Medicines & Healthcare products Regulatory Agency

CPRD

Completion of the CPRD ISAC Application Form and Protocol Information

This guidance document has been produced to help applicants complete the ISAC Protocol Application Form. Applications not completed in accordance with the guidance will be returned as invalid.

Key information

- By submitting a protocol application, applicants are:
  - declaring that they have read and understood the guidance on completing a protocol application form and required protocol information;
  - confirming that the submitted protocol application, protocol information, and all supporting documents are accurate;
  - confirming that the research team is suitably qualified and/or experienced to conduct the proposed study;
  - agreeing to abide by all contractual guidelines in relation to access to CPRD data and any other linked data where applicable;
  - agreeing to publication of summary information, should the protocol be approved by the ISAC, in accordance with CPRD's Transparency Policy;
  - agreeing to inform CPRD of the final outcome of the proposed study, including publication, prolonged delay, or termination;
  - agreeing to submit an amendment request to CPRD should applicants wish to deviate in any way from an approved protocol;
- Applications must be submitted by the Chief Investigator or Corresponding Applicant. Applications not submitted by the Chief Investigator or Corresponding Applicant may be returned as invalid;
- The application form and protocol information should be completed in Arial 10pt font only. Superscripts, subscripts, and footnotes are not permitted. The use of special characters and symbols should be avoided unless absolutely necessary. The use of graphs, charts, graphics, and images is not permitted within the application, but can be included in appendices if required;
- Each section of the application form should be completed in full. Any section not applicable to the proposed study should be completed as 'Not applicable' and justification must be provided. Applications with incomplete sections or no justification will be returned as invalid. Additional headings are not permitted, although sub-headings may be used within the protocol information where these clarify the respective section;
- Research protocols should be written in such a way that they can be reviewed as a standalone document. Where a protocol refers to previously published work, or another previously submitted protocol, sufficient information should be provided in the submitted protocol to allow the ISAC to adequately review the protocol without excessive reference to other studies;
- Word limits for certain sections must be adhered to, and protocols may be returned if word counts are exceeded;
- The completed protocol information should be no longer than 12 pages, not including appendices.
- Supplementary information such as code lists should be placed in one or more appendices, provided that the information is essential, and reference to these should be made within the protocol.
- Direct communication between applicants and members of the ISAC reviewing the protocol is not permitted. All communication regarding protocols should take place via the ISAC Secretariat, and only individuals named in the protocol may make enquiries regarding the protocol.
- Only applications submitted on the latest version of the ISAC Protocol Application Form (available at https://cprd.com/research-applications) will be accepted.

CPRD Transparency Policy

CPRD publish summary information of each research protocol approved by the ISAC on or after the 1st of July 2015. For more information on CPRD's Transparency Policy, please visit https://cprd.com/protocol-list.

Medicines & Healthcare products Regulatory Agency

CPRD

Part 1: Application Form

GENERAL INFORMATION ABOUT THE PROPOSED RESEARCH STUDY
Question 1: Study Title (Max. 255 characters) Please see the guidance in the Protocol Information on writing a suitable Study Title. 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. Applications with this section left blank will be returned as invalid.
Question 3: Chief Investigator The Chief Investigator will take responsibility for ensuring that the research is undertaken with full adherence to ISAC 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. Applications with missing information for the Chief Investigator will not be accepted and will be returned as invalid. The organisational affiliation of the Chief Investigator will be the sponsor of the proposed study.
Question 4: The Corresponding Applicant The name of a Corresponding Applicant, who will be the direct point of contact for the ISAC Secretariat, must also be provided for each application. It is acceptable for the Chief Investigator to be the corresponding applicant. Applications with missing information for the Corresponding Applicant will not be accepted and will be returned as invalid.
Question 5: Other investigators/collaborators Anyone who will have access to CPRD data, or who will contribute to the research study to a degree where they would be eligible for authorship, must be named in the ISAC protocol. As with the Chief Investigator and Corresponding Applicant, the name and individual institution/organisation email address of all study investigators/collaborators must be stated in the application form. **Curriculum Vitae (CV) requirements** Individuals who have previously submitted, or been included in, an ISAC protocol will have received a CV number from the ISAC Secretariat. This number should be quoted if applicable, or a new CV provided using the template which can be found at https://cprd.com/research-applications. All investigators/collaborators named on the ISAC application form must provide a CV number, or a new CV using the CPRD template.
Question 6: Experience/expertise available If applicable, specify how many previous CPRD Studies have been undertaken by members of the team, providing protocol numbers where necessary.

ACCESS TO THE DATA

Question 7: 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 should be the affiliation of the Chief Investigator.

Question 8: 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 9: 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 10: 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 licence institution name and address.

Please note that, for applicants requesting NCRAS data, CPRD will supply all primary care data regardless of whether any licence is in place.

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

Question 11: 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, by placing an 'X' in the relevant box(es), 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, in which case an 'X' should be placed in all three boxes and the organisation name and address provided. Applicants should copy and paste a new table for each additional location that stores, processes, or analyses the data.

Further guidance and information can be found on the ICO website.

Medicines & Healthcare products Regulatory Agency

CPRD

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 12: 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.

Investigators wishing to use CPRD Aurum data (either as a stand-alone or combination with Vision) must discuss this with a member of the CPRD Research Team (enquiries@cprd.com) and provide the name of the CPRD contact on the application form.

Correspondence with the CPRD should be via email to CPRD Enquiries (enquiries@cprd.com), and applicants should provide the enquiry reference number when requested on the form.

Question 13: Requests to access linked data

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 an ISAC application. Applicants must provide the name of the CPRD researcher, along with the date of contact, and enquiry reference number.

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

- NCRAS Cancer Patient Experience Survey (CPES) data
- NCRAS Systemic Anti-Cancer Treatment (SACT) data
- NCRAS National Radiotherapy Dataset (RTDS) data
- Mental Health Services Data Set (MHDS)
- Practice Level Index of Multiple Deprivation
- Practice Level Index of Multiple Deprivation Domains
- 2011 Rural-Urban Classification at LSOA level

Area level linkages refer to data held by CPRD but that require discussion with, and approval from CPRD before release and may be subject to additional risk minimisation measures.

Applicants seeking access to NCRAS data must complete a Cancer Dataset Agreement Form (available from CPRD on request) and submit this to the ISAC 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.

Note that CPRD will not supply both patient level Index of Multiple Deprivation (IMD) data and Townsend Scores for the same study. It is possible to obtain both a patient and practice level for the same measure. Please ask CPRD for advice regarding this (enquiries@cprd.com).

Question 14: Requesting non-standard data linkage

Investigators wishing to link to a dataset not listed in question 13 must have received approval for such a linkage from the CPRD prior to submitting an ISAC protocol. Applicants must provide the reference number for the approval of the linkage in their protocol application. Applications that do not include a reference number will be returned as invalid.

Applicants wishing to link to a dataset not listed in question 13 should contact the CPRD Research Team on (enquiries@cprd.com) to discuss your requirements**.**

Question 15: 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.

VALIDATION/VERIFICATION

Questions 16: 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 17: Patient or GP questionnaires or patient contact

Questionnaires for healthcare professionals or patients must be reviewed and approved by the ISAC 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 in the format in which 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 are advised to seek advice on questionnaire design by submitting an enquiry to enquiries@cprd.com, and should quote an enquiry reference number to support their protocol application where applicable.

The ISAC 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 ISAC require evidence of additional ethical approval for contact relating to patients.

Part 2: Protocol Information

A. Study Title (Max. 255 characters)

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 over the 255 character limit will be returned as invalid.

B. Lay Summary (Max. 250 words)

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 methodological research studies, there should 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, or go beyond the word count, will be returned as invalid.

C. Technical Summary (Max. 300 words)

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 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 tests to be used.

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.

Avoid vague references, 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 technical summary should also specify how linked datasets will be used, for example “Hospital Episode Statistics (HES) data will be used to determine hospitalisations.”

Technical summaries that do not adhere to these guidelines, or go beyond the word count, will be returned as invalid.

D. Outcomes to be Measured

This section should clearly list the primary and 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.

E. Objectives, Specific Aims and Rationale

A general objective should normally be provided, followed by one or more specific and related aims. Studies with a large number of specific aims may be considered too extensive and often do not describe all aims in sufficient detail.

The ISAC will carefully consider whether all the proposed aims have been addressed in later sections, particularly with regard to analysis of the data. Applicants should also provide a satisfactory statement regarding the rationale/need for the present study.
Include the following:

i. A description of the knowledge/information to be gained from the study (research objectives).
ii. A list of the measurements to be made, and any hypotheses to be tested (specific aims). The protocol should distinguish between a priori research hypotheses and hypotheses that are generated based on knowledge of the source data.
iii. An explanation of how achievement of the specific aims will further the research objectives (rationale).

F. Study Background

Applicants are expected to provide evidence that they have thoroughly reviewed the relevant scientific literature and, in particular, that they are aware of any previous studies conducted in GPRD or CPRD which have been published.

Applicants should explain the reason for the study, and include other essential background information, such as the findings of similar studies and other related research.

Ensure that you refer to any previous ISAC protocols that are related to your study. Any reference to a previous ISAC protocol should be accompanied by the ISAC protocol number e.g. 15_101 (even if this is being cited as published work).

G. Study Type

Specify whether the study will be primarily descriptive, hypothesis generating, hypothesis testing, or a methodological piece of research. The ISAC recognises 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

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.

Confusion sometimes arises with regard to matched control groups, for example leading to a comparative cohort study being described as "case-control".

I. Feasibility counts

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 who do not have access to CPRD data should request a simple feasibility count by emailing enquiries@cprd.com. Simple feasibility counts are limited to the number of patients with a particular condition or prescribing in the database or linked data source. Simple counts may be stratified by year of the event, age at event, and gender. To facilitate the counts, applicants should provide preliminary code list/s for the condition/s and/or drug/s of interest. To enable code list development, code browsers specific to the medical codes (Read codes) and product codes used in the CPRD can be requested from enquiries@cprd.com.

J. Sample size considerations

All protocols must include some consideration of whether the sample-size/power will be sufficient to meet the scientific objectives of the study.

Firstly, all protocols should include an estimate of the expected numbers of patients, exposures, or outcomes (as appropriate) that will be available. Investigators might arrive at such estimates by conducting/requesting simple feasibility counts of the approximate numbers in the CPRD during the study period.

Secondly, for a hypothesis testing study, it is necessary to demonstrate that the expected numbers are sufficient to investigate the stated hypotheses 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.

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.

K. Planned use of linked data (if applicable):

Any proposed use of linked data sets must be appropriate to the research. This will be assessed against statements made on the ISAC 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 Statistic (HES), Office of National Statistic (ONS) Mortality data, Cancer Registry data, please describe why the linkage data is necessary for the study and how it will be used.

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 an ISAC 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 ISAC application form. The ISAC will not approve a study requesting linked data 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 ISAC submission. It essential that any necessary legal/ethical approvals are in place for any non-standard linkage to take place before submitting to ISAC.

L. Definition of the Study population

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:
   - whether only permanently registered acceptable patients will be included;
   - whether only up-to-standard follow-up will be considered;

b) 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;
c) Describe the study population in terms of inclusions, exclusions, and the data used for each (clinical, referral, test, therapy, immunisation, consultation). Reference should be made to provisional code lists for inclusion & exclusions specified;
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. Considerations about how important biases may arise from the study should also be addressed.

M. Selection of comparison group(s) or controls

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 inclusions, exclusions, and the data used for each (clinical, referral, test, therapy, immunisation, consultation). Reference should be made to provisional code lists for inclusion & exclusions specified;
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

<u>Defining Exposures and Outcomes</u>
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. 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”.

It is rarely enough to provide a simple diagnosis as the basis for an exposure - the coding basis for it and/or process by which an exposure will be accepted as valid should be described. A specific sub-chapter of the British

National Formulary (BNF) may be used to define drug exposures but issues such as numbers of prescriptions, time windows, dose and whether a specific drug or class of drugs defines exposure usually need to be addressed. The same principles apply to outcome definitions.

If it is not possible at the time of the ISAC 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 key 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. This would suggest that reasonable steps to control for confounding will be taken. Operational definitions should also be provided for covariates of interest including the data source/s from which these will be derived.

Codes lists

Applicants should provide preliminary code lists for exposures and outcomes in order to demonstrate that they have an awareness of the practical issues involved in defining these, where appropriate. While code lists may be included as appendices to the protocol, ISAC will not review these in any detail.
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/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 to the ISAC.

O. Data/ Statistical Analysis

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. See below for a summary of the statistical analyses that the ISAC would expect to see 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 is expected. Together with simple graphics analysis, descriptive statistics form the basis of virtually all quantitative analyses. Hypothesis generating analyses also 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 is expected. The measures of association to be derived and statistical tests to be used must be specified. Sub-group analyses should be pre-specified, and applicants should indicate 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 (e.g. forward stepwise regression or backward stepwise regression).

Multiple testing

ISAC review includes consideration of statistical and methodological aspects of a protocol. The interpretation of p-values less than 0.05 (5%) as "statistically significant" is threatened when a large number of tests are carried out in a single study (Bland, 1995). ISAC protocols with large numbers of tests are unlikely to be approved unless they address this issue.

A number of approaches are possible, including:

- cautious interpretation;
- 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;
- other approaches if well justified and/or supported with references.

Whatever method or methods of accounting for multiple testing is proposed, it should be clearly specified in advance and not subject to the later discretion of the investigator. When multiple publications are based on one protocol, consideration should be given to accounting for multiple testing both within and across publications.

P. Plan for addressing confounding

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 they are doing in the design and/or analysis to control for confounding.

Q. Plans for addressing missing data

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.

The ISAC is not prescriptive about how missing data should be handled but, at the very least, expects some specific recognition of the likely issues to be present in the protocol, and for missing data to be reported and recognised as a potential limitation.

R. Patient or user group involvement (if applicable)

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. Where applicable, applicants should indicate whether patient/user groups will be engaged in any way.

Please see the 'Patient Involvement' statement from the ISAC on incorporating patient/user group involvement in your study. (http://www.cprd.com/_docs/Patient%20Involvement_Aug2012.pdf).

S. Plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication, and conflicts of interest

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.

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. In cases where multiple publications are likely to arise, a publication plan may be requested by the ISAC.

Where research is felt to provide important new evidence on the safety or effectiveness of a medicine or vaccine then pre-publication manuscripts can 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.

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

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

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.

U. References

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

Please list all appendices related to this research protocol and submit these as separate documents to the ISAC along with your application. Please do not embed documents in the research protocol.

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 isac@cprd.com to discuss any request for an extension.

Confidentiality of research protocols

All research applications to the ISAC 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

The 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). The ISAC is responsible for reviewing protocols for scientific quality 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 ISAC 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, the ISAC encourages voluntary registration of epidemiological research conducted using MHRA databases. This will not replace information on ISAC-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 ISAC Secretariat on registering a protocol and provide the location.

Reporting findings

When reporting the findings of an ISAC-approved protocol, authors are encouraged to indicate that the study was so 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 protocol number in journal submissions, with a statement in the manuscript declaring approval by the ISAC. 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 the CPRD Business Development mailbox (bd@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 fewer than 5 events.