



Release Notes: CPRD Aurum Sample Dataset October 2021

Summary

The CPRD Aurum sample dataset is a medium-fidelity synthetic dataset that resembles the real world CPRD Aurum with respect to the data types, data values, data formats, data structure and table relationships. This synthetic dataset can be used for multiple purposes including:

1. as a sample dataset to understand the structure and utility of the anonymised CPRD Aurum database
2. to use as a data management teaching/training resource
3. to develop/validate/test analytics tools for use with CPRD Aurum data
4. to improve bespoke CPRD Aurum application interfaces/algorithms, for e.g., a bespoke cohort selection tool, or
5. to develop machine learning workflows that can be applied to anonymised CPRD Aurum data

The table below presents summary statistics for the synthetic CPRD Aurum dataset.

Metric	Coverage
Total number of acceptable patients ¹ (including transferred out and deceased patients):	39,388
Current acceptable patients (i.e. registered at currently contributing practices, excluding transferred out and deceased patients):	13,858
Percentage UK population coverage ² (current patients only):	13,858 of 66,796,800 (0.02%)
Available follow-up time in years since 1 st January 1995 ³ (all patients including transferred out and deceased):	
Mean (Standard Deviation):	7.77 (7.53)
Median (25 th and 75 th Percentile):	4.98 (1.71 – 12.05)
Available follow-up time in years since 1 st January 1995 (current patients only):	
Mean (Standard Deviation):	11.13 (8.63)
Median (25 th and 75 th Percentile):	9.05 (3.38 – 18.44)
Total number of practices (current and historic) included in the database:	14
Currently contributing practices:	14
Percentage coverage of UK general practices (currently contributing practices only):	14 of 8,961 (0.16%)
Regional distribution of currently contributing practices ⁴	
North East:	4
North West:	1
Yorkshire And The Humber:	0
East Midlands:	1
West Midlands:	4
East of England:	0

¹ Permanent registrations only. Over 93% of permanent registrations are deemed to have 'acceptable' (or research quality) data based on CPRD metrics.

² [Based on latest UK population estimates from the Office of National Statistics.](#)

³ Follow-up time stated here does not incorporate the up-to-standard (UTS) date and the database includes records pre-dating the 1st of January 1995.

⁴ Expressed as a percentage of all practices currently contributing to CPRD Aurum.



South West:	1
South Central:	0
London:	1
South East Coast:	2
Northern Ireland:	0
Scotland:	0
Wales:	0

DOI

Please cite in any publications using this version: <https://doi.org/10.48329/hm7t-qs28>

CPRD Aurum Data Specification

CPRD Synthetic Aurum Specification is similar to CPRD Aurum Specification which is available on the CPRD website: <https://www.cprd.com/primary-care>