



CPRD Aurum Mother-Baby Link Documentation

6 December 2022

Documentation Control Sheet

Over time, 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 | Affected Areas of Summary of Change | Prepared by | Date | Reviewed by | Date |
|----------------|--|--------------------|-------------|--------------------|-------------|
| 0.1 | Initial Draft | Sonia Coton | 31/08/2022 | Jenny Campbell | 05/09/2022 |
| 0.2 | Updated template | Sonia Coton | 06/12/2022 | Jenny Campbell | 06/12/2022 |

Summary of changes

CPRD Aurum Mother-Baby Link data

Introduction

The ability to study individuals within family units in epidemiology is very useful. As well as being able to look at familial aspects of disease, increased focus is being placed on the importance of in-utero exposures with respect to conditions manifested in the offspring. The best examples are studies into the teratogenic effects of drugs, which are very hard to investigate using traditional pre-marketing methodologies such as RCTs for obvious ethical reasons.

Feasibility with CPRD Primary Care Data

CPRD hold a practice-specific household number that can be used to ascertain whether people are living within the same household. This is primarily based on residence but, the existence of this variable allows us to link members of families together, and particularly mothers with their children. Furthermore, because both pregnancy and delivery information are recorded during the long follow up time within CPRD primary care data, we are able to follow some children from conception rather than just from birth. This means the CPRD primary care data can be used to undertake the sort of studies that require information in all three trimesters as well as beyond birth.

To help users CPRD provide a mother to baby linked list, and the following sections provide details on the generation, validation, and utility of this list.

Methodology

The process involves selecting all maternal deliveries, selecting all patients born within the appropriate time period, and the linking together of these two data sets. The main points of the three components are set out below.

1. Selection of Maternal Deliveries

A list of delivery-related SNOMED, Read, and local EMIS[®] codes was used to extract all records from the Observation file relating to a delivery. Records were extracted for all female patients regardless of whether the patient is flagged as research 'acceptable' or not. Where the delivery code indicated a period of time after delivery (e.g., 6 weeks postnatal visit) this time was subtracted from the observation date to give an estimated delivery date. The mother's age at delivery was calculated to be the difference between the year of the delivery record and the mother's birth year. Delivery records were restricted to mothers between the ages of 11 and 49 years.

Records from before 1987 and those recorded more than a year before the mother registered at the practice (historical records) were dropped. Deliveries dated after the practice last collection date (or the mother's transfer out date) were also dropped. Duplicates of the same delivery date for each mother were removed, so there was one record for each delivery date per mother.

2. Selection of Children

All registered patients born after 1986 were extracted. Patients whose year of registration was before their birth year, or whose birth year was after the year of the practice last collection date were excluded. Since CPRD do not collect full date of birth, birth date was estimated using a hierarchical algorithmic approach.

Estimating infant's date of birth:

A list of infant related SNOMED, Read, and local EMIS[®] codes was used to extract and flag all records from the Observational file relating to new-born, 6-8 weeks infant examination, or other infant records. Where the infant code indicated a period of time after birth (e.g., 6 weeks infant examination) this time was subtracted from the observation date to give an estimated date of birth. Infant records flagged as new-born were used first to estimate date of birth, then 6-8 weeks infant examination flagged records, followed by other infant flagged records. If the infant's estimated date of birth was still missing, first consultation date was used. For infants with recorded month and year of birth, if the recorded month and year of birth did not match the estimated date of birth then the estimated date of birth became the 15th day of the recorded month and year of birth. For infants with a recorded year of birth only, if the recorded year of birth did not match the estimated date of birth, then the estimated date of birth became the 30th day of June of the recorded year of birth.

3. Linking Births to Deliveries

A cartesian join of mothers to infants by practice and household number was undertaken and only those records where the absolute difference between the delivery date and the estimated birth date was within 120 days were retained. Where more than one delivery record per mother baby pair was available, the record with the smallest absolute difference between the delivery date and the estimated birth date was selected.

Duplicates were handled as follows:

- Where a delivery was matched to more than one child, it was only included if the children had the same birth date and the same registration date, in an effort to isolate real siblings.
- Where a child was matched to more than one mother, the matches where the mother and child had different transfer out dates were discarded. If there were still multiple

records, all were dropped (regardless of how many records the child had matched with each mother) as it was not possible to be sure who the correct mother was.

- Where a mother had multiple delivery dates within 25 weeks of each other, these matches were dropped.

The number of children each mother was matched to was counted. Mothers with a single delivery date matched to five or more children and mothers with 16 or more child matches were dropped, to mitigate the risk of re-identification.

Backwards Compatibility

Because CPRD primary care data is a dynamic and updated data source, all new issues of the Mother-Baby Link cannot guarantee to contain all the previous mother baby pairings. Furthermore, it is likely that refinements to the algorithm will be developed over time which may well render newly created lists non-backwards compatible.

Caveats and Notes

No restrictions on the research standard 'acceptable' flag were included in this current Mother-Baby Link. Restrictions on acceptability and registration can be applied retrospectively, by the user, if needed.

Please note that the mother baby algorithm identifies only live birth - maternal pairings. If you need any further information on the Mother-Baby Link, please contact CPRD Enquiries (enquiries@cprd.com).

CPRD Mother-Baby Link: Data dictionary

| Column name | Description | Type | Format |
|---------------|---|---------|------------|
| pracid | CPRD practice identifier | INTEGER | 8 |
| mumpatid | Mother's CPRD patient identifier | INTEGER | 12 |
| babypatid | Child's CPRD patient identifier | INTEGER | 12 |
| deldate | Assumed delivery date for the baby (from mother's record) | DATE | dd/mm/yyyy |
| mumbirthyear | Mother's year of birth | INTEGER | 8 |
| babybirthyear | Child's year of birth | INTEGER | 8 |
| gender | Child's gender | CHAR | 13 |
| children | The number of children matched to this mother | INTEGER | 8 |
| match_level | Indicates the record type used to estimate the infants date of birth and the quality of matching between a mother-baby pair | CHAR | 13 |

Lookup for match_level:

- 1 = New-born record,
- 2 = 6-8 weeks infant examination record,
- 3 = Other infant record,
- 4 = First consultation date,
- 5 = Recorded month and year of birth,
- 6 = Recorded year of birth.

© Crown copyright 2022

Open Government Licence



Produced by the Medicines and Healthcare products Regulatory Agency.

<http://www.gov.uk/mhra>.

You may re-use this information (excluding logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit

<http://www.nationalarchives.gov.uk/doc/open-government-licence> or email:

psi@nationalarchives.gov.uk

Where we have identified any third-party copyright material you will need to obtain permission from the copyright holders concerned.

The names, images and logos identifying the Medicines and Healthcare products Regulatory Agency are proprietary marks. All the Agency's logos are registered trademarks and cannot be used without the Agency's explicit permission.