



Release Notes: CPRD Aurum March 2023

Summary

An updated version of CPRD Aurum database (data sourced from GP practices using EMIS® GP software) is now available, which contains the following:

Metric	Coverage
Total number of acceptable patients ¹ (including transferred out and deceased patients):	46,795,888
Current acceptable patients (i.e. registered at currently contributing practices ² , excluding transferred out and deceased patients):	2,127,536
Percentage UK population coverage ³ (current patients only):	2,127,536 of 67,081,000 (3.17 %)
Total patients eligible for linkage (Set 22):	38,377,102
Available follow-up time in years since 1 st January 1995 ⁴ (all patients including transferred out and deceased): Mean (Standard Deviation): Median (25 th and 75 th Percentile):	8.35 (8.21) 5.12 (2.01 – 12.38)
Available follow-up time in years since 1 st January 1995 (current patients only): Mean (Standard Deviation): Median (25 th and 75 th Percentile):	12.98 (10.02) 10.31 (3.90 – 22.76)
Total number of practices (current and historic) included in the database:	1,720
Currently contributing practices ² :	228
Percentage coverage of UK general practices (currently contributing practices only):	228 of 8,178 (2.79 %)

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

² Currently contributing practices are those contributing data to CPRD within 60 days of the database build being created. In this release, the 1,491 practices with the same data as the May 2022 release are classed as not currently contributing, explaining the drop in 'Current acceptable patients' and 'Currently contributing practices' compared to earlier releases. The next release of CPRD Aurum will include additional temporal coverage for a proportion of these existing practices.

³ [Based on latest UK population estimates from the Office of National Statistics.](#)

⁴ 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.

⁵ Expressed as a percentage of all practices currently contributing to CPRD Aurum.



Medicines & Healthcare products
Regulatory Agency

Regional distribution of currently contributing practices ⁵	
England:	228 (100.00 %)
Northern Ireland:	0 (0.00 %)
Scotland:	0 (0.00 %)
Wales:	0 (0.00 %)

Lookups

These files can be requested from CPRD Enquiries at enquiries@cprd.com

Other Updates

None

DOI

Please cite in any publications using this version <https://doi.org/10.48329/91gg-3834>

CPRD Aurum Data Specification

Available on the CPRD website: <https://www.cprd.com/primary-care>



Additional notes on the use of the CPRD Aurum March 2023 database

As noted in our 3 April and 19 April communications, there are known issues with missing data in records relating to immunisations, allergies, and referrals in the **Observation** table of the CPRD Aurum March 2023 database. We are providing the output from our source data verification (SDV) checks, along with this release, so that researchers can make an informed decision about whether they should use this database or wait for the next update, based on their specific study requirements. We also provide guidance below on situations where researchers might want to use the March 2023 release. Users should adopt a conservative approach when evaluating the assumptions stated prior to data use.

Please see the updated table below for SDV checks completed by the CPRD Aurum data supplier (EMIS) on data available in April 2023. Checks were based on all new practices in the planned CPRD Aurum March 2023 release, and a random sample of old practices included in the CPRD Aurum May 2022 release. Counts from both builds were compared to counts in EMIS Web (the online system where GPs enter patient healthcare data) with no time restrictions applied. More information on our data quality checks is available on our webpage: [Data quality | CPRD](#).

Checks confirm that data in the CPRD Aurum March 2023 Observation table on tests and the Drug Issue tables are reliable. Checks on the Consultations are ongoing, and we will provide an update on these once we have counts from EMIS. We advise users to exercise caution over the use of these data in the meantime.

Our checks confirm that for new practices (i.e., those newly contributing to CPRD Aurum since the migration to EMIS EXA), there was an issue with the flow of immunisations and referrals data to the CPRD Aurum database, such that there are very few recorded events. Our discussions with EMIS and data checks suggest that the missing records are a localised issue affecting immunisation and referrals data in this release. These data are now held in separate tables in EMIS EXA, and we will be able to access these data, and rectify these issues in subsequent CPRD Aurum releases. Please note that whilst we receive the data via separate tables, CPRD will incorporate them into the existing CPRD Aurum structure in any future release. The initial concern over allergy data was able to be resolved internally as this related to incorrect labelling as opposed to missing data.

Table 1: Source data verification counts comparing data from EMIS Web and CPRD Aurum

Clinical Events	Data Table	Practice Status	N Practices	Counts in EMIS Web	Counts in CPRD Aurum	% Counts in CPRD Aurum & EMIS Web
Total Cholesterol	Observation	New	228	12,575,540	12,297,199	97.79 ^a
HbA1c	Observation	New	228	10,957,956	10,622,385	96.94 ^a
COVID-19	Immunisation	New	228	5,814,326	196	0.00 ^b
MMR	Immunisation	New	228	4,419,999	3,405	0.08 ^b
COVID-19 product	Drug Issue	New	228	4,735,699	4,674,063	98.70 ^a
MMR product	Drug Issue	New	228	22,405	22,178	98.99 ^a
Nuvaxovid product	Drug Issue	New	228	755	712	94.30 ^a



Medicines & Healthcare products Regulatory Agency

Allergy	Observation	New	228	4,157,532	4,091,249	98.41 ^a
Referral	Observation	New	228	19,561,151	145	0.00 ^b
Consultation	Consultation	New	228	TBC	TBC	TBC
Total Cholesterol	Observation	Old	226	16,198,410	15,269,089	94.26 ^c
HbA1c	Observation	Old	226	15,945,500	14,145,942	88.71 ^c
COVID-19	Immunisation	Old	226	8,650,634	7,339,187	84.84 ^c
MMR	Immunisation	Old	226	5,596,370	4,755,489	84.97 ^c
COVID-19 product	Drug Issue	Old	226	6,871,848	6,002,509	87.35 ^c
MMR product	Drug Issue	Old	226	24,507	22,217	90.66 ^c
Nuvaxovid product	Drug Issue	Old	226	1,460	1,202	82.33 ^c
Allergy	Observation	Old	226	4,587,755	3,716,523	81.01 ^c
Referral	Observation	Old	226	24,999,719	23,993,060	95.97 ^c
Consultation	Consultation	Old	226	TBC	TBC	TBC

^a Discrepancies due to national opt-outs being applied to the CPRD Aurum data.

^b Discrepancies due to immunisation / referral data not flowing to CPRD Aurum (this is a localised issue, which we can rectify in subsequent builds).

^c Discrepancies due to opt-outs being applied to the CPRD Aurum data in addition to differences in temporal coverage (counts from CPRD Aurum are based on data up to May 2022, counts from EMIS Web are based on counts to April 2023)

Situations where researchers may use the CPRD Aurum March 2023 database as opposed to the CPRD Aurum May 2022 or earlier databases: potential benefits, and assumptions.

1. To access more data for the same study period than that covered, for example, by the CPRD Aurum May 2022 database.

Potential benefits: Increased sample size with accompanying increase in statistical power.

The CPRD Aurum March 2023 database includes 13% more patients (56,899,561) compared with 50,433,854 patients in the CPRD Aurum May 2022 database.

Assumptions: The data from the new practices is equivalent (similar in completeness and level of errors/bias) to that in the old practices

2. For a study that runs beyond the end of the CPRD Aurum May 2022 dataset (i.e., that uses data from old and new practices).

Potential benefits: use of the CPRD Aurum March 2023 database would increase sample size (i.e., include 16% more rows in the Observation table (16,256,561,378 vs 14,045,676,954), with accompanying increase in statistical power.

Assumptions: The data from the new practices is equivalent (similar in completeness and level of errors/bias) to that in the old practices, *and* the underlying patterns of information for the new practices in the period of new coverage is the same as that for the practices in the CPRD Aurum May 2022 dataset.



Medicines & Healthcare products Regulatory Agency

3. For a study period that starts after the end of coverage of the CPRD Aurum May 2022 database.
Potential benefits: Data for 6,430,449 patients from 228 new practices.
Assumptions: The data from the new practices are sufficiently complete and accurate for your specific study requirements.

Summary

There are marginal potential gains from using the CPRD Aurum March 2023 database in some specific circumstances (for example, when effect sizes are on the borderline of detectability or when investigating changes occurring right at the end of the coverage of the CPRD Aurum May 2022 dataset).

CPRD recommends a cautious approach to using these data, and careful consideration of the above assumptions in a study-specific context.