



Medicines & Healthcare products  
Regulatory Agency

# CPRD Aurum Pregnancy Register

**Version:** 1.0

**Date:** 23 May 2022

**Authors:** Jenny Campbell, Jessie Oyinlola and Hilary Shepherd

## 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 Summary of Change	Prepared By	Date	Reviewed By	Date
0.1	First Draft	Jenny Campbell	09/11/2021	Jessie Oyinlola	23/11/21
1.0	Final Draft	Jenny Campbell	07/12/2021		

## **Introduction**

The ability to study exposures before and during pregnancy, maternal/fetal complications, and outcomes of pregnancy in the CPRD depends crucially on the identification of pregnancy episodes. Some important examples are pharmacoepidemiology studies (drug and vaccine safety during pregnancy and maternal vaccine uptake and effectiveness). Because the precise time period during which a woman is pregnant can be difficult to ascertain in the data, an algorithm has been developed to identify and maximise the use of records relating to the timing and duration of pregnancy, the type of pregnancy outcome (live birth, stillbirth or pregnancy loss), and additional features pertaining to the pregnancy.

## **Dataset format**

The Pregnancy Register lists all pregnancies recorded in the CPRD Aurum database and includes details of each one. A single record represents a unique pregnancy episode. There may be more than one episode per woman.

## Field descriptions

A description of each of the fields in the Pregnancy Register is provided in the table below.

<i>Field name</i>	<i>Description</i>	<i>Type</i>	<i>Format</i>
patid	Encrypted unique patient identifier	INTEGER	20
pregid	Unique identifier of the pregnancy episode	INTEGER	20
totalpregs	Total number of identified pregnancy episodes (per woman)	INTEGER	2
pregnumber	Pregnancy episode number (per woman)	INTEGER	2
pregstart	Estimated start date of pregnancy	DATE	dd/mm/yyyy
firstantenatal	Date of earliest antenatal record within the pregnancy	DATE	dd/mm/yyyy
startsource	Data source used to estimate pregnancy start date: 1 = Imputed <sup>1</sup> , 2 = EDD, 3 = LMP, 4 = Gestational age at birth, 5 = Gestational age from antenatal record, 6 = EDC	INTEGER	1
startadj	Flag to indicate whether the pregnancy start date has been adjusted: 0 = Not adjusted, 1 = Due to antenatal records in the preceding 4 weeks, 2 = Due to specific conflicts between the estimated pregnancy duration and records indicating gestational age at birth (live births and stillbirths only), 3 = Both	INTEGER	1
Secondtrim <sup>2</sup>	Estimated start date of second trimester	DATE	dd/mm/yyyy
Thirdtrim <sup>2</sup>	Estimated start date of third trimester	DATE	dd/mm/yyyy
pregend	Estimated end date of pregnancy. NB: For pregnancies with unknown outcome, the date of the latest antenatal record in the pregnancy episode is provided.	DATE	dd/mm/yyyy
endsource	Data source used to estimate pregnancy end date: 1 = Delivery record, 2 = Postnatal record in the mother's medical record, 3 = Discharge date relating to a delivery, 4 = Baby's (month and) year of birth as recorded in the baby's medical record, 5 = Postnatal record in the baby's medical record, 6 = First consultation in the baby's medical record. Only completed for live births and stillbirths.	INTEGER	1
endadj	Flag to indicate whether the pregnancy end date has been adjusted: 0 = Not adjusted, 1 = Due to specific conflicts between the estimated pregnancy duration and records indicating gestational age, 2 = Due to prior adjustments to the start date, 3 = Both. Missing for deliveries based on late pregnancy records <sup>3</sup> .	INTEGER	1
gestdays	Estimated duration of pregnancy episode in days (calculated as pregend minus pregstart)	INTEGER	3
matage	Mother's age at end of pregnancy (years)	INTEGER	3
outcome	Outcome of pregnancy: 1 = Live birth, 2 = Stillbirth, 3 = 1 and 2, 4 = Miscarriage, 5 = TOP, 6 = Probable TOP, 7 = Ectopic, 8 = Molar, 9 = Blighted ovum, 10 = Unspecified loss, 11 = Delivery based on a third trimester pregnancy record, 12 = Delivery based on a late pregnancy record <sup>3</sup> , 13 = Outcome unknown	INTEGER	1

preterm_ev	Flag to indicate evidence of a premature delivery: 1=preterm, 0=no evidence of preterm, 9=not applicable (outcome not a delivery)	INTEGER	1
postterm_ev	Flag to indicate evidence of a post-term delivery: 1=post-term, 0=no evidence of post-term, 9=not applicable (outcome not a delivery)	INTEGER	1
multiple_ev	Flag to indicate evidence of a multiple pregnancy: 1=multiple, 0=no evidence of multiple. Missing for pregnancy losses.	INTEGER	1
conflict	Flag to indicate whether the pregnancy episode overlaps with another episode (within a woman): 1=overlapping, 0= non-overlapping	INTEGER	1

; EDD=estimated date of delivery; LMP= last menstrual period; EDC=estimated date of conception; TOP=termination of pregnancy

<sup>1</sup> For “Outcome unknown” pregnancies, the imputed start date is obtained by subtracting 4 weeks from the earliest antenatal record in the episode.

<sup>2</sup> The timing of trimesters is estimated using a common convention: first trimester (first day of LMP [pregstart] to 13 completed weeks), second (weeks 14 to 26), and third (week 27 to delivery [pregend]).

<sup>3</sup> Late pregnancy records refer to the period up to 3 weeks before delivery, e.g. “Baby overdue”.

## Methodology

The CPRD Aurum Pregnancy Register is based on an algorithm developed in partnership with the London School of Hygiene and Tropical Medicine. Full details of the algorithm can be found in the following publication:

Minassian, C., Williams, R., Meeraus, W. H., Smeeth, L., Campbell, O., & Thomas, S. L. (2019). Methods to generate and validate a Pregnancy Register in the UK Clinical Practice Research Datalink primary care database. *Pharmacoepidemiology and drug safety*, 28(7), 923–933.  
<https://doi.org/10.1002/pds.4811>.

A list of more than 6,000 pregnancy-related SNOMED, Read and local EMIS® codes and numunitids, categorised according to the type of code (antenatal, pregnancy outcome [including delivery and pregnancy loss codes], postnatal, etc.), is used to identify patients who had a pregnancy. Pregnancy-related records are extracted from the Observation file and the EMIS Diary Table of all female patients between the ages of 11 and 49 years.

The following steps (1-8) are taken to delineate unique pregnancy episodes. In brief, this involves first identifying the end of each pregnancy (when an outcome was recorded) and estimating the pregnancy end date. The end date is then used as a point of reference from which to estimate the start of the pregnancy, in combination with additional data (when available) on gestational age, the estimated date of delivery, conception, or first day of a woman's last menstrual period (LMP).

### **1: Identify all pregnancy outcome records**

All records relating to pregnancy outcomes of any type (live births, stillbirths and early pregnancy losses) are extracted. Records relating to deliveries are considered separately to those relating to early pregnancy losses.

### **2: Date the first pregnancy outcome**

The date of each woman's first pregnancy outcome is estimated using the records identified in step 1.

### **3: Group together records relating to the first pregnancy outcome and determine the type of pregnancy**

Additional pregnancy outcome records relating to each woman's first outcome are assigned to that outcome. Characteristics of the delivery (pre- or post-term, stillbirth, multiple birth) or the type of early pregnancy loss (miscarriage, induced abortion, ectopic, molar pregnancy or blighted ovum) are determined from these assigned records.

### **4: Date and characterise each successive pregnancy outcome**

Steps 2 & 3 are repeated to identify, date and characterise successive pregnancy outcomes sequentially for each woman.

### **5: Estimate the start of each pregnancy episode**

Records relating to the timing of the start of pregnancy (first day of LMP) are used to estimate pregnancy start dates. In the absence of such data, pregnancy start dates are imputed according to the type of pregnancy outcome (as determined in step 3).

### **6: Adjust the start and end dates of pregnancy**

Adjustments to pregnancy start and end dates are made either when antenatal records are identified in the 4 weeks before the estimated pregnancy start date (indicating that the initial estimated start date was too late), or when the estimated pregnancy duration exceeds the maximum duration for that type of pregnancy outcome.

**7: Assign antenatal records to each pregnancy episode**

Antenatal records occurring between the start and end date of an identified pregnancy episode are assigned to the pregnancy.

**8: Identify additional pregnancies with no recorded outcome**

All remaining unassigned antenatal records are extracted and categorised into distinct pregnancy episodes. When available, records relating to the late pregnancy period or third trimester are used to estimate each pregnancy's end date, and the pregnancy start date is estimated as per step 5. In the absence of such records, the date of the latest antenatal record in the episode is used as a proxy for the pregnancy end date. The pregnancy start date is estimated using information on gestational age (when available) or by subtracting 4 weeks from the earliest antenatal record in the episode.

## **Backwards compatibility**

Refinements to the pregnancy algorithm are ongoing and will continue to be developed over time. As a result of this, and because the CPRD is a dynamic and updated data source, we cannot guarantee that new releases of the CPRD Aurum Pregnancy Register will contain all previously identified pregnancy episodes.

## **Caveats and notes**

The pregnancy algorithm does not impose any restrictions relating to the timing of the data (data recorded before and after the up-to-standard (UTS) date within a practice are used) or the duration of patient registration with a practice. The reason for this is to maximise the available data. The Pregnancy Register may therefore include pregnancy episodes occurring before the practice became UTS or before the woman joined her current practice (including historical events recorded in the first few months after a woman registers with a practice). Restrictions on UTS data and registration can be applied retrospectively, by the user, if needed.

As described in step 8 above, some pregnancy episodes have no recorded outcome. These episodes are flagged by the field “outcome” (13 = Outcome unknown). The estimated start and end dates of these episodes are based on the earliest and latest antenatal records respectively (in the absence of information on gestational age) and must therefore be treated with caution.

The estimated duration of each pregnancy episode is provided in the field “gestdays”. It is derived by subtracting the estimated start date from the estimated end date, hence any imprecision in these dates, for example, due to possible delays in recording a pregnancy outcome, or incorrect recall of LMP, may result in implausible durations for some pregnancy episodes. Examples may include deliveries exceeding 42 weeks’ gestation or less than 24 weeks, or gestational age that is incompatible with a preterm or post-term flag.

Additionally, some women’s pregnancy episodes may appear to conflict with each other. For example, a pregnancy loss nested within a delivery episode may represent a threatened miscarriage recorded as an actual miscarriage, culminating in a later delivery. Overlapping episodes are flagged by the field “conflict” and will need to be considered with great care when using the Pregnancy Register.