Medicines & Healthcare products Regulatory Agency

CPRD

Requesting linked data from CPRD

Version 1.2
Date: 01 July 2021

Access to linked data

Access to linked data is dependent on either an approved protocol or feasibility study application.

Following protocol approval:

- For multi-study licence holders, all requests must be submitted through the linkage request service, please see **Appendix 1** for further information.
- For single study datasets, the Observational Research team will be in touch and the linked data will be supplied alongside the primary care data.

For feasibility studies, the Observational Research team will be in touch following approval of your application.

Submitting a request for linked data

Researchers requesting linked data must complete the CPRD Linkage Request Form and email this to CPRD (enquiries@cprd.com), together with the relevant patient/code list(s) as outlined in **Appendix 1**.

All files should be provided as tab-delimited text files (.txt) and zipped into a single file, prior to transfer to the CPRD. For zipped files >20MB, please contact CPRD (enquiries@cprd.com) for further advice.

Requests for linked data must be submitted by either the Chief Investigator (CI) or a collaborator named on the approved protocol, copying the CI. Applicants must also ensure that only linked data approved in their protocol are requested.

All linkage requests will be acknowledged by CPRD within 2 working days of receipt. Linked data will be provided, by secure transfer, within 10 working days of receipt of a valid request. If the form is incorrectly completed, the patient/code list(s) are not in the correct format, or the patients are not eligible for the linkages requested, the request will not be processed until these issues are resolved.

To ensure that linked data requests are processed in an efficient and timely manner, please follow the guidance provided in **Appendix 2** on how to identify patients eligible for linkage and how to request events based on clinical code lists for cohort identification (**Appendix 3**).

Accessing linked data not included under an annual licence

For access to linked data that is not currently covered by an existing contract, an additional data access agreement will be required prior to requesting any linked data for your study. Please contact the CPRD Contracts team (enquiries@cprd.com) for further information.

Contractual Acknowledgements at Publication

The following statements below should be included in publications arising from the use of CPRD GOLD, CPRD Aurum and/or linked data.

Any Publication arising from:

a. use of the CPRD primary care data should include the statement “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”. The Customer will ensure that the description of the CPRD Database in any such Publication is accurate and current, and agrees to request publication of a correction to any published description which CPRD deems to be inaccurate if so, requested by CPRD;

b. use of Office of National Statistics (ONS) data should acknowledge ONS as the provider of the ONS data contained within the CPRD Data and include the statement “The interpretation and conclusions contained in this study are those of the author/s alone”;

c. use of (Hospital Episode Statistics) HES data/ONS data should include the statement “Copyright © (year), re-used with the permission of The Health & Social Care Information Centre. All rights reserved”. Users should ensure that the description of the HES data/ONS data in any such publication is accurate and current, and agree to request publication of a correction to any published description which CPRD or the linked data owner deems to be inaccurate, if so requested by CPRD or the linked data owner;

d. use of Public Health England (PHE) data should include the statement “Public Health England (year): [Title]. [Version]. [Publisher]. [Resource Type] e.g. e.g. Public Health England (2015): National Cancer Registration Data. (CAS Snapshot 15.01), Public Health England (dataset)

e. use of Office of Population Censuses and Surveys (OPCS) codes should include the acknowledgement: "The OPCS Classification of Interventions and Procedures, codes, terms and text is Crown copyright (2016) published by Health and Social Care Information Centre, also known as NHS Digital and licenced under the Open Government Licence available at www.nationalarchives.gov.uk/doc/open-government-licence/open-government-licence.htm".

Appendix 1: Linkage request requirements

User Requirements	CPRD Requirements	What CPRD will provide
The study population will be identified using primary care data only, but I need data from one or more linked data sources for these patients.	Researchers should provide a single list of patient identifiers for individuals who are eligible for linkage to the data source/s approved in the protocol. See **Appendix 2** for guidance on how to identify the list of patients who are eligible for linkage in your study.	CPRD will provide data variables and event records from approved data source/s for the study populations comprising of <600K patients. For study populations comprising of ≥600K patients, data minimisation approaches may be applied prior to data release. Please contact the CPRD (enquiries@cprd.com) for further information if your study cohort comprises of ≥600K patients and you require linked data.
The study population will be identified based on events from linked data only. Primary care data may be used to apply additional inclusion and exclusion criteria.	Researchers should first provide the list of codes for identifying events of interest in the linked data source/s approved in the protocol. See **Appendix 3** for guidance on how to prepare your code lists for identifying events in linked data sources.	CPRD will initially provide only the relevant events of interest and limited data variables (patient identifiers, codes, events and event dates) to enable further cohort identification and finalisation.
	Researchers should finalise the study population and then provide a single list of patient identifiers for individuals who are eligible for linkage to the data sources approved in the protocol. See **Appendix 2** for guidance on how to identify the list of patients who are eligible for linkage in your study.	CPRD will provide data variables and event records from the approved data source/s for the study populations comprising of <600K patients. For study populations comprising of ≥600K patients, data minimisation approaches may be applied prior to data release. Please contact the CPRD (enquiries@cprd.com) for further information if your study cohort comprises of ≥600K patients and you require linked data.
The study population will be identified using events from both primary care and linked data sources.	Researchers should first provide the list of codes for identifying events of interest in the linked data source/s approved in the protocol. See **Appendix 3** for guidance on how to prepare your code lists for identifying events in linked data sources. Researchers should finalise the study population and then provide a single list of patient identifiers for individuals who are eligible for linkage to the data sources approved in the protocol. See **Appendix 2** for guidance on how to identify the list of patients who are eligible for linkage in your study.	CPRD will initially provide only the relevant events of interest and limited data variables (patient identifiers, codes, events and event dates) to enable case de-duplication, application of additional inclusion and exclusion criteria and additional case identification, alongside CPRD primary care data. CPRD will provide data variables and event records from the approved data source/s for the study populations comprising of <600K patients. For study populations comprising of ≥600K patients, data minimisation approaches may be applied prior to data release. Please contact the CPRD (enquiries@cprd.com) for further information if your study cohort comprises of ≥600K patients and you require linked data.

Appendix 1: Linkage request requirements		
User Requirements	**CPRD Requirements**	**What CPRD will provide**
The outcome(s) and/or covariate(s) will be identified based on one or more linked data sources and these are based on codes.	Researchers should provide the list of codes for identifying events of interest in the linked data sources approved in the protocol. See **Appendix 3** for guidance on how to prepare your code lists for identifying events in linked data sources.	CPRD will provide only those events with a corresponding code of interest and limited data variables (patient identifiers, codes, events and event dates).
The outcome(s) and/or covariate(s) will be identified based on one or more linked data sources and these are not based on codes e.g. hospital admission data, dates of death, patient level socioeconomic data.	Researchers should provide a single list of patient identifiers for individuals who are eligible for linkage to the data sources approved in the protocol. See **Appendix 2** for guidance on how to identify the list of patients who are eligible for linkage in your study.	CPRD will provide data variables and event records from the approved data source/s for the study populations comprising of <600K patients. For study populations comprising of ≥600K patients, data minimisation approaches may be applied prior to data release. Please contact the CPRD (enquiries@cprd.com) for further information if your study cohort comprises of ≥600K patients and you require linked data.
The outcome(s) and/or covariate(s) will be identified based on linked practice level data for my study.	Researchers should provide the list of unique practice identifiers included in the study.	CPRD will provide the required linked practice level data from the data source/s approved for your study.

Appendix 2: How to identify primary care patients who are eligible for linkage

1. Request the following files from CPRD (enquiries@cprd.com):

- The list of patient and practice files (CPRD Denominator files) for the primary care database build that you wish to use for your study (e.g. Aurum June 2021).
- Linkage eligibility files (linkage_eligibility.txt and linkage_coverage.txt) and supporting documentation for the linkage set you will use for your study (e.g. linkage set 21). Please note that for new research studies, CPRD will only provide linked data from the latest linkage set available. Earlier versions of linked data may be provided for ongoing studies conditional on adequate justification. Please contact the CPRD (enquiries@cprd.com) to confirm the latest version of linked data available.

2. Use the CPRD Denominator files to apply patient acceptability criteria for research and any relevant time constraints (e.g. removing patients that died before the start of your study).

3. Combine the list of acceptable patient identifiers from **Step 2** with the list of patient identifiers in the linkage eligibility file (*linkage_eligibility.txt*). Retain only those patients who appear in both files.

4. Further identify those patients in **Step 3** who are eligible for linkage to the data source/s approved for your study. For example, to identify patients eligible for linkage to Hospital Episode Statistic (HES) Admitted Patient Care data and Office for National Statistics Death Registration data, you should retain those patients where variables *hes_e* AND *death_e* are both equal to 1. These patients are eligible for linkage to both data sources and are acceptable for research. Save your list of patients including the relevant linkage flags (*patid, hes_e, death_e*).

5. For study populations identified using primary care data only, combine your patient list with the file created in **Step 4** to create your final list of patients who are eligible for linkage to the data sources approved for your study. For study populations that will be identified only among patients who are eligible for linkage, the patient list created in **Step 4** above can be considered as your source population. Use this source population to further identify your study population.

6. Save your final list of patient identifiers, including linkage flags (*patid, hes_e, death_e etc*) as a tab delimited text file and email this together with your completed linkage request form to CPRD (enquiries@cprd.com). Please save your file following the naming convention ‘protocol number_organisation_patientlist’ e.g. 21_100001_InstitutionA_patientlist.txt.

Appendix 3: Code lists for requesting linked data

Code lists should be provided to CPRD as tab delimited text files. Each code list type should be provided in a separate file and each code should appear on a new line. Please see the table below for the **coding frames** and **coding format** found in CPRD linked data sources. Please ensure that all code lists are provided in the coding format shown below to avoid delays. All code lists should be submitted together with the completed CPRD Linkage Request form to enquiries@cprd.com.

CPRD Linked Data Source	Coding Frame	Code Format	Code Example
ONS Death Registration data	ICD-9	NNN NNN.N XNNN.N	410 410.1 E953.0
HES Admitted Patient Care ONS Death Registration data	ICD-10¥	XNN XNN.N	G00 G00.1
HES Outpatient data HES Accident & Emergency	ICD-10¥	XNN XNNN	G00 G001
HES Admitted Patient Care HES Outpatient data	OPCS	XNN XNNN	Q07 Q071
HES Accident & Emergency	A&E diagnosis/treatment	NN NNN	01 201
HES Accident & Emergency	A&E investigations	NN	02
HES Diagnostic Imaging Dataset	Imaging Code - NICIP	XXXX XNXXX XXXXX XXXXXX	CART C4DAC CAAAG CCHESB
HES Diagnostic Imaging Dataset	Imaging Code - SNOMED-CT	NN*	10077008 1051311000000104

¥Where ICD codes in the format of XNN.NN are included in your code list, please ensure that you also provide the 3-digit or 5-digit version of the ICD-code of interest, where available as codes in the XNN.NN format may not be in use in some data sources.