



# CPRD

## Clinical Practice Research Datalink

### Internal knowledge sharing session

Please note that these slides have been created using publicly available information including the CPRD website: [www.CPRD.com](http://www.CPRD.com)

# Contents

**3**

What is CPRD




**7**

What kind of data CPRD contains



**12**

How to know if CPRD is suitable for your research




**13**

CPRD Data linkage



**15**

What CPRD platform looks like



## What is CPRD?

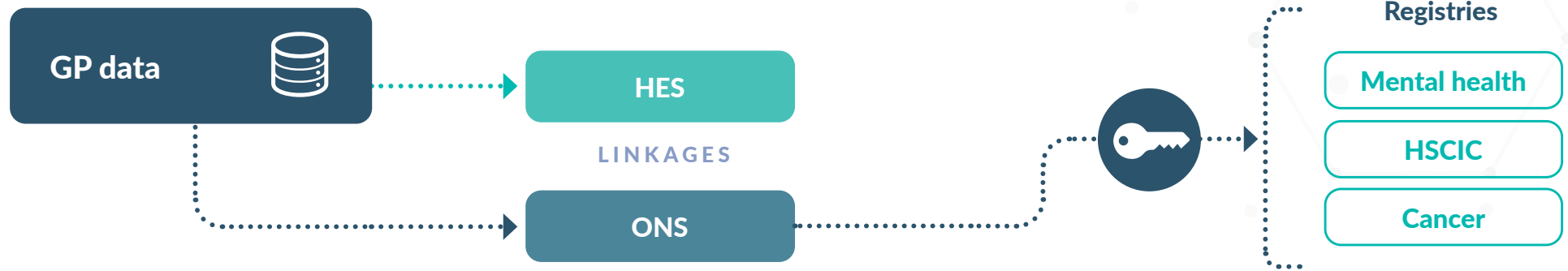


Clinical Practice Research Datalink, known as CPRD, is the UK real-world research service supporting retrospective and prospective public health studies and interventional research.

CPRD offers a range of services based on the **anonymised longitudinal** Electronic Healthcare Record (EHR) data of patients **collected from GP practices** and linked to **other datasets**.



# CPRD Primary Care Data Linkage





## CPRD UK Population Coverage



Over **60 million total patient lives** on CPRD database



**16 million** currently registered **patients** – **18%** of UK **population**



Data **representative** of UK population



**Median follow up time of 10 years** – some life long follow up



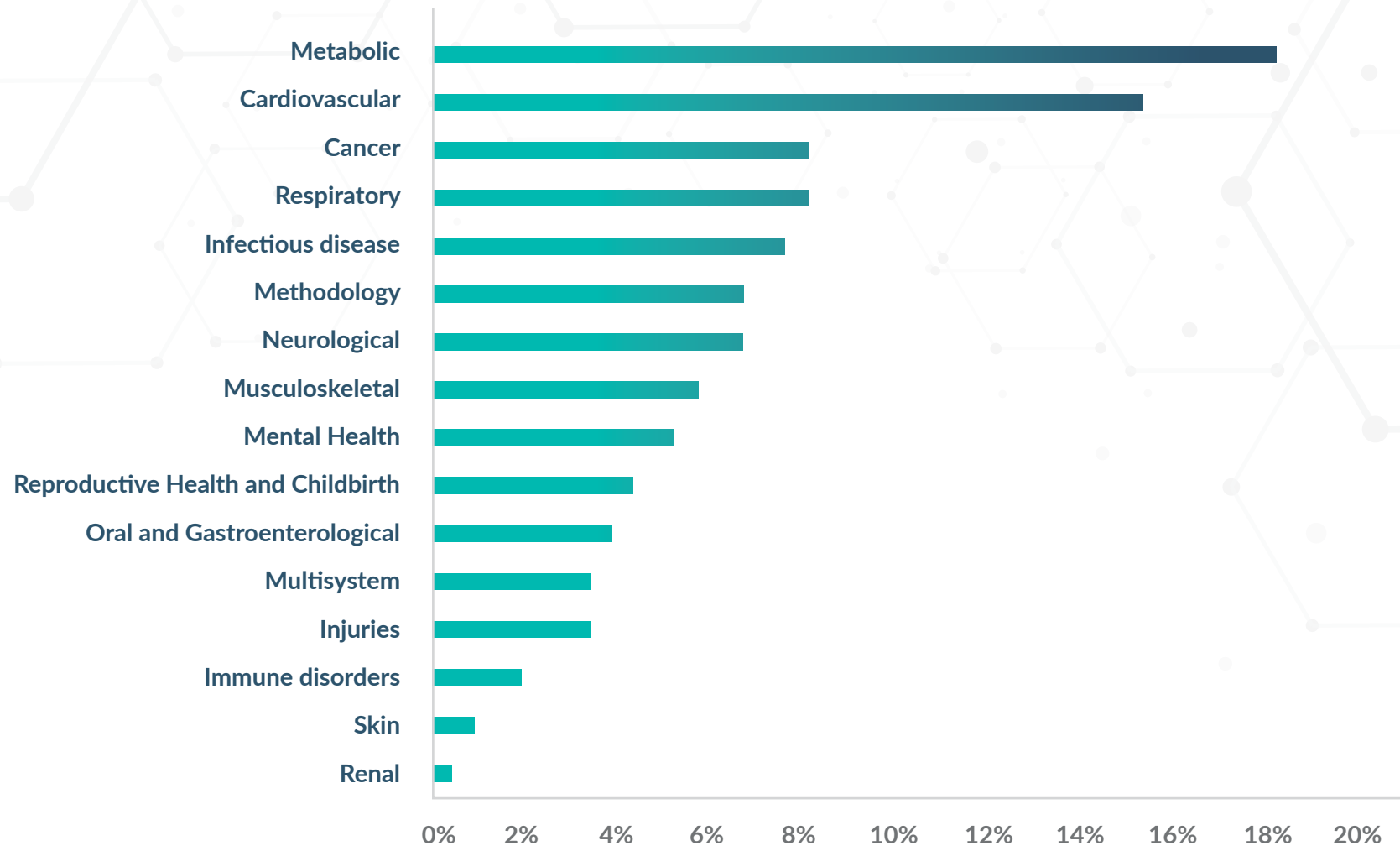
1 in every 4 GP practices in UK



In-house **quality checks** to ensure a high quality research-ready data



## CPRD research output by disease areas



CPRD bibliography disease and conditions (2016)



## What information is available in CPRD?

- **Demographic information** (year of birth, sex, weight, etc.)
- **Records of clinical events** (medical diagnoses, signs, and symptoms)
- **Referrals** to specialists and other secondary care settings
- Prescriptions issued in **primary care** (drugs and medical devices)
- Records of **immunisations/vaccinations**
- **Diagnostic testing**
- **Lifestyle information** (body mass index (BMI), smoking, and alcohol consumption)
- And all other types of care administered as part of **routine GP practice**



## Standard linkages with CPRD

### Hospital Episode Statistics (HES)

- Admitted Patient Care (ICD and Procedures (OPCS)
- Outpatient (Diagnoses (ICD) data)
- Accident & Emergency (Diagnoses (A&E codes) data )
- Diagnostic Imaging Dataset (Imaging tests data )
- PROM (Quality of life & conditions specific)

### ONS Death Registration Data

Key information (including coding/scoring system):  
Date, place, and causes of death (ICD)

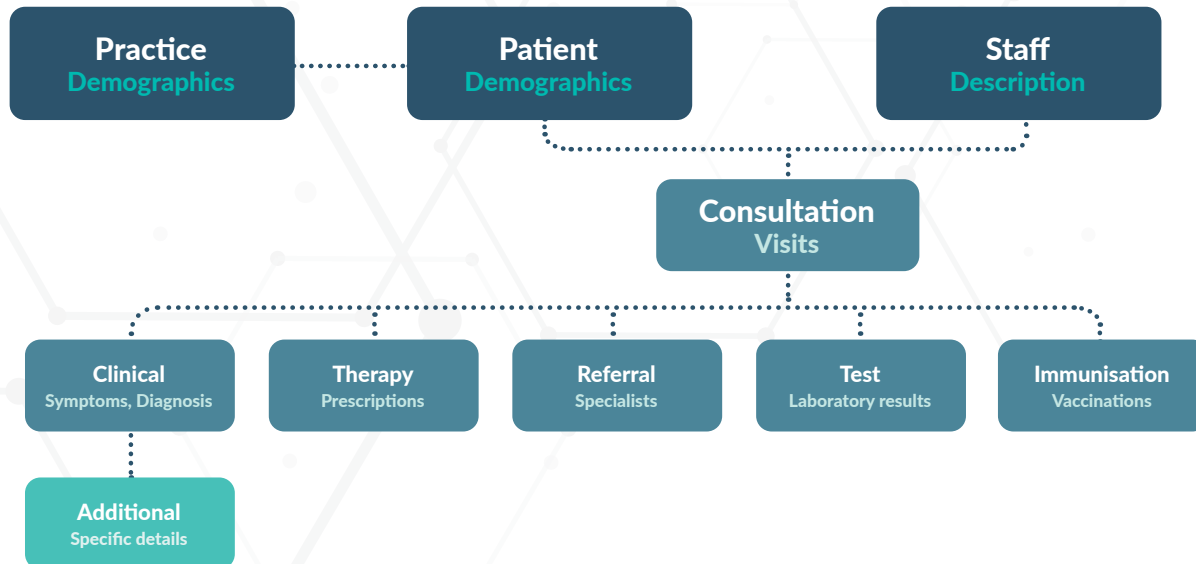
### National Cancer Registration and Analysis Service

- Cancer registration
- Systemic Anti-Cancer Treatment
- National Radiotherapy Dataset
- Cancer Patient Experience Survey (conditions specific)

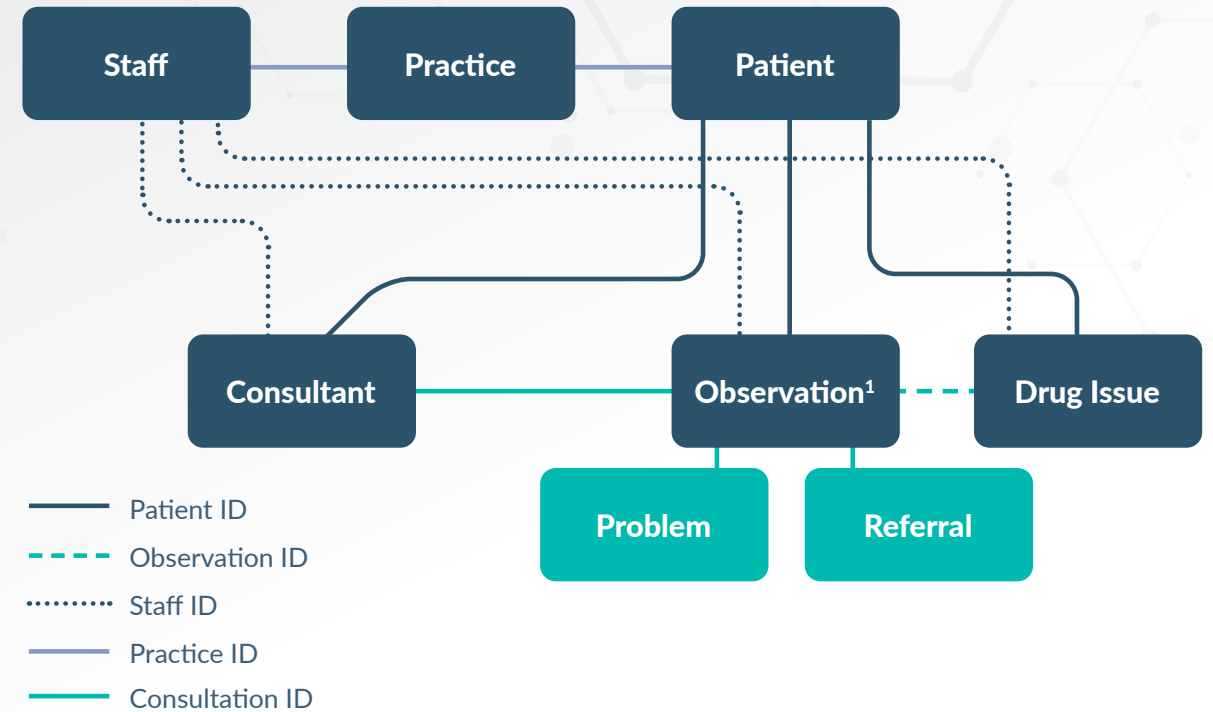


## Main database: GOLD & Aurum

**CPRD GOLD** which contains data contributed by practices using **Vision®** GP practice software



**CPRD Aurum** which contains data contributed by practices using **EMIS Web®** GP practice software.



## CPRD Gold vs CPRD Aurum

CPRD GOLD	VS	CPRD Aurum
InPS Vision® software	Clinical system	EMIS Web® software
England (5.7%) +Wales (10.2%) +Scotland (55.6%)+ Northern Ireland (28.4%)	Coverage	England (99%) +Northern Ireland (1%)
1987	Data Start date	1995
<b>20,945,522</b>	Total No of acceptable patients (at May 2022)	<b>41,200,722</b>
<b>985</b>	No of practices* (at May 2022)	<b>1491</b>
Median(IQR):5.57 (1.97 - 13.44)	Available follow-up time** (since 1995)	Median(IQR):8.74 (3.25 - 19.85)
Read code	Medical coding	Mixture of Read 2, SNOMED and local EMIS® codes
Gemscript	Product coding	Dictionary of Medicines and Devices (DM+D)
(UTS) up-to-standard date	Derived variables	UTS future development
GOLD can provide Mother-baby link	Data Linkage	-
Code Browser, Define, Refine and Extract	Online tools	Code Browser, Define and Extract

## Using both Aurum & GOLD in a study?

- Some practices have migrated from CPRD GOLD to CPRD Aurum
- Their data from GOLD time has been copied into Aurum
- CPRD provide a bridging file of these practices with dates
- Deduplication is recommended

## Is CPRD data suitable for your research needs?

Refer to relevant publications based on CPRD data in your area of research to estimate the numbers expected



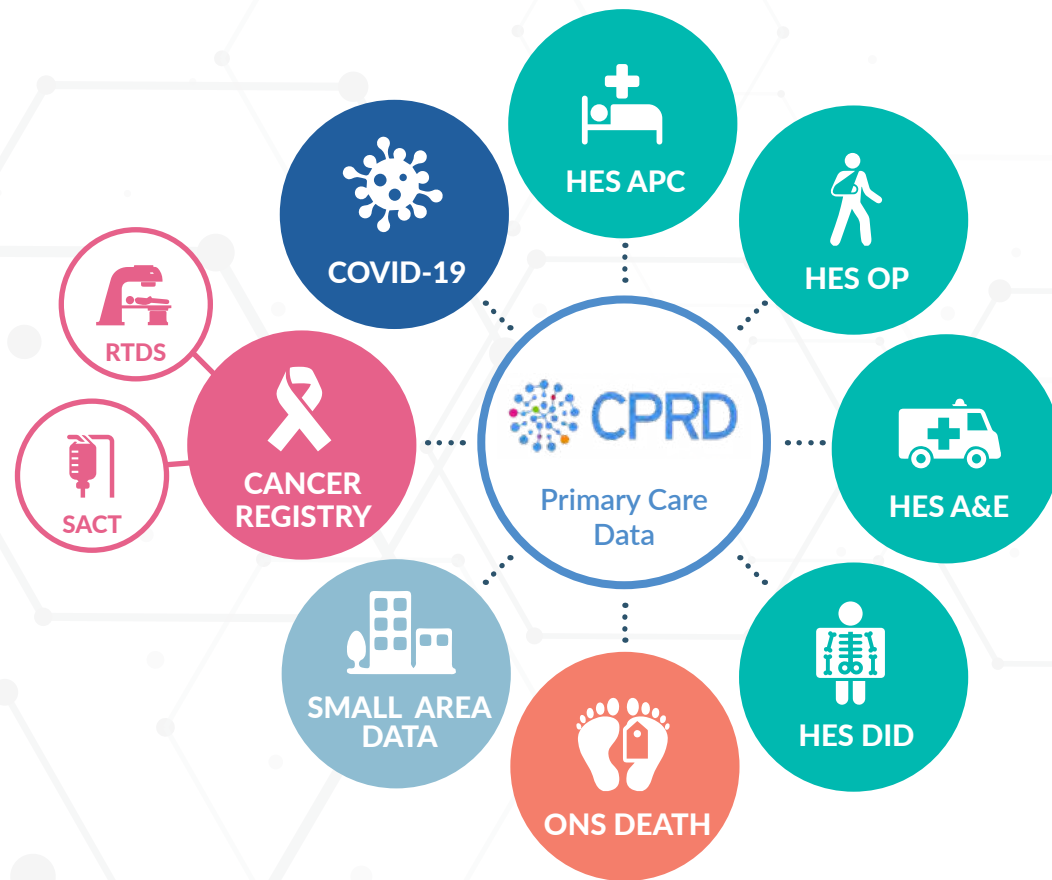
Multi-study licence holders can access the database for feasibility purposes using the online tools



If your institution does not have access to the online tools, there is insufficient data in the literature and this is your first time using the data, CPRD can provide a simple feasibility count



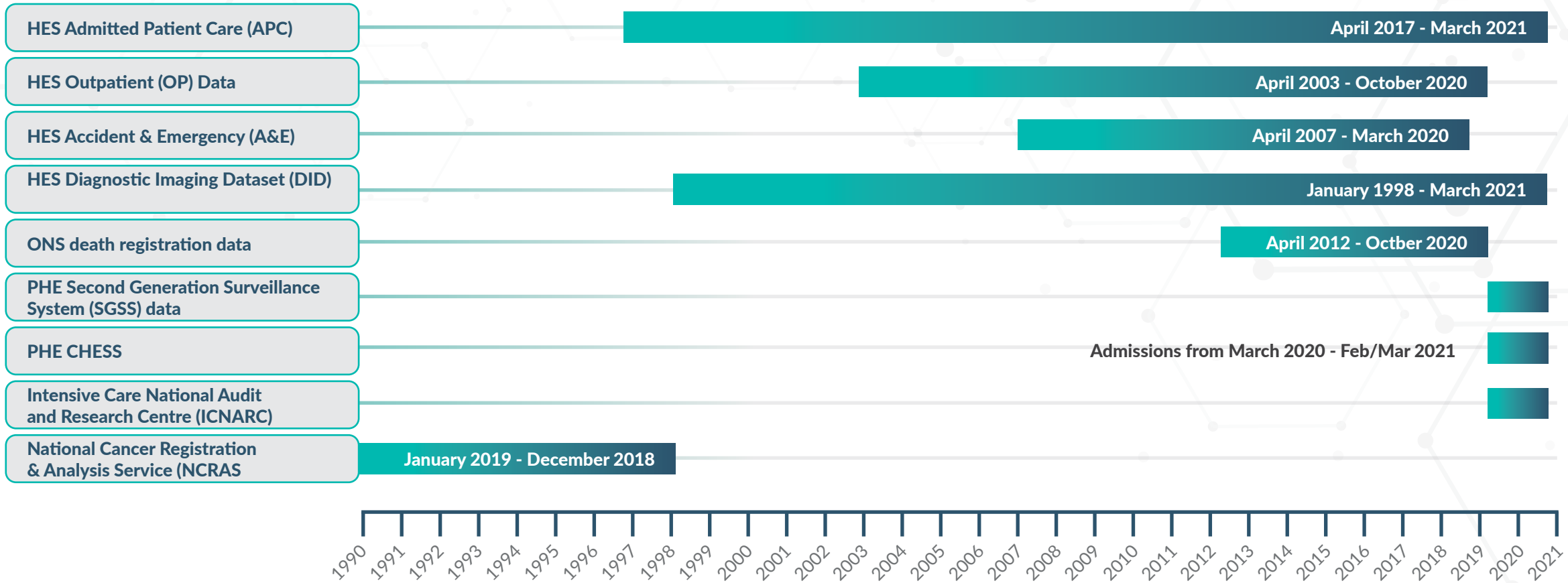
## Data linkage



These linkages enable CPRD to provide a fuller picture of the patient care journey to support vital public health research, informing advances in patient safety and delivery of care.

- Patient-level linkage is currently enabled exclusively for patients registered at practices in **England**.
- Linked data will be provided, by secure transfer, within 10 working days of receipt of a valid, approved request.
- Additional costs will be incurred for data linkages

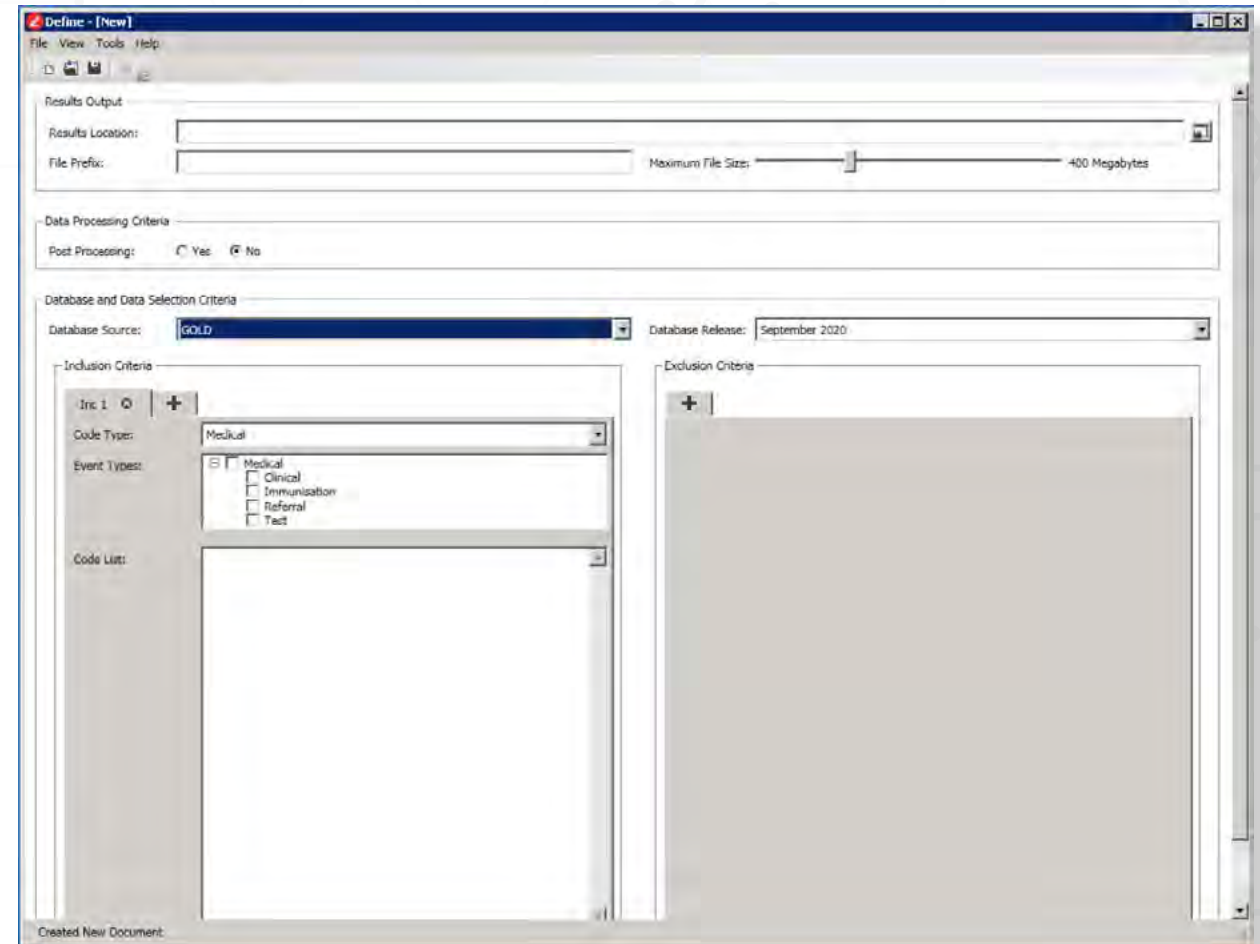
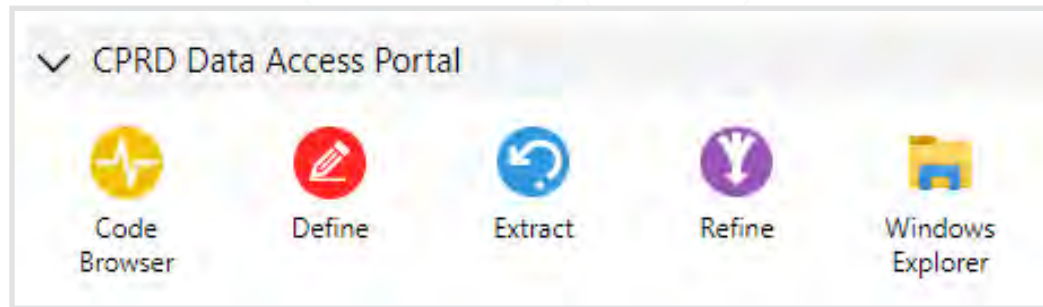
# Data linkage: update





## CPRD platform

- The Online Data portal is a closed system (on Azure) that allows you to query the CPRD database and download the necessary data.
- Application page is shown below





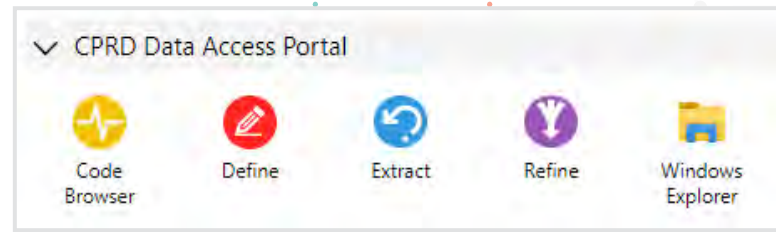
## CPRD platform – tools

### Code Browser -supports the development of code lists.

- The medical and product code dictionaries generate code lists for use in the Define tool or for sharing with collaborators.

### Extract Tool-support the extraction of data based on a list of patients from the Define tool

- 1 million patients max per extract



### Define Tool: supports the identification of a patient list based on:

- Medical and/or product codes
- Demographic characteristics, Gender, age, Relative or calendar time frames, Study period, follow-up time

### Refine Tool: Supports to define a new patient cohort on the basis of entity types or medical codes in CPRD GOLD ONLY

- Refine is not currently functional with CPRD Aurum data





## Upcoming data

CPRD is now collecting fully-coded patient electronic health records from GP practices using the SystemOne software system.

A technical project is underway to process and curate this data and eventually make it available for research. We do not yet have a date for when the anonymised data will be available for research.



## Reference

- Emily Herrett, Arlene M Gallagher, Krishnan Bhaskaran, Harriet Forbes, Rohini Mathur, Tjeerd van Staa, Liam Smeeth, Data Resource Profile: Clinical Practice Research Datalink (CPRD), *International Journal of Epidemiology*, Volume 44, Issue 3, June 2015, Pages 827–836, <https://doi.org/10.1093/ije/dyv098>
- Achim Wolf, Daniel Dedman, Jennifer Campbell, Helen Booth, Darren Lunn, Jennifer Chapman, Puja Myles, Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum, *International Journal of Epidemiology*, Volume 48, Issue 6, December 2019, Pages 1740–1740g, <https://doi.org/10.1093/ije/dyz034>
- Atkinson, R. W., Carey, I. M., Kent, A. J., van Staa, T. P., Anderson, H. R., & Cook, D. G. (2013). Long-term exposure to outdoor air pollution and incidence of cardiovascular diseases. *Epidemiology*. <http://doi.org/10.1097/EDE.0b013e318276ccb8>
- Carey, I. M., Atkinson, R. W., Kent, A. J., Van Staa, T., Cook, D. G., & Anderson, H. R. (2013). Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. *Am J Respir Crit Care Med*. <http://doi.org/10.1164/rccm.201210-1758OC>
- <https://cprd.com/>
- Training modules and reading materials



Units 1&2  
Field View  
Baynards Green  
Bicester  
OX27 7SG

+44 (0) 7804 623 339  
[info@odelletechnology.com](mailto:info@odelletechnology.com)  
[www.odelletechnology.com](http://www.odelletechnology.com)