



Medicines & Healthcare products
Regulatory Agency



CPRD Mother Baby Link Documentation

Version 1.2

Date: 3 May 2017



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.

Version	Affected Areas Summary of Change	Prepared By	Reviewed By
1.0	Initial Draft	Rachael Williams	Arlene Gallagher Tim Williams
1.1	Modified	Rachael Williams	Shivani Padmanabhan
1.2	Modified	Rachael Williams	Shivani Padmanabhan

Summary of Changes

Version 1.1

- Removed specific details of codelist

Version 1.2

- Updated references to CPRD Enquiries



CPRD 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

In CPRD primary care data there is a practice-specific family number that can be used to identify people within the same family. This is primarily based on residence, but can be edited in Vision where members of different families live at the same address. 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 is recorded during the long follow up time within CPRD primary care data, we are able to follow some children from conception rather than 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.

To help users identify the linked mothers and children 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 the identification of maternal deliveries, the identification of 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. Identification of Maternal Deliveries

All 'Clinical', 'Referral' and 'Test' records relating to a delivery (Read code list for births available on request) were extracted for female patients conforming to the standard patient acceptability criteria. Where the Read code indicated a period of time after delivery (e.g. 6 week postnatal visit) this was subtracted from the event 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 12 and 49.

Additional data from the "Additional Clinical Details" (Vision Structured Data Areas) relating to birth outcomes and postnatal examinations were extracted to generate further delivery date estimates according to the following:



<i>Additional Clinical Details</i>	<i>Calculation of Possible Delivery Date</i>
Hearing (6 weeks) Muscle tone for 6 weeks (CHS ¹) Vision CHS 6 weeks	Event date minus 42 days
Post-natal examination Post-natal visit	Event date minus number of days/weeks post-natal. A maximum of 70 days/10 weeks post-natal was allowed.
Stages of labour Delivery details CHS Apgar score at 1 minute Delivery details (CHS) Maternity infant details Perineum Maternity outcome placenta CHS Apgar score at 5 minutes	Event date
Pregnancy outcome	Discharge date minus 2 days. If discharge date was not recorded, the event date was used as a possible delivery date
Gestational age of baby	Event date if outcome of delivery was known and entered

¹ CHS = Child Health Surveillance

The ACD records were then restricted to women aged between 12 and 49 at the estimated delivery date.

All delivery record sources were combined. 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. Identification of Children Born

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 last collection year were excluded. Since CPRD do not collect full date of birth, birth date was estimated as the 15th day of the given month and year. For patients without a month of birth, the 30th June of the birth year was assumed.

3. Linking Births to Deliveries

A cartesian join of mothers to babies by practice and family number was undertaken and only those records where the absolute difference between the delivery date and the estimated birth date was within 60 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, the child was dropped (regardless of how many records the child had with each mother) as it was not possible to be sure who the correct mother was.

The number of children each mother was matched to was counted.

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

In previous issues of the Mother Baby Link, a number of key restrictions were imposed. The two major restrictions were relating to use of data from UTS time only within a practice, and a restriction that children be registered within six months of birth to be included. No such restrictions were included in this current Mother Baby Link. The reason for this was to maximise the available data. This link therefore may include children born before the practice became UTS and those who initially registered at a different practice after birth, but subsequently joined the current one. Restrictions on UTS 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

1. MBL file (mbIYYYY_MM.txt)

<i>Column name</i>	<i>Description</i>	<i>Type</i>	<i>Format</i>
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 mothers 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