



Guidance on completion of a CPRD Research Data Governance (RDG) Application

**ALL APPLICATIONS MUST BE COMPLETED AND SUBMITTED VIA THE
CPRD ELECTRONIC RESEARCH APPLICATION PORTAL (eRAP)**

www.erap.cprd.com

Part 1: Application Form

GENERAL INFORMATION ABOUT THE PROPOSED RESEARCH STUDY

Question 1: Study Title (Max. 255 characters including spaces)

It is important to ensure that the title of the study is clear, concise, easy to understand, and accurately reflects the main purpose/focus of the study.

The title should be reflective of the overarching study aim. The title of a hypothesis-testing study should give a clear indication of the primary exposure(s) and outcome(s). Ideally, the title should also refer to the study design.

Example 1: Incretin based drugs and risk of adverse renal outcomes

Example 2: Topical corticosteroids and risk of type 2 diabetes: a nested case-control study

Similarly, for a descriptive study, an example of a good title would be 'The prescribing of codeine for the treatment of pain in children: a descriptive study'.

Avoid catchy titles that are vague about the study aim. Examples of unsuitable titles would be: 'Pneumonia - the old man's friend'.

Applications with titles in excess of 255 characters will be returned as invalid.

Question 2: Research Area

Specify the research area of the proposed study. Applicants must select at least one box.

Questions 3: Purely observational Research

Approval from an NHS Research Ethics Committee may be required if the proposed study is not purely observational. However, if the research will only involve CPRD data and routine linkages, no separate ethics approval is required.

Question 4: Patient or GP questionnaires or patient contact

Questionnaires for healthcare professionals or patients must be reviewed and approved via the CPRD RDG process before being used. If available, any questionnaire should be included as an appendix to the application, otherwise the protocol should state that it will be submitted for approval prior to use. The questionnaire must be provided as it is intended to be presented to the recipients, together with any covering letter or guidance on completion which will

be provided with the questionnaire. All questionnaires must be accompanied by an appropriate explanation of the purpose of the study for the recipient.

Applicants must also seek approval of their questionnaire design and timelines by submitting an enquiry to the CPRD Interventional Research team via enquiries@cprd.com. CPRD questionnaire studies are conducted electronically via the CPRD integrated platform; fees for CPRD questionnaire studies are in addition to data fees. Fees and timelines will be confirmed by the CPRD Interventional Research team as part of the enquiry. Applicants should quote the enquiry reference number to support their protocol application.

CPRD encourages consultation and/or piloting of questionnaires with the target population (health care professionals or patient groups); evidence of which should be included in the protocol. Where validated instruments are to be used in a study, applicants should indicate whether the necessary permissions are in place to use the questionnaire/s and provide evidence of this in the study protocol.

Where patient samples are required, state what, how, and the frequency of sample collection. Note that the CPRD requires evidence of additional ethical approval for contact relating to patients.

Question 5: Chief Investigator

The Chief Investigator will take responsibility for ensuring that the research is undertaken with full adherence to CPRD RDG guidelines, and any CPRD Contracts and Terms and Conditions.

The full name, job title, organisation name, and e-mail address for correspondence of the Chief investigator must be included in the form. Applicants must indicate whether the Chief Investigator will be analysing the data.

The organisational affiliation of the Chief Investigator will be the sponsor of the proposed study.

Question 6: The Corresponding Applicant

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

Question 7: Other investigators/collaborators

Anyone who will have access to CPRD data must be named in the CPRD RDG protocol. All investigators or collaborators must have an authorised eRAP account for a protocol to be submitted.

Applicants must indicate whether each member of the research team will be analysing the data.

ACCESS TO THE DATA

Question 8: Sponsor of the study

The sponsor for the study is a company, institution, organisation, or group of organisations that takes on responsibility for initiation, management and financing (or arranging the financing) of the proposed research.

A sponsor can delegate specific responsibilities to any other organisation that is willing and able to accept them. Any delegation of responsibilities to another party should be formally agreed and documented by the sponsor.

It is the sponsor who determines what data is requested for the research study through the protocol.

The sponsor organisation is the affiliation of the Chief Investigator.

Question 9: Funding source for the study

Specify the primary funding source for the study. Any organisation, or group of organisations, providing funding for the research project should be listed, including any grants and the awarding bodies.

Question 10: Institution conducting the research

Applicants must specify the name and address for the institution that will be conducting the research using CPRD data where this is not the sponsor organisation.

Question 11: 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 licence is to be used, please indicate the licensing institution name and address.

Please note that, for applicants requesting NCRAS data, CPRD must extract and deliver all primary care and linked data for the study (including NCRAS data), regardless of whether a multi study licence is in place.

Investigators must discuss requests for CPRD to extract data with a member of the CPRD Research Team before submitting a CPRD RDG application. Please contact the CPRD Research Team on (enquiries@cprd.com) to discuss your requirements. Please also state the enquiry reference number.

Question 12: Data Processor(s).

We require information on any organisation that will be processing, accessing, or storing the data requested by the applicant.

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.

Further guidance and information can be found on the [ICO website](#).

INFORMATION ON DATA

Primary care data collected by the CPRD can be linked to a number of other patient level datasets, (including Hospital Episode Statistics, Office of National Statistic mortality data, Cancer Registry etc...) and is only available for English practices that have consented to participate in the linkage scheme.

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

Question 13: 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 GOLD primary care data. More recently, CPRD has been able to release data collected via the EMIS software system under the CPRD Aurum primary care data.

Question 14: Requests to access linked data

For all linked data requests, applicants must outline in [Section K](#) of the protocol how the main outputs of the proposed 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.

Where access to the following linked data is being requested, at least one applicant named on the CPRD RDG application form must have discussed the linkage with a member of the CPRD Research Team (enquiries@cprd.com), prior to submission of the RDG application:

- HES PROMs (Patient Reported Outcomes Measures)
- NCRAS Cancer Registration Data
- NCRAS Cancer Patient Experience Survey (CPES) data
- NCRAS Systemic Anti-Cancer Treatment (SACT) data
- NCRAS National Radiotherapy Dataset (RTDS) data
- NCRAS Quality of Life Cancer Survivors Pilot (QOLP)
- NCRAS Quality of Life Colorectal Cancer Survivors (QOLC)
- Mental Health Data Set (MHDS)
- Second Generation Surveillance System (SGSS, COVID-19)
- COVID-19 Hospitalisations in England Surveillance System (CHESS).
- Practice Level Index of Multiple Deprivation (index other than the most recent)
- Practice/Patient Level Index of Multiple Deprivation Domains

Please note that PROMs data are only available for non-commercial purposes, such as academic research, or in connection with delivering services to the NHS. Further information on access to PROMs data can be found at <https://cprd.com/linked-data>.

Applicants seeking access to NCRAS data must also complete a Cancer Dataset Agreement Form (available from CPRD on request) and submit this via the CPRD RDG process as an appendix to the protocol. Applicants must also provide consent for publication of their study title and study institution on the UK Cancer Registry website.

As a risk minimisation measure, CPRD provide only one practice and/or one patient level area linkage per study. If you require more than one practice or patient level area linkage (i.e. practice level IMD and Rural-Urban classification), this will need to be discussed with a member of the CPRD Research Team (enquiries@cprd.com), before submitting a CPRD RDG application. Applicants will be required to provide an enquiry reference number on the form.

Practice Level Index of Multiple Deprivation (index other than the most recent) refers to requests for linkages to years other than the most recent. Please be aware that these vary by UK nation. Please contact CPRD (enquiries@cprd.com) if you require further information.

Question 15: Requesting non-standard data linkage

Investigators wishing to link to a dataset not listed in question 14 must have received approval for such a linkage from CPRD prior to submitting a CPRD RDG protocol. Applicants must provide the Non-Standard Linkage (NSL) reference number for the approval of the linkage in their protocol application.

Applicants wishing to link to a dataset not listed in question 14 should review the information regarding non-standard linkage on the CPRD website (<https://cprd.com/non-standard-linkage>). CPRD RDG applications will not be accepted for studies requesting non-standard linkage that have not had the linkage approved by CPRD.

Question 16: 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 and risk management plan as an appendix and refer to it here and in the required protocol information.

ALL APPLICATIONS MUST BE COMPLETED AND SUBMITTED VIA THE CPRD ELECTRONIC RESEARCH APPLICATION PORTAL (eRAP)

www.erap.cprd.com

Part 2: Protocol Information

A. Study Title

Reviewer Assessment Criteria

In this section, reviewers will assess whether the study title clearly describes the main focus and purpose of the proposed research.

Application Requirements

It is important to ensure that the title of the study is clear, concise, easy to understand, and accurately reflects the main purpose/focus of the study.

The title should be reflective of the overarching study aim. The title of a hypothesis-testing study should give a clear indication of the primary exposure(s) and outcome(s). Ideally, the title should also refer to the study design.

Example 1: Incretin based drugs and risk of adverse renal outcomes

Example 2: Topical corticosteroids and risk of type 2 diabetes: a nested case-control study

Similarly, for a descriptive study, an example of a good title would be 'The prescribing of codeine for the treatment of pain in children: a descriptive study'.

Avoid catchy titles that are vague about the study aim. Examples of unsuitable titles would be: 'Pneumonia - the old man's friend'.

Applications with titles more than 255 characters will be returned as invalid.

B. Lay Summary (Max. 250 words)

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.

The lay summary will be published on the CPRD website for the benefit of patients and the public, to inform them of how CPRD data are being used and to what benefit.

The lay summary should provide a succinct overview of the proposed research in non-technical language. The lay summary should cover the background, purpose of the study, and the potential importance of the findings.

The lay summary should not include any technical details, such as study design or statistical methods. For all research studies, there must be a clear justification, avoiding jargon, of the expected public health benefits from the study, which must be capable of being understood by a member of the public without a scientific or medical background.

The use of the word "identify" should be avoided, or it should be made clear that it does not refer to identification of patients. Abbreviations should be clarified before use. The use of superscripts, subscripts and references is not permitted.

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

Applications with lay summaries that do not adhere to these guidelines will be returned as invalid.

C. Technical Summary (Max. 300 words)

Reviewer Assessment Criteria

Technical summaries will be evaluated for transparency in communicating the purpose, methods, and benefits of the proposed research 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 written primarily for other researchers and clinicians who may be interested in your research. This should include enough technical details to provide a clear idea of your study aim and methods. Your technical summary should be presented as 1-2 paragraphs providing a succinct overview of your research and include details on the following:

- Overarching aim and objective(s)
- Study population of interest
- Primary exposure(s) and outcome(s), where relevant,
- Data sources that will be used to achieve the aim and objectives (e.g. Hospital Episode Statistics (HES) admission data will be used to determine hospitalisations).
- Study design, methods including the main statistical tests
- Intended public health benefit of the research

You should avoid vague and broad references to methods, for example time-to-event analysis or regression models, in favour of more specific terms such as Cox proportional hazards regression or linear regression. The use of the word “identify” should be avoided, or it should be made clear that it does not refer to identification of patients. Abbreviations should be clarified before use. The use of superscripts, subscripts and references is not permitted.

Technical summaries that do not adhere to these guidelines will be returned as invalid.

D. Outcomes to be Measured

Reviewer Assessment Criteria

In this section, whether the choice of primary and key secondary outcomes will achieve the intended benefits of the research will be evaluated. An assessment of the feasibility of ascertaining the outcome(s) in CPRD may also be reviewed in this section

Application Requirements

This section should clearly list the primary and key secondary outcomes of interest in a concise list, separated by semicolons, e.g.:

“Complications of infection in primary or secondary care; Admission to Accident & Emergency; All-cause hospitalisation; All-cause mortality”

This section should not include statements relating to the study aims and objectives. For descriptive and feasibility studies, list the key variables in this section.

All definitions of the primary and key secondary outcomes should be included under the section on “Exposures, Outcomes and Covariates”

E. Specific Aims, Objectives, and Rationale

Reviewer Assessment Criteria

In this section, reviewers will evaluate the clarity, scope, feasibility, and benefits that may be achieved through the study aim(s) and objectives. Protocols should include details to enable reviewers to evaluate the public health benefits of the research, assess the methods for implementing the stated objectives, determine whether inherent public health risk(s) may arise during the conduct of the research or whether there may be risks to patient/practice confidentiality and/or privacy.

Application Requirements

A general aim should normally be provided, followed by one or more specific and related objectives. Studies with many objectives often fail to describe all objectives in sufficient detail in the protocol and may be considered too extensive for a single protocol.

Applicants should clearly state their primary and secondary study objectives.

CPRD RDG reviewers will carefully consider whether all the proposed objectives have been addressed in later sections, particularly regarding analysis of the data. Applicants should also provide a satisfactory statement regarding the rationale/need and implications for the present study.

Applicants must include the following:

- i. A description of the knowledge/information to be gained from the study, and how this will improve patient care, either directly or indirectly, by informing clinical practice guidelines or public health policy (research objectives).
- ii. The primary hypothesis to be tested.
- iii. An explanation of how achievement of the specific objectives will further the research aim (rationale).

F. Study Background

Reviewer Assessment Criteria

In this section, reviewers will assess whether the study background highlights the importance, relevance, and public health value of the research. This should be linked to relevant published literature.

Application Requirements

Applicants should explain the reason for the research aim and objectives and support this with relevant information from the scientific or other literature. This should highlight key issues that are currently unanswered or in dispute in the field. Background information may refer to previous or similar studies conducted in GPRD, CPRD or other data sources. All supporting statements should be duly referenced in this section with the full reference included in the "Reference" section of the protocol.

Ensure that you refer to any previous RDG protocols or protocols from the Independent Scientific Advisory Committee (ISAC) that may be related to your study. Any reference to a previous ISAC/RDG protocol should be accompanied by the relevant protocol number e.g., 15_101, 20_000001

G. Study Type

Reviewer Assessment Criteria

In this section, reviewers will assess whether the study type(s), as described below, are appropriate to address the stated aims and objectives of the research and to achieve the intended public health benefits.

Application Requirements

Specify whether the study will be primarily descriptive, hypothesis generating, hypothesis testing, or a methodological piece of research. We recognise that a single research study may comprise one or more of the following study types:

- Descriptive studies – These include ecological studies, cross-sectional analyses, drug utilisation studies, and case series assessment, which focus mainly on identifying patterns or trends in disease occurrence over time.
- Exploratory/ Hypothesis Generating – Exploratory or hypothesis generating studies are often descriptive studies that aim to reveal patterns associated with a specific condition or event, without an emphasis on testing pre-specified hypotheses. Thus, the emphasis of such studies is on estimation. Some quantities that can be estimated in exploratory studies are the prevalence and incidence of a disease, the resources required to treat a disease, or utilisation patterns of a product. Hypothesis generating, or exploratory studies, are acceptable within a defined framework (i.e. they do not constitute data mining), and there is a clear commitment to report the results accordingly.
- Hypothesis Testing – Hypothesis testing studies in epidemiology involves the use of data to make statistical decisions about the associations of a disease, or the degree of exposure to an agent or product and its relationship with disease. Hypothesis testing studies are therefore intended to provide results by testing hypotheses with clearly defined exposures and outcomes. Analysis of the data must therefore be based on predefined valid analysis plans.
- Methodological – Methodological studies include studies of statistical methods, comparisons of study designs, etc... The analysis of data should be based on a predefined valid analysis plan.

H. Study Design

Reviewer Assessment Criteria

In this section, reviewers will assess whether the study design(s), as described below, are appropriate and can be reliably implemented in CPRD to achieve the intended benefit of the research. Feasibility of the design may also require an assessment of the methods - numbers expected in CPRD (feasibility counts), sample size considerations, data sources to be used, study exposures and outcomes or proposed data analysis.

Application Requirements

Applicants should briefly state the overall research design, strategy, and reasons for choosing the proposed study design.

Research designs include, for example: case-control, cohort, cross-sectional, nested case-control, or hybrid designs.

Applicants should clearly outline their study design to avoid reviewer confusion arising in relation to matched control groups, for example, a comparative cohort study being described as "case-control".

I. Feasibility counts

Reviewer Assessment Criteria

In this section, reviewers will assess the feasibility of the research, that is, whether there is likely to be “adequate” number of patients in CPRD to address the main objectives of the proposed research. Where numbers may be low, reviewers will also consider the sample size calculation, study design, and data analysis sections of the protocol to assess whether the research may present patient and/or practice re-identification risks. Applicants should include a risk mitigation plan in their protocol where there may be potential re-identification risks.

Application Requirements

Applicants must provide an estimate of the expected number of patients available in the CPRD and/or linked data sets for the proposed study. Applicants may refer to relevant publications using CPRD data to gauge study feasibility or support their application with feasibility counts based on CPRD data. In some cases, feasibility counts can be requested from CPRD and for more information applicants should contact enquiries@cprd.com.

A searchable list of publications using CPRD data, which is updated monthly, can be found at <https://www.cprd.com/bibliography>. Applicants can also request code browsers for free from CPRD which will allow them to search for all medical and treatment codes that are included in the CPRD primary care database to assess whether specific conditions, treatments or other exposures are captured in the CPRD database.

Where numbers expected in CPRD are low and may present a challenge to study feasibility, applicants may wish to consider the following approaches:

- Use data from CPRD GOLD and/or CPRD Aurum to increase the sample size
- Use linked data sources for case definition/outcome ascertainment if conditions are more likely to secondary care
- Revisit the case definition
- Consider an alternative study design to increase study power e.g. matched study design

If options to increase your study population are not feasible, please outline mitigation approaches to minimise the risk of inadvertently identifying patients or practices during the conduct and/or publication of the study.

J. Sample size considerations

Reviewer Assessment Criteria

In this section, reviewers will assess whether a sample size calculation is needed and if so, whether the sample-size/power of the study will be sufficient to address the primary hypothesis of the research. Where numbers may be low, reviewers will consider your study design and proposed data analysis to assess whether your research may present potential patient and/or practice re-identification risks. Applicants should include a risk mitigation plan in their protocol where there may be potential re-identification risks.

Application Requirements

All protocols must include some consideration of whether the sample-size (and study power for hypothesis testing studies) will be sufficient to meet the primary outcome of the research.

All protocols should include an estimate of the expected numbers of patients, exposures, or outcomes (as appropriate) that will be available. Applicants may refer to relevant publications using CPRD data to gauge study feasibility or support their application with feasibility counts based on CPRD data.

For hypothesis testing studies, it is necessary to demonstrate that the expected numbers are sufficient to investigate the primary study objective with adequate power. This may be demonstrated by carrying out a formal power or sample size calculation, in which case sufficient information should be given for a statistician to be able to repeat the calculation(s), including the method and the values of numerical inputs and their sources (e.g. references). Alternatively, it may be possible to make an informal argument that the expected numbers are sufficient by comparison to previously published studies.

For hypothesis generating and descriptive studies, we typically expect demonstration that expected numbers will give reasonable precision around the effect estimates or numerical results to be calculated. For methodological studies, the appropriate approach to demonstrating that expected numbers are adequate will vary.

In all types of study, sample size/power calculations should, when relevant, reflect chosen approaches to dealing with multiple comparisons.

Where numbers expected in CPRD are low and may present a challenge to study feasibility, applicants may wish to consider the following approaches:

- Use data from GOLD and/or Aurum to increase the sample size
- Use linked data sources for case definition/outcome ascertainment if conditions are more likely to secondary care

- Revisit the case definition
- Consider an alternative study design to increase study power e.g. matched study design

If applicants wish to make a case that it is worth proceeding with a study even though the expected numbers are lower than desired – for example, in studies of extremely rare conditions – then this should be identified and clearly acknowledged as a limitation in the research protocol and addressed in a risk mitigation plan.

Please be aware that post-approval, CPRD will also review any data requests are supported by the sample size estimates stated in the approved protocol and that there is a clear justification for large sample size requests to demonstrate compliance with data minimisation principles.

While there is no specific limitation on the size of the study population, the size must be clearly justified in the protocol. Proportionate data minimisation measures will be applied when any Primary Care or linked dataset comprise of >600k patients, and will take into account feasibility counts, sample size calculation, data linkages requested (including study/coverage period), definition of the study population (including inclusion and exclusion criteria), comparison groups, exposure, outcomes and covariates definition. Please contact CPRD (enquiries@cprd.com) if you have any questions regarding data minimisation.

K. Planned use of linked data (if applicable), including the public health benefits to patients in England & Wales

Reviewer Assessment Criteria

Where applicable, reviewers will assess whether the linked data sources requested are relevant and feasible (can support cohort/comparison identification, exposure definition, outcome ascertainment or covariate definition) to address the study aims and objectives. An explicit review of how the outputs of the proposed 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 RDG application form and any other relevant information documented in the protocol. For proposals to use data sources routinely linked to CPRD data, for example, Hospital Episode Statistics (HES), Office of National Statistics (ONS) Mortality data, Cancer Registry data, practice/patient area-level data, please describe why the linkage data is necessary for the study and how it will be used.

Applications must outline how the main outputs of the proposed 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.

It is important that the relationships between the study population (e.g. with regard to dates), sample-size, and the use of linked datasets are clear within the protocol i.e. whether the entire study will be undertaken among practices which have consented to linkages or only part of it (e.g. in a sensitivity analysis). Applicants should consider how the time periods for availability of linked data might affect the study time period and censoring of patients.

Research groups which have not previously accessed CPRD linked data resources must discuss access to these resources with a member of the CPRD Research team before submitting a CPRD RDG application. Requests for access to certain linked data resources (see guidance for completing the protocol application form) must also be discussed with a member of the CPRD Research team and the evidence of this provided on the CPRD RDG application form. Study requesting linked data will not be approved unless these conditions have been met.

Studies proposing non-standard linkage of CPRD data to one or more external data sources should provide additional assurances about how the disclosure of patients and practices will be avoided in the form of a risk mitigation plan.

Any request for non-standard linkage should have received approval from CPRD prior to CPRD RDG submission. It essential that any necessary legal/ethical approvals are in place for any non-standard linkage to take place before submission to the CPRD RDG process.

L. Definition of the Study population

Reviewer Assessment Criteria

In this section, reviewers will assess whether the study population is clearly described and relevant in the context of the research; whether restricting/excluding certain patient groups from the research may disadvantage such patient groups and limit the benefits of the research; 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 protocol clearly defines the study population. The following areas listed below should be addressed in all research protocols:

- a) Describe the source/target population:
- b) State the indicative recruitment period and the definition of the start and end of follow-up for patients to allow an assessment of study feasibility.
- c) Describe the study population in terms of key inclusions, exclusions, and the data used for each (clinical, referral, test, therapy, immunisation, consultation). You should also provide justification for selecting the study population of interest.
- d) Provide a clear definition of the index date and any minimum requirements for previous follow-up time.
- e) Any reference to incidence or prevalence should be accompanied by details on how this should be defined (first record in the study period, first ever record, any record before the study end, treatment naive etc...).
- f) If any sampling from a base population is to be undertaken, provide details of sampling methods considering approaches that are likely to be free of selection bias.
- g) Also include information on the exposure window(s) of interest, where appropriate, defining clearly time which will be considered "exposed" or "non-exposed".
- h) For studies requiring linked data, please make clear the restrictions imposed by the eligibility criteria and coverage periods.

While there is no specific limitation on the size of the study population, the size must be clearly justified in the protocol. Proportionate data minimisation measures will be applied when any Primary Care or linked dataset comprise of >600k patients, and will take into account feasibility counts, sample size calculation, data linkages requested (including study/coverage period), definition of the study population (including inclusion and exclusion criteria), comparison groups, exposure, outcomes and covariates definition. Please contact CPRD (enquiries@cprd.com) if you have any questions regarding data minimisation.

For all cohort studies, the protocol should clearly define when a patient enters the cohort and when they will leave it. If there is an index date, it is important to ensure that it is clearly specified.

Please note that Researchers are not permitted to combine CPRD data with external data sources without explicit and prior permission from the CPRD. Please contact the CPRD (enquiries@cprd.com) to discuss combining or pooling of CPRD data with external data for your research. If permission is obtained to combine CPRD data with external data, please reference the Query number associated with your discussion with the CPRD on this subject and provide justification for the combining of these data in this section.

M. Selection of comparison group(s) or controls

Reviewer Assessment Criteria

In this section, reviewers will assess whether the selection of comparison groups or controls are clearly described, relevant and appropriate, and can be operationalised using the data sources and variables available in CPRD. Whether selection approaches will introduce biases that may limit the benefits of the research will also be evaluated.

Application Requirements

Where controls or comparison groups are needed to support a research question, please describe the following in the research protocol:

- a) How controls group differs from the main study population.
- b) The key inclusions, exclusions, and the data used for each (clinical, referral, test, therapy, immunisation, consultation).

- c) For studies requiring matching, type of matching (index date, calendar time, frequency, incident density sampling, high dimensional propensity score etc.) and the ratio/number of matches required should also be stated.

Applicants should also provide justification for the procedure for control selection. When making comparisons, calendar time should always be considered e.g. through use of an index date. Care should be taken to avoid the possibility of "immortal time bias". When this is a potential issue, a diagram showing how periods of time will be handled and such bias avoided is recommended.

N. Exposures, Outcomes and Covariates

Reviewer Assessment Criteria

In this section, reviewers will assess whether the exposures, outcomes and covariates of interest are clearly described, relevant in the context of the proposed research and can be operationalised in CPRD using the data sources requested. Reviewers will also assess whether there may be oversights in the selection of covariates or outcomes that may limit the public health value of the research. Potential risks to patient/practice confidentiality and/or privacy arising from the use of specific data variables or sensitive concepts will also be assessed.

Application Requirements

Defining Exposures and Outcomes

A clear description of the exposures and health outcomes of interest to the study should be provided. Operational definitions of these should also be provided to enable an assessment of feasibility. An operational definition is one that can be implemented independently using the data available in the proposed study. For example, "asthma episode" is not an operational definition; a better description would be "record of a Read code for asthma, as listed in Appendix A, and documented in the patient clinical or referral record".

If it is not possible at the time of the CPRD RDG application to provide operational definitions of exposures and/or outcomes because these will be elucidated during the course of the study, an acceptable alternative is to describe the process by which these definitions will be reached.

Data source/s

Applicants should also describe the data sources, where applicable, for determining the main exposures and health outcomes relevant to the study. Data sources might include, for example, primary care clinical records, prescription drug files, test records, administrative linked exposure/disease registries and GP questionnaires. Steps to validate exposure and outcomes are encouraged and may be suggested for diseases not previously studied in the database or for which there is commonly diagnostic uncertainty.

Covariates

A list of covariates to be included in baseline tables and statistical models as potential confounding variables and effect modifiers should be stated, including the data source/s from which these will be derived. This would suggest that reasonable steps to control for confounding will be taken.

Codes lists

Applicants should provide preliminary code lists for the main exposures and outcomes in order to demonstrate that they have an awareness of the practical issues involved in defining these, where appropriate. Code lists should be provided as appendices and not included in the body of the protocol.

Where relevant codes lists are absent, the procedure for developing them has not been described, or the use of codes from a previous study has not been proposed, protocols will be regarded as deficient in this respect. Given the nature of the medical coding system in use in UK primary care, it is advised that, where possible, a named clinician with experience of UK primary care is involved in the process of code list development.

Note that code sets must include numerical codes (Read/SNOMED-CT/CPRD Medcodes/ICD-10) and the text descriptors (Read term/ICD term). Code lists do not need to be finalised at the time of submission of the CPRD RDG application.

O. Data/ Statistical Analysis

Reviewer Assessment Criteria

In this section, reviewers will assess whether the analysis proposed are aligned with the research objectives, are broadly appropriate for the proposed research from an epidemiological and statistical point of view, and minimise potential risks of patient and/or practices re-identification or other information governance risks. For example, reviewers will assess whether analytical/statistical methods may 'single out' patients through investigation of outliers, case review by clinicians, or by conducting stratified analyses.

Application Requirements

All data management and data analysis to be performed should be covered in this section. Applicants should ensure that analytical methods proposed are consistent with all of the specific study aims and objectives listed, and with the particular study design. It is also important to ensure that this section is clear and specific about any comparisons which will be made (e.g. whether drugs classes will be compared or specific drugs). Mention of approaches to address potential problems of misclassification, bias, confounding, and missing data should be given.

Applicants should also make it clear whether sensitivity analyses will be undertaken, and outline the provisions to account for reverse causality, where this is felt to be a potential issue.

Analysis should be represented according to whether the study is hypothesis generating or testing but, in either case, the analytical methods to be used should be specified in the protocol. Please see below for suggestions on what may be included in your summary of statistical analyses for different types of studies.

Descriptive studies

Measures of central tendency (mean, median), variation, and correlation are often reported in these types of studies. Trend analysis is an important tool in descriptive studies.

Hypothesis Generating

Descriptive statistics to provide useful summaries about the sample and the outcome measures. Together with simple graphics analysis, descriptive statistics form the basis of virtually all quantitative analyses. Hypothesis generating analyses may include measures of disease frequency such as prevalence and incidence and time trend analyses.

Hypothesis Testing

Descriptive statistics to provide useful summaries about the sample and the outcome measures; measures of association to be derived and statistical tests to be conducted; pre-specified sub-group analyses including how the analysis will control for potential confounding. Where appropriate, specify the statistical modelling techniques to be used, giving some indication as to how models will be specified.

Multiple testing

Applicants are advised to consider the implications of multiple testing as the interpretation of p-values <0.05 (5%) as "statistically significant" may be threatened when many tests are carried out in a single study (Bland, 1995).

Approaches for handling missing data may include:

- cautious interpretation of findings
- Clear distinction between a pre-specified primary and several secondary hypotheses (with a commitment to caution regarding findings relating to secondary hypotheses)
- Bonferroni (Bland, 1995) or other formal statistical corrections

P. Plan for addressing confounding

Reviewer Assessment Criteria

In this section, reviewers will assess whether methods for addressing confounding have been considered, and where needed, are relevant given the study aims and objectives, study type, design and analyses proposed. Reviewers will also assess whether methods to consider confounding may increase the risk of patient and/or practice re-identification or other information governance risks during the conduct and/or publication of research findings. For example, reviewers will assess whether analytical/statistical methods may result in strata with <5 patients or 'single out' patients through other investigations.

Application Requirements

Purely descriptive studies are exempt from this requirement and can list 'Not applicable' in this section. All other studies should here provide some discussion of what will be done in the design and/or analysis to control for confounding.

Where methods to consider confounding may increase the risk of patient and/or practice reidentification or other information governance risks during the conduct or publication of your research, please outline mitigation approaches to minimise the risk of inadvertently identifying patients or practices during the conduct and/or publication of the study.

Q. Plans for addressing missing data

Reviewer Assessment Criteria

In this section and where applicable, reviewers will evaluate how missing data will be handled in the research and whether this may lead to spurious findings or incorrect conclusions that may undermine the benefits of the research.

Application Requirements

The potential for missing data is present in most studies and needs to be identified and addressed in this section of the protocol. In practice, missing data is most commonly of concern in relation to covariates, such as BMI and smoking, but would be of bigger concern if the relevant variable is an outcome or exposure.

Applicants should carefully consider their options in relation to how best to handle missing data in their research and expand on their choice and the resulting likely issues. Approaches should be considered that minimise the chance of bias, especially when data missingness is extensive and could result in a much reduced and potentially biased sample. The extent of missing data should be reported and recognised as a potentially important limitation, and applicants should state any assumptions made about the patterns of missingness for their analytical approach to be valid and outline any planned sensitivity analyses to further investigate potential selection biases due to missing exposure or covariate data

R. Patient or user group involvement

Reviewer Assessment Criteria

In this section, reviewers will evaluate whether patient or user group involvement have been considered during different stages of the research.

Application Requirements

It is expected that many studies will benefit from the involvement of patient or user groups in their planning and refinement stages, and/or in the interpretation of results, in their dissemination, and in informing plans for further work. This is particularly, but not exclusively, true of studies in which patients are to be contacted, and studies with interests in the impact on quality of life. Applicants should indicate whether patient/user groups will be engaged in any way and, if not, explain why patient/user groups will not be engaged. Applications which simply state 'Not applicable' will be returned as invalid.

S. Plans for disseminating and communicating study results

Reviewer Assessment Criteria

In this section, reviewers will evaluate whether there are any restrictions on publication of the research findings that may impact its public health value. For instance, whether the funder have a role in writing up the research or in deciding to submit the paper for publication. An assessment of whether there are potential risks of inadvertently re-identifying patients or practices during publication and dissemination of the outputs of the research will also be considered.

Application Requirements

There is an ethical obligation to disseminate findings of potential scientific or public health importance (e.g., results pertaining to the safety of a marketed medication). Authorship should follow guidelines established by the International Committee of Medical Journal Editors.

Applicants should list the following acknowledgements in publications resulting from studies using CPRD data:

- This study is based in part on data from the Clinical Practice Research Datalink obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. The data is provided by patients and collected by the NHS as part of their care and support. The interpretation and conclusions contained in this study are those of the author/s alone.
- Copyright © [YEAR], re-used with the permission of The Health & Social Care Information Centre. All rights reserved.

When reporting, applicants are advised to follow the principles outlined in the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) and any other relevant guidelines in the Enhancing the Quality and Transparency of health research (EQUATOR) network. The Consolidated Standards of Reporting Trials (CONSORT) statement refers to randomised studies, but also provides useful guidance, the principles of which may be applicable to observational hypothesis-testing studies

Where research is felt to provide important new evidence on the safety or effectiveness of a medicine or vaccine then pre-publication manuscripts may be sent by email to the MHRA at Pharmacovigilanceservice@mhra.gov.uk. Marketing Authorisation Holders should submit manuscripts for post authorisation safety studies, accepted for publication, as described in the Guideline on good pharmacovigilance practices (GVP) module VIII – Post-authorisation safety studies.

T. Conflict of interest statement

Reviewer Assessment Criteria

In this section, reviewers will evaluate applicants' conflict of interest statements to determine whether these may influence publication and/or communication of the research findings.

Application Requirements

Each applicant must provide a conflict of interest statement. The statement should be transparent about any sources of funding not already listed on the application including relevant financial interests of investigators/collaborators, and any relevant paid or unpaid positions held by investigators/collaborators.

U. Limitations of the study design, data sources, and analytic methods

Reviewer Assessment Criteria

In this section, reviewers will evaluate whether there are important limitations of the study that have not been considered or adequately considered and which may affect the conclusions drawn from the research.

Application Requirements

Limitations of the study such as issues relating to bias and confounding, misclassification, random error and generalisability etc... should be considered. Specific consideration of the potential impact on findings should be provided. For example, primary care databases contain little, if any, information about over the counter drug (OTC) usage. Applicants studying a class of drugs for which some products are available OTC should recognise which drug exposures are likely to be underestimated and discuss the expected impact on the findings.

Considerations about how important biases may arise from the study should also be addressed.

Researchers should consider situations in which certain prescriptions may not appear in the database. It should also be noted that presence of a prescription in a primary care database does not ensure that the prescription was then provided to the patient, issued by a pharmacy, and consumed by the patient. Applicants should contact enquiries@cprd.com with any queries.

V. References

Reviewer Assessment Criteria

In this section, reviewers will assess whether the supporting evidence about the research e.g. study background and methods, are linked to published scientific or other literature.

Application Requirements

Please provide a numbered list of references at the end of the protocol. The reference list should include the titles of the papers, but it is not necessary to include all the authors. A minimum of three authors is sufficient, and the Vancouver format for referencing is preferred.

List of Appendices

Reviewer Assessment Criteria

In this section, reviewers will inspect preliminary code lists for the main study exposure(s) and outcome(s) to assess their relevance and feasibility for conducting the research in CPRD data sources. Other documentation referred to in the application will also be assessed, where needed.

Application Requirements

Please provide all appendices related to this research protocol as separate documents.

Grant ID (optional)

Please provide a grant ID reference where this is applicable.

Other information

Data deletion

CPRD Dataset Agreement Terms and Conditions state that applicants will need to provide evidence that any received datasets have been deleted no later than 12 months following receipt. Applicants are required to keep a register of any copies made and will be asked to provide data destruction certificates for all copies or backups.

Applicants may apply for extensions to the 12 month period and should email rdg@cprd.com to discuss any request for an extension.

Confidentiality of research protocols

All research applications to the CPRD RDG process are held securely and confidentially at the CPRD. No information about study applicants or protocol content are released to third parties, other than in accordance with CPRD's Transparency Policy, without first seeking the agreement of the Chief Investigator of the study. Only applicants named on the research protocol can make enquiries about the protocol.

Ethical review of protocols

CPRD has obtained ethical approval from a National Research Ethics Service Committee (NRES), for all purely observational research using anonymised CPRD data; namely, studies which do not include patient involvement (which is the vast majority of CPRD studies). CPRD RDG committees review protocols for feasibility, public health benefits/risks and information governance risks, but may recommend that study-specific ethical approval is sought if ethical issues arise in relation to an individual study. Separate ethical approval will be required for any study which includes any form of direct patient involvement.

Voluntary registration of CPRD RDG approved protocols

Epidemiological studies are increasingly being included in registries of research around the world, including those primarily set up for clinical trials. To increase awareness amongst researchers of ongoing research, CPRD encourages voluntary registration of epidemiological research conducted using MHRA databases. This will not replace information on CPRD RDG -approved protocols that may be published on the CPRD website. It is for the applicant to determine the most appropriate registry for their study. Applicants should inform the RDG Secretariat on registering a protocol and provide the location.

Reporting findings

When reporting the findings of a CPRD RDG-approved protocol, authors are encouraged to indicate that the study was approved and should provide information on any deviations from the original protocol. For protocols approved from 01 April 2014 onwards, applicants are required to include the ISAC or RDG protocol number in journal submissions, with a statement in the manuscript declaring approval by the ISAC or RDG process, where applicable. If the protocol was subject to any amendments, the last amended version should be the one submitted.

Applicants are required to submit a copy of all peer-reviewed publications based on CPRD data to CPRD. Applicants should inform the CPRD of the publication outcome/s and, where appropriate, to send a copy or link of publications or a copy of funder's report summarising the research. These can be sent to CPRD enquiries (enquiries@cprd.com).

Please note that the CPRD reserves the right to audit the concordance between approved study protocols and published research.

It is essential that consideration is given to preserving confidentiality at the reporting stage. The possibility of unintentional (deductive) disclosure arises when cells with small numbers of patients are quoted. Applicants should note that, when reporting the data, CPRD policy is that no cell should contain <5 events.