform them of how CPRD's anonymised patient data are used for research.

The Corresponding Applicant is the direct point of contact for the RDG Secretariat and is authorised to submit the application on behalf of the Chief Investigator. It is also acceptable for the Chief Investigator to be the Corresponding Applicant.

All applicants must indicate whether they have statistical experience, experience of handling large datasets and/or experience practicing in UK primary or secondary care, when registering for an account on the CPRD electronic research application portal (eRAP). Research teams without statistical experience, experience of handling large datasets and/or experience practicing in UK primary or secondary care are encouraged to collaborate with other researchers with experience in the relevant areas, prior to submitting their research for RDG review. Please note that at least one member of the research team must be listed as accessing the data for the study to be conducted.

Question 4: Other investigators/collaborators

Please note that the name and affiliation of other investigators/ collaborators on approved feasibility studies are published on the CPRD website for the benefit of patients and the public, to inform them of how anonymised patient data collected by CPRD are used for research.

Anyone who will access **AND** use CPRD data to conduct the research detailed in the feasibility study application must be named in the feasibility study application. All investigators or collaborators must have an authorised eRAP account for a feasibility study to be submitted. To register for an eRAP account visit: www.erap.cprd.com

At the time of registering for an eRAP account, applicants must indicate whether they have statistical experience, experience of handling large datasets and/or practicing in UK primary or secondary care. Research teams without this experience are encouraged to collaborate with other researchers who have such experience on their proposed CPRD research. Please note that at least one member of the research team will need to be listed as accessing the data for the study to be conducted.

eRAP account requirements

All investigators/collaborators named on the study application must hold an approved eRAP account and have accepted the study invitation by logging into their eRAP accounts.

Please note that eRAP account registrations with missing information may result in delays to processing the account registration. Applicants should indicate at the time of registering for an eRAP account whether they have statistical experience, experience of handling large datasets and/or practicing in UK primary or secondary care.

To register for an eRAP account visit: www.erap.cprd.com

SECTION B: ACCESS TO THE DATA

Question 5: Data Access Arrangements

State the method that will be used to access the data for this study - a study-specific dataset agreement or an institutional multi-study licence. If a multi-study licence is to be used, please indicate the licensing institution name and address.

Datasets that will be extracted by CPRD

Investigators must discuss requests for CPRD to extract datasets for a feasibility study with a member of the CPRD Research Team **before** submitting a feasibility study application. Please contact the CPRD Research Team on (enquiries@cprd.com) to discuss your requirements. You **must** state the enquiry reference number associated with your contact with CPRD in this section.

Question 6: Site location of data

We require information on any organisation that will be processing, accessing, or storing the data requested. For each location, applicants must specify whether the organisation is processing, accessing, or storing data, and provide the organisation name, address, and processing area.

The data processing areas are – UK, European Economic Area (EEA), or Worldwide. It may be that one location stores, processes, and analyses the data.

For feasibility studies that will be conducted under a **CPRD Single Study Licence (SSL),** all data and analysis **must** be accessed and conducted on CPRD's Trusted Research Environment (TRE). If this is applicable to your study, please select "another organisation" under "Organisation" and list "CPRD Trusted Research Environment (TRE)" in the free text box. The data processing area must be set to 'UK'.

SECTION C: INFORMATION ON REQUESTED DATA

Primary care data collected by the CPRD are linked to several other patient level datasets, (including Hospital Episode Statistics, Office of National Statistic mortality data, Cancer Registry etc.) for patients at English practices that have consented to participate in the linkage scheme.

Information on the range of linked data available via CPRD can be found here https://cprd.com/cprd-linked-data.

If you have any questions about accessing linked data, please contact CPRD Enquiries (enquiries@cprd.com).

Question 7: Primary Care data

Vision and EMIS are different clinical software systems used by general practices in the United Kingdom primary care setting. CPRD has historically collected data from Vision primary care practices, which is referred to as the CPRD GOLD primary care data. CPRD also collects data via the EMIS software system under the CPRD Aurum primary care data.

Question 8: Requests to access linked data

The following linkages are permitted for feasibility studies:

- CPRD HES Admitted Patient Care, Accident and Emergency, Outpatient, and/or Diagnostic Imaging Dataset.
- CPRD ONS Death Registration Data
- CPRD Mother Baby Links
- CPRD Pregnancy Registers
- NCRAS Tumour/Treatment (excluding SACT or RTDS data)
- Practice Level Index of Multiple Deprivation (Standard)
- Patient Level Index of Multiple Deprivation (excluding Index of Multiple Deprivation domains)
- CPRD Ethnicity Record

Question 9: Requesting CPRD to extract the primary care data.

Investigators who require CPRD to extract the primary care data for their study must discuss this request with CPRD before submitting a feasibility study. Please contact enquiries@cprd.com to discuss your requirements.

A query reference number will be provided by CPRD as evidence of CPRD's agreement to extract the primary care data. This must be included in the online application when prompted.

Question 10: Commissioning CPRD to conduct Feasibility Studies

Applicants may commission CPRD to conduct feasibility studies.
Please note that the restriction on sample size does not apply to clients who commission CPRD to undertake feasibility studies on their behalf.

Applicants commissioning CPRD to conduct a feasibility study must speak to a member of the CPRD Observational Research team prior to submitting their application for RDG review. Please contact the CPRD Research Team enquiries@cprd.com to discuss your requirements. Please provide the query number relating to this discussion in your application on eRAP.

Question 11: Patient identifiers

Investigators must state whether any person named in the study has access to the data in a patient identifiable form, or any associated identifiable patient index.

If the answer to this question is 'Yes', applicants must provide a re-identification risk management plan as an appendix and refer to it here and in relevant sections of the protocol.

The re-identification risk management plan should provide a thorough and robust account of how the risk of reidentifying patients in CPRD data will be made negligible. The plan should consider the following characteristics of the environment where CPRD data will be processed, and ensure appropriate controls are in place for each:

1. **Other Data**: This relates to whether you or other members in the research team hold or have access to any information that could be linked to CPRD data, thereby enabling re-identification. This may include personal knowledge, information from publicly available sources, restricted access data sources, and other similar data releases.
2. **Agents**: This relates to the people and entities accessing, using, and interacting with the data.
3. **Governance Processes**: This relates to how agents' relationships with the data are managed. This includes formal governance such as data access controls, licensing arrangements and policies which prescribe and proscribe agents' interactions.
4. **Infrastructure**: This relates to the structures and facilities that allow CPRD data to flow and shape the data environment, including security infrastructure and wider social and economic structures. At its narrowest level, infrastructure is best thought of as the set of interconnecting structures (physical, technical) and processes (organisational, managerial) that frame the data environment. At its broadest level, infrastructure can include intangible structures, such as political, economic, and social structures, that influence the evolution of technologies for data exploitation, as well as data access, sharing and protection practices.

For more information and guidance, please refer to Section 3 of the UK Anonymisation Network's 'The Anonymisation Decision-Making Framework: European Practitioners' Guide' (https://msrbcel.files.wordpress.com/2020/11/adf-2nd-edition-1.pdf)

ALL FEASIBILITY STUDY APPLICATIONS MUST BE COMPLETED AND SUBMITTED VIA THE CPRD ELECTRONIC RESEARCH APPLICATION PORTAL (eRAP) www.erap.cprd.com

Part 2: Feasibility Study Information

Please note that during CPRD's scientific review process, **all** research seeking access to CPRD primary care and/or linked data for public health research are reviewed for **study feasibility** and **research team expertise/experience**, the **public health benefits/risks** of the research and potential **information governance risks** (risks to patient confidentiality and privacy).

Lay Summary *[1500 characters/max. 250 words]*

Please note that the lay summary is published on the CPRD website for the benefit of patients and the public, to inform them of how anonymised patient data collected by CPRD are used for research.

Reviewer Assessment Criteria

In this section, reviewers will assess whether the proposed research could be easily understood, as a standalone summary, by non-scientific readers. The importance, relevance, and implications of the research to patients, clinical practice or the health care system will also be assessed.

Application Requirements

Please provide a succinct overview of your proposed research in non-technical language.

Applicants proposing to use CPRD data for a feasibility study should clearly state the aim/s of the proposed study that the feasibility study will inform (e.g. 'the aim of this study is to assess the feasibility of undertaking a cohort study in CPRD or another data source to investigate the association between drug X and outcome Y'; or 'the aim of this study is to assess the feasibility of a randomised controlled trial assessing comparative effectiveness of different dosing schedules of treatment X within the UK primary care setting').

Lay summaries **must** focus on the proposed feasibility study (not the future study) and include the following:

- A succinct outline of what prompted the feasibility study (state the problem and indicate what you are planning to do about it by undertaking the feasibility study).
- A statement about the potential impact/ public health benefit of undertaking the feasibility study (what is likely to change for patients, clinical practice, or the wider society).
- Non-technical language, short sentences and a summary written in plain English. When writing, imagine that you are writing your research to be understood by readers 9-11 years old, the average reading age of the UK population
- Follow a logical order. This may not always coincide with a temporal order.
- Use first person and active voice ("we agreed" rather than "it was agreed").

Lay summaries **should not** include the following:

- Any technical details, such as the study design or statistical methods.

- Jargon - unless you also explain it
- Use of the word “identify” when referring to cohort definitions or determining study eligibility, to avoid giving the impression that it is possible to locate or contact patients using information contained in CPRD data.
- Abbreviations - these should be defined on first use or explained.
- The use of superscripts, subscripts, and references.

The lay summary should provide an overview of the feasibility study without the need to refer to the technical summary.

Studies with lay summaries over **250** words will be returned as invalid.

Technical Summary *[1800 characters /max. 350 words]*

Please note that the technical summary is published on the CPRD website for the benefit of patients and the public, to inform them of how anonymised patient data collected by CPRD are used for research.

Reviewer Assessment Criteria

Technical summaries will be evaluated for transparency in communicating the purpose, methods, and benefits of the proposed feasibility study to scientific readers as a standalone summary. A high-level assessment of the relevance/feasibility of the stated methods and analytical approaches will also be undertaken. Reviewers will also assess whether the benefits of the proposed methods to achieve the objectives of the research outweigh potential risks, including information governance risks such as patient or practice confidentiality issues.

Application Requirements

The technical summary is primarily written for other researchers and clinicians. There should be enough technical detail to allow another researcher to obtain a clear idea of your feasibility study aim and methods.

The technical summary should provide a succinct overview of the overarching study aim and objectives, primary exposure(s), and outcome(s), if relevant, study design, and methods including the main statistical analyses to be conducted.

The technical summary should also specify the primary care and linked data sources requested will be used, for example, “Hospital Episode Statistics (HES) admission data will be used to determine any cause or cause hospitalisations for condition X”.

The technical summary must include a statement about the public health benefit/s of the proposed feasibility study.

Studies with technical summaries more than 300 words will be returned as invalid.

Outcomes to be Measured *[600 characters/max. 100 words]*

Please note that study outcomes listed in this section are published on the CPRD website for the benefit of patients and the public, to inform them of how anonymised patient data collected by CPRD are used for research.

Reviewer Assessment Criteria

In this section, reviewers will assess the feasibility of ascertaining the outcome(s) in CPRD.

Application Requirements

This section should clearly **list** the key outcome variables, separated by semicolons. For example: "Ischaemic stroke; Composite of all bleeding; Major bleeding; Gastrointestinal bleeding; Clinically significant non-major bleeding; Myocardial infarction; All-cause mortality; Intracranial haemorrhage"

This section should not include statements relating to the study aims and objectives.
All definitions of outcomes should be included under the section on "**Information on the Study Population"**

Planned use of linked data (if applicable), including the public health benefits to patients in England & Wales *[1200 characters/max. 200 words]*

Reviewer Assessment Criteria

Where applicable, reviewers will assess whether the linked data sources requested are relevant and feasible (can support cohort/comparison definition, exposure definition, outcome ascertainment or covariate definition) to address the feasibility study aims and objectives. An explicit review of how the outputs of the proposed feasibility study using linked data will benefit patients in England & Wales will also be evaluated.

Application Requirements

Any proposed use of linked data sets must be appropriate to the research. This will be assessed against statements made on the CPRD feasibility study application form. For proposals to use data sources routinely linked to CPRD data, for example, Hospital Episode Statistics (HES), Office of National Statistics (ONS) Mortality data, practice/patient area-level data, please describe why each linked data source is necessary for the study and how it will be used.

Applications **must** outline how the main outputs of the proposed feasibility study will benefit patients in **England and Wales**. You may base your justification on how the study findings would improve patient care either directly or indirectly by informing clinical practice guidelines or public health policy.

Information on the Study Population *[3500 characters/max. 600 words]*

Reviewer Assessment Criteria

In this section, reviewers will assess whether the study population is clearly described and relevant in the context of the feasibility study; whether restricting/excluding certain patient groups from the research may disadvantage such patient groups and limit the benefits of the study; whether research to combine data from CPRD with other external non-CPRD data sources may present potential patient and/or practice re-identification risks or other information governance risks.

Application Requirements

It is important to ensure that the application clearly defines the study population. The following areas listed below should be addressed as relevant:

a) Describe the source/target population:
   - cohort definition (age, sex, key condition/exposure of interest)
   - whether only permanently registered acceptable patients will be included
   - whether only up-to-standard follow-up will be considered

- state the recruitment period and state the definition of the start and end of follow-up for patients, including whether the CPRD death date should be used in defining the end of follow-up.

b) Describe the study population in terms of inclusions, exclusions, and the data used for each (clinical, referral, test, therapy, immunisation). Reference should be made to provisional code lists for inclusion & exclusions specified
c) Provide any minimum requirements for previous follow-up time
d) Information on the exposure window(s) of interest, where appropriate, clearly defining time which will be considered "exposed" or "non-exposed"
e) For studies requiring linked data, please make clear the restrictions imposed by patients' eligibility for linkage and the coverage period for each linked dataset requested. Further information on patient's linkage eligibility and the coverage periods of linked data sources may be requested from the CPRD (enquries@cprd.com).
f) For studies requiring linked data, please make clear whether the study population should be defined among all eligible patients in CPRD or only among patients eligible for linkage to your linked data source of interest.

Information Governance Risks

The possibility of unintentional (deductive) disclosure arises when cells with small numbers of patients or events are reported. CPRD policy is that no cell should contain <5 patients or events during reporting. It is therefore essential that consideration is given to preserving confidentiality at the reporting stage using secondary suppression methods. In this section please also include mitigation approaches to minimise the risk of inadvertently identifying patients or practices during the publication of your research.

Appendix 1: Examples of feasibility studies that may be approved under the Feasibility Study application route

EXAMPLE 1

Future Study: To investigate the association between LABA/LAMA prescribing and rate of decline of lung function in patients with moderate to very severe Chronic Obstructive Pulmonary Disease (COPD) (GOLD stages II to IV).

Feasibility study: To explore definitions of COPD severity based on diagnosis, treatment and/or spirometry data

Identify the target population

- Number of patients with a diagnosis of COPD during the study period
- Number of patients with spirometry data recorded
- Number with valid spirometry data recorded
- Proportion of patients with FEV1/FVC < 70
- Patients treated with COPD therapy
- Patients with a diagnosis of asthma

Implement algorithm for classifying COPD severity (GOLD stages II to IV) based on treatment and spirometry data.

- Distinguish pre- and post-bronchodilator spirometry data
- Implement GOLD categories using data using spirometry and other data recorded in primary care.

EXAMPLE 2

Future Study: Prediction of exacerbation onset in Chronic Obstructive Pulmonary Disease (COPD) patients

Feasibility Study: To examine the recording of COPD exacerbations in primary care and understand the extent to which sub-group analyses may be possible.

Identify the target population

- Number of patients with a diagnosis of COPD during the study period
- Numbers eligible for linkage, number with at least 12 months of prior UTS follow-up
- Numbers with current/ex-smoker status,
- Proportion of patients with FEV1/FVC < 70 at diagnosis and 6-month post diagnosis.
- Numbers with a record of asthma and COPD

Develop algorithm for identifying COPD exacerbations

Proportion of patients treated with combined therapy (dual, triple therapy) for at least 3 months/6 months

- Proportion of patients receiving add-on therapy and cumulative dose
- Proportion of patients switching therapy, type of therapy, dosage, and duration
- Proportion of patients with at least 2 exacerbations during dual therapy
- Proportion of patients with at least 2 exacerbations during triple therapy

Frequency of exacerbations among patients with COPD

Proportion of patients treated with combined therapy (dual, triple therapy) for at least 3 months/ 6 months who have a record of an exacerbation

Identify COPD exacerbations in the hospital setting.

EXAMPLE 3

Future Study: To investigate the association between Chronic Obstructive Pulmonary Disease (COPD) medication adherence and resource utilisation.

Feasibility Study: To assess the completeness of recording of COPD prescribing data and explore methods for estimating treatment adherence in light of the data that are available.

Identify the target population

- Number of patients with a diagnosis of COPD during the study period
- Numbers eligible for linkage to Hospital Episode Statistics data
- Numbers with at least 12 months of Up-to-standard (UTS) follow-up prior to COPD diagnosis
- Numbers with at least 24 months of follow-up following COPD diagnosis
- Numbers initiating maintenance medication during the first 365 days of the 24-month post-index period (have at least one prescription for COPD maintenance - long-acting muscarinic antagonists [LAMA] and inhaled corticosteroid-long-acting beta agonist [LABA] fixed-dose combinations)
- Evaluate dose of maintenance therapy by treatment type

Explore approaches for measuring COPD treatment adherence

- Proportion of days covered (PDC)
- Medication Possession Ratio (MPR)

Examine the characteristics of adherent/non-adherent

- Proportion adherent/non-adherent
- Mean PDC/MPR
- Baseline characteristics of adherent and non-adherent groups (Mean, median, and interquartile range)