Medicines & Healthcare products Regulatory Agency

CPRD

CPRD Aurum
Frequently asked questions (FAQs)

Version 1.0

Date: 6 December 2017

Author:

Helen Booth and Daniel Dedman, CPRD, UK

NHS
National Institute for Health Research

Documentation Control Sheet
During the course of the project it may be necessary to issue amendments or clarifications to parts of this document. This form must be updated whenever changes are made and should be filed inside the front cover of the new or amended document.

Version	Summary of Change	Prepared By	Date	Reviewed By	Date
1.0		Helen Booth & Dan Dedman	06/12/2017		

Summary of Changes

Contents

Introduction

What is CPRD Aurum?

Clinical Practice Research Datalink (CPRD) is implementing the phased roll-out of a new database of de-identified coded primary care records for use in public health research. These data are contributed by general practices that use the Egton Medical Information Systems electronic patient record system (EMIS Web) software and the product will be called CPRD Aurum (the Latin word for gold). The first database version will be released on 11th December 2017.

How does CPRD Aurum differ from the CPRD GOLD database?

CPRD Aurum contains data contributed by practices using EMIS Web software, whilst CPRD GOLD holds data from a different software provider named Vision. Due to differences in the structure and coding of the data between the two systems the research databases will be released as separate data offerings. There are no plans to integrate the databases during the phased roll-out period. The basic structure and content of the research tables in the CPRD Aurum database is similar to the CPRD GOLD database. For example, CPRD Aurum has patient, practice, consultation observation, referral and drug issue tables. There are unique identifiers in each table that enable them to be linked in various ways according to researcher requirements. The main differences are outlined in later sections of this document.

What is the population coverage of CPRD Aurum?

The first release of the CPRD Aurum database (December 2017) contains data from:

- 186 practices
- 5,544,687 patients (permanently registered historic and current patients)
- 1,958,684 patients (permanently and currently registered patients)

For further information on the data included in each monthly build of the CPRD Aurum database, please contact enquiries@cprd.com.

Currently, all practices contributing data to the CPRD Aurum database are in England. CPRD is exploring options for the inclusion of data from EMIS Web practices in the devolved nations.

What will the CPRD Aurum database contain?

When an EMIS Web practice agrees to contribute data to CPRD Aurum, CPRD receives a full collection of the coded part of their electronic health records; this includes data on deceased patients and those who have left the practice. Where a practice has switched from another software system, such as Vision, CPRD will receive all historical data, including information that was recorded using the old software system, in addition to prospectively added data. Consequently, there is some overlap between the CPRD GOLD and CPRD Aurum databases where practices have contributed via different software systems. Further information on dealing with this in studies intending to use data from both databases is provided in the section on 'Duplication between the CPRD GOLD and CPRD Aurum databases'.

What will happen during the phased roll-out of CPRD Aurum?

The initial CPRD Aurum database release will include data from a sub-group of practices using EMIS Web software that have already agreed to contribute data to the CPRD database. The large volume of data that CPRD is on-boarding from these practices and implementation of software updates in the practices means that incorporating all practices will take some time. Additional practices will be added over the subsequent monthly builds. As these numbers are changing rapidly, further details on the numbers of practices and patients will be released with the monthly database build release notes that can be obtained by contacting enquiries@cprd.com.

How can I access CPRD Aurum?

The ISAC application process for CPRD Aurum data will be the same as for CPRD GOLD, but applicants should discuss their proposals with a CPRD researcher before submitting an application, to ensure an understanding of the data structure and implications for study design. Please see the section 'Guidance on applying to ISAC' for further advice on ISAC applications to use the CPRD Aurum data. The CPRD Aurum data is a separate product to the CPRD GOLD data, and will be charged on a dataset basis. The data will be released to clients as study specific datasets only following ISAC approval. CPRD Aurum will not be accessible via the CPRD GOLD online tools during the phased roll-out. Options for client access to CPRD Aurum are being explored during the phased roll-out.

What is the cost for accessing CPRD Aurum data?

During the phased roll-out, CPRD Aurum datasets will be priced the same as CPRD GOLD datasets. This may be subject to change over time, but we will endeavour to update clients promptly should this be the case. Please contact enquiries@cprd.com for a quote for your specific study.

How will I receive CPRD Aurum data?

The process of dataset delivery will be the same as for CPRD GOLD study-specific datasets. Once the research protocol has been granted ISAC approval the dataset will be assigned for processing and delivery by the CPRD Observational Research team. CPRD researchers will agree a dataset specification with the client. Following execution of an appropriate licence agreement between the client and CPRD and payment of any associated fees, the data will be extracted and sent to the client via secure file transfer protocol (SFTP). During the phased roll-out period, the proposed timeline for dataset delivery will be 30 working days starting from agreement of the data specification between the client and CPRD. Should there be any particular concerns about a dataset due to its complexity, CPRD will communicate this to the client at the earliest possible opportunity.

Are there differences in the ISAC application for CPRD Aurum?

The ISAC application process for CPRD Aurum data will be the same as for the CPRD GOLD data. Please see the 'Guidance on applying to ISAC' section for further information.

How will I know if the CPRD Aurum data are suitable for my research needs?

CPRD will be providing simple feasibility counts free-of-charge to clients.

- Simple feasibility requests are limited to counts of patients or events recorded in a specified period.
- Simple counts should include no more than three medical and/or prescribing definitions combined, in a single request.
- Counts may be restricted to one or more of: study period, patient's age, gender and period of follow-up in CPRD Aurum.
- Counts may be stratified by calendar year, gender or age-band only.
- Users are expected to provide the relevant medical codes (Read, SNOMED, ICD-10 or OPCS codes) or therapy codes to identify events of interest in the respective data sources (code browser facilities will be provided to users).
- Examples of simple feasibility counts based on 1-3 criteria are outlined in the table below.
- No denominators will be provided as part of the simple feasibility count service i.e. CPRD can provide the numerator (prevalent or incident counts), but not prevalence or incidence of an exposure or disease.

Examples of simple feasibility counts include:

Examples of counts based on one criterion
The total number of metformin prescriptions recorded in CPRD Aurum data during 01/01/2004 - 31/12/2015, stratified by calendar year.
Examples of counts based on two criteria
Separate counts of the total number of patients with at least one prescription for metformin or sulfonylureas recorded in CPRD Aurum data during 01/01/2004 - 31/12/2015, stratified by calendar year of first prescription. Two separate counts will be provided.
Examples of counts based on three criteria
Separate counts of the total number of patients with at least one prescription for metformin, sulfonylureas or thiazolidinediones recorded in CPRD Aurum data during 01/01/2004 - 31/12/2015, stratified by calendar year of first prescription. Three separate counts will be provided.

To request a free simple feasibility count you will need to send CPRD a code list of medical events or prescriptions that can be used with the CPRD Aurum dictionaries. Please see the section 'Guidance on generating code lists' for further information. We are happy to advise you on this process. Clients can request more sophisticated feasibility counts and are advised to discuss their needs with a CPRD researcher so that a service quote can be provided if relevant. Please email enquiries@cprd.com to discuss your needs further.

What are the major differences between the CPRD Aurum and CPRD GOLD databases?

The table below outlines some differences between the CPRD GOLD and the CPRD Aurum data that you may find useful if you are considering using the CPRD Aurum data alone, or in combination with CPRD GOLD, for a study. Differences that are temporary have been highlighted. Further information can be found in the 'Points for consideration' section.

Difference	Context	CPRD Aurum	CPRD GOLD	Further information
Medical coding*	The NHS is moving to universal coding using SNOMED-CT	Source coding in CPRD Aurum uses a mixture of Read 2, SNOMED and local EMIS codes.	Source-coding in CPRD GOLD is based on Read coding. SNOMED coding will be added to the GOLD database, and this will be mapped 1:1	Advice on producing code sets in the CPRD Aurum (& CPRD GOLD) data is provided here.
Product coding		Product coding in CPRD Aurum uses DM+D	Product coding in CPRD GOLD uses Gemscript	Advice on producing code sets in the CPRD Aurum (& CPRD GOLD) data is provided here.
Test and value recording*		In CPRD Aurum test and value results are recorded in the Observation table	In CPRD GOLD test and value results are recorded in the Additional Clinical Details and Test tables	Advice on finding measurements in CPRD Aurum can be found here.
Vaccination recording		In CPRD Aurum vaccinations are recorded in the Observation table	In CPRD GOLD vaccinations are recorded in a separate immunisation table	
Linkages		At the initial release, linkages will not be available for CPRD Aurum. All standard linkages will become available in 2018.	No change	This is a temporary difference.
Derived variables	CPRD offers a number of derived variables to facilitate research in the CPRD GOLD database, such as a derived death date, acceptable flag, up-to-standard date	The initial CPRD Aurum database release versions will not include derived variables. These will be added from 2018.	No change	Some advice is provided on variables to be introduced, and alternative approaches that CPRD users may wish to adopt in the interim here. This is a temporary difference.
Consultations		Clinical observation recording may be added outside of the context of a consultation. Consultation identifiers may not be present	Observations are all linked to a consultation by a consultation identifier	

* See 'Points for consideration' for further information

Points for consideration

Duplication between the CPRD GOLD and CPRD Aurum databases

A number of GP practices that previously contributed data to CPRD GOLD are now supported by EMIS Web software and have agreed to contribute data to CPRD Aurum. In this situation, CPRD will hold duplicate historical data for such practices in the CPRD GOLD and CPRD Aurum databases. If you are planning to use data from both databases for a study, CPRD can provide a bridging file to identify the overlapping practices and dates. We are also able to remove the relevant practices from a CPRD Aurum or CPRD GOLD dataset if preferable.

Medical and Drug dictionaries

EMIS Web software enables clinicians to record some observations using local codes, rather than Read or SNOMED CT codes. Where possible, local EMIS codes have been mapped to SNOMED CT, but you may still find items in the medical dictionary that are not mapped to either Read or SNOMED codes. To add value to the CPRD Aurum drug dictionary, CPRD has mapped it to the Dictionary of Medicines and Devices (DM+D).

For advice on producing code sets for CPRD Aurum, see the section Guidance on generating code sets.

Recording of coded clinical information

EMIS Web software offers greater opportunity for GPs to use free text rather than coding to record clinical observations. CPRD does not receive free text due to information governance restrictions which may mean that there are systematic differences in the recording of observations between CPRD GOLD and CPRD Aurum. CPRD researchers have conducted a preliminary evaluation of data sourced from EMIS Web as compared to data sourced from Vision for research and have found similar prevalence estimates for common conditions, including heart failure and chronic kidney disease, between the two databases. As CPRD increases its understanding of these issues, we will share our findings with clients. However, this is a difference that you should be aware of when preparing definitions and code lists to be used with the CPRD Aurum data.

Identifying clinical measurements

In CPRD Aurum, clinical measurements such as blood pressure, height and weight are recorded in the observation table. Relevant measurements should be identified via a medical code list and presence of a value to filter observations. See 'Example research questions' for further information.

How are referrals recorded

Referral information is recorded in two separate tables: the Observation table contains details about the reason for the referral (as a medical code) and event date; the Referral table contains details about the source and target organisation, referral urgency and service type. The complete Referral record can be reconstructed by linking the Observation and Referral records using the observation identifier ('obsid') which is present in each table.

What is the problem table?

GPs are able to assign 'problem' status to observations in the EMIS Web software. This is a way of enabling GPs to view a patient's medical history by clinical issue rather than in chronological order. For instance, classifying a patient's diabetes as a problem would allow them to link observations, such as diabetes medication reviews and blood tests, in order to better monitor their diabetes management. This table may contain valuable information in addition to the observation table but it is important to note that there could be variation in the way that different GPs use the 'problems' recording option. Problem information is recorded in two separate tables: the Observation table contains details about the nature of the problem (as a medical code) and event date; the Problem table contains further details including duration, clinical significance, whether the problem remains active. The complete Problem record can be reconstructed by linking the Observation and Referral records using the observation identifier ('obsid') which is present in each table.

Linked data

Linkage of CPRD Aurum to all the standard patient-level linked datasets available for the CPRD GOLD database will be commenced from 2018. The process of linking CPRD Aurum to other datasets will be the same as for CPRD GOLD.

Guidance on applying to ISAC

The ISAC application process for the CPRD Aurum data will be the same as for the CPRD GOLD data. The form offers a tick-box option to request 'EMIS' (CPRD Aurum) data. If you are considering a study using CPRD Aurum we would expect you to speak to a senior researcher at CPRD for advice on the feasibility of your proposed study. Enquiries can be sent to enquiries@cprd.com and will be directed to the appropriate person. If you are planning on conducting a study using both CPRD GOLD and CPRD Aurum data, you should consider potential differences in the databases that may impact your results. An understanding of these potential differences and the implications for your study conduct and findings should be demonstrated in your ISAC protocol.

Guidance on generating code sets

The CPRD Aurum dictionaries are more complex than those for the CPRD GOLD database, as described previously. In the first instance, dictionaries will be provided as text files that can be imported into standard statistical software to enable code searching. The dictionaries will also be available through the CPRD code browser. The CPRD code browser and a user guide can be requested by contacting enquiries@cprd.com. If you are already using the code browser to search the CPRD GOLD dictionaries you will still need to contact us to download the browser containing the CPRD Aurum dictionaries.

At CPRD, our experience of working with both CPRD GOLD and CPRD Aurum databases has indicated that developing reusable search strategies for code list generation is preferable to maintaining static code lists. These strategies may combine searches of descriptor terms (for example, CKD or chronic kidney disease) and hierarchical classifications such as Read, to identify codes for inclusion or exclusion. The benefit of this strategy is that it can be re-used at a later date to update code lists. Further, it can be applied to generate code lists for both the CPRD GOLD and CPRD Aurum databases simultaneously rather than having to replicate a code list developed in one database. For large and complex code lists replication of an existing code list may be difficult.

Further advice is available from a CPRD researcher via enquiries@cprd.com.

Derived variables

The initial database releases of CPRD Aurum will not include derived variables, some of which are routinely included with CPRD GOLD. These include region (Practice table), CPRD consultation type (Consultation table) and CPRD death date (Patient table) as well as the data quality metrics, acceptable patient flag (Patient table) and up-to-standard date (Practice table). Fields for these variables will be included in the data tables, but will either not be populated or may contain data without a lookup. Please see the Data Specification document for further information.

CPRD researchers will be working on developing data quality metrics for CPRD Aurum data which will be added in later release versions.

Acceptable patient flag

A CPRD algorithm is being implemented that will be similar to that used to define acceptable patients in the CPRD GOLD database. Only ‘regular’ patients (patienttypeid=3) are flagged as acceptable for

research as they are more likely to have a complete patient record and longitudinal follow-up data. In the interim, we suggest you include regular patients in studies using CPRD Aurum data unless there is a clear justification for you to include temporarily registered patients for your study.

Up-to-standard date

CPRD GOLD data includes a practice-level data quality metric, the 'up-to-standard (UTS) date'. The UTS date for CPRD GOLD data is calculated based on the assurance of continuity in data recording (gap analysis), and avoidance of use of data for which transferred out and dead patients have been removed (death recording).

Derived death date

The death date as recorded in EMIS data (*emis_ddate*) is known to be inaccurate in some cases. For example, it may reflect the date that the death was recorded on the system or the GP notified, rather than the date of occurrence.

A CPRD algorithm is being developed to provide a more realistic estimate of the date of death (*cprd_ddate*), using all available information in the patient record. As far as feasible, the CPRD Aurum algorithm will be consistent with the algorithm used for the CPRD GOLD database. Until this data item is available, researchers are advised to treat the *emis_ddate* with great caution. They should combine the *emis_ddate* with information from the Observation records where a clinical code indicates that a death has occurred.

Example research questions

This section will be updated with additional problems and solutions as our understanding of the data increases. If there is a particular question on which you would like further information, please email enquiries@cprd.com and we will be happy to advise you.

Finding results of tests, investigations and other clinical measurements

Numeric results of tests, investigation and other clinical measurements are recorded in the Observation table, as a combination of:

- A medical code [medicalcodeid] which describes the parameter being record. The text description [term] associated with the code can be obtained from the medical dictionary.
- A numeric (integer) value [value]
- A unit of measurement [numunitid]

- Optionally there may be two values to define the lower limit [numrangelow] and upper limit [numrangehigh] of the 'normal range' for the measurement.

Examples of numeric results for tests, investigations and clinical measurements:

1. Initially, the medical dictionary should be searched for Read terms that could be used to record the measurement of interest. For blood pressure, a search using the terms ‘systolic blood pressure’ and ‘diastolic blood pressure’ can be used to produce a code list to identify relevant observations.
2. The codelist can then be applied to the Observation table to identify observations that include measurements for blood pressure in the ‘*value’* field.
3. The identified observations can then be cleaned by checking whether a value has been recorded, and using the *‘numunitid’* to check the measurements have the appropriate unit (mmHg).

Additional documentation

The following documents may be useful for background information.

NHS Digital: SNOMED CT resource
https://digital.nhs.uk/snomed-ct

Code set generation
Clinical code set engineering for reusing EHR data for research: A review.
http://www.sciencedirect.com/science/article/pii/S1532046417300801