



# Quantifying second trimester pregnancy loss rates in the UK: Evaluating the suitability of the CPRD Pregnancy Register

---

Prof Lucy Smith

Dr Jonny Broomfield



# Background

---

While we collect robust reliable data on stillbirth (via MBRRACE-UK) there is no mandatory data collection on miscarriage and so the burden of pregnancy loss is unknown

Miscarriage data are much more challenging to gather as they occur in different settings

- Gynaecology, early pregnancy unit, maternity, at home etc
- Often without a hospital stay
- Some may go completely unreported to healthcare services



# Context

---

Second-trimester pregnancy loss has major psychological and social impacts on parents, similar to stillbirth

Stillbirth data collection has helped to quantify the burden and identify risk factors

Earlier pregnancy loss is many years/decades behind this

- No countries are reporting these losses



# Aims

---

Aim to estimate rates of miscarriage (second trimester loss) for the first time in the UK by linking general practice data with hospital episode statistics

- Using the Clinical Practice Research Database (CPRD) Pregnancy Register
- Focusing on 2000-2020

Key for future development of policy and practice in this area

Try and identify risk factors for loss, such as understanding any socioeconomic or ethnic inequalities



# What is the CPRD?

---

Electronic health record with primary and secondary care records

Available from consenting UK-based general practices

First established as the General Practice Research Database (GPRD)

- Renamed CPRD in 2012

Expanded over time to include linkage to hospital records, cancer registries, ONS mortality...

GOLD and Aurum



# What is the CPRD Pregnancy Register?

---

Derived from CPRD GOLD and Aurum

Algorithm-based identification that uses medical codes for pregnancy-related events (e.g. antenatal visits, delivery, miscarriage)

Estimates pregnancy start and end dates using recorded gestational age and outcome data

Captures pregnancies with and without explicitly recorded outcomes

Includes uncertain pregnancy episodes (e.g. missing outcomes, overlapping pregnancies)



# Pregnancy Register documentation

<i>Field name</i>	<i>Description</i>
patid	Encrypted unique patient identifier
pregid	Unique identifier of the pregnancy episode
mblbabies	Number of babies the pregnancy is linked to in the MBL
babypatid <sup>1</sup>	Encrypted unique patient identifier (linked baby)
babymob	Baby's month of birth as recorded in the baby's medical record
babyyob	Baby's year of birth as recorded in the baby's medical record
totalpregs	Total number of identified pregnancy episodes (per woman)
pregnumber	Pregnancy episode number (per woman)
pregstart	Estimated start date of pregnancy
firstantenatal	Date of earliest antenatal record within the pregnancy
startsource	Data source used to estimate pregnancy start date: 1 = Imputed <sup>2</sup> , 2 = EDD, 3 = LMP, 4 = Gestational age at birth, 5 = Gestational age from antenatal record, 6 = EDC

gestdays	Estimated duration of pregnancy episode in days (calculated as pregend minus pregstart)
matage	Mother's age at end of pregnancy (years)
outcome	Outcome of pregnancy: 1 = Live birth, 2 = Stillbirth, 3 = 1 and 2, 4 = Miscarriage, 5 = TOP, 6 = 4 or 5, 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>4</sup> , 13 = Outcome unknown



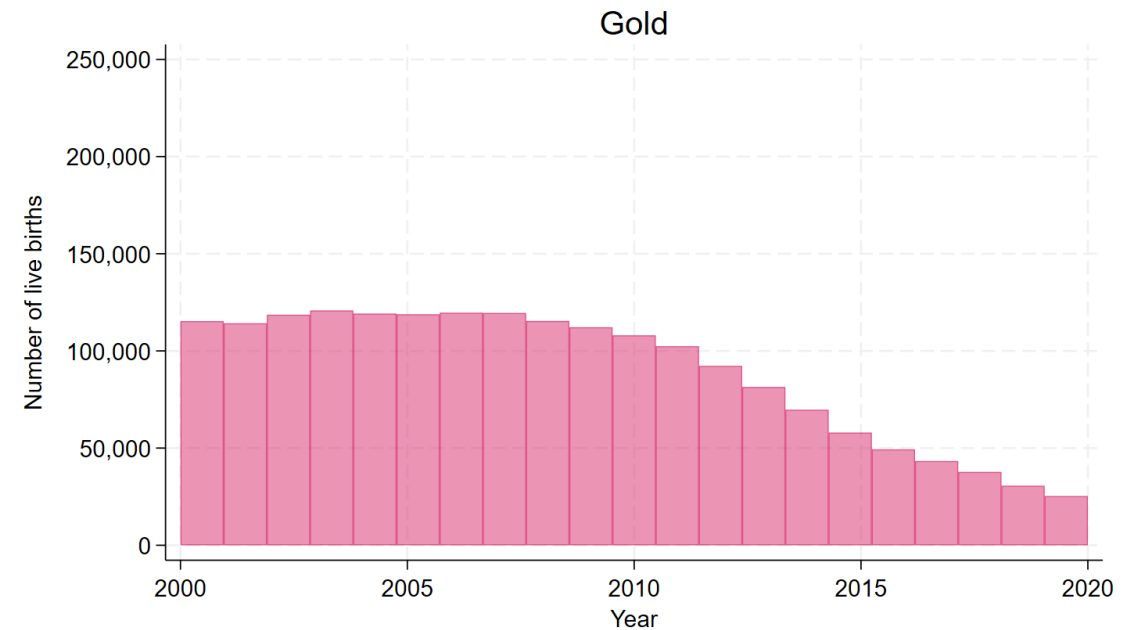
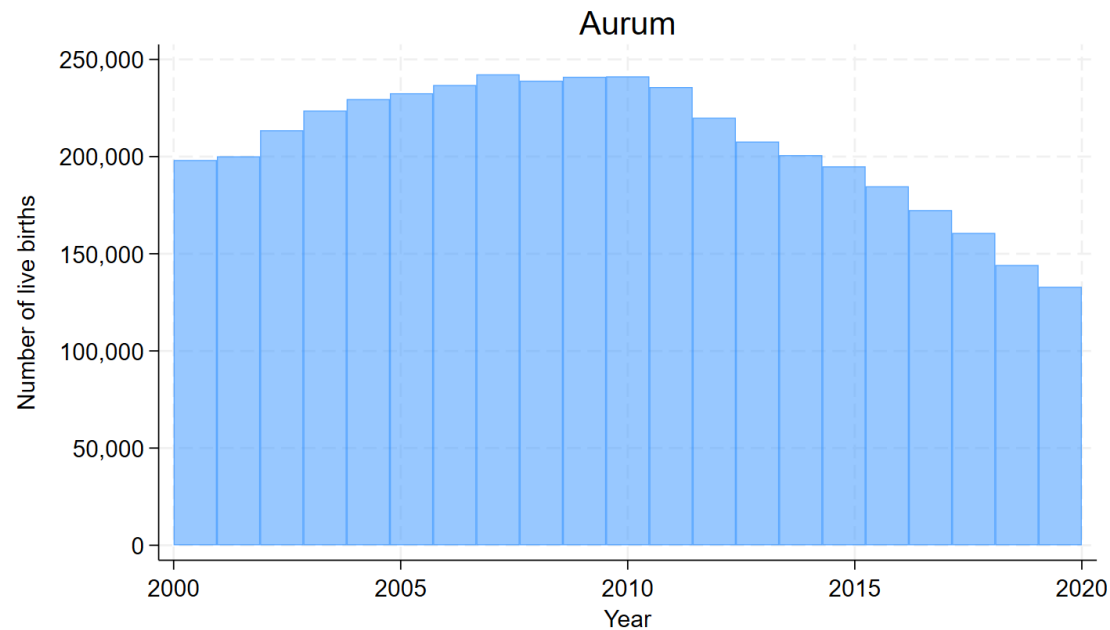
# outcome variable

- 1 = Live birth
- 2 = Stillbirth
- 3 = 1 and 2
- 4 = Miscarriage
- 5 = Miscarriage or termination of pregnancy (TOP)
- 6 = 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 (up to 3 weeks before delivery)
- 13 = Unknown

outcome	Freq.	Percent
Live birth	6,224,421	48.58
Stillbirth	29,551	0.23
Live stillbirth	417	0.00
Miscarriage	1,226,537	9.57
TOP	237,704	1.86
Miscarriage or TOP	1,171,626	9.14
Ectopic	114,130	0.89
Molar	23,776	0.19
Blighted ovum	3,739	0.03
Unspecified loss	81,878	0.64
Delivery - third trimester	137,787	1.08
Delivery - late	40,788	0.32
Unknown	3,520,210	27.47



# Number of live births by year

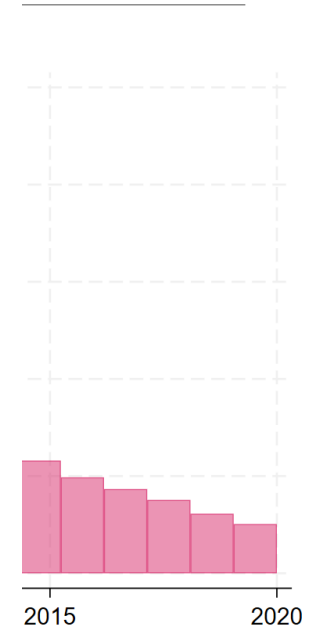
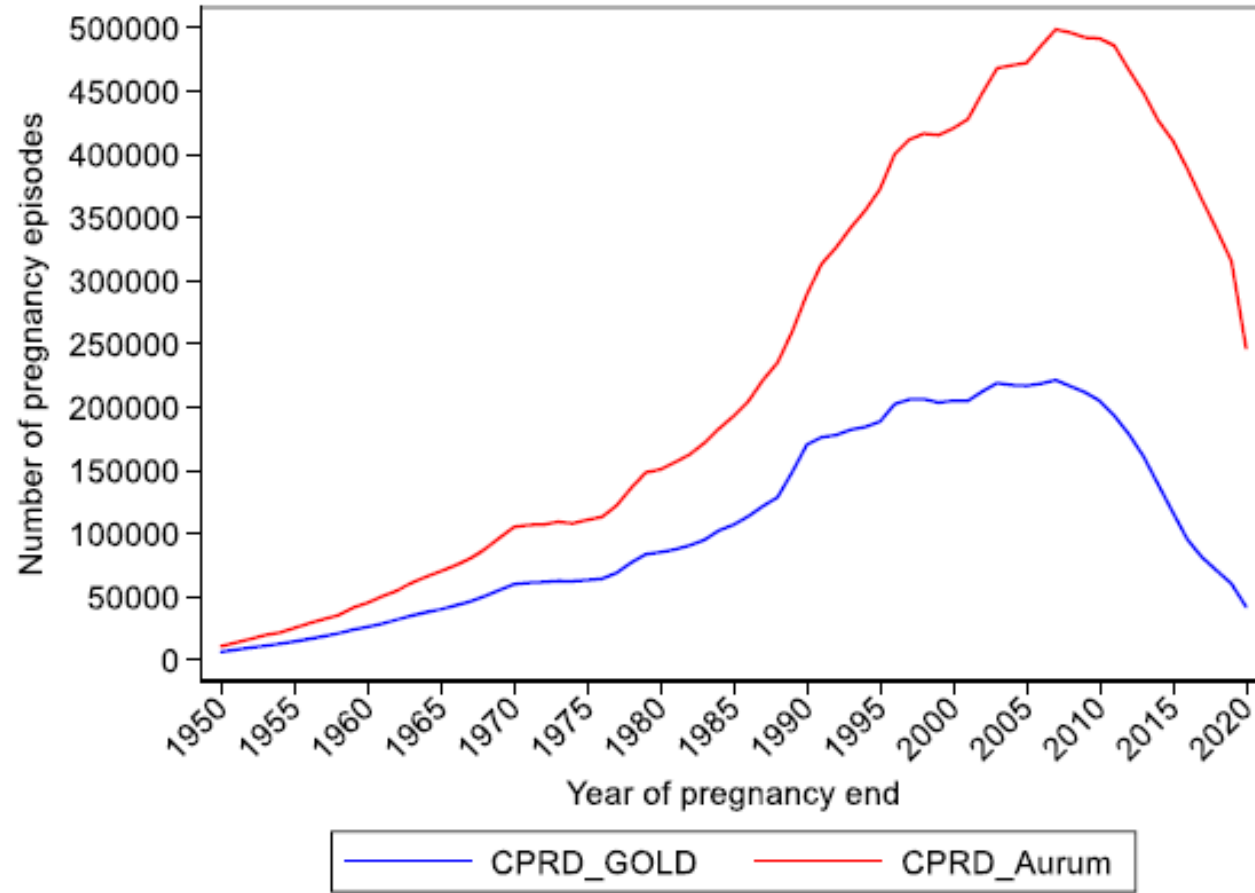
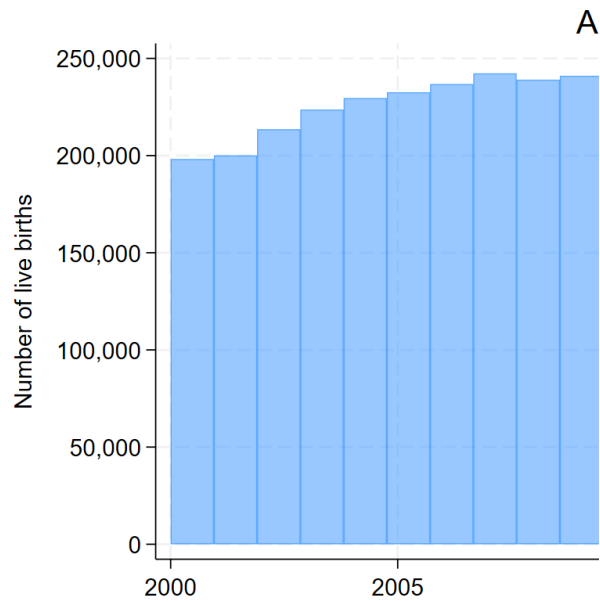


These plots show the number of live births observed in the CPRD Pregnancy Register for Aurum and Gold by year. The number of live births in Aurum decreases from 2010 onwards, while the number of women aged 16-50 registered to Aurum practices continues to increase. The number of live births in Gold also decreases, at a much faster rate than the number of registered women decreases.

At first we were concerned but...



# Number of live births



**FIGURE 1** Comparison of the number of pregnancy episodes by year in the Clinical Practice Research Datalink (CPRD) Aurum and CPRD GOLD Pregnancy Registers

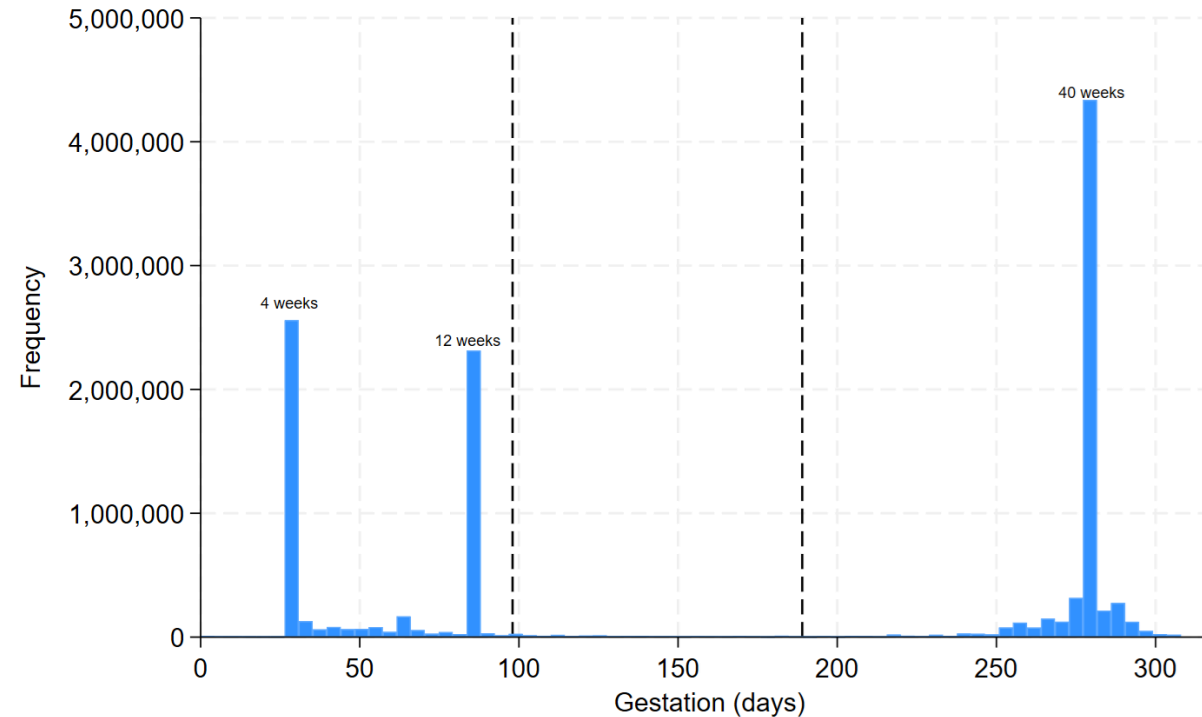
These plots show the number of live births in Aurum decreases from 2000 to 2020. The number of live births increases.

At first we were concerned but...

The number of live births continues to decrease.



# Gestation length in CPRD Pregnancy Register



Clear peaks in recording of gestation length at 4 weeks, 12 weeks and 40 weeks

These correspond to unknown outcomes, imputed miscarriages and imputed live births respectively



# startsource variable

Data source used to estimate pregnancy start date

1 = Imputed

2 = Estimated date of delivery

3 = Last menstrual period

4 = Gestational age at birth

5 = Gestational age from antenatal record

6 = Estimated date of conception

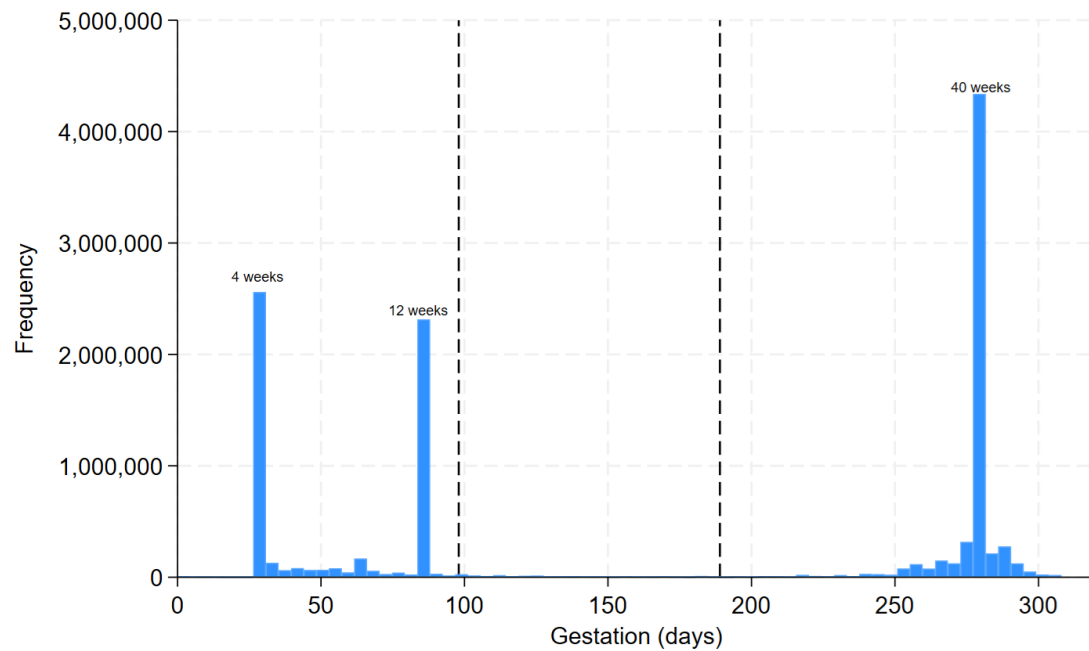
Missing = unknown (outcome==13)

startsource	Freq.	Percent	Cum.
1	6,770,255	52.84	52.84
2	1,079,878	8.43	61.27
3	568,767	4.44	65.71
4	181,277	1.41	67.12
5	691,662	5.40	72.52
6	515	0.00	72.53
.	3,520,210	27.47	100.00
Total	12,812,564	100.00	

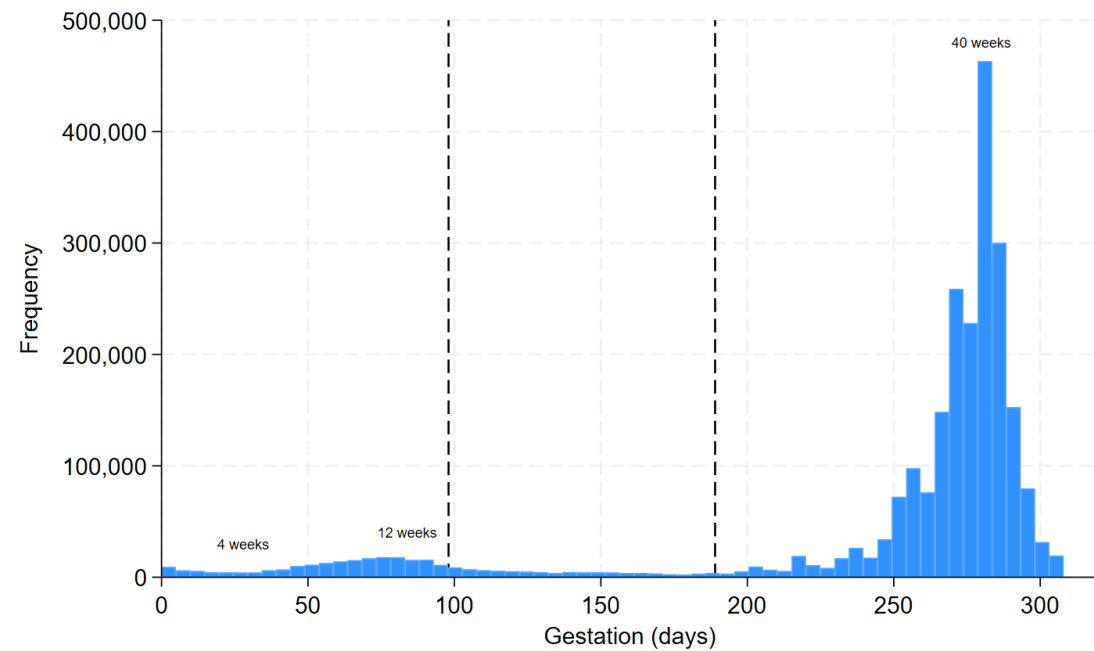


# Gestation length in CPRD Pregnancy Register

Including imputed start date/unknown outcomes

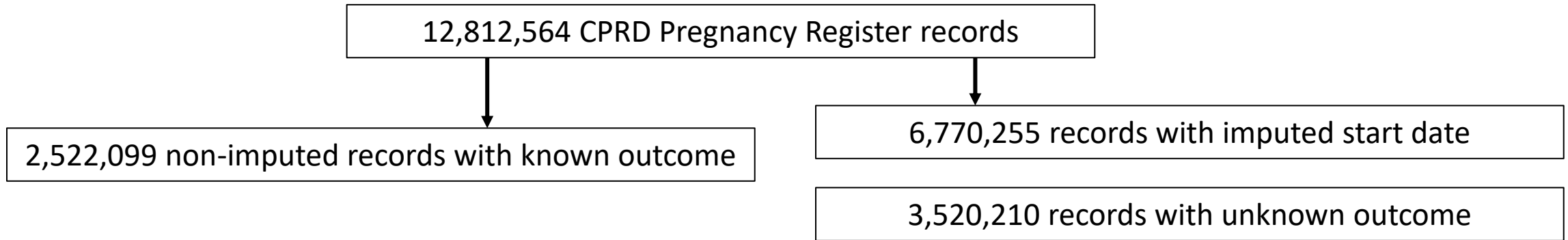


Excluding imputed start date/unknown outcomes

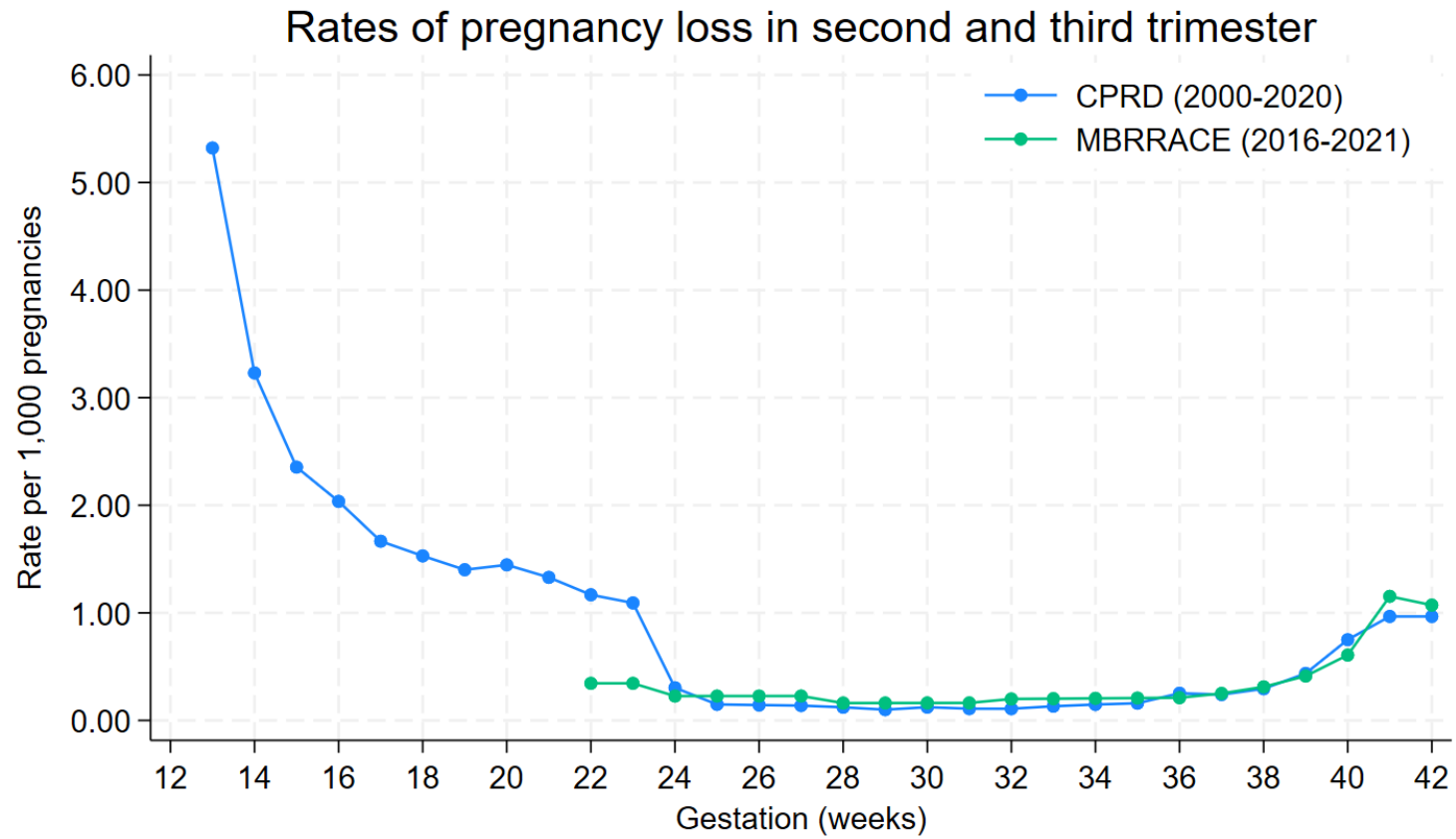


# Current analysis – CPRD

---



# Analysis on non-imputed records only





# HES pregnancy records

We have HES records for all mothers in CPRD Pregnancy Register from 1/1/2000 – 31/12/2020, including ICD-10 codes (within Diagnosis file) for:

- Z37 – outcome of delivery
- O03 – spontaneous abortion including miscarriage

<b>Z37</b>	<b>Outcome of delivery</b>
<b>Z37.0</b>	<b>Single live birth</b>
<b>Z37.1</b>	<b>Single stillbirth</b>
<b>Z37.2</b>	<b>Twins, both liveborn</b>
<b>Z37.3</b>	<b>Twins, one liveborn and one stillborn</b>
<b>Z37.4</b>	<b>Twins, both stillborn</b>
<b>Z37.5</b>	<b>Other multiple births, all liveborn</b>
<b>Z37.6</b>	<b>Other multiple births, some liveborn</b>
<b>Z37.7</b>	<b>Other multiple births, all stillborn</b>
<b>Z37.9</b>	<b>Outcome of delivery, unspecified</b>

## **O03** Spontaneous abortion

*Incl.:* miscarriage

- .0 Incomplete, complicated by genital tract and pelvic infection**  
With conditions in O08.0
- .1 Incomplete, complicated by delayed or excessive haemorrhage**  
With conditions in O08.1
- .2 Incomplete, complicated by embolism**  
With conditions in O08.2
- .3 Incomplete, with other and unspecified complications**  
With conditions in O08.3-O08.9
- .4 Incomplete, without complication**
- .5 Complete or unspecified, complicated by genital tract and pelvic infection**  
With conditions in O08.0
- .6 Complete or unspecified, complicated by delayed or excessive haemorrhage**  
With conditions in O08.1
- .7 Complete or unspecified, complicated by embolism**  
With conditions in O08.2
- .8 Complete or unspecified, with other and unspecified complications**  
With conditions in O08.3-O08.9
- .9 Complete or unspecified, without complication**



# HES pregnancy records

We have HES records for all mothers in CPRD Pregnancy Register from 1/1/2000 – 31/12/2020, including ICD-10 codes (within Diagnosis file) for:

- Z37 – outcome of delivery
- O03 – spontaneous abortion including miscarriage

## Z37 Outcome of delivery

ICD	Freq.	Percent
Z37.0	4,459,281	98.09
Z37.1	21,200	0.47
Z37.2	62,014	1.36
Z37.3	1,710	0.04
Z37.4	340	0.01
Z37.5	975	0.02
Z37.6	118	0.00
Z37.7	35	0.00
Z37.9	306	0.01

## O03 Spontaneous abortion

*Incl.:* miscarriage

ICD	Freq.	Percent
O03.0	1,217	0.42
O03.1	14,972	5.18
O03.2	42	0.01
O03.3	2,524	0.87
O03.4	144,216	49.94
O03.5	774	0.27
O03.6	4,195	1.45
O03.7	20	0.01
O03.8	985	0.34
O03.9	119,834	41.50



# HES pregnancy records

---

We have HES records for all mothers in CPRD Pregnancy Register from 1/1/2000 – 31/12/2020, including ICD-10 codes (within Diagnosis file) for:

- Z37 – outcome of delivery
- O03 – spontaneous abortion including miscarriage

Z37 (and small number of O03) records can be linked to gestational age

- For those that can't, an imputed start date will be assumed when matching to CPRD records (next slide)
  - 40 weeks before HES episode start date for live births and stillbirths (Z37), 24 weeks before for miscarriages (O03)



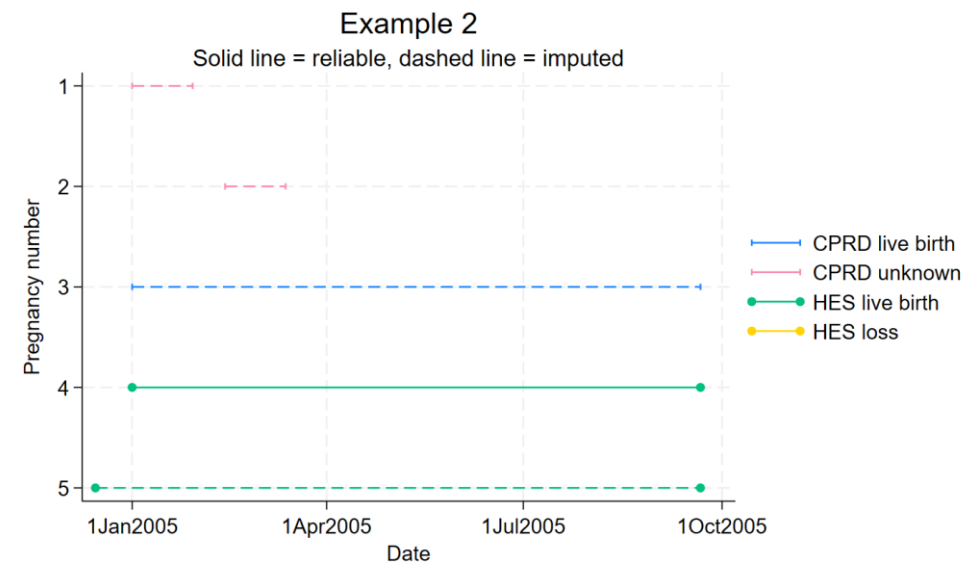
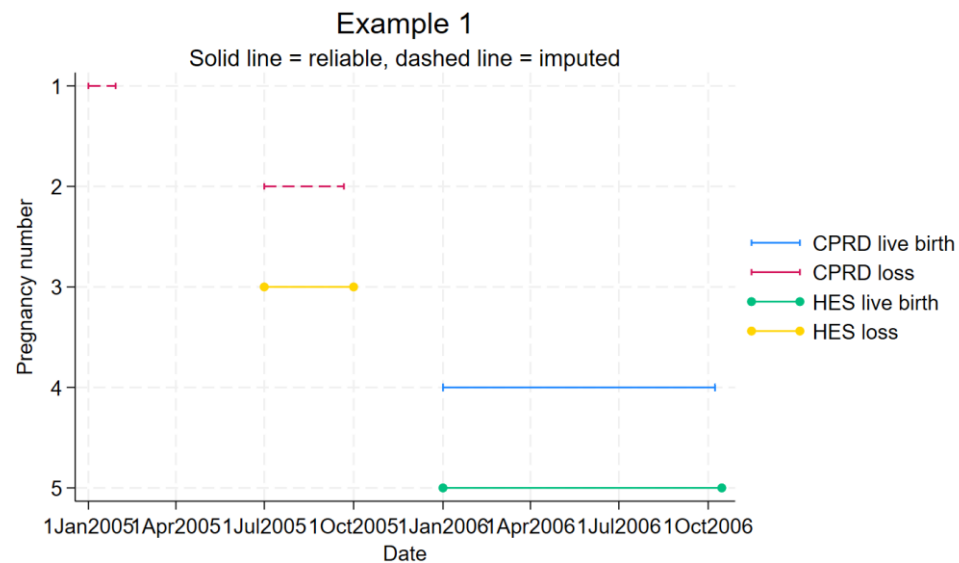
# Validating CPRD records with HES records

Reliable CPRD records with non-imputed start date and known outcomes

- Ready for analysis

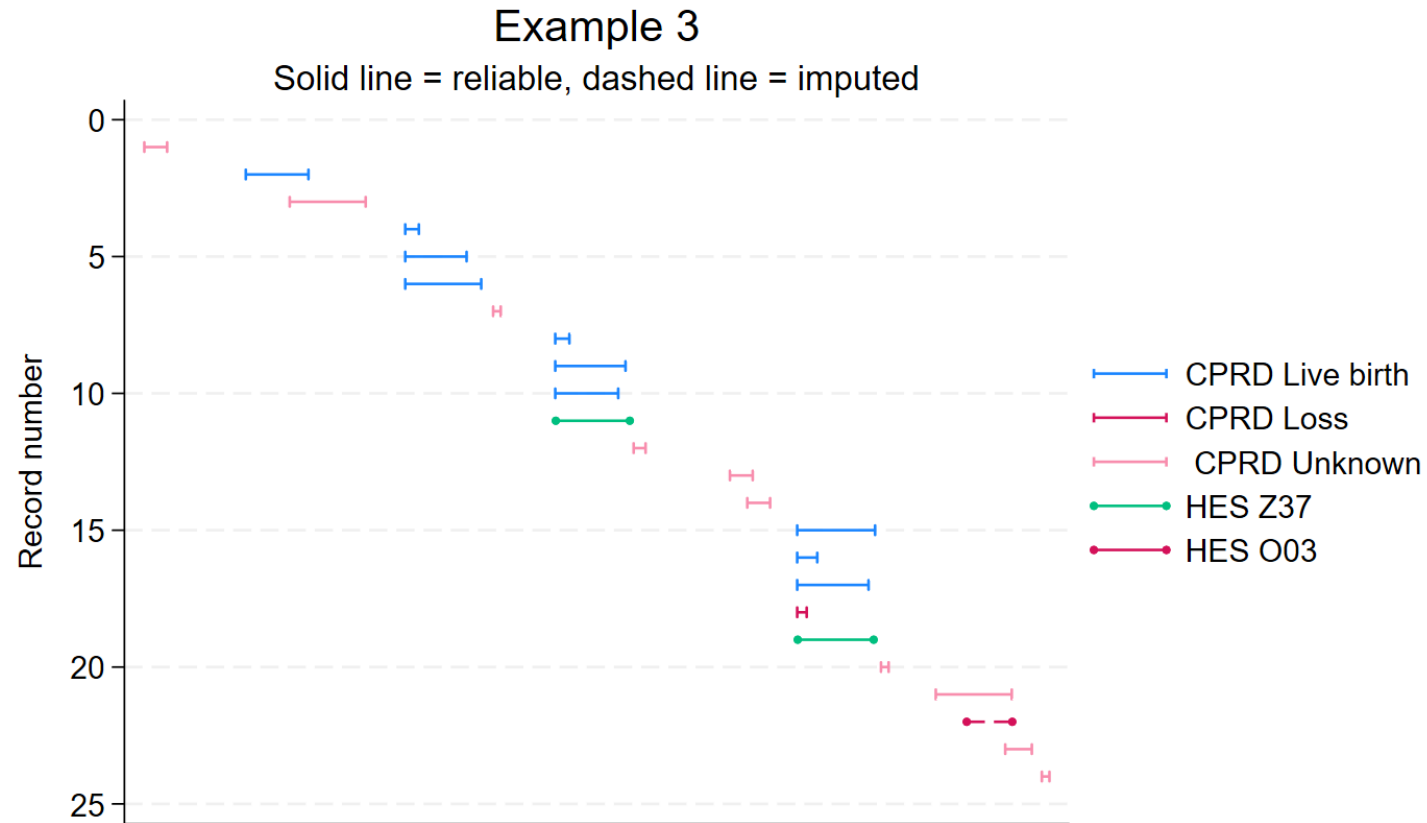
Unreliable CPRD records with imputed start date and/or unknown outcomes

- Match to HES records to determine which are accurate and which conflict



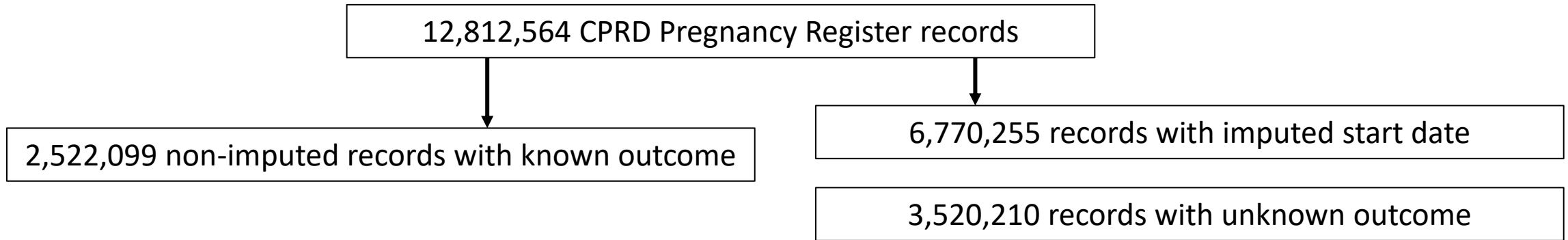


# Validating CPRD records with HES records



# Current analysis – CPRD

---



# Current analysis – CPRD + HES

---

12,812,564 CPRD Pregnancy Register records

2,522,099 non-imputed records with known outcome

6,770,255 records with imputed start date

3,520,210 records with unknown outcome

288,779 HES O03 records

4,545,979 HES Z37 records

?? non-imputed records with known outcome

+





# Overlapping pregnancy records

---

`conflict` variable in CPRD Pregnancy Register dataset denoting whether the pregnancy episode overlaps with another episode for the same woman

We will also potentially introduce some conflicts when validating with HES data (see example 2)

We have defined a set of rules to determine which records, if any, are reliable to include in the final analysis

- Prioritise non-imputed records and records with known outcomes
- Prioritise biologically plausible records (e.g. over a live birth recorded before 22 weeks)
- If records are too conflicting (e.g. between CPRD and HES), do not include in final analysis

Following these, we will include non-overlapping, non-conflicting (reliable) CPRD records with a non-imputed start date and a known outcome (live birth, stillbirth or miscarriage) in the final analysis



# Final analysis

---

Main analysis will estimate crude rate of pregnancy loss across second and third trimesters (13<sup>+0</sup> to 42<sup>+0</sup> weeks)

- Second trimester rates (13<sup>+0</sup> to 21<sup>+6</sup>) are of main interest
- Third trimester rates (22<sup>+0</sup> to 42<sup>+0</sup>) will be compared to MBRRACE-UK stillbirth rates for external validation

Secondary analysis will investigate variability in rates by demographic factors

- Maternal age
- Birth cohort (5 categories from 2000-2020)
- Maternal ethnicity
- Socioeconomic status (Index of Multiple Deprivation, quintiles)
- Birth number (first pregnancy or not)



# Summary

---

Robust UK-based estimates of second-trimester pregnancy loss, while acknowledging limitations of the data

Balance between needing estimates to quantify the burden and addressing the quality of the available data

We would appreciate any thoughts on striking this balance or suggestions on how best to combine CPRD and HES records into a final analysis cohort

[jb781@le.ac.uk](mailto:jb781@le.ac.uk)

With thanks to Dr Jenny Campbell, Dr Clare Gillies, Dr David Lo, Dr Elpida Vounzoulaki and Dr Freya Tyrer for their comments and suggestions so far